

PREVENTING DIARRHEAL INFECTIONS WITH HOUSEHOLD WATER TREATMENT:
LESSONS FROM SIMULATION MODELS

By

Kyle Scott Enger

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Fisheries and Wildlife

2012

ABSTRACT

PREVENTING DIARRHEAL INFECTIONS WITH HOUSEHOLD WATER TREATMENT: LESSONS FROM SIMULATION MODELS

By

Kyle Scott Enger

Diarrheal disease kills two million children per year in developing countries. Diarrhea is controlled in industrialized countries by systems to remove sewage and distribute clean water. These systems are difficult to fund, build, and maintain in developing countries, so simpler technologies are promoted: e.g., household water treatment (HWT), handwashing, and latrines. These technologies, however, require consistent effort by individuals for proper use and maintenance, defined here as 'compliance'. Measuring compliance is difficult, and often neglected.

It is important to understand how the extent and pattern of compliance within communities affects the prevention of diarrhea by HWT, while accounting for biases in field trials and characteristics of natural transmission systems, such as the presence of multiple pathogens, transient spikes of contamination, and multiple transmission routes. This question was answered by: reanalyzing and generalizing results from a HWT field trial, using a quantitative microbial risk assessment (QMRA) model to adjust for bias (chapter 3); examining the joint effects of HWT antimicrobial efficacy and compliance on prevention of diarrhea (chapter 4); and using a model of diarrheal infection transmission incorporating multiple routes of infection to further examine efficacy and compliance issues with HWT (chapter 5).

The QMRA model of the field trial found that compliance greatly affected HWT effectiveness: with low compliance, 10% of diarrhea was prevented; with high compliance, 90% was prevented. It also estimated source water pathogen concentrations source water that were

consistent with measurements from other developing countries. The model found that the effect of an imperfect placebo device used during the field trial depended on the assumed level of compliance during the field trial.

The QMRA model was modified to examine how HWT compliance and antimicrobial effectiveness jointly altered diarrheal disease risk. Given perfect compliance, increasing antimicrobial effectiveness always lowered risk. If compliance was incomplete, increasing antimicrobial effectiveness eventually ceased to lower risk, except in a few scenarios with high incidence, high compliance, or large water contamination spikes. The pattern of compliance by communities also influenced risk; e.g., risk was lower if 90% of people used HWT perfectly and 10% never used HWT, than if 100% of people used HWT 90% of the time.

A preliminary transmission model simulated a community in which infected people shed pathogens, which were ingested by other people via exposure to land, drinking water, their household environment, or visits from other households. It found similar results to the QMRA models regarding compliance. Transmission of diarrheal pathogens by household visits appeared unimportant compared to other routes; however, visits consisted of exchanges of pathogens between household environments. Other types of visits (e.g., shared child care) might lead to greater transfer of pathogens and a greater influence of visits on illness. The model also inferred that viruses and protozoa were attenuated (removed from the system, e.g., by decay or sequestration) ~10 times faster than bacteria. Future sensitivity and uncertainty analyses will highlight important aspects of the model and its parameters that contribute to its results.

The amount and pattern of compliance strongly affects diarrhea prevention by HWT. Further research should be conducted on improving compliance with HWT in developing countries.

© Copyright by

Kyle Scott Enger

2012

Some rights reserved.

This work is licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported License.

To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/3.0/> or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.



Computer code in this work is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

Computer code in this work is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

<http://www.gnu.org/licenses/>

The full GNU General Public License is also reproduced in the appendix of this work (chapter 10), with the computer code.



ACKNOWLEDGEMENTS

I appreciate the assistance and oversight of my graduate committee members, who reliably ask and answer difficult questions: Joseph N. S. Eisenberg, Anne E. Ferguson, and Cheryl A. Murphy, led by my major professor, Joan B. Rose.

I thank Joseph N. S. Eisenberg and Kara L. Nelson for substantial guidance on model design and manuscript editing; Bryan T. Mayer for reviewing the program code for chapters 3, 4, and 5; Sophie Boisson and Thomas Clasen for discussion and interpretation of the LifeStraw® randomized controlled trial (Boisson et al., 2010) and access to original data; Mark H. Weir for discussions of modeling and programming; and Ian M. Dworkin, Joseph N. S. Eisenberg, Daniel B. Hayes, James D. A. Millington, and Carl P. Simon for their excellent instruction in graduate courses concerning modeling, programming, and mathematics. I am also grateful to the Michigan State University High Performance Computing Center (HPCC) and the Institute for Cyber Enabled Research (iCER), for providing hardware and software to run the simulations described in chapters 4 and 5, as well as technical assistance (particularly Dirk J. L. Colbry).

My work was financially supported by: a University Distinguished Fellowship from the Michigan State University Graduate School; a research assistantship from Joan B. Rose and the Center for Advancing Microbial Risk Assessment (CAMRA); and a grant from the Vestergaard Frandsen Corporation, which produces the LifeStraw® Family filtration device; that grant's principal investigators were Joseph N. S. Eisenberg and Kara L. Nelson. Although this represents a potential conflict of interest, we attest that the Vestergaard Frandsen Corporation did not influence study design, results, or the decision to publish the results.

In conclusion, I am particularly grateful to Joan B. Rose for many insightful and thought-provoking conversations; research oversight; generous funding for my stipend, travel, and equipment; and a great deal of editing.

TABLE OF CONTENTS

LIST OF TABLES.....	xi
LIST OF FIGURES.....	xii
1. INTRODUCTION	1
1.1. Brief summary of diarrheal disease and control	1
1.2. Scientific questions	2
1.3. General assumptions within this dissertation	4
2. REVIEW OF DIARRHEAL INFECTION: EPIDEMIOLOGY, INTERVENTIONS, AND MODELING	5
2.1. Introduction	5
2.2. Definition and measurement of diarrhea	6
2.2.1. Persistent diarrhea	8
2.3. Multiple infections	9
2.4. Asymptomatic infections	9
2.5. Burden of diarrheal disease	10
2.5.1. Morbidity	10
2.5.2. Mortality	10
2.5.3. Disability-adjusted life years (DALYs) attributable to diarrhea	11
2.6. Changes in diarrheal morbidity over time in individuals	13
2.7. Diarrhea-malnutrition vicious cycle	14
2.8. Cyclic or recurring changes ('seasonality') in diarrhea incidence	16
2.9. Crowding (urban vs. rural)	17
2.10. Bias in epidemiological studies of diarrhea	17
2.11. Relative contributions of pathogens to diarrheal etiology	19
2.12. Survey of diarrheal pathogens	22
2.12.1. Viral diarrheal pathogens	22
Rotavirus.....	23
Norovirus.....	25
2.12.2. Bacterial diarrheal pathogens	27
Pathogenic Escherichia coli.....	27
Campylobacter species.....	31
2.12.3. Protozoan diarrheal pathogens	33
Cryptosporidium parvum and Cryptosporidium hominis.....	33
Giardia species.....	35
2.12.4. Metazoan pathogens (helminths)	37
2.13. Interventions to prevent transmission of diarrheal pathogens	38
2.14. Effectiveness trials of interventions	38
2.14.1. Measures of effect in intervention trials	38
2.14.2. Nature of bias in intervention trials	40
2.15. Compliance with interventions, and long-term sustainability	42
2.15.1. Costs of compliance (monetary and otherwise)	43

2.15.2.	Psychological, social, and cultural aspects of compliance	44
2.15.3.	Examples of compliance measurements in the field	45
2.16.	Interaction of intervention effects	46
2.17.	Descriptions of individual interventions	49
2.17.1.	Sanitation	49
2.17.2.	Water supply improvement	50
	Water quantity improvement.....	51
	Water quality improvement.....	51
2.17.3.	Hygiene	52
	Handwashing.....	52
	Diapering and open defecation.....	54
	Anal cleansing.....	55
	Food preparation.....	55
	Fly control.....	56
2.17.4.	Household water treatment (HWT), or point-of-use (POU) technology	56
	Safe storage.....	57
	Boiling.....	57
	Solar disinfection (SODIS).....	58
	Chlorination.....	59
	Filtration.....	60
	Biosand filters.....	61
	Ceramic filters.....	62
	Advanced filter technologies.....	62
	Recent controversy surrounding HWT.....	65
2.17.5.	Other interventions	66
2.17.6.	Gaps in knowledge about diarrheal disease interventions	67
2.18.	Infection transmission modeling and its application to diarrheal disease	68
2.18.1.	Types of models	68
2.18.2.	Model verification and validation	70
2.19.	Modeling transmission of diarrheal infections	71
2.19.1.	Simple mathematical example of a transmission model	72
2.19.2.	Environmental infection transmission models	73
2.19.3.	Quantitative microbial risk assessment (QMRA) models	75
2.20.	Conceptual model of diarrheal disease transmission	76
2.21.	Theoretical issues regarding interventions	77
2.22.	Tools and information useful for modeling diarrhea transmission	79
2.23.	Published models relevant to endemic diarrhea transmission	81
2.23.1.	Mechanistic model of diarrheal infection: Eisenberg et al., 2007	81
2.23.2.	Empirical model of diarrheal disease: Schmidt et al., 2009	82
2.23.3.	Modeling indirect effects of interventions: Halloran et al., 2002	83
2.23.4.	Environmental infection transmission system (EITS) models: Li et al., 2009	84
2.24.	Conclusion	85
3.	LINKING A QUANTITATIVE MICROBIAL RISK ASSESSMENT MODEL TO A HOUSEHOLD WATER TREATMENT FIELD TRIAL	86
3.1.	Abstract	86
3.2.	Introduction	87

3.3. Materials and Methods	89
3.3.1. Conceptual framework linking QMRA models to epidemiological studies	89
3.3.2. Model description	90
Step 1: Parameter entry.....	93
Step 2: Parameter values inferred through calibration.....	94
Step 3: Initiating each run; establishing equilibrium waterborne infection.....	96
Step 4: Calculating daily doses of marker pathogens.....	97
Step 5: Dose response functions.....	98
Step 6: Assignment of infection.....	99
Step 7: Assignment of infection duration, recovery, and immunity.....	99
Step 8: Surveying the population about reported diarrhea.....	100
Step 9: Determining model outcomes corresponding to the Lifestraw RCT... ..	101
Step 10: Repetition of calibration runs.....	102
3.3.3. Example runs of the model	103
3.3.4. Analytical process	105
3.4. Results	106
3.4.1. Calibration step	106
3.4.2. Estimation step	109
3.5. Discussion	113
3.5.1. Predicting pathogen concentrations in drinking water sources	116
3.5.2. Calibration of microbial risk assessment models	116
4. THE JOINT EFFECTS OF EFFICACY AND COMPLIANCE IN HOUSEHOLD WATER TREATMENT EFFECTIVENESS	119
4.1. Abstract	119
4.2. Introduction	120
4.3. Materials and methods	123
4.3.1. Compliance	124
4.3.2. Baseline incidence and etiologic fraction	125
4.3.3. Short-term contamination spikes	125
4.3.4. Calibration step	126
Determination of etiologic fractions.....	129
4.3.5. Estimation step	130
4.3.6. Replication of the WHO model	130
4.4. Results	131
4.4.1. Calibration step	131
4.4.2. Estimation step	134
Comparison with the WHO QMRA model.....	134
Effect of LRVs given imperfect compliance.....	135
4.4.3. Charts of median incidences from all estimation scenarios	143
4.4.4. Assessing impact of high LRVs: significance testing & classification trees	150
4.5. Discussion	154
4.5.1. Information needed to inform models of diarrheal infection transmission	155
Pathogen concentrations in source waters.....	156
Etiology of diarrheal disease.....	156
Routes of transmission other than drinking water.....	156
4.5.2. Conclusions	157

5. TRANSMISSION MODEL OF DIARRHEAL INFECTION	158
5.1. Abstract	158
5.2. Introduction	159
5.3. Materials and methods	160
5.3.1. General description of the model	160
Summary of the steps in each model run.....	165
5.3.2. Technical description of the model	169
Step 1: Enter parameters.....	171
Step 2: Create random number tables and output logs.....	171
Step 3: Set up the simulated community.....	171
Step 4: Start daily loop and tally people in all states.....	174
Step 5: Defecation.....	174
Step 6: Transfer pathogens from land to surface water.....	175
Step 7: Transfer pathogens between households via visits.....	175
Step 8: Resupply stored water & apply household water treatment (HWT).....	176
Step 9: Pathogen transfer from household environment to drinking water.....	177
Step 10: First attenuation of pathogens.....	177
Steps 11 & 12: Decrement all status counters and apply status shifts.....	178
Step 13: Calculation of daily pathogen doses and dose response.....	178
Step 14: Assignment of baseline exposures.....	180
Step 15: Second inactivation of pathogens.....	180
Step 16: Continue to next day, or end simulation.....	181
5.3.3. Calibration and estimation	181
Determination of calibration parameter ranges.....	182
5.4. Results	183
5.4.1. Calibration	183
5.4.2. Estimation	191
5.5. Discussion	196
5.5.1. Calibration step	196
5.5.2. Estimation step	197
5.5.3. Limitations of the EITS model	197
5.5.4. Insights gained during the model construction process	200
5.5.5. Future applications of the model	202
Readily achievable applications.....	202
Applications requiring substantial modifications to the model.....	203
5.5.6. Conclusions	205
6. CONCLUSIONS	207
6.1. Summary of research	207
6.2. Implications for future diarrheal research and prevention efforts	209
6.2.1. Conduct and description of field trials	209
6.2.2. Compliance with interventions that prevent diarrheal infections	211
7. APPENDIX A: DISCUSSION OF PARAMETER VALUES USED IN THE MODELS	214
7.1. Water ingestion rate	214
7.2. Hand-mouth contacts per day	215
7.3. Log10 reduction values (LRVs) attributable to interventions	215
7.3.1. LRVs attributable to sanitation	215

7.3.2. LRVs attributable to intervention and placebo filters in the Lifestraw RCT	216
7.4. Dose response functions	218
7.4.1. Dose response for E. coli infection	219
7.4.2. Dose response for rotavirus and Giardia	219
7.5. Incubation periods	219
7.6. Morbidity ratios	220
7.6.1. Morbidity ratios in young children	220
7.6.2. Morbidity ratios in adults	221
7.7. Durations of illness and infection	222
7.7.1. Duration of diarrheagenic E. coli infection and illness	223
7.7.2. Duration of rotavirus infection and illness	223
7.7.3. Duration of Giardia infection and illness	223
7.8. Duration of immunity	224
7.9. Incomplete recall of diarrheal disease	224
7.10. Fecal excretion of pathogens	225
7.10.1. Concentrations of pathogens in feces	225
7.10.2. Amount of feces excreted	226
7.11. Inactivation, removal, or attenuation of pathogens in the environment	227
7.12. Pathogen movement from land to surface water	228
7.13. Demographic parameters	228
7.14. Summary table of all parameter values	228
8. APPENDIX B: ADDITIONAL INTERVENTIONS FOR MITIGATING DIARRHEA	235
8.1. Nutritional interventions	235
8.1.1. Breastfeeding	235
8.1.2. Zinc supplementation	236
8.1.3. Other nutrients	237
8.2. Treatment of diarrheal illness	237
8.2.1. Oral rehydration salts/solution/therapy (ORS)	237
8.2.2. Access to care	238
8.3. Vaccination	238
9. APPENDIX C: SOURCE CODE FOR THE MODELS	239
9.1. Overview of the source code for the models	239
9.2. GNU General Public License	239
9.3. The QMRA model simulating the Lifestraw field trial (chapter 3)	254
9.4. The QMRA model investigating compliance and LRVs (chapter 4)	303
9.5. The EITS model (chapter 5)	375
10. APPENDIX D: GLOSSARY	446
11. REFERENCES	453

LIST OF TABLES

Table 2.1. Log10 reduction values (LRVs) for various interventions.....	64
Table 2.2. Log10 reduction values: standards for household water treatment (HWT).....	65
Table 2.3. Terms commonly used to describe or categorize models.....	69
Table 3.1. Fixed parameter values used in the QMRA model of the Lifestraw RCT.....	93
Table 3.2. Ranges for stochastically varying parameters.....	96
Table 3.3. Longitudinal prevalence measures from the Lifestraw field trial.....	103
Table 4.1. Compliance among individuals in each model run, given compliance type β	124
Table 4.2. Criteria for the calibration step of the QMRA model.....	125
Table 5.1. Summary of calibration parameters.....	169
Table 5.2. Ranges over which calibration parameters were sampled.....	169
Table 5.3. Description of matrices tracking households and people.....	173
Table 5.4. Criteria for calibrating the transmission model.....	182
Table 5.5. Association of calibration parameters with incidence.....	188
Table 6.1. Important community characteristics for measurement in field trials.....	210
Table 7.1. Summary table of all parameter values.....	230

LIST OF FIGURES

Figure 2.1. Diarrhea mortality by WHO region.....	11
Figure 2.2. Diarrhea DALYs by WHO region.....	13
Figure 2.3. Estimated etiology of childhood diarrhea worldwide.....	20
Figure 2.4. Simple F-diagram of transmission of diarrheal pathogens.....	49
Figure 2.5. Schematic of the Susceptible-Infectious-Removed (SIR) model.....	72
Figure 2.6. Simple environmental infection transmission system model.....	74
Figure 2.7. Expanded diagram of diarrheal pathogen transmission.....	77
Figure 3.1. Conceptual model for simulation of a randomized controlled trial.....	90
Figure 3.2. Simulation model flowchart.....	92
Figure 3.3. Comparison of dose response functions.....	99
Figure 3.4. Example run of the model, with higher infection levels than the Lifestraw RCT.....	104
Figure 3.5. Example run of the model, infection levels consistent with the Lifestraw RCT.....	105
Figure 3.6. Distributions of LPRs consistent with the Lifestraw RCT.....	108
Figure 3.7. Distributions of simulated microbial concentrations.....	109
Figure 3.8. LPR distributions for differing compliance assumptions and placebo behavior.....	111
Figure 3.9. Longitudinal prevalence distributions under differing compliance assumptions.....	112
Figure 3.10. Longitudinal prevalence of waterborne infection in the estimation step.....	113
Figure 4.1. Calibration results assuming medium incidence of diarrhea.....	128
Figure 4.2. Mean baseline pathogen concentrations from calibration.....	133
Figure 4.3. Comparison with WHO model.....	135
Figure 4.4. Effect of compliance with HWT on the incidence ratio of diarrhea, by LRV.....	136
Figure 4.5. Effect of compliance and spikes on the IR of childhood diarrhea, by LRVs.....	137
Figure 4.6. Effect of dose response function nonlinearity at high doses.....	138

Figure 4.7 Incidence ratio of diarrhea by compliance level & type, spikes, and LRVs.....	139
Figure 4.8. Effect of compliance on IR if large contamination spikes occur.....	141
Figure 4.9. Effect of compliance on IR with extreme pathogen concentrations.....	143
Figure 4.10. Detailed estimation results (low incidence).....	145
Figure 4.11. Detailed estimation results (medium incidence).....	146
Figure 4.12. Detailed estimation results (high incidence).....	147
Figure 4.13. Detailed estimation results, WHO/EPA recommended LRVs.....	149
Figure 4.14. Detailed estimation results, extremely high baseline pathogen concentrations.....	150
Figure 4.15. Classification tree for incidence difference (ID) criterion.....	152
Figure 4.16. Classification tree for incidence ratio (IR) criterion.....	153
Figure 5.1. Simplified overview of the simulated community.....	162
Figure 5.2. Structure of each household within the simulated community.....	163
Figure 5.3. Daily progression of the EITS model of diarrheal infections.....	165
Figure 5.4. Flowchart of the operations of the EITS model.....	170
Figure 5.5. Calibration output from EITS model, transfer parameters.....	185
Figure 5.6. Calibration output from EITS model, scatterplots & histograms.....	186
Figure 5.7. Simulated diarrhea incidence in children (calibration step).....	187
Figure 5.8. Distributions of transfer calibration parameters.....	190
Figure 5.9. Distributions of attenuation calibration parameters.....	191
Figure 5.10. Estimation step, incidence by LRV of HWT.....	193
Figure 5.11. Comparison of EITS and QMRA results.....	195

1. INTRODUCTION

1.1. Brief summary of diarrheal disease and control

Diarrheal disease is a major cause of illness and death in developing countries, particularly among young children (World Health Organization, 2008). Diarrheal infections are generally transmitted by the fecal-oral route: ingesting water, food, or soil that has been contaminated by feces. However, transmission can be influenced (overtly or subtly) by a wide variety of factors, including: availability of water infrastructure (e.g., piped treated water systems, or other improved water sources such as carefully constructed wells); sanitation infrastructure (latrines or sewer systems); personal hygiene; types and quantities of pathogens present; nutritional status; climate; socioeconomic status; cultural factors; and many others. Therefore, the effectivenesses of interventions to reduce diarrhea can vary greatly depending on the characteristics of the community which applies them.

Reduction of diarrheal disease is an important public health goal. This can be accomplished by many different interventions that improve or modify some of the factors listed above. Some examples include: water supply improvements (for quality, quantity, or both); hygiene education (particularly handwashing); a wide array of household water treatment (HWT) interventions (e.g., boiling, chlorination, filtration); and sanitation (latrine or sewer construction). Interventions are most effective if they are used and maintained independently by communities over many years; this is referred to as 'compliance', or sometimes 'adherence'. High compliance is difficult to attain in the developing world, where money, materials, skilled personnel, and good governance are often in short supply. It is unclear how best to measure and maintain compliance within communities.

The effectiveness of interventions that prevent diarrhea is difficult to measure. Conducting research investigations in developing countries is inherently challenging due to lack of resources

and infrastructure, and researchers are often faced with linguistic or cultural barriers. Field trials of interventions are usually impossible to blind; blinding prevents study participants from knowing whether they are receiving an active intervention or an inactive 'placebo' intervention. Blinding is difficult because interventions are usually visually obvious (e.g., a large filter unit, or people attending handwashing education sessions). Therefore, field trials are likely to be biased by the expectations of investigators or participants, or by other factors (known or unknown). Published field trials often lack key contextual information about the communities that participated in the study, which impedes interpretation of the results.

Although the body of scientific literature concerning diarrhea is enormous and continues to grow, many aspects of diarrheal disease in developing countries remain poorly understood. There is a need to synthesize the available information to determine where to prioritize future research. Chapter 2 summarizes key ideas in this body of literature. Although meta-analyses concerning various aspects of diarrheal disease continue to be published and are useful, they do not clarify interrelationships between these aspects. Mechanistic modeling is a useful tool for describing and simulating complex systems problems such as disease transmission within human communities. This dissertation uses modeling methods (coupled with published data) to simulate diarrheal disease transmission. The models allow conclusions to be drawn about appropriate diarrhea control methods, and identify aspects of diarrhea transmission and control that require further research.

1.2. Scientific questions

Goal 1: Develop a method to link information from epidemiologic studies with simulation models (chapter 3). Field studies of interventions that prevent diarrhea are difficult and time-consuming to perform, and are subject to numerous biases (see page 17 for further discussion). Models based on field studies can be used to infer the effect of

interventions on risk under circumstances that were not actually studied.

Hypotheses: Biases that were observed during a field trial can be corrected by constructing a model that simulates the trial as closely as possible, and then altering the model structure (or its parameters) to remove the bias. In a similar fashion, the outcomes of counterfactual field trials can also be estimated.

Method: Construct a model to simulate an actual field trial. Calibrate the model to the outcomes of the trial in order to infer the values of unobserved parameters. Use the sets of inferred parameters in subsequent estimation steps, in which the model has been modified to simulate desired situations (e.g., differing compliance levels, elimination of biases).

Goal 2: How does noncompliance with interventions affect diarrheal disease levels (chapters 4 and 5)?

Hypotheses:

- a) Perfect compliance with an intervention by X% of the population is more effective than 100% of the population using the intervention X% of the time.
Or: the level of effectiveness of an intervention drops more rapidly when consistent use of the intervention is decreased, as opposed to when overall adoption of the intervention is decreased.
- b) What level of noncompliance renders a specific intervention ineffective (i.e., <10% decrease in longitudinal prevalence) under particular conditions? Trials are seldom powered to detect a decrease this small.
- c) Linear QMRA models and more complex EITS models incorporating feedback loops show similar effects of imperfect compliance and differing compliance patterns on diarrheal disease risk.

Method: Construct QMRA and EITS models of diarrheal disease transmission, apply a simulated intervention, alter the levels of the two types of noncompliance, and observe the results in the form of their effects on diarrheal longitudinal prevalence. Compare the results returned by both model types.

1.3. General assumptions within this dissertation

Control of diarrhea requires sustainable use of interventions over many years by the communities (or individuals) that currently have high risk of diarrhea. Therefore, the effectiveness of an intervention is best described by changes in the average level of endemic diarrheal illness. Although diarrheal epidemics are undeniably important (particularly cholera epidemics), they represent short-term changes (days or weeks) in diarrheal disease risk. Any intervention that reduces endemic diarrhea consistently over several years is also likely to decrease the likelihood or severity of diarrheal epidemics during that time. Therefore, epidemics are given little consideration in this dissertation.

Although the scientific literature concerning diarrhea in developing countries is vast, the precise quantitative values of many important parameters for models that describe diarrhea are unclear. The detailed datasets necessary for validating such models are likewise lacking. Human communities are also idiosyncratic, and diarrheal disease transmission in a particular community might be greatly altered by characteristics that are unmeasured or difficult to measure (e.g., cultural practices, social structure, education, socioeconomic status, local geology, local climate, etc.). Therefore, the models described in this dissertation cannot make firm predictions for specific communities. However, they can still provide general insights about the transmission of diarrheal infections and how interventions can affect their transmission. They can also indicate aspects of diarrhea that require further scientific investigation.

2. REVIEW OF DIARRHEAL INFECTION: EPIDEMIOLOGY, INTERVENTIONS, AND MODELING

2.1. Introduction

Diarrhea is a common disease, particularly among children. Although diarrheal infections are nearly always transmitted by ingestion of feces, the details of transmission and control are complicated. Diarrheal pathogens are extremely diverse, including many distinct bacteria, viruses, and protozoa. Furthermore, these pathogens can be transmitted by different routes, such as drinking water, food, soil, or household objects. Many different interventions are available, such as sanitation (latrines), handwashing, or household water treatment; these interventions impede transmission of diarrheal pathogens on different routes. The transmission and prevention of diarrheal disease has been extensively studied, giving rise to a vast body of published scientific literature. Nonetheless, many important gaps in our knowledge remain.

The central goal of this manuscript is to examine the issue of compliance with interventions to prevent diarrhea, particularly household water treatment (HWT). People in developing countries are often encouraged to adopt HWT methods, but they may use these methods inconsistently, or not at all. Although compliance affects estimates of diarrhea prevented by interventions in field trials, compliance is difficult to measure. Chapters 3, 4, and 5 use risk assessment models and infection transmission models to simulate field trials and examine the effect of differing levels and patterns of compliance on prevention of diarrhea. These models must incorporate existing information about diarrheal pathogens, their transmission, characteristics of infection and disease, and antimicrobial efficacy of HWT.

This chapter summarizes available knowledge regarding diarrheal infections and their prevention in developing countries. It begins by concentrating on the epidemiology of endemic diarrhea in young children (generally under 5 years of age). Diarrheal illnesses which are

strongly epidemic (e.g., cholera) are only briefly mentioned. Interventions for preventing transmission of diarrheal pathogens are described, beginning on page 38. Application of simulation models to understand transmission and prevention of childhood diarrhea is discussed beginning on page 68.

2.2. Definition and measurement of diarrhea

Acute diarrhea is often considered to be 3+ loose or watery stools in 24 hours, not counting normal soft stools from breastfed babies (USAID et al., 2005). However, the number of episodes/day and the number of days that are considered to separate episodes vary. Diarrhea lasting for 14+ days is generally considered 'persistent' (Ejemot et al., 2008).

Dysentery is a diarrheal syndrome in which the stools are bloody (T. F. Clasen et al., 2006). *Shigella* species are a common cause of this type of diarrhea.

Investigation of childhood diarrhea usually relies on the recall of the parents, which may be incomplete. Recall of diarrhea in infants by Guatemalan mothers was found to be unchanged for the two days prior to the interview, but dropped by 37% on the third day (Zafar et al., 2010). Recall for 4th and prior days was about 50% of that for the first two days. Severe illness was recalled more reliably. These results were similar to those reported in prior studies (Zafar et al., 2010). In a review (Kosek et al., 2003) including 27 field studies of diarrhea, 10 interviewed caregivers weekly or monthly, indicating that diarrhea may be underestimated in those studies. Fortunately, the remaining 17 studies interviewed caregivers two or more times per week (Kosek et al., 2003). Use of daily diaries may mitigate some of the effects of infrequent followup.

Differing methods are used to measure changes in diarrheal risk. Incidence rates (the number of disease episodes over a given time in a certain number of people) are most commonly used, which may be appropriate when disease transmission is the outcome under study (Morris et al., 1996). However, the association of longitudinal prevalence (the number of person-days ill

over the total number of person-days observed) with growth faltering and death was stronger than the association of incidence with growth faltering and death (Morris et al., 1996) in a rural area of Ghana that was underserved with health services (Ghana VAST Study Team, 1993). Changes in longitudinal prevalence do not necessarily change incidence; for example, a treatment regime might reduce longitudinal prevalence by reducing the duration of illness, but incidence would remain unchanged. Longitudinal prevalence is also appealing because it more directly measures the extent of illness; a child with persistent diarrhea throughout most of the study period might contribute only 1 bout of illness to the community's diarrheal incidence, but is in actuality very sick (Morris et al., 1996). Point prevalence (the proportion of individuals ill at a given point in time) is sometimes used, but is far from ideal because the burden of diarrheal disease changes with time. Ratios of the above measures with the intervention group in the numerator and the non-intervention group in the denominator are used to estimate the magnitude of effect of the intervention.

Odds ratios are also used to estimate the effectiveness of interventions. Unlike the previously discussed measures of risk, which are population-based and measure ill people as a proportion of total people, the odds is the number affected divided by the number not affected. The odds of illness in a treatment group can then be divided by the odds of illness in a control group to yield an odds ratio. Although the odds of illness is similar to the risk of illness when the disease is rare (i.e., the number of non-ill people nearly equals the entire population), this is not the case for diarrheal illness in many developing countries, and 'risk' as measured using odds therefore appears larger. Odds ratios also tend to be larger than risk ratios. Odds ratios are commonly used in regression analysis where the effect of an intervention is controlled for confounding factors, since the parameter estimates are convenient to represent mathematically as odds ratios. They are also used in case-control studies, where population-based risk cannot be

determined because the study population is not necessarily representative of the actual population (Hennekens & Buring, 1987).

2.2.1. Persistent diarrhea

Persistent diarrhea is generally considered to last for 14 days or more. This is mainly based on convenience, and a better definition may be needed that takes nutrient deficiency and growth faltering into account (Bhutta et al., 2008). Enteroaggregative *E. coli*, *Cryptosporidium*, and *Giardia* are commonly implicated in persistent diarrhea, but their relative contributions are unknown (Bhutta et al., 2008). Factors associated with reduction of persistent diarrhea are mostly unknown, although early feeding of non-human milk to children and multiple acute diarrheal episodes are important risk factors (Bhutta et al., 2008).

A recent review (Abba et al., 2009) of 20 studies from the 1980s and 1990s in various developing countries describing pathogens associated with childhood persistent diarrhea found no particular differences in the types of pathogens isolated from children with persistent diarrhea and children with no diarrhea. There was also no evidence to conclude that certain pathogens were more/less common in different regions of the world (most studies were from India and Bangladesh, though Latin America, southeast Asia, and sub-Saharan Africa were also represented), although individual studies varied. Enteropathic *E. coli* (particular types could not be distinguished) was detected in 25% to 33% of children, and no other pathogen was found in more than 10% of children. Children with persistent diarrhea were more likely to have at least 1 pathogen detected than children without diarrhea (75% vs. 43%). However, studies tested for different suites of pathogens, sample size was often small, laboratory procedures were often poorly documented, and studies commonly did not cover a whole number of years, so seasonality may have influenced some results.

2.3. Multiple infections

Simultaneous infections by multiple diarrheal pathogens are common. In a review focusing on norovirus (Patel et al., 2008), 4 of 5 studies of children in developing countries hospitalized for severe sporadic acute gastroenteritis showed that <7.5% had a mixed infection, although a Peruvian study was an outlier with 24%. It was not stated which organisms were tested for.

A study focusing on diarrheal parasites of hospitalized patients of all ages in Kolkata, India (A. K. Mukherjee et al., 2009) showed that mixed infections were the norm. 101/147 patients infected with *Giardia* and 66/84 patients infected with *Cryptosporidium* were coinfecting with at least 1 other pathogen. *Vibrio cholerae* was most commonly (25% of coinfections) associated with *Giardia*, and rotavirus and other parasites (33% and 30% respectively of coinfections) were most commonly associated with *Cryptosporidium*. *E. coli* was also identified in 13% of *Giardia* and *Cryptosporidium* cases.

2.4. Asymptomatic infections

Particularly in developing countries, asymptomatic infection with various diarrheal pathogens is common (Wennerås & Erling, 2004). Asymptomatic infections can be described by morbidity ratios, i.e., the proportion of infections that yield illness. Morbidity ratios can be estimated by stool surveys, where a random sample of a community is chosen and one or more stools from each person is collected and analyzed, regardless of whether the person has diarrhea or not. Morbidity ratios vary greatly by pathogen and setting; morbidity ratios tend to be lower in developing countries than in developed countries due to more frequent exposure, since immunity is often developed to disease rather than infection (R H Gilman et al., 1988; Cravioto et al., 1990; Valentiner-Branth et al., 2003; A. H. Havelaar et al., 2009).

Because multiple infections and asymptomatic infections with diarrheal pathogens are common, it is often difficult or impossible to reliably attribute an episode of diarrhea to a

particular pathogen, even if detailed laboratory results are available.

2.5. Burden of diarrheal disease

Diarrheal disease is common throughout the world. It is usually mild and self-limiting in healthy, well-nourished people in clean environments, but is an exceptionally serious problem among children in developing countries (World Health Organization, 2008).

2.5.1. Morbidity

In the late 1980s to mid-1990s, it was estimated that children under 5 years of age in developing countries suffered from a median of 3.2 diarrheal episodes per child-year of life (Kosek et al., 2003). Children aged 6-11 months were most at risk, with a median of 4.8 episodes per child-year. Diarrheal risk may vary drastically depending on location and the degree of (under)development of the community. Age-specific incidences of diarrhea in developing countries have not decreased from 1955 to 2000, according to three major reviews that were conducted sequentially during that time period (Jamison et al., 2006). Gamma distributions can satisfactorily describe the number of episodes per child and the duration of episodes (Schmidt & Cairncross, 2009); the gamma distribution has a flexible shape (which can resemble a mound with a long right tail, or a monotonic decrease), and is defined by two parameters.

2.5.2. Mortality

The approximate mortality rate due to diarrhea in children under 5 years of age in developing countries is 4.9 per 1000 per year (Kosek et al., 2003), for a total of about 2 million deaths per year (World Health Organization, 2008; Boschi-Pinto et al., 2008). In contrast to morbidity, mortality due to diarrhea has substantially decreased since the 1950s-1970s, when the rate was about 13.6 per 1000 per year (Kosek et al., 2003). This trend appears to be continuing, at least in part due to promotion of oral rehydration solution and other good practices for care of children with diarrhea (Jamison et al., 2006). However, the relatively constant levels of incidence

argue that the root problem of diarrheal disease transmission remains to be addressed, and diarrheal illness is still a major killer of children. In children under 5 years of age, 17% of deaths are due to diarrhea; for comparison, 17% of deaths are due to pneumonia, and malaria accounts for 7% of deaths (World Health Organization, 2008).

About 35% of mortality in children under five years of age due to diarrhea is thought to be from acute diarrhea, 45% from persistent diarrhea, and 20% from dysentery (R E Black, 1993).

Huge disparities are seen when diarrhea mortality is separated by WHO region (Figure 2.1). The situation was worst in Africa by far (World Health Organization, n.d.).

Figure 2.1. Diarrhea mortality by WHO region

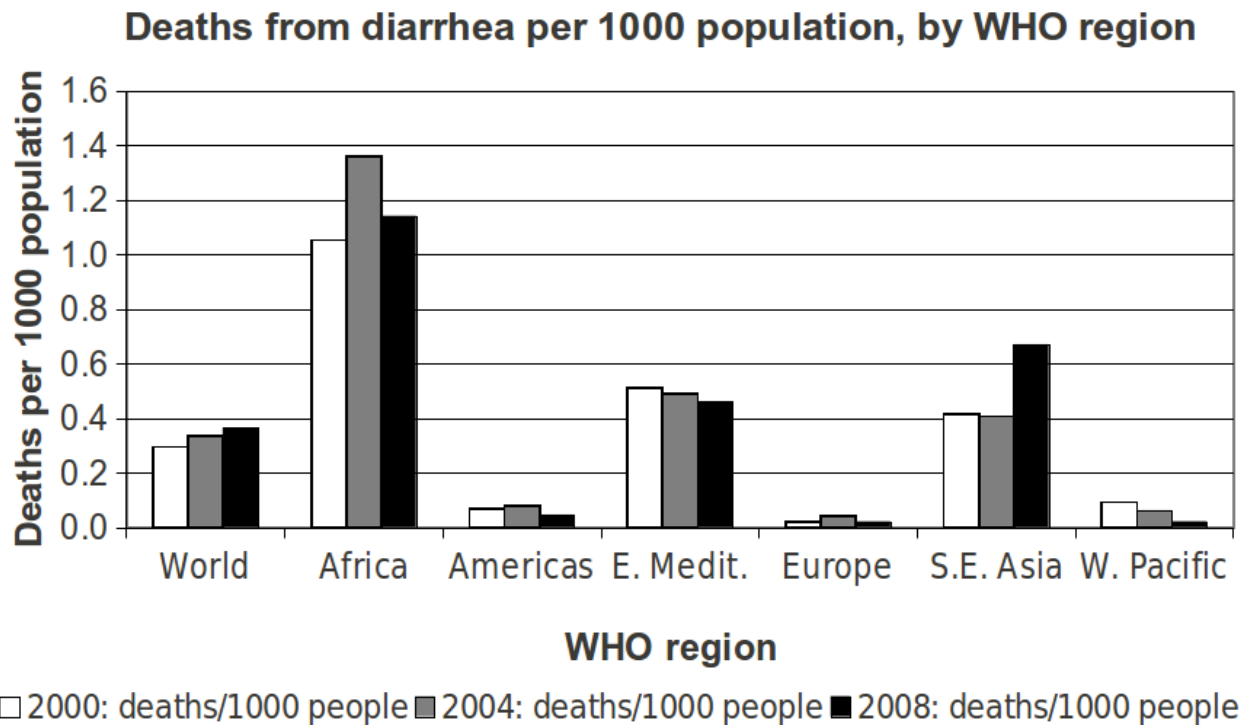


Chart depicts publicly available data (World Health Organization, 2012).

2.5.3. Disability-adjusted life years (DALYs) attributable to diarrhea

Diarrhea morbidity and mortality can also be measured using DALYs per person lost due to diarrhea (Figure 2.2). DALYs describe years of healthy life that are lost to disease, and they

are used to compare morbidity between widely varying diseases (Murray & A. D. Lopez, 1996). Africa loses the most DALYs per person to diarrhea, and this increased from 2000 to 2004 (WHO DALY estimates for 2008 were not yet available at this writing). Since diarrheal disease is often brief and self-limiting, mortality is responsible for nearly all of the DALYs; even though death is unlikely for a given diarrheal episode, it contributes many DALYs when it occurs (because fatalities nearly always occur in children, resulting in the loss of many future years of life). It has been argued that DALYs for diarrhea should consider long-term sequelae of chronic gastrointestinal infection (e.g., reduced intelligence and reduced physical fitness); this increases the number of DALYs lost to diarrhea by two to six times, depending on the assumptions used (Guerrant et al., 2002). However, such sequelae are not included in World Health Organization DALY calculations.

Figure 2.2. Diarrhea DALYs by WHO region

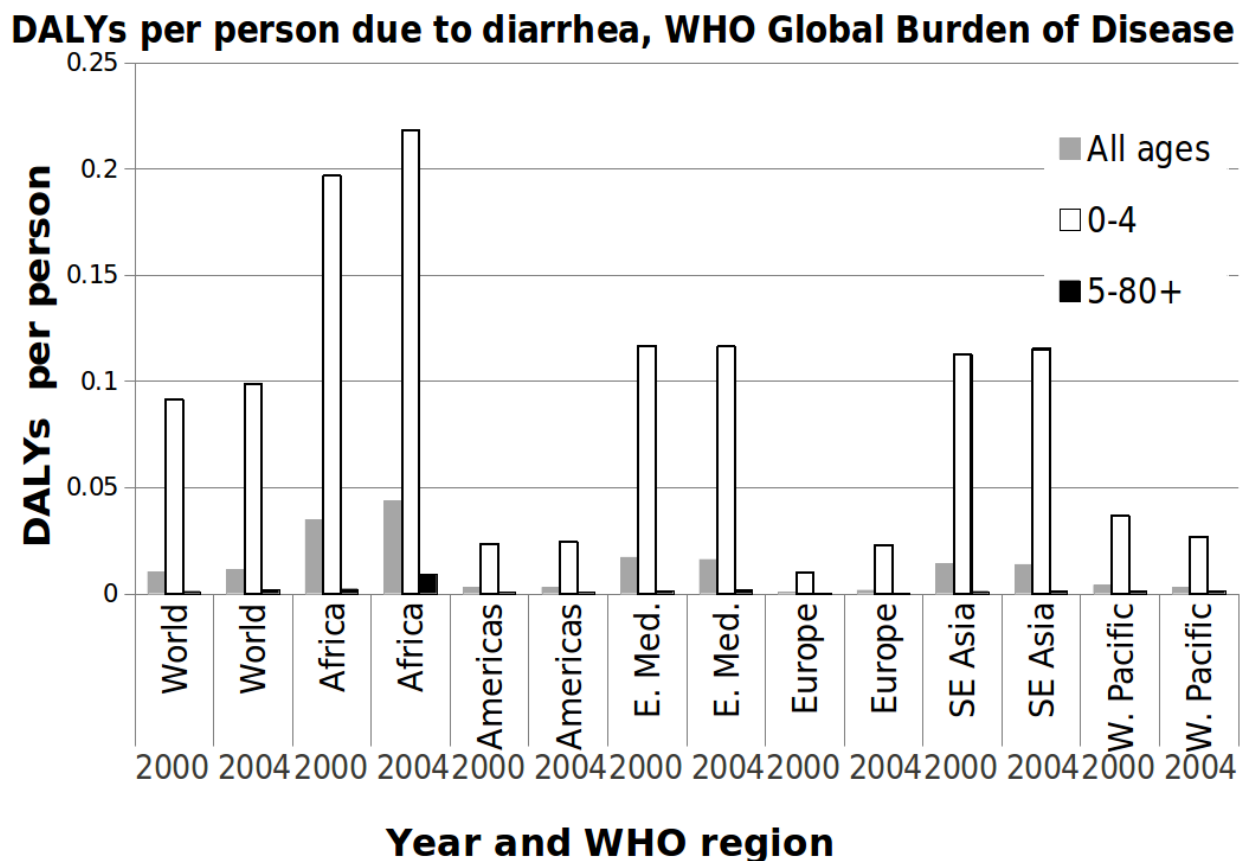


Chart depicts publicly available data (World Health Organization, 2012).

2.6. Changes in diarrheal morbidity over time in individuals

Risk for diarrheal morbidity approximately doubles in the 6th-11th month of age, compared with the first 6 months of life. This is largely because weaning is a critical time for breast-fed infants, increasing exposure to pathogens; when weaning commences, diarrheal infections and risk of death increase markedly (Motarjemi et al., 1993). The risk drops off sharply in one-year-olds and declines gradually thereafter (Bern 1992). Diarrheal mortality is highest in the first year of life and drops by roughly a factor of 4 among ages 1-4 years (C Bern et al., 1992).

Although immunity to certain diarrheal pathogens can be acquired (e.g., effective vaccines against rotavirus and poliovirus exist), its importance is somewhat unclear. Diarrheal pathogens

are very diverse, and immunity from one pathogen serotype does not necessarily confer immunity to other serotypes of the same pathogen. Furthermore, immunity to diarrheal disease does not necessarily confer immunity to infection (A. H. Havelaar et al., 2009). Infected people without diarrhea can still shed pathogens and may still suffer important health effects; for example, children with asymptomatic cryptosporidiosis grow more slowly than uninfected children (Checkley et al., 1997). Finally, although diarrheal risk decreases as children age, they are acquiring immunity at the same time that they are learning to behave more hygienically, and it is unclear whether improved hygiene or increased immunity is more important for preventing diarrhea.

Evidence for approximately twofold increased risk of a diarrheal episode following recovery from a previous episode has been found in some datasets; this risk gradually declines over several weeks (Schmidt et al., 2009). A similar effect with risk declining more than twofold over about 13 weeks was seen in an observational study in urban Brazil, and remained even after adjusting for the age at the time of an episode and age upon enrollment in the study (Genser et al., 2006). This increased risk was only related to the timing of the prior episode, and not to the number of previous episodes, indicating that relapsing or intermittent symptoms from the same infection might be the cause (Genser et al., 2006). This general phenomenon has also been reported from rural Zaire (Tonglet et al., 1999).

2.7. Diarrhea-malnutrition vicious cycle

The evidence on how diarrhea and malnutrition compound each other has been recently reviewed (Guerrant et al., 2008). Malnourished children tend to suffer more and longer bouts of diarrheal illness, approximately doubling the amount of time spent ill. This is partially explained by damage to the absorptive capacity of the small intestine by enteric illness, as well as by malnutrition itself which further limits the resources available to repair that damage. Even

asymptomatic gastrointestinal infections appear to compound this feedback loop. However, the pathophysiology of these relationships is not well understood. Increased burden of enteric infection in the first 2 years of life is associated with decreased work productivity, earning capacity, and cognitive impairment in later childhood and adulthood (Guerrant et al., 2008).

Exclusive breastfeeding in children aged < 6 months protects them from malnutrition if they develop diarrhea, but exclusive breastfeeding cannot be sustained past approximately 6 months of age (Motarjemi et al., 1993). Any breastfeeding is likely to mitigate the effect of diarrhea on malnutrition; under some circumstances, partially weaned children can have more diarrhea or slower growth than completely weaned children, which could be due to energy, protein, or micronutrient deficit from overreliance on breast milk (Tonglet et al., 1999; McDade & Worthman, 1998).

It is unclear how diarrhea and malnutrition are linked quantitatively. Both of these syndromes are broad aggregations of many factors. Although many studies have linked anthropometric measures of malnutrition to increased diarrheal risk, such associations can disappear after adjusting for age, sex, time of enrollment (seasonal factors), and the caretaker's assessment of the child's growth (Tonglet et al., 1999).

It may be possible to clarify the action of the diarrhea-malnutrition vicious cycle by calibrating an infection transmission model to data describing changes in diarrhea incidence among children who frequently suffer from diarrhea. Children who have suffered a recent bout of diarrhea are more likely to suffer subsequent bouts (Schmidt et al., 2009), and children who are stunted or wasted due to repeated episodes of diarrhea are also more likely to suffer further from diarrhea (Guerrant et al., 1992). Increased susceptibility to new diarrheal infections after resolution of a previous infection might be modeled by an increase in the probability that a single pathogen causes disease, which is the interpretation of the k parameter in the exponential dose

response equation (see Equation 3.2, page 98) (Haas et al., 1999). This increase could be estimated by calibrating a diarrhea transmission model to field data. The duration of time during which the probability of infection is increased would also need to be specified; it would probably be greater than the duration of disease, and might occur even for asymptomatic infection.

2.8. Cyclic or recurring changes ('seasonality') in diarrhea incidence

The word 'seasonality' is often used to describe regular changes in disease incidence over time, often on an annual scale. However, seasons themselves are often distal influences on a more proximal factor such as temperature or humidity that better predicts the 'seasonal' effect (Fisman, 2007). 'Seasonal' changes in disease can be tenuously related (or unrelated) to seasons or climate, such as children coming together at the beginning of a school term leading to increased measles transmission (Grassly & Fraser, 2006). The obvious nature of seasons might lead to overattribution of disease fluctuations to them, or failure to measure other potentially more informative factors (Fisman, 2007). Therefore, seasonality should be considered as regular changes in disease incidence or prevalence, rather than changes related to particular seasons of the year.

Seasonality of diarrhea is variable and depends on local climate as well as the organism. For example, rotaviral disease tends to increase in cooler, drier weather (K. Levy, A. E. Hubbard & J. N. S. Eisenberg, 2009), while enterotoxigenic *E. coli* (ETEC) tends to peak at warm, wet times (Estrada-Garcia et al., 2009). Seasonal effects can differ depending on the community; for example, a community whose preferred water source dries up during the the dry season might experience increased diarrhea at that time due to use of poorer quality source water. In other areas, the onset of the rainy season might increase diarrheal disease for various reasons, such as rain washing pathogens from land into source water, or relatively poor nutritional status because food stores from the previous harvest have been depleted (S. Sutra et al., 1990).

Apparent 'seasonal' (or simply oscillatory) cycles can arise in infectious disease models that include acquired immunity. This can arise from accumulation of susceptible individuals as they are born, followed by an increase in disease mediated by an increased proportion of susceptibles, with a subsequent decline in illness until sufficient susceptible individuals are born (Grassly & Fraser, 2006). Combining this behavior with seasonal factors can lead to complex dynamics, including chaos (Grassly & Fraser, 2006).

2.9. Crowding (urban vs. rural)

Overcrowding is associated with increased risk of diarrhea, particularly when coupled with poor sanitation. This phenomenon has been widely reported during military campaigns and in refugee camps (Lim & Wallace, 2004). Crowding may occur within the household (many people in a small dwelling) or in the community as a whole (periurban slums). Crowding within households is easily measured (number of people sleeping in a room or occupying a hut) and is associated with diarrhea (Chacín-Bonilla et al., 2008); however, the effect of this type of crowding on diarrhea is likely to be confounded with many other risk factors, such as low socioeconomic status and malnutrition. The precise effect on diarrheal risk by crowding at the community level is more difficult to measure, and also difficult to disentangle from other determinants of diarrheal risk, such as sanitation and malnutrition. Studies comparing diarrhea in districts that are more or less densely populated but have otherwise similar populations do not appear to be available in the published literature.

2.10. Bias in epidemiological studies of diarrhea

Many epidemiological studies concerning diarrheal disease have been published. A recent systematic review (T. Clasen, I. G. Roberts, et al., 2009) of water quality intervention studies found 68 studies. Unfortunately, methodological problems are frequent in such studies (V. A.

Curtis & Cairncross, 2003; T. Clasen, I. G. Roberts, et al., 2009). This is partly because of the challenging nature of research in developing countries: cultural differences, language barriers, and infrastructure limitations complicate work in these communities. In field trials of interventions to prevent diarrhea, bias is a serious problem that is difficult to quantify.

There are two broad types of bias that are particularly important in field trials: selection bias and information bias. Selection bias occurs when people who are selected to participate in a study are not representative of the larger population that the study is meant to describe (Last, 1995). For example, investigators might choose communities for an intervention trial that they think would be particularly likely to comply with the intervention; this would increase the apparent effectiveness of the intervention. Selection bias hinders generalization of results to larger populations.

In contrast to selection bias, which generally occurs at the beginning of a study and is unrelated to the experimental groups within a study, information bias refers to informational error within the experimental groups. There are several types of information bias that are relevant to field trials. Probably the least serious is nondifferential misclassification bias, where some participants are in the wrong experimental group. If misclassification is similar in all experimental groups (hence nondifferential), then the effect size cannot be exaggerated and a conservative estimate of the effectiveness of an intervention would be obtained (Rothman, 1986). However, there are many other types of information bias that are more serious because they yield different effect measurements depending on the experimental group. For example, courtesy bias may arise when participants tell the experimenters what they think they want to hear; therefore, users of a HWT device might report less diarrhea, even if the device is completely ineffective. This might be deliberate or unconscious behavior by the participant. Recall bias could occur if participants in different groups differ in how they remember (and report) diarrheal episodes; for

example, presence of a HWT device in the home might prompt frequent consideration of clean water and diarrhea, and therefore diarrhea might be more completely recalled. Hawthorne effects, in which participant behavior is influenced by the knowledge that they are being observed (McCarney et al., 2007), are another type of information bias. This could act similarly to selection bias by boosting compliance with interventions, but it might also have different effects by experimental group. For example, recipients of an intervention might receive more follow-up during the study, which could alter their participation of the study or their perception of diarrheal illness. Interviewer bias (Last, 1995) can also arise when investigators know which experimental group they are studying; for example, interviewers might unconsciously interview control or intervention participants differently because of their expectations about the effectiveness of an intervention.

Two important strategies to reduce bias are: 1) random assignment to study groups; and 2) blinding of the study group assignment to the participants (single-blind) or, preferably, to participants and investigators alike (double-blind) (Hennekens & Buring, 1987). However, these strategies are often problematic in field trials because the intervention is visually obvious and cannot be concealed. A meta-analysis of a wide variety of intervention studies indicates that lack of blinding can exaggerate a protective effect by about 30% (L. Wood et al., 2008). Therefore, an unblinded study of an intervention that in fact has no effect would be expected to show a relative risk of about 0.7, which is similar to the effect size reported by many household water treatment studies (Hunter, 2009).

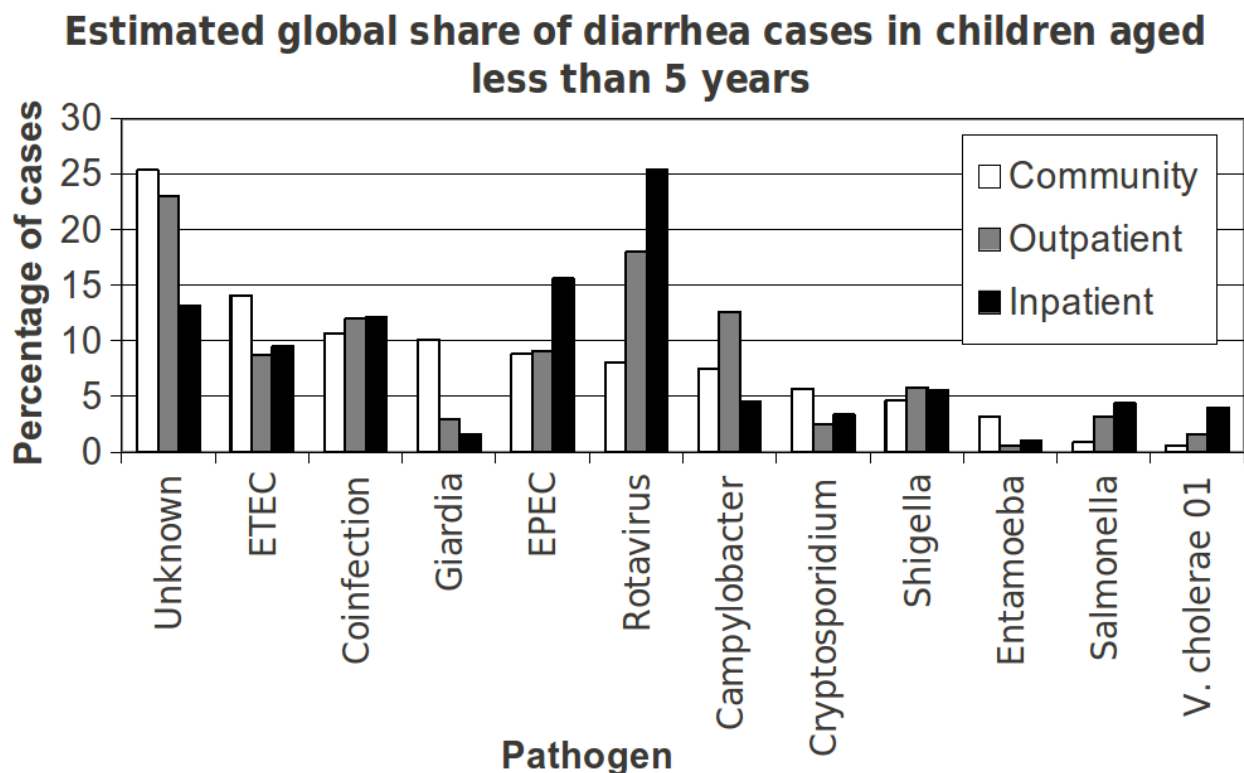
2.11. Relative contributions of pathogens to diarrheal etiology

Although dozens of different pathogens can cause childhood diarrhea in developing countries, pathogenic *E. coli*, rotavirus, norovirus, *Shigella*, *Campylobacter jejuni*, *Giardia*, and *Cryptosporidium* are particularly important. They contribute differently to acute and persistent

diarrhea. Enterotoxigenic *E. coli* (ETEC), rotaviruses, and noroviruses are particularly important in acute diarrhea, but enteroaggregative *E. coli* (EAEC), *Cryptosporidium*, and *Giardia* seem particularly important in persistent diarrhea (Guerrant et al., 2008).

Estimates of the proportion of disease caused by particular pathogens are difficult to obtain. However, a review (Lanata & W. Mendoza, 2002) of 266 studies from 1990-2002 indicates that pathogenic *Escherichia coli*, *Giardia*, and rotavirus are probably the most common pathogens in community settings worldwide (Figure 2.3). Rotavirus is remarkable for the large numbers of inpatient and outpatient visits attributed to it; in contrast, *Giardia* accounts for few outpatient and inpatient cases compared to community cases. Coinfections and diarrheal illnesses of unknown etiology accounted for a large proportion (25-35%) of all cases.

Figure 2.3. Estimated etiology of childhood diarrhea worldwide



Adapted from Lanata and Mendoza (2002), graphs 1-3.

Examination of four WHO subregions (AfroD, AfroE, AmroB, and SearoD) which had

five or more community studies covering a wide array of pathogens showed reasonably consistent proportions of disease attributable to bacteria, protozoa, and viruses. Considering cases where an etiology was identified, the proportion of cases where bacteria are isolated is approximately 60% in those four regions (Lanata & W. Mendoza, 2002). Protozoa accounted for 16-34%, and rotavirus accounted for 9-22%. However, since only rotavirus was included, the contribution of viruses is probably underestimated. Also, it was common for no pathogen to be isolated (30-55% of cases in those four regions).

There are limitations to the Lanata and Mendoza (2002) review; in particular, the studies examined might not be representative, since investigators may have chosen to work in regions with known problems or on their particular pathogen(s) of interest. Pathogens that are more difficult to detect might be underrepresented.

A more recent review (Abba et al., 2009) examined the etiology of persistent diarrhea (lasting >14 days) in young children (age ranges varied, but all were under six years), using four studies from Bangladesh, six from India, five in Central & South America, two in Zambia, one in Thailand, and one in Viet Nam. Overall, pathogenic *E. coli* was found in 31%-41% of children with persistent diarrhea, and in 22%-30% of children without diarrhea. Rotavirus, enteric adenovirus, *Campylobacter*, *Salmonella*, *Vibrio cholerae*, enterohemorrhagic *E. coli* (EHEC), *Giardia*, *Cryptosporidium*, and *Entamoeba* were all found, but at 10% or less.

Much less information is available concerning the etiology of diarrhea in people older than five years. Although a recent review (Fischer Walker et al., 2010) found 22 studies, 15 were inpatient and 7 were outpatient. Of the outpatient studies, only two of them studied more than one pathogen. Only one study was community-based, and it examined ETEC only. Information about asymptomatic infection in older children and adults is similarly lacking, and information about coinfection was usually absent (Fischer Walker et al., 2010). However, the outpatient

studies indicated similar proportions of diarrhea attributable to bacteria, protozoa, and viruses as the community studies described in Lanata and Mendoza (2002): 60%, 21%, and 19%, respectively (Fischer Walker et al., 2010).

2.12. Survey of diarrheal pathogens

Common pathogens causing endemic diarrhea are discussed below, with particular reference to developing countries. Rotavirus, diarrheagenic *E. coli*, and *Giardia* are given special attention because they are used as model organisms in the simulation models in chapters 3, 4, and 5 of this dissertation. Numerous parameters are used to describe these organisms' behavior in the models; they are briefly mentioned here, but are described in detail in the appendix.

2.12.1. Viral diarrheal pathogens

A wide variety of enteric viruses are known to cause diarrhea. In general, they are small (25-100 nm) nonenveloped RNA viruses that are environmentally stable, although they are usually vulnerable to chlorine. The combination of flocculation/sedimentation and chlorination in standard water treatment usually reduces enteric viruses by four \log_{10} or more. Due to their small size, they cannot be removed by most filtration methods.

Rotavirus and norovirus (formerly known as Norwalk virus) appear to be most important with regard to diarrhea. Among hospital-based studies of children in 11 different developing countries, rotavirus prevalence was 34.9% on average, and calicivirus (the family that includes norovirus) prevalence was 10.3% on average, while adenovirus accounted for 6.3% and astrovirus for 3.5% (Ramani & Gagandeep Kang, 2009). It is unclear whether these proportions also apply to disease in the community setting. Both rotavirus and norovirus tend to cause more severe disease than other enteric pathogens (Ramani & Gagandeep Kang, 2009), they are both highly potent with extremely high shedding, and they are responsible for high proportions of

severe childhood diarrheal illness in developing and developed countries, with 12% due to norovirus (Patel et al., 2008) and 33% due to rotavirus (Cook et al., 1990). Norovirus also accounted for 12% of mild to moderate diarrheal illness in all ages (Patel et al., 2008).

Rotavirus

Transmission and ecology

Rotavirus is primarily transmitted by the fecal-oral route but may also be transmitted person-to-person, and in some circumstances it might be inhaled (Heymann, 2004). Although improved sanitation and hygiene greatly inhibit transmission of many other diarrheal pathogens, rotavirus generally infected all children before four years of age in both developing and developed countries before rotavirus vaccine was available (Cook et al., 1990).

Although there are several different serotypes of rotavirus, allowing repeated infection, risk of further episodes declines with each additional episode, and the severity of subsequent episodes also declines (Velázquez et al., 1996).

Rotavirus disease is strongly seasonal in temperate regions, peaking in the winter in the Americas and in spring or fall in other areas, but is seen year-round with less seasonal change in tropical regions (Cook et al., 1990). Even within tropical areas, however, rotavirus incidence tends to be higher under cooler and drier conditions, although there is much heterogeneity (K. Levy, A. E. Hubbard & J. N. S. Eisenberg, 2009). It is not clear why this is, but drier conditions might facilitate airborne suspension of droplet nuclei and promote inhalation of rotavirus, and differing sanitary conditions and episodic local events (such as floods) likely modify any effect of humidity or temperature (K. Levy, A. E. Hubbard & J. N. S. Eisenberg, 2009).

Dose response

Rotavirus is extremely potent, with an ID₅₀ of approximately 6 focus-forming units (FFU) (Anon, 2012) as measured in a human feeding study (Ward et al., 1986).

Symptomology

The incubation period for rotavirus infection is short, 1-3 days (Blaser et al., 2002). Rotavirus disease consists of fever, vomiting, and watery diarrhea; fever and vomiting usually last 2-3 days, while diarrhea may continue for 8 days. However, rotavirus infection is frequently asymptomatic; approximately 60% of young rural children who were shedding rotavirus in Guinea-Bissau, Mexico, and Argentina did not have diarrhea (Cravioto et al., 1990; Vergara et al., 1996; Fischer et al., 2002).

Burden of disease

Rotavirus is responsible for about 6% of diarrhea cases in children aged less than five years in the developing world and 20% of the childhood diarrhea deaths according to a review (de Zoysa & Feachem, 1985) of 7 studies from the early 1980s. A subsequent review (Parashar et al., 2003) of studies published from 1986 to 2000 yields a similar estimate of rotavirus involvement in 8% (IQR 4 to 12) of in-home cases for children under five years of age, and approximately 20% for outpatient and inpatient cases. Approximately 20% of diarrhea deaths in low-income countries were due to diarrhea, similar to the earlier estimate. This corresponded to a 1/205 risk of dying from rotavirus by age 5 years, compared to a 1/49,000 risk in high-income countries (Parashar et al., 2003). The proportion of severe diarrheal disease attributed to rotavirus has increased over time (Harry B Greenberg & Mary K Estes, 2009). This may be because of reductions in diarrhea caused by other pathogens; high shedding and high potency means that rotavirus is particularly difficult to control (Harry B Greenberg & Mary K Estes, 2009).

Persistence in the environment

Rotavirus resists inactivation in the environment. On aluminum or ceramic at 4°C or 20°C in high or moderate relative humidity, approximately 90-99% is inactivated within the first week, but further inactivation is much slower; even at 60 days approximately 1% of the original

amount of virus remained (Abad et al., 1994).

Control

Rotavirus disease is very difficult to control due to high levels of shedding and its extremely low ID₅₀ of ~6 virions (Anon, 2012). Two live-virus vaccines have been developed, RotaTeq and Rotarix, which are both highly effective: 74% against diarrhea and 100% against severe diarrhea for RotaTeq, and 95% against severe diarrhea for Rotarix (Harry B Greenberg & Mary K Estes, 2009). It is unclear whether these vaccines will be as effective in severely underdeveloped environments, although trials are underway (Harry B Greenberg & Mary K Estes, 2009). Before rotavirus vaccine was available, rotavirus generally infected all children before the age of four years in both developing and developed countries (Cook et al., 1990). Correct use of oral rehydration solution (ORS) in diseased children is very effective in preventing rotavirus mortality (Blaser et al., 2002).

Norovirus

Transmission and ecology

Norovirus is highly potent and is shed in large quantities by infected individuals, and it can easily be transmitted by person-to-person contact as well as by contaminated food or fomites (Mattison, 2011).

A substantial fraction of the population lacks a receptor necessary for infection with Norwalk virus (the first norovirus discovered); these 'secretor-negative' (Se-) individuals appear innately immune (Lindesmith et al., 2003). In a study population of 77 individuals (49% male, 71% white, 23% black) 28.6% were Se- (Lindesmith et al., 2003). It is unclear how much Se+/- status varies among human populations. Even if an individual is Se+, they may still be resistant due to acquired immunity, if they have been exposed to norovirus previously. However,

noroviruses are diverse, and important characteristics of immunity (duration and cross-protection among strains) are poorly understood (Lindesmith et al., 2003); furthermore, acquired immunity following norovirus infection only lasts a few months, after which a previously infected individual can be reinfected by the same serotype (Carter, 2005).

Peak shedding of norovirus occurred during (31%) or after (69%) disease in 16 experimentally infected volunteers, at approximately 7×10^{10} genome copies / mL of stool (Atmar et al., 2008). Virus was detectable for 1 to 9 weeks in stool; however, peak levels only lasted for one to three days (Atmar et al., 2008). The duration of illness and the incubation period both averaged two days (Atmar et al., 2008).

Dose response

Like rotavirus, norovirus appears highly potent; however, dose response relationships are inconclusive. Although some dose response models have been published (Peter F M Teunis et al., 2008), much of the data used came from feeding studies using viral stocks that had been stored for long periods and were highly aggregated. The extent to which noroviruses are naturally aggregated in the environment is unclear.

Symptomology

Norovirus disease is unpleasant, but relatively mild and seldom dangerous (Blaser et al., 2002). It has an incubation period of about two days, and lasts about two days (Atmar et al., 2008). Vomiting or diarrhea may onset suddenly; however, the extent of these symptoms varies greatly among diseased persons (Blaser et al., 2002).

Burden of disease

Norovirus had been primarily considered an epidemic gastroenteritis, but evidence is mounting that it is an important cause of sporadic gastroenteritis as well. The annual incidence of norovirus disease in children less than 5 years of age in developing countries is approximately

197 inpatient cases per 100,000 (Patel et al., 2008). This figure was 118 in industrialized countries; by contrast, the incidence of outpatient cases was estimated to be 1665 per 100,000 (Patel et al., 2008). It is unclear what the outpatient or community-level incidence would be in developing countries but it is likely to be substantially larger than the inpatient incidence.

Persistence in the environment

Norovirus is very stable in the environment, although disinfection studies commonly use surrogate caliciviruses because it is not currently possible to grow human norovirus in cell culture (Mattison, 2011).

Control

Good hygiene and disinfection practices are essential. No vaccine is available.

2.12.2. Bacterial diarrheal pathogens

Many different bacteria can cause diarrheal illness, including numerous species and strains from the genera *Salmonella*, *Vibrio* (notably *Vibrio cholerae*, the cause of cholera), *Clostridium*, *Streptococcus*, *Yersinia*, and *Bacillus*. However, pathogenic *Escherichia coli* and *Campylobacter* appear to be responsible for particularly large shares of bacterial diarrhea (Figure 2.3, page 20), and these are discussed in detail below.

Pathogenic Escherichia coli

Important pathotypes of E. coli

Escherichia coli usually resides in the mammalian large intestine, benefiting itself as well as the host. However, there are several well-established pathotypes of diarrheagenic *E. coli* (Kaper 2004, Nataro 1998). With respect to diarrhea, the most important are enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), and enteroinvasive *E. coli* (EIEC), all of which produce watery diarrhea. Although enterohemorrhagic *E. coli* (EHEC) is much more potent than EPEC, ETEC, or EIEC and can cause severe or lethal disease, EHEC is relatively uncommon,

and is not a major cause of diarrhea. *Shigella* species are highly potent, like EHEC, but they are less dangerous, although they are an important cause of dysentery; *Shigella* are very closely related to *E. coli*, and are probably technically conspecifics (James B Kaper et al., 2004). Although EIEC is the *E. coli* pathotype most closely related to *Shigella*, EIEC is much less potent (Blaser et al., 2002; Anon, 2012).

Transmission and ecology

Shedding of ETEC may vary by strain. For *E. coli* H10407, the geometric mean number of bacteria per gram of feces was 4.03×10^7 , and it did not differ greatly whether subjects were symptomatic or not (Levine et al., 1980). For *E. coli* 214-4, the geometric mean was 5.23×10^8 , over tenfold higher (Levine et al., 1980).

ETEC can often be detected in apparently healthy people. In developing countries among healthy 0-11 month olds, and 1-4 year olds, 11.7% and 7.1%, respectively, are estimated to be colonized with ETEC (Wennerås & Erling, 2004).

ETEC disease tends to be more prevalent in warm, wet weather, which aids its multiplication in the environment (Qadri et al., 2005; J P Nataro & J B Kaper, 1998). In Bangladesh, its seasonality is similar in shape and magnitude to that of cholera, with one peak at the beginning of the hot season (spring) and another in autumn just after the monsoons, when more fecal material is entering surface water (Qadri et al., 2005).

A household-level case-control study (R E Black et al., 1981) of ETEC was conducted in the rural Matlab area of Bangladesh. Extensive culturing of water, food, domestic animals, and family members was carried out. Compared with households testing negative for ETEC, risk of infection was only slightly higher (10.0% vs. 8.3%) in households where a water source or a domestic animal was positive for ETEC. In contrast, risk of infection increased 3.5 fold, to

29.0%, among households for which ETEC was found in stored water or cooked food, highlighting the importance of exposure within the household. Within households having an infected member, the proportion of household contacts becoming infected, and the proportion of those infected contacts actually acquiring disease, declined with age. This is consistent with development of immunity, since contamination of food and water within the household indicates that all members are likely to be similarly exposed.

Dose response

Pathogenic *E. coli* is much less potent than the other pathogens described in this chapter. Many studies with human volunteers have used disease as the outcome, and their ID₅₀s range from 2×10^5 to 1×10^8 . The only available experiment with infection as the outcome used EIEC (H L DuPont et al., 1971) and had an ID₅₀ of 2×10^6 CFU (Anon, 2012)

Feeding studies of ETEC or EPEC in healthy volunteers typically give 2-3g of NaHCO₃, which neutralizes stomach acid and reduces the infectious dose (Levine et al., 1977). However, it has been suggested that food as a vehicle would have a similar acid-neutralizing effect (Levine et al., 1977). ETEC and EPEC do not appear to be transmitted person-to-person, partly because of their low potency; a study of ETEC-infected volunteers co-housed with uninfected volunteers did not result in any transmission of infection (Levine et al., 1980). Food was all served individually to the volunteers over the course of the experiment, so there was no opportunity for ETEC to spread via that route (Levine et al., 1980).

Symptomology

ETEC disease generally consists of watery diarrhea without fever, although its severity can vary widely, depending partly on the set of toxins that the strain is carrying (Blaser et al., 2002). The incubation period lasts from one to two days (Blaser et al., 2002), and ETEC diarrhea in 11

experimentally infected adult volunteers lasted 82.1 (SD 50.7) hours, with 12.0 (SD 9.29) stools over the course of the illness (R E Black et al., 1982). These were compared with other volunteers who were treated with antibiotics; immediate treatment cut diarrheal duration, stool volume, and hours ill by half.

EPEC disease can be more severe than ETEC, and there are many documented examples of lethal epidemics in children (Blaser et al., 2002). However, experimental infections in adult volunteers showed a relatively mild syndrome of less than two days of watery diarrhea after an incubation period of less than 1 day (Blaser et al., 2002).

Burden of disease

In a recent review (Abba et al., 2009), pathogenic *E. coli* was by far the most common pathogen isolated from children with persistent diarrhea in various developing countries; it was detected in 31% to 41% of children, compared to <10% for other pathogens. Pathogenic *E. coli* also predominated similarly among pathogens isolated from asymptomatic cases (Abba et al., 2009).

Enterotoxigenic *Escherichia coli* (ETEC) is the most common type of diarrheagenic *E. coli* (Qadri et al., 2005). It is probably also the most common cause of childhood diarrhea in the developing world, responsible for approximately 1/7 of diarrheal episodes in children aged less than 1y and almost ¼ of diarrheal episodes in 1-4 year olds (Wennerås & Erling, 2004). It can also cause severe dehydrating cholera-like disease in adults (Qadri et al., 2005). Diagnosis is complicated since many other Gram-negative bacteria produce similar toxins, so both toxins as well as the *E. coli* bacterium must be tested for in order to yield accurate results (Wennerås & Erling, 2004).

Persistence in the environment

Laboratory studies of *E. coli* in unfiltered river water in the dark indicate exponential

decay at a rate of 1.15 day⁻¹ (in water taken from above a sewage outfall) to 0.64 day⁻¹ (in water from below a sewage outfall) (Flint, 1987). Filtering or autoclaving the water before adding *E. coli* enhanced survival (Flint, 1987). A summary of inactivation rates of *E. coli* published by other workers indicates high variability, spanning over 2 orders of magnitude (0.009 to 2 days⁻¹) at 15-20°C (Pond et al., 2004).

Control

In addition to sanitation and hygiene, proper food preparation and handling (particularly of weaning foods), is important (Motarjemi et al., 1993). ETEC vaccines are being actively researched (Harro et al., 2011).

Campylobacter species

Transmission and ecology

In humans, *Campylobacter jejuni* and, less commonly, *Campylobacter coli* can cause gastroenteritides (A. H. Havelaar et al., 2009). Although *Campylobacter* can be spread by contaminated food and water, campylobacteriosis is mainly a zoonosis, being primarily associated with the gastrointestinal tract of birds, especially poultry (Dechesne et al., 2006). *Campylobacter* does not grow in water and (like *E. coli*) are an indicator of post-treatment contamination in water distribution systems. However, immunity appears to protect against disease rather than infection, and asymptomatic shedding is common (A. H. Havelaar et al., 2009). In a comparison of Mexican children aged less than four years and Swedish patients (ages not given), Swedish patients tended to carry only one *Campylobacter* serotype, while mixed serotypes were carried by 42% of Mexican children tested (Sjögren et al., 1989).

Campylobacter epidemiology varies greatly between the developed and underdeveloped world, probably due to development of immunity early in life. Illness is rare after about five years of age in developing countries, but occurs among adults in industrialized countries,

probably because they avoided exposure (and therefore immunity) in childhood (A. H. Havelaar et al., 2009).

Dose response

In human volunteers, *C. jejuni* has an ID₅₀ of approximately 900 CFU (R E Black et al., 1988; Anon, 2012). Neutralization of stomach acid by food or sodium bicarbonate could increase its potency (Miliotis & Bier, 2003).

Symptomology

Campylobacter can cause acute self-limited watery diarrhea lasting 2-6 days in healthy humans with an incubation period of about three days (range of 1-7 days), often with fever and nausea, but seldom with vomiting (Miliotis & Bier, 2003). In developing countries, campylobacteriosis is typically a disease of very young children; after about age two years, immunity has been acquired (A. H. Havelaar et al., 2009).

Burden of disease

Campylobacter is a substantial contributor to childhood diarrhea in developing countries; although it might not generally contribute as much morbidity as diarrheagenic *E. coli* (Blaser et al., 2002). Illness is rare after early childhood, due to development of immunity from early infections.

Persistence in the environment

Campylobacter survival increases with decreasing temperature, and it may survive for weeks in water samples in the laboratory at 4-10°C (Buswell et al., 1998). It is vulnerable to acid (although it can pass through the stomach if protected by food), but it can tolerate freezing (1-2 LRVs by freezing and thawing) and salting (about three weeks in 6.5% NaCl at 4°C) (Miliotis & Bier, 2003).

Control

Limiting contact with bird feces and proper preparation of meat and eggs is important. It is somewhat unclear how best to manage household poultry to limit *Campylobacter* infection in developing countries. A randomized trial (Oberhelman et al., 2006) of chicken corralling in a Peruvian periurban community showed similar incidences (about three infections per person-year) of asymptomatic infections in children less than six years of age between households with corrals and households without. However, it also found 1/3 higher incidence of diarrhea in general and twofold higher incidence of symptomatic *Campylobacter* infection in households with corrals.

2.12.3. Protozoan diarrheal pathogens

Protozoa are more difficult to culture than bacteria. However, they are larger, and are generally detected and counted under the microscope once they have been concentrated and stained with immunofluorescent dyes. They are more easily filtered from water due to their size, but tend to be more resistant to chlorine and are more potent than many bacterial pathogens. *Cryptosporidium* is used as a de facto standard for evaluating water filtration since it is smaller (5 microns) and more environmentally resistant than *Giardia*; if *Cryptosporidium* is undetectable, *Giardia* should be as well (American Water Works Association, 1999).

Cryptosporidium parvum and *Cryptosporidium hominis*

Transmission and ecology

Cryptosporidium belongs to the phylum Apicomplexa and reproduces both sexually and asexually in the intestinal tract, where it is an obligate intracellular parasite. It has a broad host range, occurring in most (perhaps all) vertebrates, but it does not cause diarrhea in adult dogs, cats, or horses, perhaps meaning that these animals are low risk for transmission to humans. Its oocysts are the infective stage, which are transmitted by the fecal-oral route. Infections in ruminants seem to be the biggest source of environmental contamination (Miliotis & Bier, 2003).

Cryptosporidium hominis appears to only infect humans; it was recently shown to be a separate species from *Cryptosporidium parvum*, which primarily infects cattle but can also infect humans (Hunter & R. C. A. Thompson, 2005). Many other species of *Cryptosporidium* exist, but do not appear to infect humans (Hunter & R. C. A. Thompson, 2005).

Cryptosporidium is an intracellular parasite within epithelial cells in the small intestine, which shields it from the host immune response and limits the effects of chemotherapies (Miliotis & Bier, 2003).

Dose response

Several dose response datasets are available for *C. parvum* infection in humans. They provide ID₅₀s ranging from 12 to 455 oocysts (median 165) (Anon, 2012). The two largest datasets (8 doses each) fit the exponential dose response equation, with ID₅₀s of 165 and 132 oocysts (H L DuPont et al., 1995; Messner et al., 2001; Anon, 2012). Only one feeding experiment has been done with *C. hominis* in humans (Cynthia L Chappell et al., 2006); the response measured was disease (rather than infection), and the ID₅₀ was 17 oocysts (Anon, 2012). Therefore *C. hominis* may be roughly 10 times more potent than *C. parvum* in humans.

Symptomology

Cryptosporidiosis had a mean incubation period of 5 days and lasted for a mean of 6 days in 13 volunteers experimentally infected with *C. hominis* (Cynthia L Chappell et al., 2006). It is generally short and self-limited, causing watery diarrhea in healthy people, but is particularly dangerous to people with AIDS because there is no effective treatment (Miliotis & Bier, 2003). This can lead to infections lasting months or years, with severe damage to the gut. Infections can often be asymptomatic in apparently healthy children and adults (Blaser et al., 2002). Disease duration can be substantially longer in malnourished developing-country children compared with

people in industrialized countries; mean diarrhea durations of 21 and 13 days have been reported for *C. hominis* and *C. parvum*, respectively, in children under five years of age in a Brazilian shantytown (Bushen et al., 2007).

Burden of disease

Although *Cryptosporidium* is ubiquitous throughout the world (Blaser et al., 2002), cryptosporidiosis prevalence is relatively high (20-27%) (Miliotis & Bier, 2003) in children in certain underdeveloped contexts. However, it causes little or no symptomatic disease in older children and adults in developing countries.

Persistence in the environment

Cryptosporidium oocysts are about 5 microns in diameter, and are generally even hardier than the durable cysts of *Giardia*. Oocysts are inactivated more quickly at warmer temperatures (Erickson & Ortega, 2006). Ultraviolet light is effective, yielding 1 to 5 LRVs depending on the strain and the type/intensity of UV (Erickson & Ortega, 2006). They can survive for months in cold lakes or streams. They are highly resistant to chlorine. They can be inactivated by heat >64.2°C for 2+ minutes, or drying at 18-28°C for >4h.

Control

Effective control of *Cryptosporidium* is generally achieved in municipal drinking water treatment by filtration yielding nonturbid water (< 1 NTU) (USEPA, 2012). No effective anti-cryptosporidial medications exist. HWT is particularly important for HIV-infected people to strictly limit exposure to *Cryptosporidium* in drinking water, even in industrialized countries.

Giardia species

Transmission and ecology

The currently accepted name for the *Giardia* species that primarily affects humans is *Giardia duodenalis*, although many papers refer to it as *G. lamblia* or *G. intestinalis*. Even within

G. duodenalis, certain assemblages are generally restricted to humans, and earlier beliefs that giardiasis is a single zoonosis now appear to be incorrect (Cacciò et al., 2005). Assemblages within *G. duodenalis* may later prove to be separate species (Cacciò et al., 2005).

G. duodenalis is a flagellated protozoan that attaches to the small intestinal wall and absorbs nutrients from the gut lumen. Shedding of cysts varies greatly over time (0 to 2.5×10^7 cysts/g of feces), even within the same symptomatic individual (Porter, 1916). Because of this, diagnosis is often based on three specimens collected over several days (Miliotis & Bier, 2003).

Reinfection with *Giardia* following drug treatment can be extremely rapid; 98% of 44 Peruvian children who were treated had reacquired infection within 6 months (R H Gilman et al., 1988). Immunity is to symptomatic giardiasis, rather than to *Giardia* infection.

Dose response

Giardia is a potent pathogen, with an ID₅₀ of 35 cysts based on a human feeding study; it fits the exponential dose response function (Rendtorff, 1954; Anon, 2012).

Symptomology

Giardiasis often presents with diarrhea and flatulence, with foul-smelling foamy stools (Miliotis & Bier, 2003). In a study of experimentally infected humans (Rendtorff, 1954), it had a mean incubation period of 14 days; the mean duration of infection was 18 days, not counting two participants whose infection lasted > 100 days but did not shed during much of that time.

Giardia infection is noninvasive, but can lead to villous atrophy and nutrient malabsorption (Miliotis & Bier, 2003)

Partial immunity appears to develop, and infections are often asymptomatic (Miliotis & Bier, 2003); reports exist of 3% of infections being symptomatic among fathers and their children in Pakistan (Ensink et al., 2006) with a similarly low symptomatic proportion in urban

Brazilian 6-45 month olds (M S Prado et al., 2005). However, a report (Peréz Córdón et al., 2008) from a periurban area of Peru with poor sanitation describes 60% of infections among children aged 1 month to 9 years as being symptomatic.

Burden of disease

Giardia infection is common (prevalence of 2-5%) in the developed world but is 20%-30% prevalent elsewhere; some communities have been documented with much higher prevalences (Blaser et al., 2002). Weaned children are more susceptible than adults; prevalence declines somewhat after adolescence (Blaser et al., 2002). Since asymptomatic infections are so common, the true health impact of *Giardia* remains unclear.

Persistence in the environment

Pooling data from two studies (Wickramanayake et al., 1985; deRegnier et al., 1989) yields an estimate of inactivation at a rate of 0.55/day in water at 20°C. Cysts degrade more quickly in soil or cattle feces than in water (Olson et al., 1999). The cysts do not tolerate freezing (Olson et al., 1999).

Control

Effective water treatment and good hygiene limit *Giardia* transmission. In addition, several curative chemotherapies for *Giardia* infection are available (Blaser et al., 2002).

2.12.4. Metazoan pathogens (helminths)

Intestinal tapeworms or roundworms can aggravate the diarrhea-malnutrition vicious cycle, although light infections are thought to have negligible nutritional consequences. Worms themselves do not appear to cause much diarrhea (Hall et al., 2008). In addition, poor sanitary conditions are likely to simultaneously lead to infection with helminths and diarrheal pathogens. An observational study (Genser et al., 2006) in an urban Brazilian area with poor sanitation tested children once for *Ascaris lumbricoides*, *Trichuris trichiura*, and *Giardia* species, finding

that only *Giardia* was associated with diarrhea incidence.

2.13. Interventions to prevent transmission of diarrheal pathogens

Diarrheal infections nearly always have a major fecal-oral component to their transmission, and this is the major target of interventions to control diarrhea. Nonetheless, many different types of interventions are being used and investigated throughout the world. Unfortunately, it is difficult to ascertain how effective most of these interventions are in actual practice.

2.14. Effectiveness trials of interventions

The scientific literature is replete with results of effectiveness trials of various interventions in developing countries, with some assessing multiple interventions used simultaneously. In general, these trials usually take place over a few weeks or months, or perhaps a full year. They frequently show diarrheal disease reductions of approximately 30%, regardless of the type of intervention. Although children under 5 years of age are usually studied because they bear the greatest morbidity and mortality, other groups (all ages, <15 years of age, <3 years of age, etc.) are sometimes studied. Effectiveness may be measured with incidence ratios (number of new cases) or prevalence ratios (the number of persons infected at a given time); these two measures may differ because incidence and duration of illness are not tightly linked (Schmidt et al., 2009). Unlike clinical trials for determining drug effectiveness, intervention trials in communities are seldom blinded and may not be randomized, which can lead to bias. However, practical considerations (e.g., obvious visibility of intervention materials, expense of conducting research in multiple communities) make clinical trials of community interventions impractical.

2.14.1. Measures of effect in intervention trials

Field trials of interventions commonly use some type of risk ratio to describe the effectiveness of the intervention. Many different measures of risks are used, with corresponding

risk ratios (RRs; also called relative risks), in which the risk in an intervention group is divided by the risk in a control group. An RR of 1 generally indicates no effect, while an $RR < 1$ suggests that the intervention prevents disease, and an $RR > 1$ would mean that there was more disease in the intervention group than in the control group. In general, the preventable fraction (PF) is given by $1 - RR$, and describes the proportion of disease that the intervention prevents. Brief descriptions of common types of risk follow.

Point prevalence (sometimes simply called prevalence) is the proportion of people who are affected at a single point in time. Its relative risk measure is the prevalence ratio (PR). Since the amount of diarrhea present in communities can vary greatly over time, point prevalence is not a very good measure of risk.

Incidence (sometimes called 'incidence rate') is the number of new cases of a disease that arise in a population over a period of time (often expressed in terms of 1 year). Its relative risk measure is the incidence ratio (IR).

The longitudinal prevalence (LP) is the number of person-days affected, divided by the number of person-days observed. Person-days are the product of the number of people observed and the mean time of observation per person. Its relative risk measure is the longitudinal prevalence ratio (LPR).

The odds is the number of instances where the disease occurred divided by the number of instances where the disease did not occur. Its relative risk measure is the odds ratio (OR). Odds are often used in case-control studies where the association of a particular factor with disease is being assessed.

All of these measures of risk are specific to a particular population (such as a particular age group studied in a certain community). The risk of diarrhea in general may be measured, or the risk of infection or diarrhea from a particular pathogen.

Although the primary goal of interventions is to lower the risk of diarrheal disease, it is also useful to measure the antimicrobial effectiveness of these interventions. If an intervention cannot remove microorganisms in the laboratory, there is no reason to conduct a field study of that intervention to assess its impact on diarrheal disease. The \log_{10} reduction value (LRV) is often used to assess water treatment methods; it is calculated by taking the \log_{10} of the number of a certain microorganism detected in a certain volume of treated water, and subtracting the \log_{10} of the number of those microorganisms detected in untreated water. Thus, an LRV of 1 removes 90% of microorganisms; an LRV of 2 removes 99%, and so on. Very high LRVs (5 or more, 99.999% reduction; see Table 2.1, page 64 for examples) are often desirable because some pathogens are both extremely infectious and extremely abundant; thus it might be necessary to eliminate nearly all of a pathogen in order to reduce risk to an acceptable level. If the LRV is low (< 2), antimicrobial effectiveness is sometimes described as a simple percentage of a certain type of microorganism removed or inactivated.

2.14.2. Nature of bias in intervention trials

Effectiveness of interventions as measured by trials is likely to be higher than effectiveness in actual practice for several reasons:

- The intervention may be discontinued by the population after the study is completed and the investigators depart from the community.
 - Cost to families or the community (in time or money) may become prohibitive (Stockman et al., 2007).
 - The population may not believe that the intervention is worthwhile, or the intervention may not be culturally appropriate (Paul, 1955)
 - The community might fail to maintain the intervention in good working order because

they lack the necessary skills, equipment, or money; or maintenance may simply be forgotten.

- Blinding is difficult or impossible to implement in community-based trials (B. F. Arnold & Colford, 2007), which can introduce bias:
 - Respondents, knowing that they are receiving an intervention, may consciously or unconsciously underreport diarrheal illness, or comply more effectively with the intervention, e.g., the Hawthorne effect, or ‘courtesy bias’ (Luby et al., 2006).
 - Investigators’ observations may be biased by their expectation that the intervention will be effective.
- Surveys themselves can alter behavior:
 - Question-behavior effects: the survey causes the respondent to think about the topic more deeply than usual, potentially changing the response or the behavior (Spangenberg et al., 2008). Question-behavior effects may themselves be altered, for example, by differing frequency of surveying (Zwane et al., 2011).
- The investigators conducting a trial may themselves introduce bias:
 - A trial showing high efficacy might be more likely to be published than a trial that shows little/no efficacy (publication bias) (Hunter, 2009).
 - Investigators receiving outside funds may (consciously or unconsciously) preferentially report findings that are aligned with the funder’s desires. This may occur through publication bias or through other means.

Bias might also lead to understatement of effectiveness. For example, a family that has poor hygiene practices might nonetheless report regular handwashing because it is a socially desirable answer. In a hygiene education trial, this would overestimate apparent handwashing in

the absence of an effect on disease, understating effectiveness of the intervention.

Although bias is seldom quantified, a recent meta-epidemiological study (L. Wood et al., 2008) compared a wide variety of intervention studies (e.g., caesarian sections, smoking cessation, cancer treatment outcomes, etc.) with good/poor blinding to determine how these factors influence the reported effect of an intervention. When considering trials with subjectively reported outcomes, comparison of 104 unblinded trials with 205 blinded trials showed that unblinded trials tended to have odds ratios that were 25% lower (indicating a bias toward greater perceived effectiveness) compared to blinded trials (L. Wood et al., 2008). There was a similar effect regarding proper randomization of participants to treatment groups (allocation concealment). However, unblinded trials with objective outcomes did not show evidence of bias (L. Wood et al., 2008). This would seem particularly relevant to interventions to prevent diarrhea in developing countries since diarrhea reporting can be highly subjective, although such interventions were not the focus of this meta-epidemiological study.

A few trials report on microbiological outcomes, such as the log reduction of microorganisms attained by a particular water treatment method. However, most trials consider reduction of a diarrheal syndrome as the main outcome, without reference to particular pathogens. Since the ultimate goal of an intervention is to improve health, and it is easier to measure symptoms in people than to identify pathogens in feces, disease outcomes tend to be preferred in the literature.

2.15. Compliance with interventions, and long-term sustainability

Compliance with an intervention refers to the extent that people or communities actually use an intervention. It is often measured in terms of the proportion of people who are using (or claim to use) an intervention at a particular time. The word 'adherence' is sometimes used in place of compliance (Aronson, 2007). Sustainability is a related concept that refers to

maintenance of an intervention in a community over many years (essentially forever) without outside assistance, and it is notoriously difficult to attain (Shediac-Rizkallah & Bone, 1998). Compliance is necessary, but not sufficient, for sustainability. An intervention that people comply with and initially seems sustainable may become unsustainable if conditions change, such as deterioration of the infrastructure, economy, environment, or political situation.

Household water treatment (HWT) trials tend to be short, rendering sustainability unmeasurable; only 4 of 35 household interventions reviewed (T. Clasen et al., 2007) lasted one year or more, while 4 of 6 trials of interventions at the water source lasted three or more years. These water source interventions included well digging and installation of public taps, and did not include connection of individual households to the water distribution network (T. Clasen et al., 2007).

2.15.1. Costs of compliance (monetary and otherwise)

Monetary cost to the user is a key issue in compliance, and this is somewhat controversial. Some nonzero price, however small, may reduce waste since families willing to use the intervention would presumably also be willing to pay for it. Establishment of local for-profit businesses that sell interventions (such as HWT devices) are likely to be sustainable if they are profitable (Joe Brown et al., 2007); however, even seemingly tiny prices (e.g., \$0.08/month) can impede compliance among very poor people (Stockman et al., 2007). It is also often politically difficult to allot government resources to marginalized or underprivileged populations (Batterman et al., 2009).

Like money, time is a limited resource that must be carefully managed, and competing demands on time (particularly for women) in developing countries mean that the time required for a new task must be taken from other tasks (Awumbila & Momsen, 1995). Since interventions to prevent diarrhea generally require daily effort to maintain (e.g., consistent handwashing, or

daily treatment of drinking water), interventions that require little time and effort to operate should have higher compliance and a better likelihood of sustainability.

2.15.2. Psychological, social, and cultural aspects of compliance

Even if an intervention's effectiveness can be significantly demonstrated in a trial and is adopted initially by the community, the intervention may still fail if the magnitude of the disease reduction is too small to be perceived easily by individuals, who may therefore fail to recognize its importance (E. M. Rogers, 2003; Mäusezahl et al., 2009). A mother might not notice any difference between three illnesses per year in her child as opposed to four illnesses per year, even though this represents a 25% reduction in incidence. Therefore she might not be motivated to comply with the intervention. Even though the basics of the germ theory of disease are widely understood in many developing countries, 'germs' are invisible and abstract, and may not effectively motivate behavior change (V. A. Curtis et al., 2009). Interventions that yield visible daily benefits, such as savings in time and energy due to improved water supply or construction of conveniently available and comfortable latrines, may be maintained better by the community (Waddington et al., 2009).

Habits (a behavior that occurs 'automatically' following a particular stimulus) appear important for handwashing compliance, and habits are typically established in childhood (V. A. Curtis et al., 2009). If a habit has not been established, it might be learned or encouraged by particular motivators. In a review of structured observation studies of handwashing in several developing countries (V. A. Curtis et al., 2009), disgust from feces and dirty hands was a key motivator for handwashing. Affiliation (essentially, doing what everybody else is doing) was also important; although affiliation can promote handwashing if it is common, it can also discourage handwashing if it is uncommon. Fear was not an effective motivator, except in the context of a severe immediate threat like a cholera epidemic (V. A. Curtis et al., 2009). These motivations

might apply similarly to other interventions to prevent diarrhea, such as sanitation or HWT.

Compliance with interventions also depends on cultural factors. Local beliefs about water or hygiene impact the acceptability of an intervention. For example, women's cleanliness is often highly regarded socially, but in some settings cleanliness implies attempting to attract other women's husbands (V. A. Curtis et al., 2009). Belief that disease is pre-ordained or that people are powerless to affect disease (fatalism) can also impede the acceptance of interventions. In addition, diarrhea is often perceived to be a normal part of child development rather than a disease (V. A. Curtis et al., 2009). Communities also have different capacities to adapt to changing conditions, which is based partly on available resources but also on cultural factors (Batterman et al., 2009).

2.15.3. Examples of compliance measurements in the field

In practice, high compliance is seldom attained. HWT chlorination trials provide informative illustrations of compliance, because measuring chlorine residuals in stored water during unannounced home visits is a rapid and objective measurement. As part of a large meta-analysis of HWT trials, 16 chlorination trials were reviewed (T. Clasen, I. G. Roberts, et al., 2009), finding mixed evidence for effectiveness (e.g., a significant rate ratio of 0.61 for all ages across four studies, but a nonsignificant longitudinal prevalence ratio [LPR] of 0.91 for children under 5 years of age across 5 other studies). Eleven of these trials measured chlorine residuals. Three of these trials estimated compliance at 49% (Chiller et al., 2006), 44% (Crump et al., 2005), and 61% (Crump et al., 2005). A fourth trial of 20 families had apparently perfect compliance because a health worker treated families' water daily, but found no protective effect (Kirchhoff et al., 1985). ; a possible explanation might be consumption of untreated water outside the home. A fifth trial (Doocy & Burnham, 2006) was an outlier that differed from the other trials by its remarkably low LPR of 0.09; it was carried out in refugee camps and reported

95% compliance. The compliance levels during the remaining trials were unclear. More accurate measurements of compliance are needed to improve quantitative understanding of its impact on effectiveness. With respect to HWT, the amount of untreated water consumed appears particularly important.

The challenges in implementing sustainable interventions are illustrated by the following example. An evaluation (B. Arnold et al., 2009) of a large HWT and handwashing promotion campaign in 90 Guatemalan villages, during which families with children under three years of age were visited for 30 minutes monthly or bimonthly by health educators who promoted various HWT methods, handwashing with soap, and good nutrition. At the end of the campaign it was estimated that 70% of participating households were using HWT regularly. However, in an evaluation of 600 households 6 months after the conclusion of the intervention, various health measures were no different than in control villages. Longitudinal prevalence of diarrhea, 'highly credible gastroenteritis', cough, or difficulty breathing were all no different, as were anthropometric measures of malnutrition, and only 37% of intervention households still self-reported as using HWT. Furthermore, hygienic conditions, soap use, and self-reported handwashing behavior were no different. Statistically significant, but small, positive differences were seen in confirmed HWT use, which was 26% using any HWT method.

2.16. Interaction of intervention effects

Interaction of interventions means that their joint effectiveness differs from their combined individual effectivenesses. Interaction between interventions is common, but may be positive or negative. In a review (Fewtrell et al., 2005) including 5 studies combining water supply improvements with sanitation and hygiene education, and found that the effect on childhood diarrhea was similar in those studies as the effect seen in other studies for water supply, sanitation, or hygiene alone. There was a similar effect when combining water treatment and

handwashing in Karachi squatter neighborhoods; there was no additional benefit to combining the interventions, although each intervention alone cut diarrhea prevalence by about 50% (Luby et al., 2006).

However, a meta-analysis (Gundry et al., 2004) of 7 household water treatment (HWT) intervention studies found that the effectiveness of interventions increased as sanitation improved. Positive interaction between sanitation and source water quality has been observed for diarrhea (VanDerslice & Briscoe, 1995); furthermore, increased water use and latrine possession positively interacted to improve infant weight and length in rural Lesotho (Esrey et al., 1992) (presumably by reducing diarrhea). In general, though, interaction between interventions appears to be negative, with sanitation an occasional exception. Since good sanitation removes feces from the environment and therefore drastically reduces the amount of pathogens available, it may enable greater effectiveness of other interventions by allowing them to act on a region of the dose-response curve where disease risk is more rapidly declining. A possible explanation for negative interaction may be found in bias due to difficulty in blinding field trials. For example, the effectiveness of an intervention in a field trial may be overestimated if the trial is unblinded and the outcome is subjective (L. Wood et al., 2008). Such a bias would probably be similar whether a single intervention or two joint interventions are being applied, which would give results resembling negative interaction.

Since diarrheal illness is transmitted by multiple pathogens over multiple routes, an intervention that can successfully impede transmission might not reduce pathogen exposure enough to detectably impact disease in a highly contaminated environment; enough transmission may occur by alternate routes to maintain high endemicity (Briscoe, 1984). Once conditions are somewhat improved, interventions which seemed initially to have low effectiveness might yield larger reductions in disease by further reducing the amount of pathogens into a zone of a dose-

response curve where disease risk rapidly decreases. This is especially likely if the relationship between the number of pathogens ingested and the development of disease is nonlinear (Briscoe, 1984). Unfortunately, preexisting characteristics of the area that might affect an intervention trial are often omitted or not clearly reported; many 'single interventions' might be considered 'multiple interventions' if prior community development is considered. Furthermore, meta-analyses attempt to generate overall effectiveness measures for particular interventions, but they generally lump trials together to obtain a pooled result with little regard for characteristics of the community before the intervention was implemented.

Even if such characteristics are considered, power may be lacking to determine if heterogeneity between studies is due to community characteristics. A meta-analysis (Gundry et al., 2004) examined sanitation, urban/rural setting, type of water source, water storage, and blinding to see if they affected household water treatment effectiveness, but found no relationships other than increasing effectiveness with increasing sanitation. Sanitation explained 1/3 of the heterogeneity between those studies, but the remaining 2/3 could not be explained (Gundry et al., 2004).

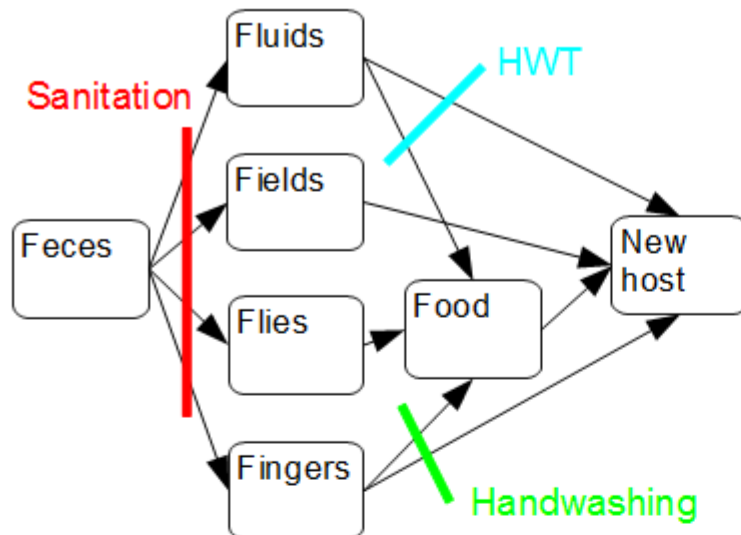
Interventions may also interact simply by facilitating each other in the household. For example, water supply improvements facilitate handwashing simply by making more water available (Curtis 2000). They might also facilitate activities such as breastfeeding and food preparation by increasing the amount of time available to the mother, such as by eliminating the necessity of walking long distances to collect water.

Natural conditions may also interact with interventions. For example, HWT chlorination was ineffective during abnormally heavy July rains in Karachi, but was effective for most of the rest of the year (Luby et al., 2006).

2.17. Descriptions of individual interventions

In general, diarrheal disease should be prevented by preventing feces from entering the mouth, as illustrated by the classic 'F-diagram' (Figure 2.4) (V. A. Curtis et al., 2000); sometimes 'fomites' are also added, denoting an object contaminated with pathogens. Anything that removes feces or pathogens from the environment has the potential to reduce infectious diarrhea; different interventions are possible, which act on different parts of the diagram. Interventions are described in detail below and will be referred to briefly in the context of particular pathogens.

Figure 2.4. Simple F-diagram of transmission of diarrheal pathogens



The 'F-diagram' (so called because nearly all compartments begin with 'F') illustrates how diarrheal pathogens can be transferred by different vectors within a community. HWT means household water treatment. Modified from V. A. Curtis et al. (2000). See Figure 2.7 (page 77) for a more detailed diagram. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation (at Michigan State University, proquest.com, Google Scholar, or openthesis.org).

2.17.1. Sanitation

Sanitation appears particularly important because it can restrict all routes connecting feces to other compartments (Figure 2.4). Although developed-country urban sanitation infrastructure including flush toilets, piped sewer systems, and sewage treatment plants are a highly effective

ideal, they are prohibitively expensive in many locations. Nonetheless, in crowded urban areas, this may be the only feasible solution due to the lack of space for building latrines.

Although sanitation seems important intuitively because of its key role in removing feces from the environment, there have been relatively few rigorous trials establishing its effectiveness on diarrhea; those available tend to be observational and unrandomized (Barreto et al., 2007; Henry & Rahim, 1990; Esrey, 1996). Large sanitation construction initiatives tend to be accompanied by other infrastructure improvements as well as hygiene education, making it difficult to isolate the effect of sanitation alone (Barreto et al., 2010). A large study (Barreto et al., 2007) comparing longitudinal prevalence of diarrhea in 0-3 year olds in the city of Salvador in Brazil before and after the implementation of a sewerage construction program showed reduction of diarrhea by about 43% in areas of the city with prevalence above 8 days of diarrhea per child per year. Most areas with lower prevalence of diarrhea did not have significant reductions in diarrhea, leading to an overall decrease of about 22% in the longitudinal prevalence of diarrhea. A later study of 0-4 year olds in the same area (Barreto et al., 2010) found a reduction in prevalence of *Giardia* infection from 14.1% to 5.3%; the new sewerage connections alone appeared to account for the greatest share (about 25%) of this decline, with improved cleanliness in/near the house accounting for an additional 17% (Barreto et al., 2010).

Given available space, several inexpensive and effective latrines can be built with local materials, such as the ventilated improved pit (VIP) latrine, pour-flush latrines connected to a hand-dug septic pit, and composting toilets (e.g, SkyLoo). These are considerable improvements over crude pit latrines, which may be avoided due to stench and risk of collapse. Latrines must also be carefully located in order to prevent fecal contamination of groundwater.

2.17.2. Water supply improvement

Improvement of the water supply at the source is an intuitively appealing intervention,

perhaps because it often involves building large tangible objects like wells, water tanks, or distribution systems. Provision of a piped connection and tap for each household is ideal but very expensive, costing approximately \$100 per person just for the initial investment (Hutton & Laurence Haller, 2004). If a tap is absent, water must be gathered outside the household and stored within the household, where it can be easily recontaminated even if source water quality is excellent. A recent comprehensive review (T. Clasen, I. G. Roberts, et al., 2009) found variable results among six source-based interventions (improved wells or public taps, not including private taps); four studies showed reductions in diarrhea, including an incidence ratio of 0.83 in 6-23 month olds in a rural Bangladeshi setting and a risk ratio of 0.45 in all ages in a rural Chinese setting.

Although water supply improvements commonly improve both the quantity of water available to the family as well as its quality, these two improvements have also been studied separately. Either alone generally yields similar benefits as both together (Esrey et al., 1991).

Water quantity improvement

Improving the quantity of available water, even if it remains contaminated, can be beneficial by facilitating more frequent washing and therefore improved hygiene. Time and labor required to gather water indirectly reduce the amount of water available. Basic guidelines for disaster response (Sphere Project, 2011) have recommended 15 liters per person per day, with a water point < 500 meters from the house, and < 15 minutes waiting time at the water point. However, these standards are frequently unmet in developing countries.

Water quality improvement

Improvements in water quality primarily yield benefits through reduced ingestion of pathogens in drinking water, although it may also reduce the amount of pathogens introduced into food. This may be particularly important if incompletely cooked weaning foods are made

with contaminated water (Motarjemi et al., 1993).

The relationship between measured water quality and diarrhea risk is unclear; although presence of common indicators of poor water quality (turbidity, thermotolerant coliforms, high heterotrophic plate count) strongly suggest that water is unsafe to drink, the levels of these indicators are poorly correlated with the amounts of actual pathogens that are present. Furthermore, apparently clean water may still be contaminated with pathogens. A meta-analysis of 11 studies measuring *E. coli* or thermotolerant coliforms found no relationship between childhood diarrhea risk and the level of the bacterial indicator (Gundry et al., 2004).

2.17.3. Hygiene

Approximately 2 to 6 liters of water are needed daily per person for basic hygiene practices (Sphere Project, 2011). The best-studied aspect of hygiene with respect to diarrhea is handwashing, although aspects such as diapering and food preparation are also important.

Handwashing

Failure (or inability due to lack of water) to wash hands provides a route for pathogens to be transmitted from person to person, as well as between people and their environment. Curtis and Cairncross (2003) describe 9 studies of handwashing in developing countries finding that 0 to 20% of people (median 13%) washed their hands after defecation or after cleaning a child who had defecated. Dirty hands can also introduce bacteria such as pathogenic *E. coli* into food, where they may multiply (Motarjemi et al., 1993), and mothers are usually responsible both for cleaning children after defecation and preparing food for the family (V. A. Curtis & Cairncross, 2003). Handwashing may be particularly important in preventing transmission of highly potent pathogens, such as *Shigella* sp. (Motarjemi et al., 1993). Accordingly, promotion of handwashing with soap in developed country daycare facilities has shown significant disease reduction (relative risk of approximately 0.5), even though the environment would seem to be much less

contaminated than in a developing country setting (V. A. Curtis & Cairncross, 2003).

Handwashing works best when soap is used, which is fortunately inexpensive. Transient bacteria are removed equally well by handwashing with ordinary soap and water as with antiseptics, and resident bacteria on the skin are relatively unaffected by handwashing with soap and water (Lowbury et al., 1964). Diarrheal pathogens contaminating hands would generally be transient bacteria. Although antibacterial agents may have residual activity on bacteria that remain on the skin following handwashing (Lowbury et al., 1964), one formulation of antibacterial soap was no more effective than ordinary soap in preventing diarrhea or pneumonia in a Karachi shantytown (Luby et al., 2005). LRVs measured in studies of handwashing are found in Table 2.1 (page 64).

Highly variable (relative risk from 1 to about 0.25) effectiveness of handwashing on diarrhea has been reported in a meta-analysis of 20 studies of various age groups, both observational and intervention-based (V. A. Curtis & Cairncross, 2003). Overall, the relative risk of diarrheal incidence was 0.57, although many studies were of poor quality.

A meta-analysis (Ejemot et al., 2008) considered five community-based, randomized, controlled intervention trials of hygiene education that included handwashing; an overall incidence rate ratio of 0.68 was estimated. This measure was consistent across studies of young children as well as children aged less than 15 years.

Despite the usefulness of handwashing in preventing diarrhea and other diseases, perpetuating good handwashing behavior remains a challenge. A 53% reduction in longitudinal prevalence of diarrhea effectiveness due to handwashing was estimated in squatter settlements in Karachi (Luby et al., 2006). However, no effectiveness was seen against childhood diarrhea during a ~13 month followup period that began 18 months after the earlier study concluded (Luby et al., 2009). Although intervention households were more likely to have a place to wash

their hands and were more likely to demonstrate better handwashing technique, there was little difference in longitudinal prevalence of diarrhea between intervention and control households. Furthermore, intervention and control households had similar spending on soap and were equally likely to have soap in the house. The authors suggest that frequent reinforcement of handwashing behavior may be necessary for sustainability; although continued home visits would be prohibitively expensive, mass media messages might be helpful. However, the similarities in purchasing and having soap between the intervention and control groups indicate that the price of soap may be a barrier to sustaining effective handwashing behavior.

Even if water and soap are available, good handwashing practices are uncommon. In 11 developing countries, handwashing with water after defecation was only observed in 45% of caregivers, and just 17% used soap (V. A. Curtis et al., 2009). Handwashing is largely habit-driven, meaning an unplanned reaction to something in the environment, such as touching something unclean (V. A. Curtis et al., 2009). Since habits are largely acquired in childhood, changing handwashing behavior in a community might require many years; once a community improves handwashing, it can be very stable over time, because conforming to social norms is a powerful influence on behavior. However, if the social norm is lack of handwashing, as it is in many areas, this impedes improvement of handwashing behavior; increased promotion via broadcast media or visual cues like posters may help redefine social norms while providing cues that reinforce habits (V. A. Curtis et al., 2009).

Diapering and open defecation

Fecal contamination around the home has repeatedly been linked to increased diarrheal disease risk (V. A. Curtis et al., 2000), although this relationship disappeared in a study in Myanmar when adjusted for education and socioeconomic status (Han & K. Moe, 1990). Latrines do not necessarily remove all fecal contamination, and diarrhea reductions attributed to

latrines are likely due to removing stools from the environment (V. A. Curtis et al., 2000). Many cultures do not consider infant stools to be hazardous (Motarjemi et al., 1993), and diapering was associated with absence of diarrheal illness in children under two years of age in a case-control study in Nicaragua (Gorter et al., 1998).

Even if diapers or potties are available, feces may still enter the environment if they are not properly disposed of, such as in a latrine. Yeager et al. (1999) describe conditions in a dry shantytown area of Peru. Potty-training can be a long and difficult process, during which defecation in clothes or on the ground is common (Yeager et al., 1999). Even if child feces are noticed on the floor or near the house, the mother is often too busy to dispose of them immediately, and disposal in places such as the garbage dump or the street is common; nonetheless, feces in potties are more likely to be disposed of, and even incomplete disposal may be helpful (Yeager et al., 1999). Effective potty-training may therefore yield important reductions in fecal contamination of the environment.

Anal cleansing

Few published scientific papers have studied anal cleansing. Effective soft materials similar to toilet paper may be unaffordable or scarce, which can lead to increased exposure to feces (McMahon et al., 2011). Focus groups of Kenyan schoolchildren reported lack of instruction from their parents and teachers regarding proper anal cleansing, and parents did not perceive the benefits of toilet paper as worth the cost (McMahon et al., 2011). A study of school toilets and hygiene in Colombia suggested that simply providing toilet paper, towels, and soap could be an important method for preventing diarrhea, since that composite factor was associated with more cases of diarrhea than the number of toilets available (Koopman, 1978).

Food preparation

Diarrheal risk substantially increases following 6 months of age, when infants who were

exclusively (or mostly) breastfed begin consuming weaning foods, which are often more heavily contaminated than other foods eaten by the family (Motarjemi et al., 1993). Weaning foods are often thin porridges which are not thoroughly heated since long cooking makes a porridge that is too thick for infants, and organisms such as *Bacillus cereus* and pathogenic *E. coli* can multiply in such foods at ambient temperature (Motarjemi et al., 1993). The necessity of feeding infants several times per day means that food may be prepared in advance and stored at ambient temperature to save time (Motarjemi et al., 1993). Storage of hot food in vacuum flasks (which are durable and relatively inexpensive) reduces the rate of cooling, allowing food to be stored safely for up to 12 hours (Mensah & A. Tomkins, 2003).

Certain traditional grain food preparations that are fermented with lactobacilli can limit growth of, or actually kill, pathogenic bacteria, principally through the creation of acidic conditions (pH 3.5 to 4.5) (Adams & Nicolaides, 1997). Production of other compounds, such as bacteriocins (polypeptide ‘antimicrobials’), hydrogen peroxide, and CO₂, may also aid pathogen inhibition (Adams & Nicolaides, 1997). It is unclear to what extent fermentation inactivates viruses; based on limited research, rotavirus appears to persist in fermented foods (Mensah & A. Tomkins, 2003).

Fly control

Although measures such as insecticide treatment can yield reductions in diarrheal illness due to fly killing, such interventions are not likely to be sustainable (V. A. Curtis et al., 2000). Sanitation interventions are probably more effective, since they remove feces from the environment, preventing flies from accessing them (V. A. Curtis et al., 2000).

2.17.4. Household water treatment (HWT), or point-of-use (POU) technology

This is a particularly active area of investigation. Since HWT can be carried out by

individual families, it can be effective even if infrastructure or community cooperation is limited. Water is frequently collected outside the home and stored in the home, and even if the source water is perfectly clean, the stored water may become contaminated when things fall into it, or when dirty hands or cups are placed in it (A. J. Pickering et al., 2010). HWT can therefore be more effective than interventions at the water source (T. F. Clasen et al., 2006). HWT can also be implemented far more cheaply per person than more infrastructure-intensive interventions like latrines and wells (Hutton 2004). Although HWT seems appealing, it can require substantial behavioral change or monetary investment, which impede sustainability. There are several different methodologies, which are described below. LRVs for various HWT methods are given in Table 2.1, and antimicrobial standards for HWT are given in Table 2.2.

Safe storage

Safe storage refers to use of a storage container that does not allow anything to touch the water inside, preventing recontamination. This commonly takes the form of a vessel with a covered top (or a narrow neck) and a spigot at the bottom. It is often combined with other HWT methods.

A less expensive alternative to safe storage is the ‘two-cup method’, which attempts to prevent contamination of stored water by using a single, clean, cup which is supposed to be the only object that touches the stored water. Water is decanted from the storage vessel with the cup and then poured into another cup to drink. However, it may be difficult (or impossible) to ensure that the cup remains clean, or to prevent other objects from entering the water.

Boiling

Boiling is the most widely used HWT method, and it is the only method that can reliably destroy all pathogens (T. Clasen, 2009). However, it requires fuel, which is frequently expensive or unavailable. Depending on how it is used, boiled water may require cooling, which is time-

consuming. Stored boiled water is also subject to recontamination in the household.

Solar disinfection (SODIS)

Solar disinfection is accomplished by filling clear 1 to 2 liter polyethylene terephthalate (PET) bottles with water and placing them in the sun for at least 5 hours, or for two days under cloudy conditions (EAWAG, 2004). Oxygenating the water (e.g., by shaking the bottles) increases effectiveness (EAWAG, 2004). The method is primarily suited for treating drinking water; it is difficult to produce enough water by SODIS for washing or other purposes (EAWAG, 2004). Pathogens in the water are inactivated by a combination of heating and irradiation by ultraviolet A (UV-A). Relatively clear water of <30 NTU is recommended for irradiation to be effective (EAWAG, 2004); nonetheless, highly turbid water (>300 NTU) can still be made safer by SODIS due to solar heating of the water.

Simply heating water using solar energy, for example in a solar cooker or using black or metal vessels, is a similar method to SODIS. Under sunny conditions temperatures over 60°C can be attained, which essentially pasteurizes the water, inactivating most microorganisms (Sobsey, 2002). Even if a solar cooker is used, at least three hours in full sun is required to inactivate 99% of viruses in 3.8 L of water (Sobsey, 2002).

Disadvantages of these methods include the substantial time and effort needed to fill bottles and place them in the sun, as well as having to wait for several hours for the method to work. The method is much less effective on cloudy days.

Although some community trials have shown effectiveness of SODIS in reducing diarrhea (T. Clasen, I. G. Roberts, et al., 2009; A. Rose et al., 2006), a recent community-randomized trial in rural Bolivia (Mäusezahl et al., 2009) found no effect of SODIS on diarrhea incidence, longitudinal prevalence, severe diarrhea, or dysentery, despite provision of bottles and repeated educational sessions demonstrating the method. However, the study was only powered

to detect a 33% reduction in diarrhea incidence, and the communities lacked sanitation, so other infection pathways remained open and may have swamped any effect of SODIS. Furthermore, only about 32% of intervention households were observed to be using SODIS on any given day, and its use tended to be lower during the cultivation season when families were busy. These observations highlight the difficulty of implementing SODIS sustainably.

Chlorination

Chlorination can be performed in the home by adding sodium hypochlorite (household bleach) to water (attaining about 5 mg/L free residual chlorine) (Centre for Affordable Water and Sanitation Technology, 2008), shaking the container to mix the water, and letting it stand for 30 minutes. This will kill most microorganisms, although *Cryptosporidium* is a common pathogen that is resistant. Commercial bleach may be used, or a custom solution may be produced and sold. Such a solution usually corresponds to one capful of solution for the particular size of storage vessel that is common in the community; the solution bottles are custom-designed to accomplish this. Increasing the pH of the solution above 11.9 with NaOH lengthens the shelf life to 12-18 months, though it should be used within 60 days of opening (Lantagne & Gallo, 2008). A double dose of hypochlorite solution is often recommended if the water is turbid (Lantagne & Gallo, 2008), although this may increase the risk of exposure to carcinogenic disinfection byproducts from the reaction of hypochlorite with organic material in the water. Chlorination is sometimes combined with a flocculant (e.g., PuR packets) which facilitates settling of particulate matter, improving its appearance and aiding removal of pathogens. Certain plants such as *Moringa oleifera* and *Opuntia* species can also provide flocculants (S. M. Miller et al., 2008). Some people find the taste of chlorinated water unappealing, although properly treated water should not taste strongly of chlorine. CDC's Safe Water System (SWS) combines chlorination with safe storage (Lantagne & Gallo, 2008). Devices containing polymer beads that bind

chlorine or bromine and release it into contaminated water are more expensive than adding chlorine solution to water, but may be easier to use because they do not require measurement of a dose; furthermore, bromine has a milder taste than chlorine (Dunk, 2007). Such devices are effective at inactivating bacteria and bacteriophages (McLennan et al., 2009; Coulliette et al., 2010).

In a meta-analysis of the effectiveness of HWT chlorination against childhood diarrhea, a pooled relative risk estimate of 0.71 (0.56-0.89) was determined for children under five years of age (B. F. Arnold & Colford, 2007). Three studies in urban or peri-urban areas showed a larger risk reduction (0.63: 0.50-0.80) than the 5 rural studies (0.89: 0.71-1.13). However, the trials were not carried out over multiple years (the longest one covered 87 weeks), and longer trials appeared to show lower effectiveness (although this trend was nonsignificant), raising questions about sustainability.

Recurring costs and interruption in chlorine supply are major obstacles to sustainability of HWT chlorination. For example, in Malawi, where a chlorination solution (WaterGuard) has been marketed nationwide, 64% of a nationwide sample (Stockman et al., 2007) of mothers had heard of the solution, and 12% of those said they were currently using it. Among mothers who had used WaterGuard in the past but were not using it at the time of the survey, 39% said that they couldn't afford it and 34% said that it was "currently unavailable". The price of WaterGuard was about \$0.08 for one month's supply. Even if rural families are willing to pay for chlorination solution, distance (or the cost of transportation) may make it impossible for them to obtain it.

Filtration

The two major types of filters are ceramic filters and sand filters. They are relatively expensive (initial cost of \$20 or more) (Sobsey et al., 2008) and require frequent maintenance. Ceramic filters need to be scrubbed to remove trapped material that slows flow through the filter;

sand filters require cleaning of the upper layer of sand to remove sediment and improve water flow through the filter. Filtration trials have shown relative risks for diarrhea of about 0.4 when comparing households with filters to those without (Sobsey et al., 2008; T. Clasen, I. G. Roberts, et al., 2009). Ceramic filters appear more effective at removing bacterial and protozoan pathogens from drinking water than biosand filters (Sobsey et al., 2008). Filters have the important advantage of improving the appearance (and often the taste) of the treated water. Chlorination may also be used on stored filtered water to further improve disinfection and prevent recontamination.

Biosand filters

The biosand filter (BSF) is a recent improvement on older slow sand filter designs. It consists of a tall bucket containing a layer of gravel with several feet of sand atop it. The outlet pipe originates at the bottom of the filter (in the gravel layer), and the mouth of the pipe (where filtered water exits) is slightly higher than the top of the sand layer. This ensures that the entire filter column remains wet, allowing an aerobic biofilm to establish itself near the top of the sand. The biofilm is believed to enhance deactivation of pathogens (Manz, 2009); this is consistent with results showing increasing deactivation of pathogens (including some viruses) while the filter ‘matures’ over several weeks (Elliott et al., 2008), but could also be explained by increased residence time in the filter due to decreased flow rate as pores close over time (Elliott et al., 2008). Inactivation is drastically lowered (by 1-3 LRV of *E. coli*) if the amount of water passed through it daily exceeds its pore volume (Elliott et al., 2008). Although BSFs can often be constructed in developing-country communities according to instructions that are freely available, manufacture of properly functioning BSFs requires great attention to detail and consideration of local conditions (e.g., the characteristics of the sand/gravel available, quality of cement/concrete, etc.) (Manz, 2007). Although BSFs can produce water on demand, a separate

container is usually needed to store filtered water (Manz 2007), which is subject to recontamination if not stored safely.

Ceramic filters

Ideally, ceramic filters have pores that are too small (about 0.2 μm) to allow most bacteria and protozoa to pass through, although they cannot block viruses (Joe Brown et al., 2007). They can sometimes be made out of local materials, though quality control can be challenging (Sobsey et al., 2008). The filters need to be cleaned, and cracks in the filter reduce (or eliminate) its effectiveness. The two main ceramic filter designs are ‘pot’ (a large ceramic pot-shaped filter nested inside another vessel that captures the filtered water) and ‘candle’ (a cylindrical filter, often used inside a plastic receptacle to capture filtered water).

An evaluation of a program to produce, distribute, and promote ceramic pot filters in Cambodia (Joe Brown et al., 2007) showed that the number of filters in use decreased linearly over time, with about half of 506 households still using the filter after 24 months. Breakage accounted for 65% of the disused filters, so replacement filters must be purchased frequently (they cost \$2.50 to \$4.00 to produce).

A pilot project in a small Bolivian community (T. F. Clasen et al., 2006) evaluating imported candle filters indicated that they were effective in reducing diarrhea, but identified several problems: many users reported inadequate quantity or flow of water, replacement filters were unavailable, and approximately 8% of the filters were broken 9 months after they were provided.

Advanced filter technologies

Nanofilters with a pore size small enough to remove viruses have also been built into HWT devices for use in developing countries (T. Clasen, Naranjo, et al., 2009; Boisson et al., 2010). Such filters require advanced manufacturing techniques and cannot be produced locally like

ceramic pot filters. However, they are effective at removing pathogens, and in general should perform comparably to ceramic filters against bacteria and protozoa, while removing more viruses than ceramic filters.

Table 2.1. Log₁₀ reduction values (LRVs) for various interventions

Intervention	Laboratory*			Actual community*			Citations	Notes
	Bact- erial	Viral	Proto- zoan	Bact- erial	Viral	Proto- zoan		
HWT chlorination	6+	6+	5+	3	3	3	(Sobsey et al., 2008)	<i>Cryptosporidium</i> resists Cl; LRV doesn't apply to it.
HWT coagulation & chlorination (PUR [®] sachets)	9	6	5	7	2-4.5	3	(Sobsey et al., 2008)	
Chlorinated polymer beads (HaloPure [®])	6	3					(McLennan et al., 2009; Coulliette et al., 2010)	3 LRVs against <i>Clostridium</i> .
Brominated polymer beads (HaloPure [®])	6	5					(McLennan et al., 2009; Coulliette et al., 2010)	3 LRVs against <i>Clostridium</i> .
Ceramic filtration	6	4	6	2	0.5	4	(Sobsey et al., 2008)	
Biosand filtration	3	3	4	1	0.5	2	(Sobsey et al., 2008)	
Sari filtration (4 layers of ordinary cloth)	2						(Huq et al., 1996)	<i>V. cholerae</i> attached to particulates.
Nanofiltration (LifeStraw Family [®])	6	4	3				(T. Clasen, Naranjo, et al., 2009)	Pore size of 20 nm.
Solar disinfection (SODIS)	5.5+	4+	3+	3	2	1	(Sobsey et al., 2008)	
Handwashing with soap	3			0.5			(Lowbury et al., 1964; Luby et al., 2001)	
Handwashing without soap	1	1		0.3			(Ansari et al., 1989; A. J. Pickering et al., 2011)	

*LRVs are commonly higher in the laboratory (careful implementation and maintenance) compared with actual communities (might be improperly used or poorly maintained).

Table 2.2. Log₁₀ reduction values: standards for household water treatment (HWT)

Standard	Bact- erial	Viral	Proto- zoan	Citation
WHO 'highly protective' target	4+	5+	4+	(Sobsey & Joe Brown, 2011)
WHO 'protective' target	2+	3+	2+	(Sobsey & Joe Brown, 2011)
WHO 'interim' target	Meets 2/3 of 'protective' targets above.			(Sobsey & Joe Brown, 2011)
USEPA HWT standard	6+	4+	3+	(USEPA, 1987)
USEPA primary drinking water standard for treatment facilities	*	4+	3+†	(USEPA, 2012)

*<500 colonies on heterotrophic plate count; ≤5% positive samples for total coliforms monthly.
†LRV of 2+ for *Cryptosporidium*.

Recent controversy surrounding HWT

Three recent reviews (Schmidt & Cairncross, 2009; Waddington et al., 2009; Hunter, 2009) have questioned whether there is really enough evidence to warrant widespread promotion and scaling up of HWT, despite substantial investment by WHO, the Gates Foundation, and others to expand it (Schmidt & Cairncross, 2009). WHO has referred to chlorination and safe storage as “consistently the most cost-effective” water, sanitation, and hygiene intervention (World Health Organization, 2002). However, Schmidt and Cairncross (Schmidt & Cairncross, 2009) argue that the decision of whether to promote HWT primarily rests upon demonstration of a health effect, since other benefits from HWT are small and the scalability and acceptability (which are critical to sustainability) of HWT remain unclear. The health effect is difficult to establish because results from trials vary greatly, and publication bias may be a large problem, particularly regarding smaller trials (Hunter, 2009). These reviews also conclude that the published health effect estimates attributed to HWT might be wholly explained by bias (reporter or observer), since it is difficult or impossible to blind studies of HWT interventions. They therefore call for

larger, long-term, blinded studies to be carried out to establish whether beneficial health effects truly exist, before making large investments in expansion of these interventions.

Hunter's recent review (Hunter, 2009) also showed that intervention effectiveness was lower among studies with longer follow-up; similar findings have been reported (B. F. Arnold & Colford, 2007) regarding household water chlorination. This suggests difficulty with sustainability. However, other explanations might include declining of reporting bias over time or by a tendency for longer-term studies to be better designed and therefore less subject to bias. Hunter's analysis (Hunter, 2009) also shows that effect sizes shown in most HWT trials over longer periods are roughly within the range that might be explained by bias in unblinded studies with subjective outcomes (risk ratio of ~ 0.7) (L. Wood et al., 2008). Ceramic filtration is a notable exception, showing substantial reductions in diarrhea even in long-term trials (risk ratio of 0.44, 95% CI 0.28 to 0.70) (Hunter, 2009). However, Wood's (L. Wood et al., 2008) quantification of bias due to lack of blinding drew on a widely varying set of clinical trials, nearly all of which were drug trials or therapeutic intervention trials. However, there seems little reason to believe that HWT trials would be any less susceptible than drug trials to bias due to lack of blinding.

These assertions have been contested by several noted researchers in the HWT field, who want current HWT promotion efforts to continue even while additional needed research is carried out (T. Clasen, Bartram, et al., 2009). Patterns of publication suggesting publication bias in HWT have been noted elsewhere, but have been interpreted as possibly being caused by trials of different methods in drastically differing settings (T. Clasen, I. G. Roberts, et al., 2009).

2.17.5. Other interventions

Many other important methods are available to mitigate the impact of diarrheal disease, such as: nutritional interventions (e.g., encouragement of breastfeeding, zinc supplementation);

vaccination against rotavirus; and treatment methods (particularly oral rehydration solutions). However, these interventions are not directly addressed in the research described in chapters 3, 4, and 5 of this dissertation. The reader may refer to page 235 in the appendices for a discussion of these methods.

2.17.6. Gaps in knowledge about diarrheal disease interventions

There is relatively little information about real-world effectiveness of interventions outside of field trials, and high-quality field trials lasting for a year or more are scarce. Use of an intervention may decline as equipment or infrastructure deteriorate and outside encouragement of healthy behaviors decreases. People may also forget how to apply the intervention effectively. This is difficult to investigate because the same communities must be studied over many years.

Intervention trials commonly consider diarrhea as the outcome, without regard to etiology. Given the differing characteristics of common diarrheal pathogens, it is likely that interventions will affect different pathogens in distinct ways. The relative abundances of different pathogens in different areas may therefore affect intervention effectiveness, as well as the interactions of interventions with each other. However, information concerning antimicrobial effectiveness of some interventions (particularly HWT methods) is available (Table 2.1, page 64).

The effectivenesses of interventions are likely to differ depending on a community's level of development, because improvements in sanitation, water supply, or hygiene in an area are likely to restrict certain routes that pathogens travel in that community. Therefore, an intervention might prevent a smaller or larger fraction of disease depending on the interventions that have preceded it (Briscoe, 1984; VanDerslice & Briscoe, 1995). Existing intervention effectiveness estimates from field trials are likely to be incomplete because they do not take into account previous interventions (or protective characteristics) that are impacting the community at the time of the study. Effectiveness studies should carefully describe the status of the community

in detail before the intervention took place, in order to place the resulting effectiveness estimates in the context of the community. This would yield a large number of effectiveness estimates narrowly defined to particular settings. Although they would be difficult to summarize, they would be useful for validating diarrheal transmission models and the effect of interventions on the transmission network.

2.18. Infection transmission modeling and its application to diarrheal disease

Classical epidemiological methods (e.g., regression analysis of risk factors associated with a particular health outcome) are powerful within certain contexts, such as determining factors associated with illness in a point-source outbreak of disease. However, these methods often incorporate inappropriate assumptions with respect to infectious disease transmission. For example, classical statistical methods often assume that outcomes experienced by different individuals are independent (i.e., are not influenced by other individuals). This is clearly violated when pathogens are passed from person to person (Koopman, 2004). Modeling allows formalization of relationships between individuals, as well as between individuals and their environment, avoiding the false assumption of independence.

Models of infectious disease transmission have been used to guide public health efforts, particularly with regard to choosing among different control strategies. These include polio eradication by vaccination (K. M. Thompson et al., 2006), smallpox preparedness planning (Keeling & Rohani, 2008), and slowing pandemic influenza spread (Cooper et al., 2006). However, modeling methods have seldom been applied to common childhood diarrhea in developing countries (but see Eisenberg et al., 2007), despite the magnitude of the problem.

2.18.1. Types of models

In general, models attempt to represent the behavior of a real system in a simplified fashion according to a defined set of criteria. Models and simulations are used in many disciplines and

the nomenclature can be confusing and inconsistent. The general classification in Table 2.3 is based on Haefner (2005) with some modifications:

Table 2.3. Terms commonly used to describe or categorize models

Model descriptor	Key question	Notes
Mechanistic (yes) Empirical (no)	Does the model explicitly include the inner workings of the system being studied?	Empirical models are sometimes called descriptive or phenomenological models, and often constitute simple equations.
Dynamic (yes) Static (no)	Does the modeled system change over time?	
Continuous (yes) Discrete (no)	Is time (or other quantities of interest, e.g., population size) measured continuously?	Systems of differential equations usually describe time continuously, while a system of difference equations could measure time as a discrete integer number of days.
Spatially heterogeneous (yes) Spatially homogeneous (no)	Are spatial relationships explicitly represented?	A spatially heterogeneous model must also represent space in a continuous or a discrete manner (discrete space is divided into cells or zones).
Stochastic (yes) Deterministic (no)	Are random events included?	Models based on matrices of probabilities (Markov chain models) are sometimes called stochastic models because they track expected long-run behavior of systems where behavior of individual units is uncertain, even if they do not directly incorporate random events.
Model descriptor	Definition	Notes
Compartmental	Quantities (e.g., pathogens, water) flow between storage compartments (e.g, people, tanks).	Commonly used in models of infection transmission.
Agent-based	The model represents discrete units whose characteristics change; relationships between these units are explicitly defined.	Also called individual-based models.
Network	Describes links between units ('nodes'); e.g., person A might contact persons B & C, but B & C do not contact each other directly.	

Any model will incorporate some combination of the characteristics described in Table 2.3. Examples might be a mechanistic dynamic discrete-time spatially homogeneous deterministic model, or a mechanistic dynamic continuous-time discrete-space stochastic model. A larger model might also contain a sub-model of a different type. An advantage of mechanistic models is their ability to explicitly describe interdependence between individuals and groups. The ability to represent nonlinearities, such as feedback loops, is an important feature as well.

2.18.2. Model verification and validation

Any model must be carefully verified. Model verification means ascertaining that the model does what its designer intends; i.e., there are no bugs in the algorithm/program (Haefner, 2005). This is separate from whether the designer's intentions are ill-informed or mistaken. Once the model has been verified, validation statistically compares the behavior of the model to real systems to assess whether the model can explain or predict the behavior of those systems (Haefner, 2005). Thorough validation thus requires detailed data collected by carefully designed experimental or observational studies.

Although model validation may be possible for very simple systems, it is probably impossible to thoroughly validate a model of a complex or complicated system (Oreskes et al., 1994). It is unlikely that we can observe everything of importance within a system; furthermore, any measurement we make of a system contains error (Oreskes et al., 1994). Many aspects of diarrheal infections are poorly understood; for example, although we know that malnutrition and diarrhea exacerbate each other, we do not know exactly how, the mechanism is likely to vary depending on the particular pathogen, and it is unclear how much diarrhea is attributable to particular pathogens in particular communities. Nonetheless, models of complex, poorly-understood systems can still provide insight about how the real systems operate, and models can always be tested to determine whether they approximate certain essential features of the system;

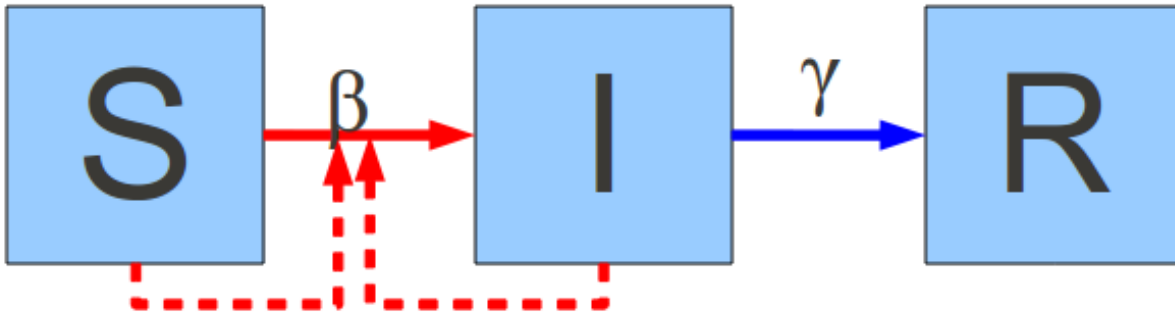
this might be termed 'weak validation' or 'confirmation'. However, confirmation does not necessarily mean the model accurately reflects the workings of the real system; further investigation of the system might reveal inconsistencies (Oreskes et al., 1994).

Although models have limitations (as do all forms of human knowledge), constructing and assessing them remains useful because they allow us to formalize available knowledge and apply it to understanding and addressing problems. The process of model building often provides guidance about the kind of information that needs to be gathered by studies in order to more fully understand the system. This can occur even if scant data are available for validating or confirming the model. Modeling is a process that attempts to improve understanding of a complex system through rigorous description of the system's characteristics and their interrelationships, followed by comparison of the model against reality, followed by revision of the model, and so on in a repeating cycle. Analysis of the model and its results may suggest experiments that could allow more rigorous validation, or provide information to improve the structure of the model (e.g., by carefully measuring a particular value that greatly alters the results of the model).

2.19. Modeling transmission of diarrheal infections

Infection transmission models are typically dynamic models, describing the behavior of a system of susceptible and infected hosts over time. They explicitly describe changes in transmission depending on the number of people that are infectious. These models are commonly represented by box-and-arrow diagrams, in which boxes (sometimes called 'compartments') represent a 'stock' and solid arrows represent a 'flow'. For example, the classic SIR (Susceptible – Infectious – Removed) model (Keeling & Rohani, 2008) represents hosts (e.g., people) flowing in and out of 'susceptible', 'infectious', and 'removed' stocks, where 'removed' could mean immunity or death (Figure 2.5).

Figure 2.5. Schematic of the Susceptible-Infectious-Removed (SIR) model



Solid arrows represent hosts flowing between stocks (which represent infection states); dashed arrows represent the influence of a stock upon a flow. Red and blue arrows represent infection and recovery, respectively. All letters represent terms in equation 2.1 below. Modified from Keeling and Rohani, 2008.

2.19.1. Simple mathematical example of a transmission model

Simple infection transmission models can be described deterministically by systems of differential equations, where the rate of change of the number of persons in a particular stock is described using first-order rates of individuals entering or leaving the stock. For example, in the SIR model (Figure 2.5), S, I, and R represent the proportion of the population in the susceptible, infectious, or immune stocks at any given time (thus, the three stocks sum to 1) (Keeling & Rohani, 2008):

$$dS/dt = -\beta SI \qquad dI/dt = \beta SI - \gamma I \qquad dR/dt = \gamma I \qquad (2.1)$$

Transmission is often described by a parameter (β) that indicates how likely it is for a contact between an infected and uninfected person to result in a new infection. Assuming that every member of the population is equally likely to contact every other member, the product of S and I is proportional to the number of possible contacts where an infectious person can infect a susceptible person. The rate of recovery (γ) is the reciprocal of the infectious period and governs the recovery of infected people (i.e., the rate of their movement from I to R). The SIR model

structure can be modified to represent many different systems, e.g.: susceptible or infected (SI) if immunity is not an issue; susceptible, exposed, infected, or removed (SEIR) where the ‘exposed’ state accounts for the incubation period, etc.

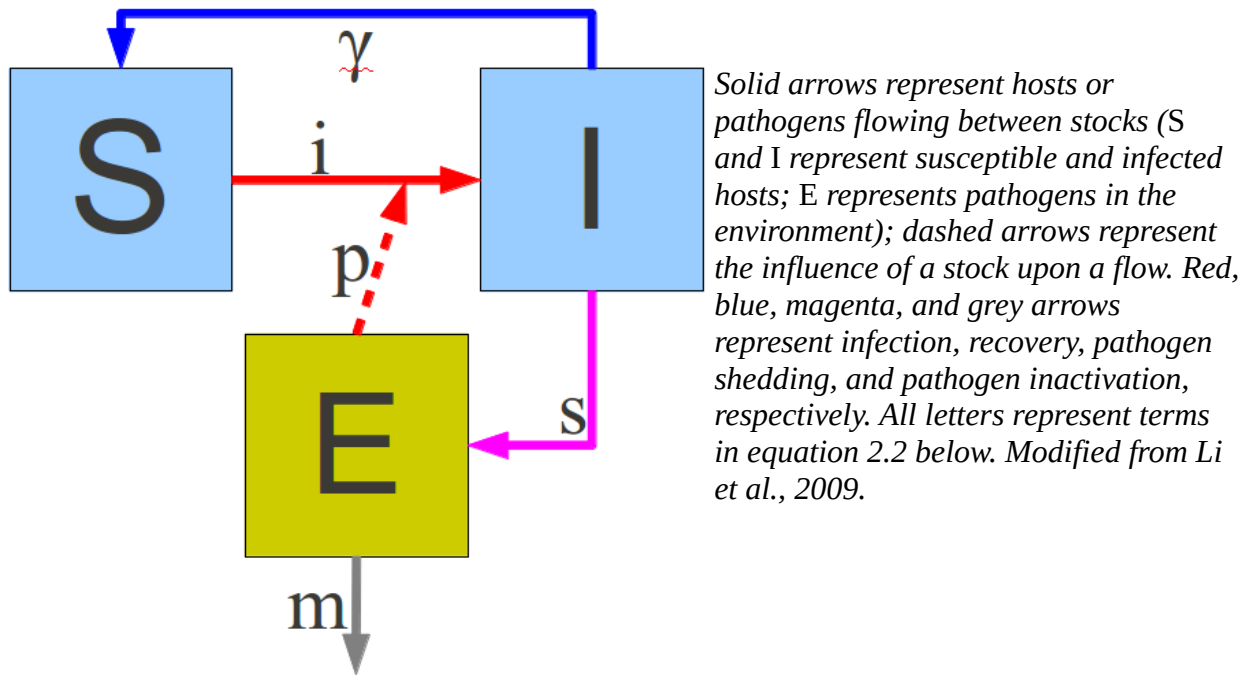
The basic reproduction ratio, R_0 , is a particularly important concept in infection transmission modeling. It is the average number of individuals that an infected individual directly infects if it enters a completely susceptible population (Anderson & May, 1991). The infection must die out if R_0 is less than 1, because the pathogen would not replace itself (Anderson & May, 1991). Since populations are seldom completely susceptible, an R_0 substantially greater than 1 might be necessary for the disease to remain endemic in a population. The general concept of R_0 can also be applied to subpopulations within a larger population, describing the number of new infections in one subgroup due to transmission from one other subgroup (M. G. Roberts & Heesterbeek, 2003). Epidemics may still occur if R_0 is < 1 at the community level if conditions become favorable for transmission within some subgroup, effectively establishing a local $R_0 > 1$; this can be simulated in stochastic models (Halloran et al., 2002). Stochastic models can also allow random extinction of the disease in the population, rendering the population disease-free unless the pathogen is reintroduced.

2.19.2. Environmental infection transmission models

The SIR model and related models described above assume that pathogens are only transmitted through person-to-person contact. However, many pathogens remain viable in the environment, and can be transmitted between hosts in many ways (e.g., food, water, or fomites). Mechanistic models describing the transmission of pathogens through the environment and their effects on hosts are called environmental infection transmission systems (EITS) models (Li et al., 2009). These models facilitate the simulation of disease prevention interventions because they

explicitly represent the numbers of pathogens in the system, which can then be directly modified by an intervention. The action of an intervention to reduce a particular flow of pathogens reduces risk to the people who ingest the pathogens. A simple EITS model that could represent transmission of diarrheal infections can be created (Figure 2.6) from the SIR model described previously (Figure 2.5), by adding an 'environment' compartment that describes pathogens in the environment; infectious hosts release pathogens, which can be picked up by other hosts. Since diarrheal infections seldom confer complete immunity, the 'removed' compartment has been removed. Note that the two blue stocks represent susceptible or infectious hosts, and the yellow environment stock represents pathogens. The number of pathogens in the environment stock influences the rate by which susceptible hosts become infectious.

Figure 2.6. Simple environmental infection transmission system model



The EITS model in Figure 2.6 is represented mathematically as follows (modified from Li et al., 2009):

$$dS/dt = -piSE + \gamma I \qquad dI/dt = piSE - \gamma I \qquad dE/dt = sI - mE \qquad (2.2)$$

As in the SIR model, γ is the rate of recovery of infectious people, although in this model they become susceptible again instead of becoming 'removed'. The rate of shedding into the environment (pathogens per infectious host per day) is represented by s ; the inactivation rate of pathogens in the environment is represented by m ; the pickup rate of pathogens by susceptible people is p ; and the probability of infection per pathogen is i . Interventions to reduce transmission could be simulated by increasing m , or by reducing p . This EITS model is highly simplified; it can be made more realistic by using quantitative microbial risk assessment (QMRA) techniques to represent the relationship between environmental pathogens and infection.

2.19.3. Quantitative microbial risk assessment (QMRA) models

QMRA models are a widely used method for understanding the risk of disease (Haas et al., 1999). Typically, these models use an exposure step to estimate the dose of pathogens ingested, followed by a dose response step in which the mean dose of pathogens entering the host is translated into a probability of infection (or illness, or death) by a dose response equation. These equations (often called dose response models) assume that a single pathogen has some probability of causing infection; several possible equations exist, and the most commonly used are the exponential equation and the beta-Poisson equation (Haas et al., 1999). The parameters of dose response equations are determined by fitting them to data from experimental studies in which hosts (sometimes humans) are given differing doses of pathogens by a particular route (e.g., oral, inhaled, or parenteral), and the proportion of hosts who become infected or diseased by each dose are recorded. However, it is unclear whether these dose-response relationships apply to (possibly malnourished or ill) children, or to developing country settings. For ethical reasons, dose response studies in humans must include only healthy volunteers, who are nearly

always adults. Although some live attenuated rotavirus vaccine trials with healthy children have provided dose response data (Vesikari et al., 1985; Pichichero et al., 1990), the vaccine strains of these pathogens probably behave differently than wild-type pathogens. It is not clear (and perhaps unknowable) how dose response equations might differ in malnourished (or otherwise unhealthy) children.

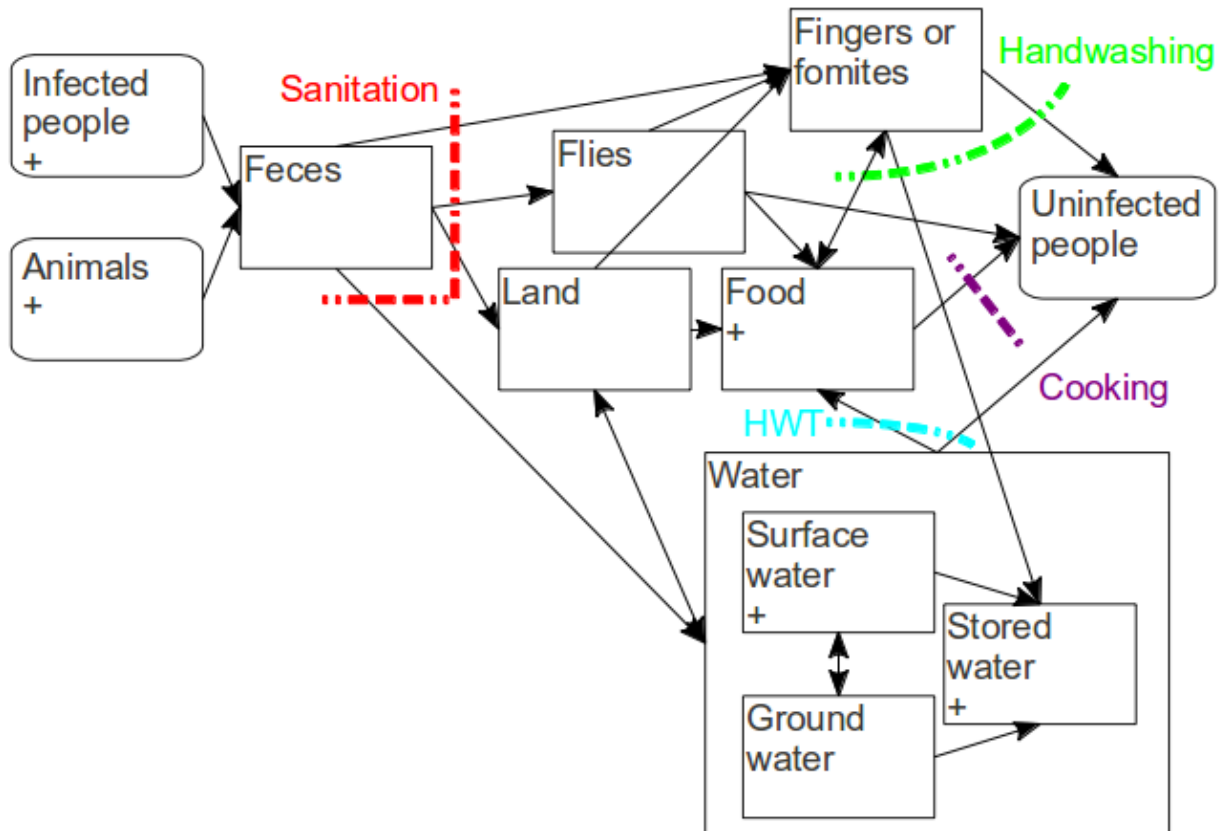
QMRA models are often relatively simple to construct and use (e.g., in an Excel spreadsheet). However, they cannot fully describe secondary transmission of infection, such as when a person who becomes infected from the initial exposure passes the infection to additional people. Therefore, they are particularly good for describing risk from exposures to pathogens where secondary transmission of infection to additional uninfected people is low. QMRA models can describe a static response to a particular dose in a particular population, or they can dynamically describe changes in risk with changing exposure over time. QMRA techniques can also be used as components of infection transmission models incorporating transmission of pathogens through the environment.

2.20. Conceptual model of diarrheal disease transmission

Diarrheal disease is characterized by diverse, simultaneous, interdependent modes of transmission that differ among many different organisms. This makes it a challenging syndrome to model compared to other infectious diseases. Transmission routes for diarrheal infections can be described by using stocks to represent pathogens moving between different locations. A good starting point for this is the 'F-diagram' (V. A. Curtis et al., 2000) shown earlier (Figure 2.4, page 49). The F-diagram can be expanded to more faithfully represent the complexities of transmission and control points for interventions (Figure 2.7). The boxes in each diagram represent stocks of pathogens, and the arrows represent flow of pathogens between them. Stocks marked with a + denote areas where some pathogens can multiply. Colored lines indicate where

interventions remove pathogens, limiting transmission to new hosts. These lines are broken because no intervention can always inactivate all pathogens using any particular route.

Figure 2.7. Expanded diagram of diarrheal pathogen transmission



The symbol '+' denotes places where some pathogens might multiply. HWT: household water treatment. Produced by the author, using V. A. Curtis et al. 2000 as a starting point (see Figure 2.4, page 49).

Figure 2.7 shows that diarrheal infections are transmitted in a variety of complicated ways. An EITS model with so much detail would be extremely difficult to construct and interpret. However, the diagram remains useful for considering which routes to include or omit from the model, and it was used to guide construction of the EITS model described in chapter 5 of this dissertation.

2.21. Theoretical issues regarding interventions

When an intervention is applied to a disease transmission system, it has a direct effect on

the individual receiving it by reducing their risk of disease. In addition, it has an indirect effect on individuals not receiving the intervention by reducing disease transmission within the community. This has been clearly described (Halloran et al., 2002) from the point of view of communities Y and N, in which some community Y residents are immunized, but no community N residents are immunized. Unimmunized people in community Y still benefit indirectly from immunization because reduction of disease spread lowers their chance of contacting a diseased person. The overall difference between the two communities is the community-level effect of the immunization intervention. The difference between the immunized and unimmunized in community Y is the direct effect of immunization, while the difference between the unimmunized in community Y and the unimmunized in community N is the indirect effect. The total effect is the difference between the immunized persons in community Y compared with the unimmunized in community N. The intervention need not be an immunization; interventions where adoption is variable in the community, such as handwashing or latrine use, should operate in a similar way. A comparison of two similar rural Zimbabwean villages, one of which had partial latrine coverage and the other having no latrines, showed decreased diarrhea among households lacking latrines in the village with latrines, compared with the village without latrines (Root, 2001). Systems modeling allows explicit representation of the mechanisms of indirect effects, and allows investigation of how incomplete participation (termed 'compliance' or 'adherence') in an intervention decreases the indirect effects of that intervention.

Interaction of control measures against diarrheal illness is common, but may be positive or negative. A review (Fewtrell & Colford, 2005) considering five studies combining water supply improvements with sanitation improvements and hygiene education found that the effect on childhood diarrhea (32% reduction in diarrhea) was similar in those studies as the effect seen from other interventions alone: water supply (two studies, 33% reduction), sanitation (1 study,

24% reduction), HWT (eight studies, 34% reduction) or hygiene (seven studies, 46% reduction). A similar effect was found when combining water treatment and handwashing in Karachi squatter neighborhoods; there was no additional benefit to combining the interventions, although each intervention alone cut diarrhea prevalence by about 50% (Luby et al., 2006). However, a meta-analysis (Gundry et al., 2004) of seven point-of-use intervention studies found that the effectiveness of point-of-use interventions increased as sanitation improved. Increased water use and latrine possession positively interacted to improve infant weight and length in rural Lesotho (Esrey et al., 1992), and positive interaction has been observed between sanitation and source water quality (VanDerslice & Briscoe, 1995). Sanitation is one of the best studied interventions with regard to interaction with other interventions, and given its key role in removing feces from the environment at the point where they are produced, it seems likely to interact particularly strongly with other interventions. In particular, it seems likely that improved sanitation is necessary to realize further benefits from other interventions. By mechanistically modeling transmission of diarrheal infection and the effects of interventions, it should be possible to predict conditions under which interaction between two interventions is positive or negative.

2.22. Tools and information useful for modeling diarrhea transmission

Intervention trials in communities throughout the world have shown the extent of effectiveness of interventions under favorable conditions, in terms of effectiveness against diarrhea in children under 5 years of age without reference to particular pathogens. Important characteristics of the communities that could impact diarrhea transmission, such as the use of latrines in the village, nutritional status of the population, or the prevalence of breastfeeding, are often not given, which makes it difficult to interpret the results. Nonetheless, the available trials provide indications of the effectiveness of various interventions.

Much useful work has been done within the framework of quantitative microbial risk

assessment. Dose response equations that translate dose of a pathogen into the probability of developing disease have been developed for many diarrheal pathogens (Haas et al., 1999). Distributions have also been developed that describe likely exposure to pathogens on the basis of how they adhere differentially to different surfaces (e.g., hands compared with a cloth), which could theoretically allow estimation of a dose.

Basic infection transmission models like the examples described above (Figures 2.5 and 2.6, page 72) commonly assume that individuals mix evenly, i.e., every person has the same chance of contacting any other person. However, even mixing is unrealistic. People most frequently contact other people within their own household, people mix preferentially within their own age group (Mossong et al., 2008), and the nature of the contact affects its intensity, e.g., contacts at home are more likely to be physical than contacts at work (Mossong et al., 2008). Home, school, workplace, and leisure contacts accounted for over 80% of all contacts in a survey carried out in several European countries (Mossong et al., 2008). Even if such factors are accounted for, people are also likely to have stable (i.e., nonrandom) patterns of connections to other people over time (Keeling & Eames, 2005). Several other studies (Wallinga et al., 2006; Bates et al., 2007; Ogunjimi et al., 2009) of how frequently people contact each other in various ways have been conducted recently, providing information on opportunities for pathogen transmission. However, with the exception of Bates et al. (2007), those studies were carried out in industrialized countries.

Epidemiological data concerning diarrheal pathogens also assists model construction. Factors such as incubation period, period of communicability, immunity, asymptomatic carriage, and persistence of the organism in the environment are important for describing disease transmission. Diarrheal disease risk is also known to change greatly with age, with greatest risk after weaning and decreasing risk thereafter.

Demographic characteristics are also important since susceptibility to diarrhea varies with age and young children have poor hygiene and sanitation habits. Transmission within households is particularly important due to frequent contacts between individuals sharing a household (Mossong et al., 2008). Information about household composition is available for various countries through the Demographic and Health Surveys (DHSs) (USAID, 2012).

Many meta-analyses have summarized the effects of drinking water interventions, hygiene interventions, and nutritional interventions using relative risks of disease comparing groups with an intervention to groups without an intervention (Gundry et al., 2004; B. F. Arnold & Colford, 2007; Waddington et al., 2009; Ejemot et al., 2008; Lazzerini & Ronfani, 2008; Hunter, 2009). However, these meta-analyses and the studies that they summarize seldom report useful information for mechanistic modeling, such as pathogen concentrations in drinking water or the interventions already used in communities before the studies began.

2.23. Published models relevant to endemic diarrhea transmission

Although transmission models are widely used in infectious disease epidemiology, few consider diarrhea in humans. Several that model diarrhea directly or provide useful insights to modeling infectious diarrhea and its prevention are discussed below.

2.23.1. Mechanistic model of diarrheal infection: Eisenberg et al., 2007

A model of diarrheal infection transmission within a hypothetical community of a single pathogen with complete immunity following infection, considered five transmission routes: within-household, between-household, household-to-water, water-to-household, and introduction of pathogens from outside the community (J. N. S. Eisenberg et al., 2007). This allowed interdependencies between transmission routes; for example, reduction of transmission between water and households would also secondarily reduce within-household and between-household transmission. Parameters governing the strength of the transmission routes were varied, and the

results of a hypothesized water treatment intervention that eliminated all risk from contaminated drinking water subsequently varied, giving the following results from the model:

1. Low water contamination levels resulted in low preventable fractions of diarrhea from water treatment.
2. If between-household transmission was low, the preventable fraction increased as within-household transmission increased, since that route created secondary cases within households, and these secondary cases could then be prevented by water treatment.
3. If within-household transmission was low and between-household transmission was increased, the preventable fraction increased at first but decreased again at high levels of between-household transmission (which created an important alternate route for disease transmission aside from contaminated water). If within-household transmission was then increased (i.e., both within- and between-household transmission were high), the preventable fraction decreased further. In this case transmission could be sustained by contacts between people, and water treatment therefore prevented little diarrhea.

Analysis of this model showed that the effectiveness of an intervention could vary drastically depending on transmission characteristics within a community. Since additional interventions could preferentially alter different transmission routes, it also partially explains how interactions between various interventions might arise.

2.23.2. Empirical model of diarrheal disease: Schmidt et al., 2009

Schmidt et al. (2009) constructed a model describing diarrheal illness in a population by drawing the number of illness episodes from a distribution for each individual and then assigning an onset day and a duration to each episode. Individuals had a subject-specific error assigned to them to simulate differential susceptibility by individuals. Assignment of episodes was partly determined by increasing risk during the days following a previous episode, to simulate

autocorrelation (‘clumping together’) of disease episodes. Individuals in the model were independent, and transmission between individuals was not modeled. The model’s primary intended use was to generate simulated datasets to assess different disease surveillance schemes and explore effects of likely sources of error in surveillance data. Although it is an empirical model that does not mechanistically simulate disease transmission, the authors provide gamma distribution parameters drawn from real datasets for episode duration and number of episodes per year.

An important issue raised by Schmidt et al. (2009) is the likely existence of autocorrelation of disease episodes within individuals. Some datasets show evidence for increased risk of an episode during the few weeks following a previous episode. This is reasonable given the fact that it may take several weeks for the full nutrient absorptive capacity of the small intestine to regenerate after an episode of acute diarrhea (Chen et al., 1983). In addition, people living in unhygienic environments often have chronic alterations (lasting for years) of the small intestine (termed tropical sprue, tropical malabsorption, or tropical enteropathy) which is associated with diarrhea and similar intestinal abnormalities that lead to malabsorption of nutrients; it is thought to have a root bacterial cause, partially because it can be treated with antibiotics (Lunn, 2000; Blaser et al., 2002). Autocorrelation of diarrheal episodes could also be explained by the existence of intermittent or relapsing infections (Schmidt et al., 2009).

2.23.3. Modeling indirect effects of interventions: Halloran et al., 2002

Modeling non-diarrheal infections can still provide important insights into diarrhea prevention. A detailed stochastic model of influenza transmission by contact between individuals in simulated communities has been described (Halloran et al., 2002). Contacts between individuals within their communities depended on neighborhood and level of school (and therefore also age). To simulate an intervention, the proportion of individuals vaccinated was

then increased in some communities. The model was used to investigate the direct and indirect effects of increased immunization, finding large indirect benefits to unvaccinated individuals. For example, if half of the population in the intervention community was given a vaccine that was 70% effective, the unvaccinated population received benefit equivalent to directly receiving a vaccine that was 40% effective. Indirect benefits to unvaccinated people increased as the proportion of vaccinated people increased. This scenario is analogous to many other interventions, in which a particular behavior (e.g., handwashing) is already practiced in a community at a certain level, and is increased by means of an intervention. However, it is unclear whether adoption of interventions preventing diarrhea by some individuals would similarly protect noncompliant individuals indirectly.

2.23.4. Environmental infection transmission system (EITS) models: Li et al., 2009

Much infectious disease modeling work has examined person-to-person transmission, but diarrheal illness is largely mediated by the environment (Figure 2.4, page 49; Figure 2.7, page 77). An SIR model of influenza transmission with an added environmental (E) compartment, modeled as a system of differential equations, has been described (Li et al., 2009); a simplified version of this model is shown in Figure 2.6, page 74. The model used susceptible, infected, and removed (i.e., immune or dead) compartments to track hosts, while an environment compartment contained pathogens shed into the environment by infected hosts. Hosts can then pick up pathogens from E and become infected; pathogens in E are also inactivated at a particular rate. The model used three different parameter sets to simulate differing transmission modalities: frequently touched fomites, infrequently touched fomites, and airborne transmission. However, the R_0 s from the three parameter sets were identical; this allowed comparison of the intervention effectiveness under differing infection transmission conditions. Two interventions were

considered: 1) decontamination, increasing the inactivation rate of pathogens in the environment by 25%; or 2) an 'avoidance measure' (perhaps analogous to improved hygiene) decreasing the pick-up rate of pathogens from the environment by 25%. When transmission occurred via frequently touched fomites, neither intervention was effective. However, if transmission was airborne or via infrequently touched fomites, each intervention reduced the incidence of infection by about a third, with decontamination being more effective than 'avoidance'. Thus the effectiveness of an intervention depends on the nature of the transmission route it affects, even if transmission is identical pre-intervention.

2.24. Conclusion

Diarrheal disease in developing countries is characterized by diverse pathogens, multiple routes of transmission, and numerous interventions to prevent it. A clearer understanding of the action of interventions on the transmission of diarrhea would be helpful for planning and implementing diarrhea control programs. Modeling techniques are useful for synthesizing the large amount of published information regarding transmission and control of diarrhea, and they can yield further insight about diarrhea transmission and control. Chapters 3, 4, and 5 of this dissertation describe the development and application of such models.

3. LINKING A QUANTITATIVE MICROBIAL RISK ASSESSMENT MODEL TO A HOUSEHOLD WATER TREATMENT FIELD TRIAL

This chapter consists of previously peer-reviewed and published content (including supplementary material) that has been reformatted and reorganized:

Enger, K.S., Nelson, K.L., Clasen, T., Rose, J.B., Eisenberg, J.N.S., 2012. Linking quantitative microbial risk assessment and epidemiological data: informing safe drinking water trials in developing countries. *Environmental Science and Technology* 46, 5160–5167.

3.1. Abstract

Intervention trials are used extensively to assess household water treatment (HWT) device efficacy against diarrheal disease in developing countries. Using these data in policy, however, requires addressing issues of generalizability (relevance of one trial in other contexts) and systematic bias associated with design and conduct of a study. A published randomized controlled trial (RCT) of the LifeStraw® Family Filter in the Congo was used as the basis for a quantitative microbial risk assessment (QMRA) model, to demonstrate the application of models to water safety and health issues. The QMRA model accounted for bias due to 1) incomplete compliance with filtration, 2) unexpected antimicrobial activity by the placebo device, and 3) incomplete recall of diarrheal disease. Effectiveness was measured using the longitudinal prevalence ratio (LPR) of reported diarrhea. The Lifestraw RCT observed an LPR of 0.84 (95% CI: 0.61, 1.14). The model predicted LPRs, assuming a perfect placebo, ranging from 0.50 (2.5-97.5 percentile: 0.33, 0.77) to 0.86 (2.5-97.5 percentile: 0.68, 1.09) for high (but not perfect) and low (but not zero) compliance, respectively. The calibration step provided estimates of the concentrations of three pathogen types (modeled as pathogenic *E. coli*, *Giardia*, and rotavirus) in drinking water consistent with the longitudinal prevalence of reported diarrhea measured in the trial constrained by epidemiological data from the trial. The QMRA model demonstrated the

importance of compliance in HWT efficacy, the need for pathogen data from source waters, the effect of quantifying biases associated with epidemiological data, and the usefulness of generalizing the effectiveness of an HWT trial to other contexts.

3.2. Introduction

The randomized controlled trial (RCT) is considered the gold standard study design in epidemiology; it is the study design with the least systematic bias, and therefore the highest internal validity. Two important components of RCT design for internal validity are: the randomization of subjects to the intervention and the non-intervention groups; and blinding of the subject and investigator to group assignment. It is difficult to blind HWT interventions because these devices are visually obvious and cannot be concealed from participants or investigators. It is also difficult to develop a placebo HWT filter that does not remove pathogens, but improves the appearance of water like an effective filter (Boisson et al., 2010). Other biases may also affect the internal validity of an estimate derived from the trial, such as recall bias, incomplete compliance with the intervention, or unexpected difficulties conducting the trial (Boisson et al., 2010).

In a recent RCT (Boisson et al., 2010) in rural communities in the Democratic Republic of the Congo (DRC) using the LifeStraw® Family Filter (LFF; Vestergaard Frandsen Corporation, Lausanne, Switzerland), investigators attempted to blind the intervention. The LFF is an ultrafilter with a 20 nm pore size that was shown to remove 99.99999% of *Escherichia coli*, 99.998% of MS2 coliphage, and 99.97% of *Cryptosporidium* oocysts from challenge water in the laboratory (T. Clasen, Naranjo, et al., 2009). For the Lifestraw RCT, investigators developed a placebo filter resembling the LFF in appearance, weight, operation, and flow rate (Boisson et al., 2010). The placebo was tested in the laboratory for three weeks against the same three organisms, and no removal was observed. In the field, however, the intended placebo removed

on average 91% (95% CI: 88-93%) of thermotolerant coliform bacteria (TTC), a group that includes *E. coli* and indicates fecal contamination, from source water (Boisson et al., 2010). Therefore, the study could only compare a highly effective filter with a poorly effective filter. Although 65% of people reported using the filter, most filter users also reported drinking unfiltered water (Boisson et al., 2010). The proportion of unfiltered water that people consumed was not quantified. The Lifestraw RCT did not find a statistically significant ($P < 0.05$) effect of the LFF against diarrhea (Boisson et al., 2010).

Quantitative microbial risk assessment (QMRA) models can examine and account for biases associated with environmental intervention trials (e.g., imperfect compliance, recall bias, or an imperfect placebo) and can explore risks associated with contexts different than those observed in empirical studies. Such models can provide a conceptual framework for understanding systems that are difficult to explore in the real world. QMRA models have been used to quantify disease risk in many contexts (J. N. S. Eisenberg et al., 2008; Haas et al., 1999; Parkin, 2008). The analytic framework for linking QMRA and epidemiological data described here consists of: 1) a calibration step using a QMRA model to produce results consistent with the epidemiological study; and 2) an estimation step that examines counterfactual scenarios that adjust for biases within the study and explores how altered contexts affect risk. The effectiveness of an intervention in those contexts can then be estimated, even if it was never directly studied under such conditions.

This chapter describes a counterfactual causal inference framework using a QMRA model to evaluate the impact of biases on estimates of intervention efficacies. This is illustrated by simulating the Lifestraw RCT (Boisson et al., 2010) and adjusting for some of its biases, to estimate the effectiveness of the LFF compared with a perfect placebo under differing levels of LFF compliance.

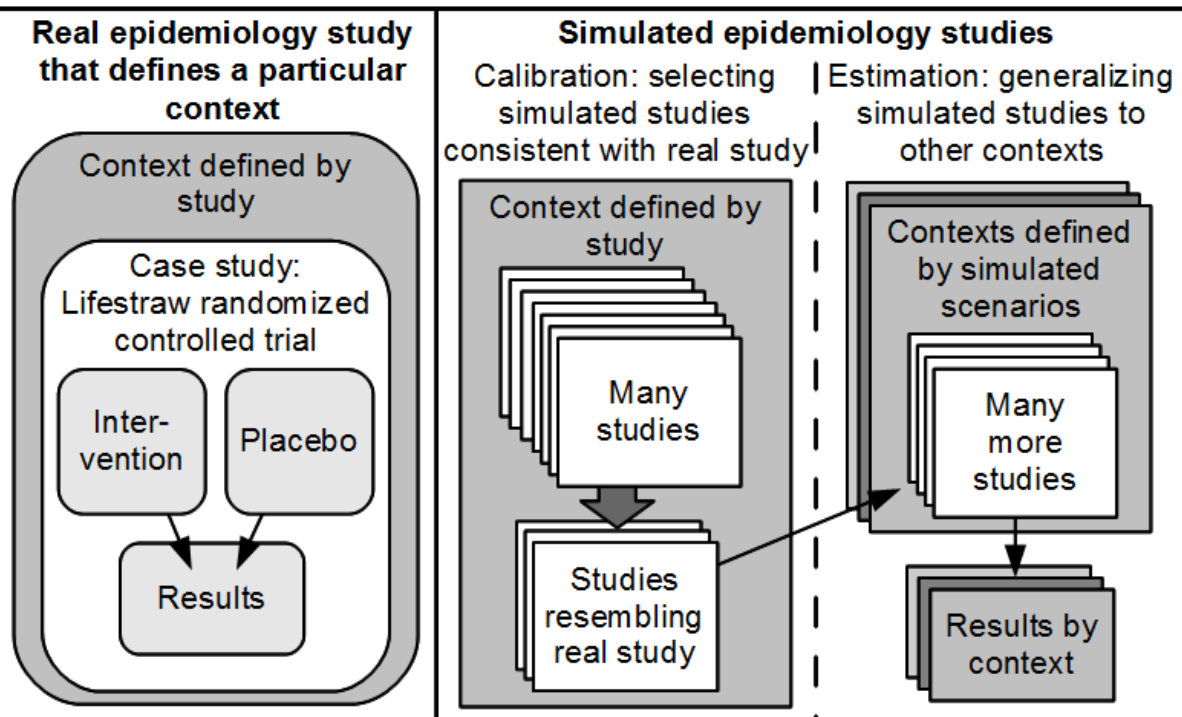
3.3. Materials and Methods

3.3.1. Conceptual framework linking QMRA models to epidemiological studies

Quantitative microbial risk assessment (QMRA) uses environmental contamination data as inputs to models used to predict risk of infection and/or disease. Epidemiological studies provide data on patterns of disease measured by incidence or prevalence, and measures of relative risk. This chapter illustrates a framework for the calibration of risk models by using epidemiological data from a particular study that describes the risk in a particular context, where the context is defined by a particular time in a particular geographic setting (Figure 3.1). The calibration process involved simulating a risk model many times using different input and parameter values. The parameter sets (or parameter distributions) representing those simulations that were consistent with the epidemiological study comprised the calibrated model; the parameter distributions represented the context in which the epidemiological study was conducted. Using this calibrated model, the epidemiological study was generalized to other contexts in the estimation step. The estimation step consisted of a set of simulations in which specific parameter values were varied to describe different contexts, such as alternative intervention strategies or different ecological or social settings.

Figure 3.1. Conceptual model for simulation of a randomized controlled trial

Linking QMRA models to epidemiological studies: a conceptual model



The results from an actual epidemiology study defined a context (e.g., the LifeStraw® Family Filter randomized controlled trial [Lifestraw RCT] in rural Congo). The calibration phase provided a set of simulated studies that are consistent with this defined context. Calibration also estimated values for parameters that were not observed during the real study, thus inferring unobserved context of the real study. The estimation phase provided simulated studies that were generalized to other contexts (e.g., higher or lower compliance than was observed during the Lifestraw RCT).

3.3.2. Model description

For the research described in this chapter, a QMRA model was developed that simulates the following chain of events:

1. Determination of the concentrations of three pathogen types (bacteria, protozoa, and viruses) in drinking water, sampled from gamma distributions
2. Calculation of daily doses of pathogens based on their concentrations and the amount of water consumed
3. Use of dose response functions: convert daily pathogen doses to probabilities of infection

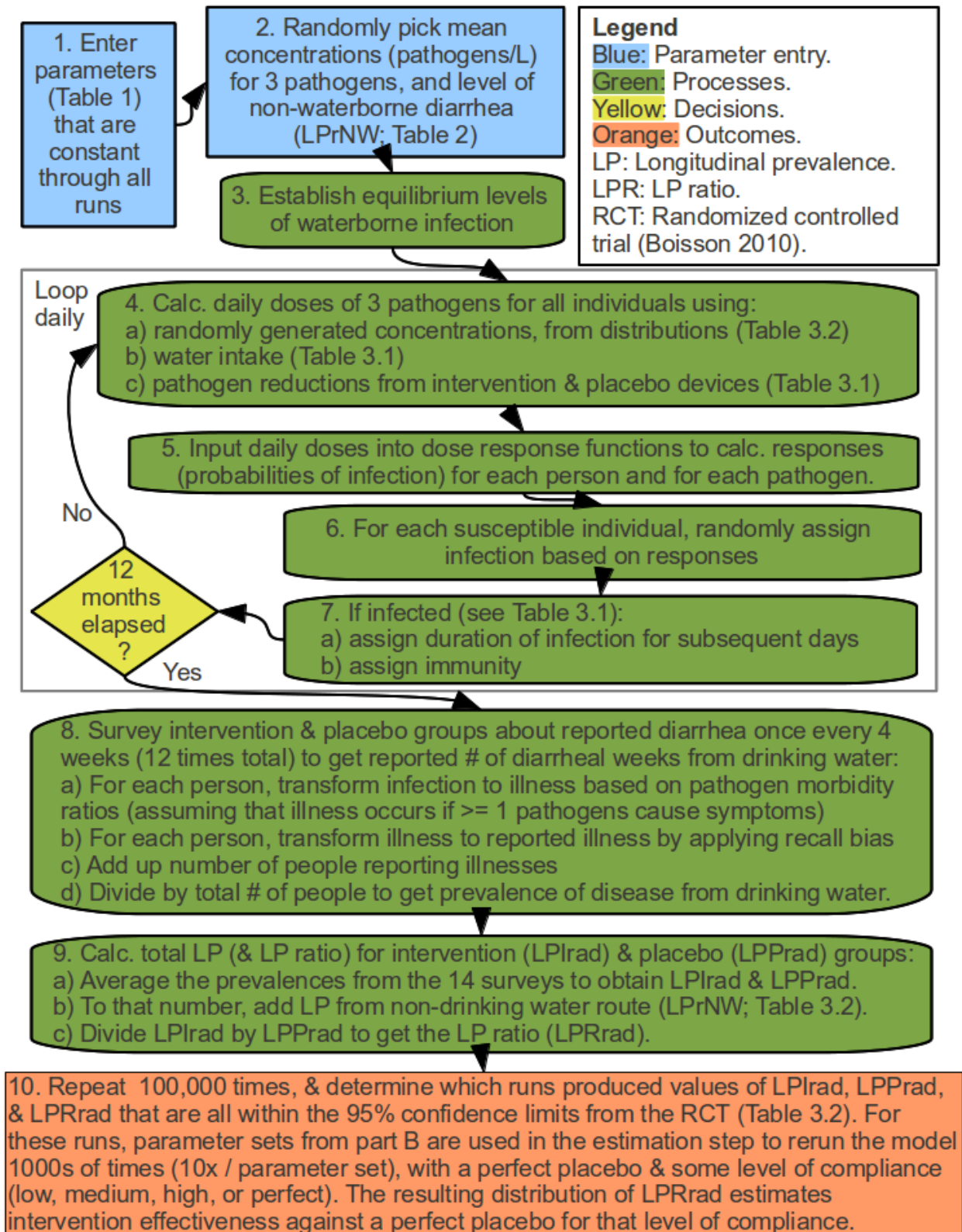
4. Assignment of infection to individuals, based on the probability of infection
5. Assignment of diarrheal illness, based on morbidity ratios

The same conceptual approach illustrated in Figure 3.1 could also be applied to more complex models including processes such as transmission dynamics (J. N. S. Eisenberg et al., 2005; J. N. S. Eisenberg et al., 2007) or environmental fate and transport dynamics (J. N. S. Eisenberg et al., 2006).

The model describing the Lifestraw RCT conducted in the Congo (Boisson et al., 2010) follows a simulated population of children under five years of age for 12 months using a time unit of 1 day. The population was surveyed about their diarrheal symptoms every four weeks, similar to the Lifestraw RCT. The simulated children ingested bacteria, protozoa, and viruses in their drinking water, respectively represented by diarrheagenic *Escherichia coli*, *Giardia* cysts, and rotavirus. These three pathogens were chosen because they are major causes of diarrheal disease in much of the developing world, and they represent the three main taxa of waterborne pathogens. (Lanata & W. Mendoza, 2002) A child was either susceptible to, immune to, or infected by each of these three pathogens; we assumed that the infective processes of each pathogen were independent of each other, and a child could therefore be infected with 0, 1, 2, or 3 types of pathogens simultaneously. Children were divided into two groups; the first group received the intervention filter, whose \log_{10} removal values (LRVs) for *E. coli*, *Giardia*, and rotavirus were 6.9, 3.6, and 4.7 respectively based on laboratory testing (T. Clasen, Naranjo, et al., 2009). The other group received the placebo filter, whose LRVs were set to 1.05 for all three pathogens based on the field trial for thermotolerant coliform removal (Boisson et al., 2010).

The model is more completely described by the flowchart in Figure 3.2; the steps of the flowchart are explained in detail below.

Figure 3.2. Simulation model flowchart



Step 1: Parameter entry

A simulation run begins by reading 28 model parameters (Table 3.1), which were estimated from the published scientific literature and are discussed in greater detail in chapter 7. They remain constant for every simulation run.

Table 3.1. Fixed parameter values used in the QMRA model of the Lifestraw RCT

Description of parameter values	Value	Reference
Morbidity ratios (proportion of infected who are symptomatic)		
<i>Escherichia coli</i>	0.214	(Vergara et al., 1996)
<i>Giardia</i>	0.590	(Pérez Córdón et al., 2008)
Rotavirus	0.397	(Fischer et al., 2002)
Duration of infection		
<i>Escherichia coli</i> (gamma distribution, mean 3 days)	shape = 1.775 scale = 1.690	(Estrada-Garcia et al., 2009)
<i>Giardia</i> (gamma distribution, mean 11 days)	shape = 3.206 scale = 3.431	(Kent et al., 1988)
Rotavirus (uniform distribution; mean 2.5 days)	Range 1-4 days	(Kapikian et al., 1983)
Shape parameter for all gamma distributions of pathogen type concentrations ^a	1.85	(Boisson et al., 2010)
Period of immunity for all pathogens	7 days	
No. children under 5 years of age, intervention group	85	(Boisson et al., 2010)
No. children under 5 years of age placebo group	105	(Boisson et al., 2010)
Longitudinal prevalence of reported diarrhea for each group		
Intervention (LP_{Irad})	0.0749	(Boisson et al., 2010)
Placebo (LP_{Prad})	0.0896	(Boisson et al., 2010)
Longitudinal prevalence ratio of reported diarrhea (LPR_{rad})	0.836	(Boisson et al., 2010)
Water ingestion	1.178 L/day	(Akpatá, 2004)
Log ₁₀ reduction values (LRVs), intervention group		
<i>Escherichia coli</i>	6.9	(T. Clasen, Naranjo, et al., 2009)
<i>Giardia</i>	3.6	(T. Clasen, Naranjo, et al., 2009)

Table 3.1 (cont'd)

Description of parameter values	Value	Reference
Rotavirus	4.7	(T. Clasen, Naranjo, et al., 2009)
LRVs, placebo group, all 3 pathogens		
Calibration step	1.05	(Boisson et al., 2010)
Estimation step	0	
Dose response function parameters		(Anon, 2012)
<i>E. coli</i> (enteroinvasive); beta-Poisson parameters	$\alpha = 0.155$ $N_{50} = 2.11 \times 10^6$	(H L DuPont et al., 1971)
<i>Giardia</i> ; exponential k parameter	0.0198	(Rendtorff, 1954; J B Rose et al., 1991)
Rotavirus; beta-Poisson parameters	$\alpha = 0.2531$ $N_{50} = 6.171$	(Haas et al., 1993; Ward et al., 1986)
Chance of remembering diarrhea >2 days in the past	0.54	(Zafar et al., 2010)
Compliance with device use: chance of using device on a given day		
Calibration step	0.65	(Boisson et al., 2010)
Estimation step	0, 0.65, or 1.00	
Compliance with device use: If using device on a given day, proportion of water treated		
Calibration step	2/3 or 1/3	

^a The scale parameters for the gamma distributions of pathogen types are determined by the mean concentration of pathogen types, which is randomly sampled from a uniform distribution during calibration (Table 3.2).

Step 2: Parameter values inferred through calibration

There are 4 other parameters (Table 3.2) which are unknown and were therefore inferred during the calibration process:

1. Baseline longitudinal prevalence of reported non-waterborne diarrheal disease in 'children' (under five years of age; LP_{rNW}). This value was assumed to be the background LP of diarrhea due to the sum of all transmission pathways except drinking water, as well as all non-communicable causes of diarrhea;
2. Mean concentration of diarrheagenic *Escherichia coli* (representing bacteria) in untreated

drinking water (bacteria/L);

3. Mean concentration of *Giardia* (representing protozoa) in untreated drinking water (cysts/L);
4. Mean concentration of rotavirus (representing viruses) in untreated drinking water (virions/L).

These values are necessary for the model, but no pathogen data exist for these water sources, nor are there clinical data describing the etiology of diarrheal disease where the Lifestraw RCT was conducted. The drinking water sources used in the Lifestraw RCT study area were unimproved, and consisted mainly of surface water and unprotected springs; however, the water was abundant and naturally clear (Boisson et al., 2010). Diarrheagenic *E. coli*, *Giardia*, and rotavirus were chosen to represent bacteria, protozoa, and viruses because they are common causes of diarrheal disease in developing countries (Lanata & W. Mendoza, 2002), and they represent the three major taxa of diarrheal pathogens. The prior distributions for the mean concentrations of these pathogen types were defined as uniform distributions, with ranges given in Table 3.2. Posterior distributions for these values were obtained by running multiple simulations with varying mean concentrations. Model runs returned three key outcomes: the longitudinal prevalence of reported diarrhea for the intervention ($LP_{I_{rwd}}$) and placebo ($LP_{P_{rwd}}$) groups, and their ratio, the longitudinal prevalence ratio (LPR_{rad}). These three outcomes were also estimated by the Lifestraw RCT (Table 3.3). If the outcomes from a model run fell within the 95% confidence intervals for the outcomes estimated by the Lifestraw RCT, the mean pathogen type concentrations and the LP_{rNW} were retained for use in the estimation step.

Table 3.2. Ranges for stochastically varying parameters

Uniform distributions used to determine the values of the stochastically varying parameters for each simulation model run during calibration

Description	Lower limit	Upper limit (low calibration compliance)	Upper limit (medium calibration compliance)
Mean concentration per L, pathogenic <i>E. coli</i> in untreated drinking water	0	7.0×10^4	8.0×10^4
Mean concentration per L, <i>Giardia</i> cysts in untreated drinking water	0	0.95	1.3
Mean concentration per L, rotavirus in untreated drinking water	0	0.14	0.18
Baseline non-waterborne diarrhea longitudinal prevalence (LP_{rNW})*	0	0.0972	0.0972

* The upper limit for baseline diarrhea longitudinal prevalence is the upper limit of the 95% CI for LP in the <5 year old intervention group in the Lifestraw RCT (Boisson et al., 2010).

Step 3: Initiating each run; establishing equilibrium waterborne infection

Once all parameters were available, the simulation run could begin. It was assumed that at the beginning of a simulated intervention study, each participant was in one of three states (susceptible, infected, or immune), and that the proportions of participants among these states were in equilibrium. In the model, this equilibrium was established by initially infecting every child with all three pathogen types and assigning infection durations randomly from a uniform distribution ranging from 0 to 50 days ('equilibration infections'). This range was chosen in order to prevent extreme oscillations in the proportion infected. In earlier versions of the model, these oscillations arose because the entire population was initially susceptible, leading to immediate infection of much of the population, followed by simultaneous recovery, followed by simultaneous infection. By using a wide range of infection durations for the equilibration infections, children become susceptible at different times during the equilibration period, thereby preventing large oscillation artifacts. After children recovered from an equilibration infection,

they were immune for 7 days, after which they could be reinfected through exposure to contaminated drinking water; infection durations were subsequently assigned from the distributions described in Table 3.1. Equilibrium was reached when the temporal trend of the infection prevalence was flat; the infection prevalence at equilibrium stochastically oscillated around the mean prevalence. The model reached equilibrium in approximately 60 days (e.g., Figure 3.4 and Figure 3.5); the precise time depended on the mean concentrations of pathogens. The simulated intervention study begins on day 128.

Step 4: Calculating daily doses of marker pathogens

The model included 85 children in the intervention group and 105 in the placebo group, consistent with the Lifestraw RCT (Boisson et al., 2010). The concentration of each pathogen type in the source water was sampled from a gamma distribution (Table 3.1) for each child on each day. The shape parameter of these distributions was always 1.85; the scale parameter was the randomly selected mean concentration for a particular pathogen, divided by the shape parameter. The shape parameter was obtained by fitting a gamma distribution to the thermotolerant coliform (TTC) counts measured by the Lifestraw RCT in untreated source water, excluding high outliers (over the detection limit of 30,000 CFU / 100 mL; 3.8% of the data). Each child’s daily dose of a particular pathogen type was determined using the LRVs (Table 3.1) attributable to the device that child is using:

$$\text{Daily dose} = cd[(1 - w) + w10^{-r}] \tag{3.1}$$

where c is the concentration per liter of a pathogen type in untreated water (sampled from a gamma distribution), w is the proportion of water treated (which varies depending on compliance), r is the LRV (which varies depending on whether the intervention filter, the placebo filter, or no filter is being used), and d is the liters of water consumed daily.

Step 5: Dose response functions

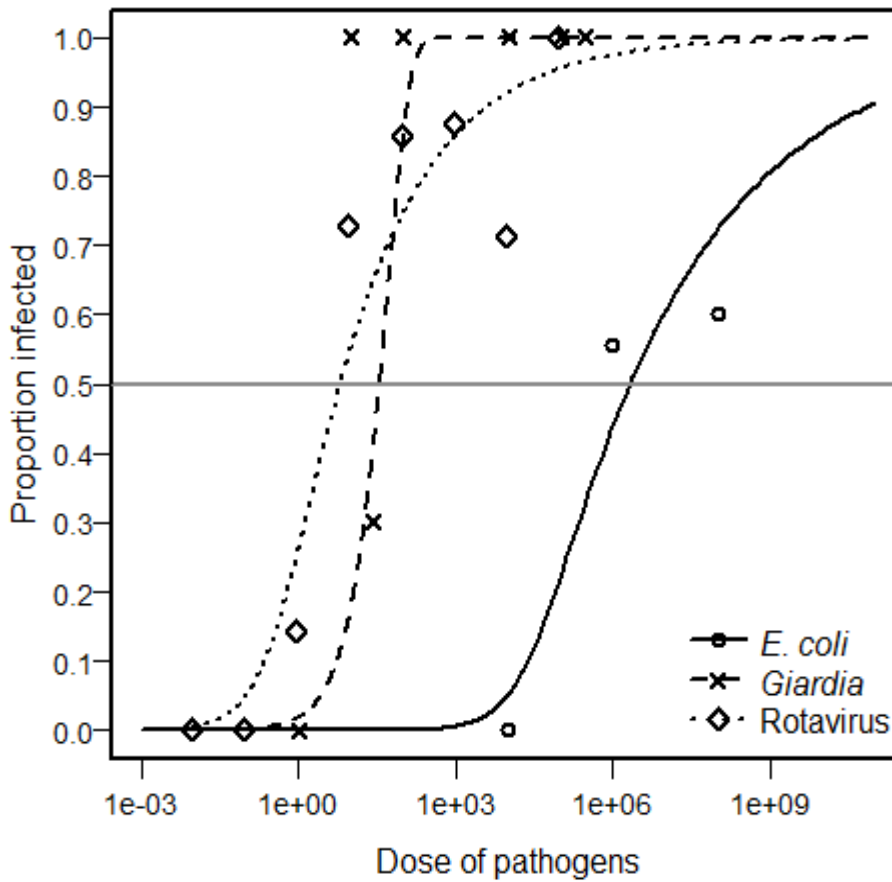
The daily doses of pathogen types were converted to responses (i.e., daily probabilities) per susceptible person of becoming infected using dose response functions (Anon, 2012; Haas et al., 1999). These functions were obtained using results from studies in which adult volunteers were fed widely varying doses of pathogens, and monitored for development of infection. An exponential dose response function was used for *Giardia*, and a beta-Poisson dose response function was used for *E. coli* and rotavirus:

$$\text{Exponential:} \quad \text{Response} = 1 - e^{-kd} \quad (3.2)$$

$$\text{Beta-Poisson:} \quad \text{Response} = 1 - (1 + [d / N_{50}][2^{1/\alpha} - 1])^{-\alpha} \quad (3.3)$$

where d is the dose (number of pathogens ingested per day in drinking water), k is the parameter for the exponential model, and α and N_{50} are the parameters for the beta-Poisson model (N_{50} is the dose at which 50% of the population exhibits the response). As α approaches infinity, the beta-Poisson dose response function approaches the exponential dose response function (Haas et al., 1999). A graph of the dose response functions is found in Figure 3.3 (Figure 4.6a, page 138, presents the same information in log-log scale).

Figure 3.3. Comparison of dose response functions



Step 6: Assignment of infection

New infections were randomly assigned to susceptible children each day, according to the response probabilities obtained from the dose response functions.

Step 7: Assignment of infection duration, recovery, and immunity

If a child is infected, they were assigned an infection duration (Table 3.1, page 93) sampled from a gamma distribution (*E. coli* or *Giardia*) or a uniform distribution (rotavirus). Infections due to the 3 pathogens were tracked independently within each child. Functionally, this was done using a matrix with 1 row per child and 1 column per pathogen. The entries of the matrix were numbers of days; positive numbers denoted time remaining until recovery and negative numbers denoted elapsed time since recovery. Each day, 1 was subtracted from all entries, and an

individual recovered when an entry reaches 0. Following recovery, the individual was immune for 7 days; after this, they are once again susceptible and may be reinfected.

The children in the placebo group were identical to the children in the intervention group, except that the \log_{10} reductions attributable to the placebo device were lower (Table 3.1, page 93), and they therefore ingested higher doses of pathogens.

Step 8: Surveying the population about reported diarrhea

Although the model explicitly tracked infection, the Lifestraw RCT measured reported disease. Since some infections are asymptomatic or unreported, the model had to simulate the process of people reporting diarrhea in order to output disease measures comparable to the Lifestraw RCT.

Reporting of diarrhea was simulated via 12 monthly surveys of the population. The mean of the 12 monthly estimates of prevalence during the year-long study period estimated the longitudinal prevalence (LP) of diarrhea. More generally, an LP is the proportion obtained by dividing the person-time affected by the total person-time observed. Each simulated survey estimated the prevalence of diarrhea by allowing each person to report whether any diarrheal illness occurred over the previous 7 days, corresponding to the actual survey process during the Lifestraw RCT. Reporting of diarrhea by people in the simulated community followed these rules:

1. the most recent infection for a particular child was determined, i.e., the infection with the longest duration remaining, or if not currently infected, the infection that resolved most recently;
2. that infection was stochastically determined to be symptomatic or asymptomatic, using the morbidity ratio (Table 3.1, page 93) as the probability of symptoms given infection;

3. asymptomatic infections were never reported;
4. symptomatic illness on that day or the previous 2 days was always reported;
5. symptomatic illness during the previous 3-7 days had a 54% chance of being remembered (Table 3.1, page 93); if the illness was remembered, it was reported.

The 5th rule came from published recall bias measurements (Zafar et al., 2010). The prevalence of diarrhea reported by a particular survey was the proportion of people reporting diarrhea according to the rules above. Averaging the results from the 12 simulated surveys gave an estimate of the LP.

Step 9. Determining model outcomes corresponding to the Lifestraw RCT

To clarify this process, some terminology is described here. There were several different longitudinal prevalences (LPs) which were tracked or output by the model. LPs were determined (and subscripted) according to 2 categories: 1) intervention group (I) or placebo group (P); and 2) waterborne infection (_{wi}), or waterborne diarrhea (_{wd}), or any diarrhea (_{ad}). In addition, reported diarrhea was prefixed with _r. For example, the longitudinal prevalence in the intervention group of any reported diarrhea was LP_{Irad} . Another parameter, LP_{rNW} , was defined as the longitudinal prevalence of reported diarrhea acquired by nonwaterborne routes. LP_{rNW} was not affected by water treatment. Total longitudinal prevalence of reported diarrhea for the intervention group (LP_{Irad}) or the placebo group (LP_{Prad}) was obtained by adding LP_{rNW} to LP_{Irad} or LP_{Prad} . Since LPs are proportions and a person might simultaneously carry infection acquired from waterborne or non-waterborne routes, addition was carried out in this manner to avoid double-counting of the intersection between the two types of routes:

$$LP_{Irad} = LP_{Irdw} + LP_{rNW}(1 - LP_{Irdw}) \quad \text{or} \quad LP_{Prad} = LP_{Prdw} + LP_{rNW}(1 - LP_{Prdw}) \quad (3.4)$$

Table 3.1, continued

The effectiveness of the LFF was measured by a longitudinal prevalence ratio (LPR). To correspond with the Lifestraw RCT, the primary outcome measure was the LPR of any reported diarrhea (LPR_{rad}), calculated as follows:

$$LPR_{rad} = LP_{Irad} / LP_{Prad} \quad (3.5)$$

Analogous LPs and LPRs could also be calculated for waterborne infection, waterborne diarrhea, or reported waterborne diarrhea. For example, since the model tracked waterborne infection daily, the LPs for waterborne infection in the intervention (LP_{Iwi}) and placebo (LP_{Pwi}) groups could be used to generate LPRs for waterborne infection (LPR_{wi}), as if the entire population was observed with perfect accuracy (Figure 3.10, page 113).

Step 10: Repetition of calibration runs

During the calibration step, the simulation model was run 100,000 times. Each run randomly selected different pathogen concentrations and a background diarrheal longitudinal prevalence (LP_{rNW}) from a uniform distribution (Table 3.2, page 96). The upper limits of these uniform distributions were obtained from a simplified calibration process carried out before the actual calibration step. This process used the idea that the maximum possible concentration of a particular pathogen type is the concentration that yields the maximum LP_{Prad} consistent with the Lifestraw RCT if other two pathogen types are absent. For a particular pathogen type, these values were estimated by examining results from 12 successive model runs. Each of the 12 runs progressively increased the concentration of a single pathogen type, with concentrations of the other 2 pathogens set to 0. The smallest concentration yielding an LP_{Prad} above the 95% confidence limit for the placebo group (i.e., an $LP_{Prad} > 0.11$; Table 3.3) provided the upper limit of the uniform distribution for that pathogen type's concentration (Table 3.2, page 96). The

appropriateness of these upper limits were checked by visually examining scatterplots of pathogen concentration by LP_{Prad} after the full calibration step had completed (not shown).

Parameter values and the results from all runs in the calibration step were saved. The time course of infection and reported diarrhea for two example simulation runs are in Figures 3.4 and 3.5 (page 104). Parameter combinations are selected that yield LP_{Irad} , LP_{Prad} , and LPR_{rad} values within the 95% confidence intervals reported in the RCT trial (Table 3.3). These parameter combinations are then reused in the estimation step of the process.

Among the three pathogen types, high concentrations of one pathogen were associated with lower concentrations of the other pathogens. This occurred because the model was calibrated to match the level of reported disease seen in the Lifestraw RCT, without reference to the particular pathogen types. Therefore higher levels of disease from one pathogen type must be balanced by lower levels of disease from other pathogen types.

Table 3.3. Longitudinal prevalence measures from the Lifestraw field trial

Measure	Estimate	Lower limit (95% CI)	Upper limit (95% CI)
Longitudinal prevalence, intervention group (LP_{Irad})	0.0749	0.0526	0.0972
Longitudinal prevalence, placebo group (LP_{Prad})	0.0896	0.0673	0.112
Longitudinal prevalence ratio, intervention/placebo (LPR_{rad})	0.84	0.61	1.14

* During calibration, for a simulation model run to be considered consistent with the Lifestraw RCT, all 3 measures must fall between the lower and upper limits.

3.3.3. Example runs of the model

Graphical output from two runs of the model is shown in Figure 3.4 and Figure 3.5. They display the time courses of two calibration runs of the QMRA model. Figure 3.4 has higher

waterborne infection and reported waterborne disease levels than observed in the Lifestraw RCT, and Figure 3.5 is consistent with the Lifestraw RCT. The simulated surveys of diarrhea are shown by the purple × symbols; equilibration occurred during the first 128 days, before the first simulated survey. Each survey asked whether any diarrhea was remembered during the previous 7 days. In contrast, the lines signify daily prevalence of infection, as simulated daily over the course of the model run.

Figure 3.4. Example run of the model, with higher infection levels than the Lifestraw RCT

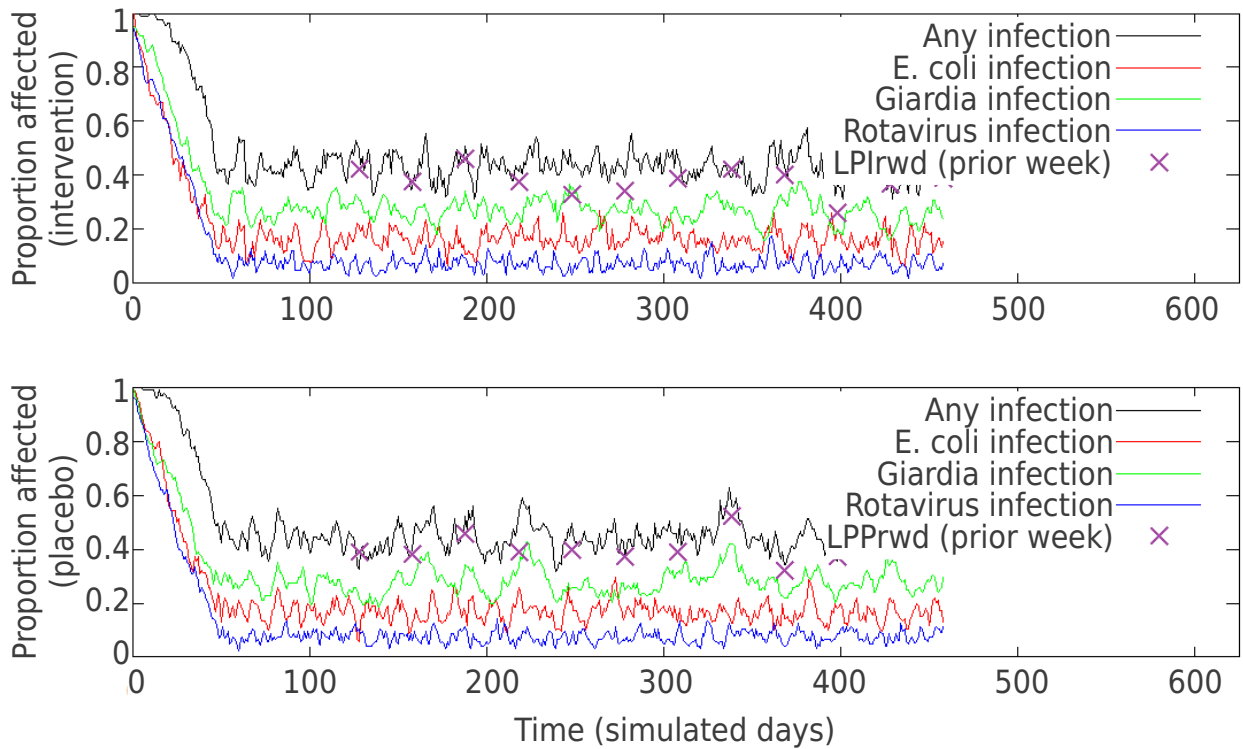
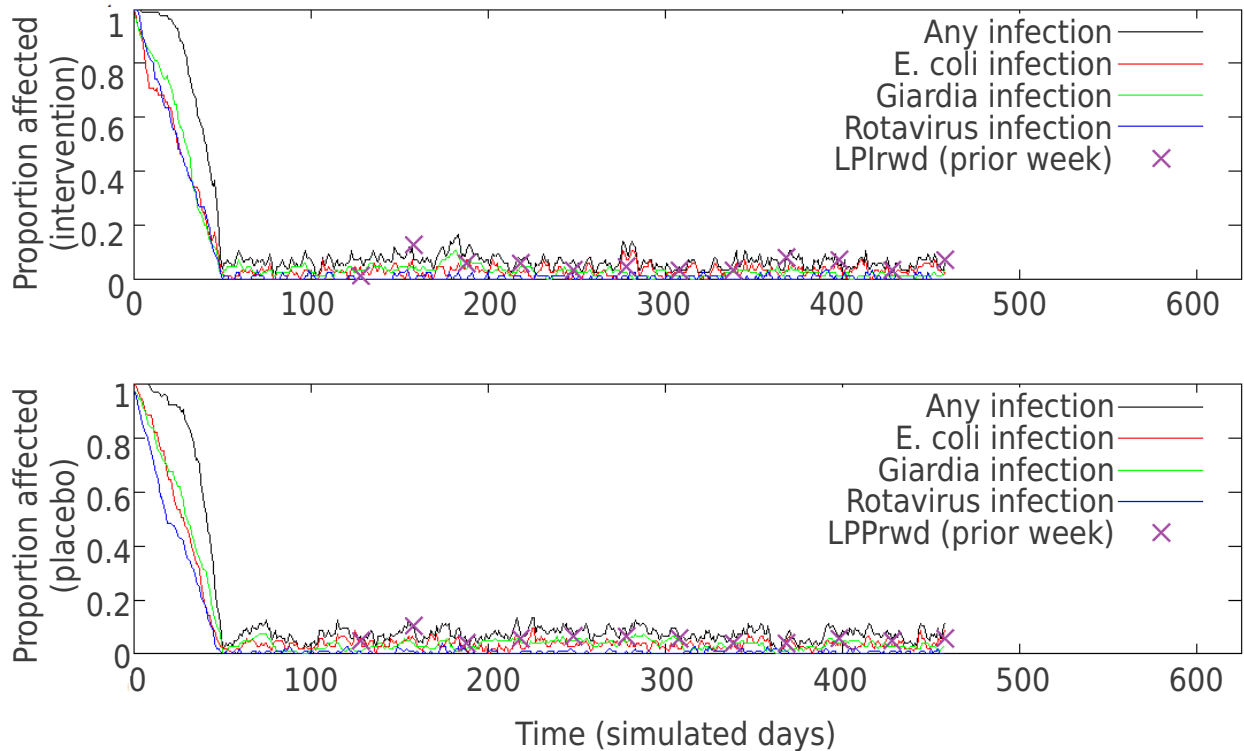


Figure 3.5. Example run of the model, infection levels consistent with the Lifestraw RCT



3.3.4. Analytical process

The simulation model described above was implemented in two steps: calibration and estimation (Figure 3.1, page 90).

The calibration step estimated the four unknown parameter values (Table 3.2, page 96) by constraining the model outputs to the results of the Lifestraw RCT (Table 3.3, page 96). The calibration step included 100,000 runs. It was conducted for two calibration compliance conditions, low and medium, because the proportion of water treated by filter users during the Lifestraw RCT was unknown but believed to be substantially less than 1. Additionally, the parameters in Table 3.2 (page 96) were randomly sampled from uniform distributions for each simulation run. If a model run yielded results consistent with the Lifestraw RCT, its set of four parameter values (Table 1) was used in the estimation step. A run was considered consistent if the LP_{Irad} , LP_{Prad} , and LPR_{rad} all fell within the 95% confidence limits reported from the Lifestraw

RCT (Table 3.3, page 103).

The estimation step determined effectiveness given: 1) a perfect placebo, and 2) low, medium, high, or perfect estimation compliance. It consisted of ten model runs for each parameter set that was consistent with the Lifestraw RCT (totaling >2000 runs) for each of four estimation compliance levels, assuming a perfect placebo.

The estimation step simulated measurements of LP_{Irad} and LP_{Prad} , and their ratio LPR_{rad} , which were calculated in the same way as in the calibration step (number of monthly person-surveys reporting diarrhea during the previous 7 days, divided by the total number of person-surveys).

Differing compliance values were used in the calibration and estimation steps. In the calibration step, ‘calibration compliance’ referred to a set of compliance values describing what probably occurred during the actual Lifestraw RCT; these were necessary to calibrate the model to the four parameter values in Table 3.2 (page 96). In the estimation step, ‘estimation compliance’ referred to a larger set of compliance values that allow the model to make predictions for several different scenarios. Calibration compliance and estimation compliance were considered simultaneously in the results because different calibration compliance levels led to different results in the estimation step.

The QMRA model was programmed in Octave 3.2; the code also runs in MATLAB 7.11. The program code is in the appendices (chapter 9, page 239). Results were analyzed using R 2.11; the two-tailed Wilcoxon rank sum test ($\alpha = 0.05$) was used to compare distributions.

3.4. Results

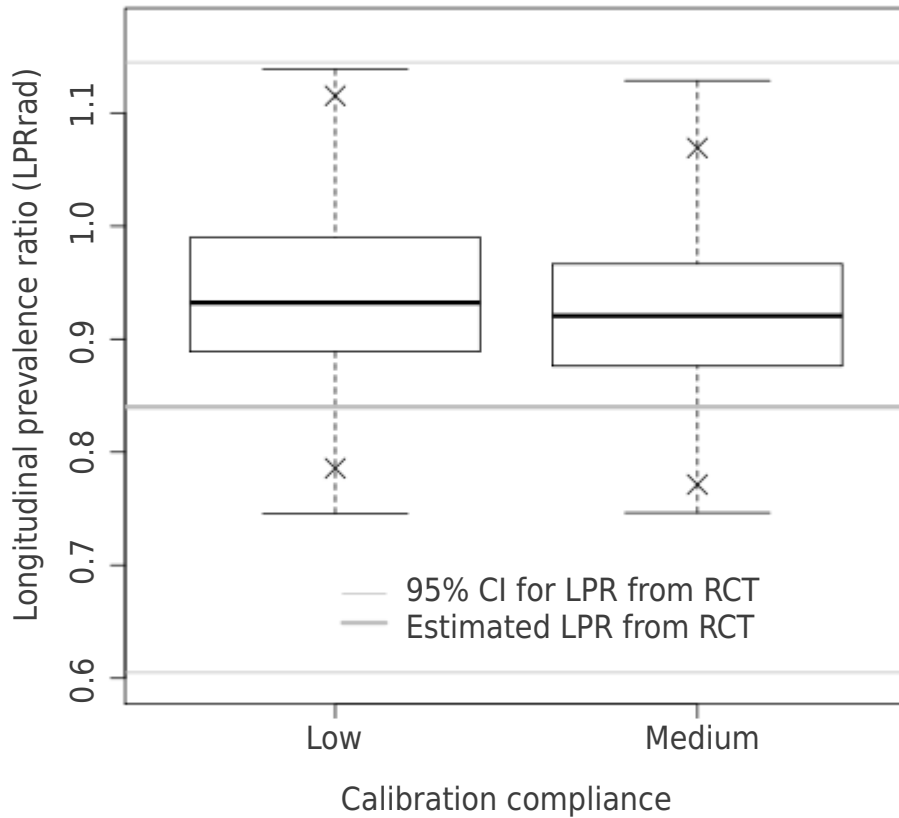
3.4.1. Calibration step

Out of the 100,000 simulation runs in the calibration step, 210 were consistent with the Lifestraw RCT based on the criteria in Table 3.3 (page 103) and assuming low calibration

compliance with water filtration. Repeating the calibration step assuming medium calibration compliance yielded 258 consistent runs. Calibration estimated distributions for two outputs:

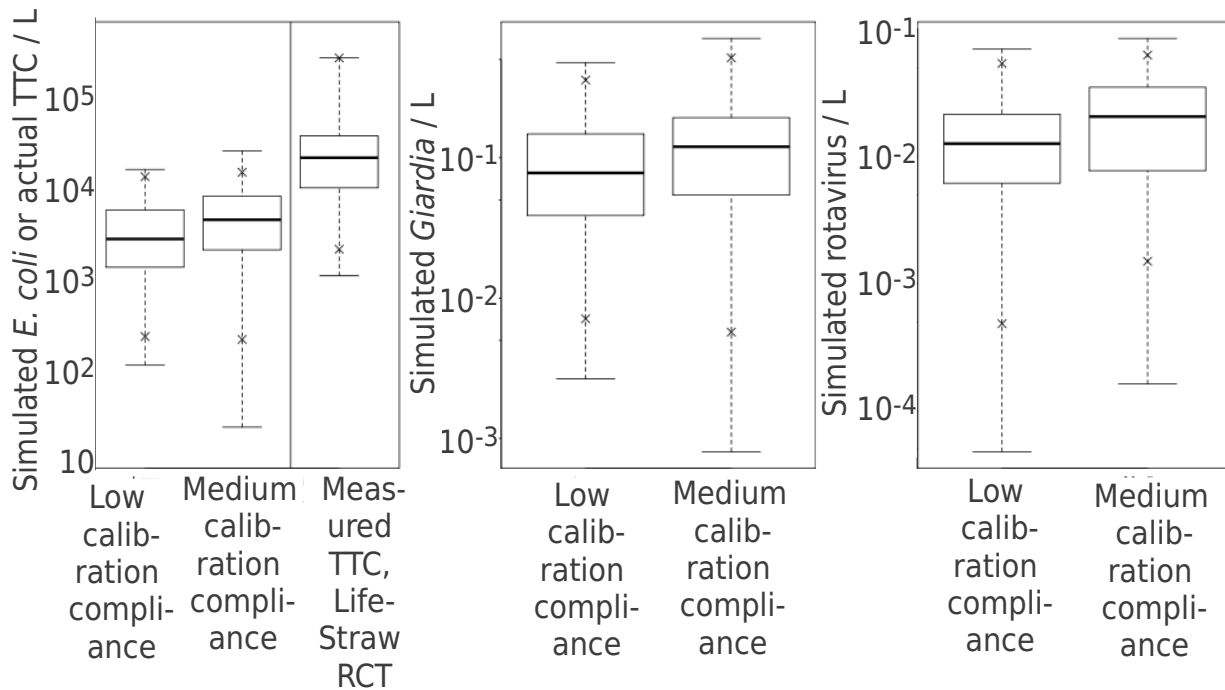
1. The longitudinal prevalence ratio (LPR_{rad}) distributions were similar by level of calibration compliance (Figure 3.6). The estimate from the Lifestraw RCT falls within the central 95% of the distributions, suggesting consistency between the model and the Lifestraw RCT. The median LPR_{rad} estimated by the model differed from the Lifestraw RCT because the Lifestraw RCT was a single experiment, whereas each distribution of LPR_{rad} represented over 200 simulated experiments.
2. Simulated concentrations of pathogen types in untreated water (Figure 3.7) were higher for medium calibration compliance compared with low calibration compliance, which was necessary to produce LP_{Irad} and LP_{Prad} values consistent with the RCT. In individual calibration runs, higher concentrations of one pathogen type were associated with lower concentrations of the other two pathogen types. The median diarrheagenic *E. coli* concentration predicted by the model is lower than the median thermotolerant coliform (TTC) concentration measured in untreated drinking water in the Lifestraw RCT (Figure 3.7). This is plausible since *E. coli* are a subset of TTC, and not all *E. coli* are pathogenic.

Figure 3.6. Distributions of LPRs consistent with the Lifestraw RCT



Distributions of longitudinal prevalence ratios from simulation runs consistent with the Lifestraw RCT from the calibration step for low (65% of children treat 1/3 of their drinking water) and medium (65% of children treat 2/3 of their drinking water) calibration compliance. These distributions differed significantly (Wilcoxon rank sum test, $p = 0.02$). Boxplots include: median (heavy line), 25th and 75th percentiles (lower and upper limits of the box), 2.5th and 97.5th percentiles (X symbols), and range (whiskers).

Figure 3.7. Distributions of simulated microbial concentrations



Simulated distributions of microbial concentrations per liter of untreated water, consistent with the Lifestraw RCT. Distributions were obtained from the calibration step assuming low (65% of children treat 1/3 of their drinking water) or medium (65% of children treat 2/3 of their drinking water) calibration compliance. Thermotolerant coliforms (TTC) measured by the Lifestraw RCT are also shown for comparison with simulated E. coli. For all three pathogen types, the concentration distributions differed by calibration compliance (Wilcoxon rank sum test, $p < 0.001$).

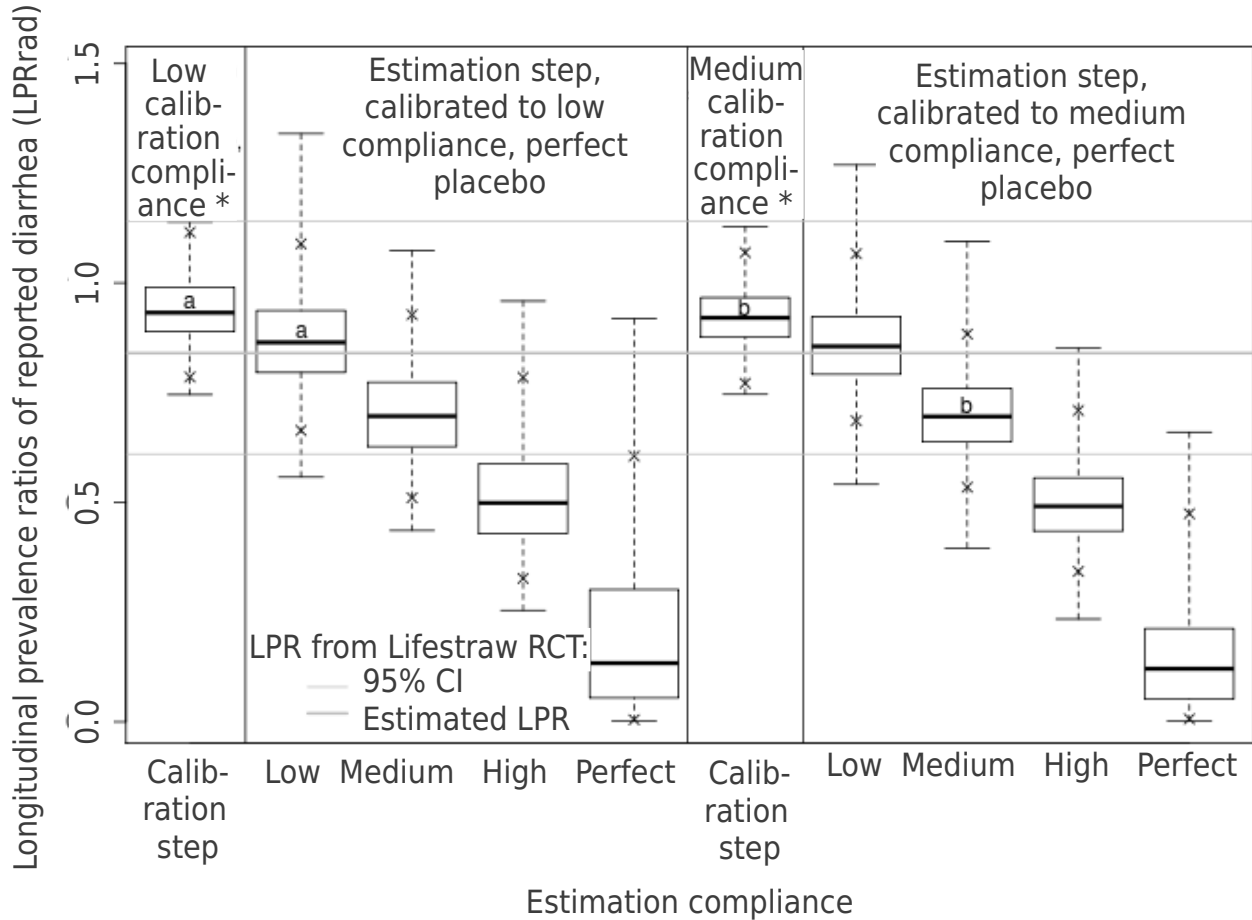
3.4.2. Estimation step

This step estimated LFF effectiveness compared to a perfect placebo for low, medium, high, and perfect estimation compliance, given low or medium calibration compliance.

Estimation compliance was a major driver of effectiveness. For example, under low, medium, high, and perfect estimation compliance, the median LPR_{rad} was 0.86, 0.70, 0.50, and 0.13 respectively, regardless of calibration compliance (Figure 3.8). Additionally, LP_{Irad} was significantly greater with medium calibration compliance (compared to low calibration compliance), for all levels of estimation compliance except perfect (Figure 3.9). This difference

occurred because both calibration steps (low and medium calibration compliance) were constrained to the same RCT result; if calibration compliance decreases, the pathogen concentrations must also decrease for the model to remain consistent with the Lifestraw RCT. During the estimation step, the higher LP_{Prad} for medium calibration compliance was due to lack of protection from the perfect placebo; therefore, the LP_{Irad} values were also higher. The differences between low and medium calibration compliance decrease as estimation compliance increases.

Figure 3.8. LPR distributions for differing compliance assumptions and placebo behavior

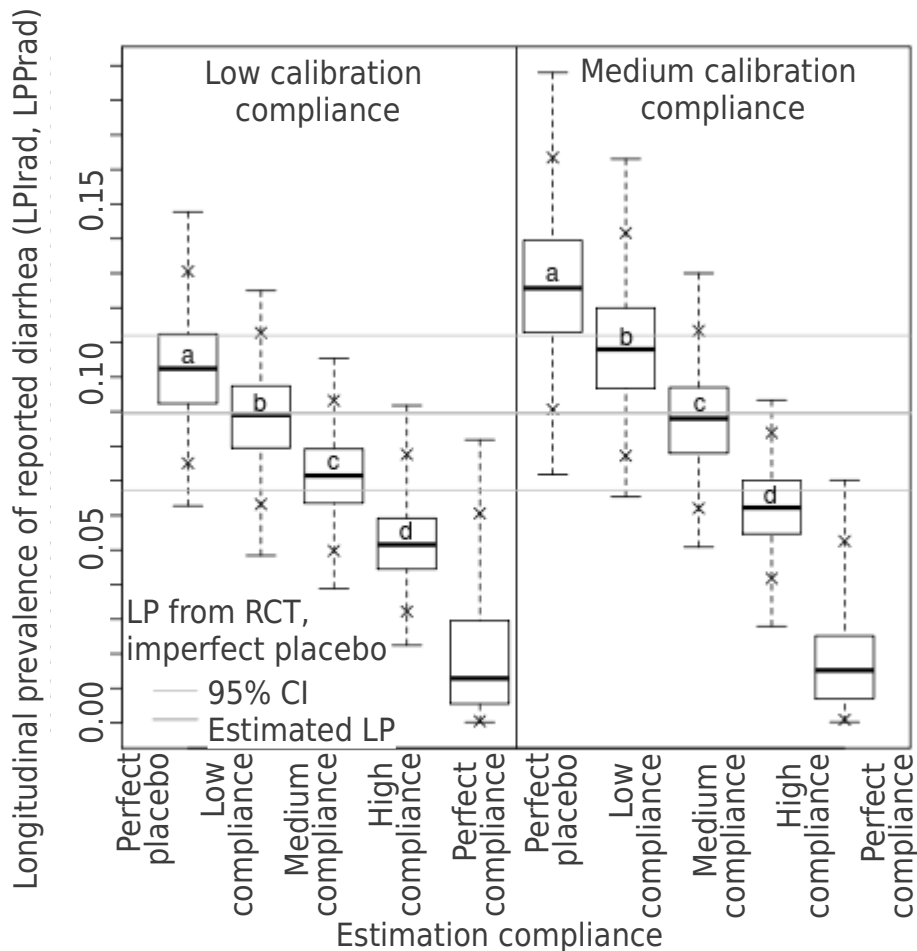


Distributions of longitudinal prevalence ratios of reported diarrhea (LPR_{rad}), estimation step for the intervention group, by calibration compliance, estimation compliance, and placebo type.

*: LPRs from the calibration step, and therefore with an imperfect placebo.

a, b: These pairs of distributions illustrate the effect of the imperfect placebo on LPR_{rad} , depending on low (a) or medium (b) calibration compliance. They differed significantly by imperfect vs. perfect placebo (Wilcoxon rank sum test, $p < 2 \times 10^{-16}$).

Figure 3.9. Longitudinal prevalence distributions under differing compliance assumptions



Distributions of reported longitudinal prevalence (LP) of diarrhea in the estimation step for the intervention group, by calibration compliance and estimation compliance.

a, b, c, d: Each of these pairs of distributions differed significantly by calibration compliance (Wilcoxon rank sum test, $p < 2 \times 10^{-16}$).

Compliance levels: Perfect placebo (100% of children treat 0% of their drinking water), low compliance (65% of children treat 1/3 of their drinking water), medium compliance (65% of children treat 2/3 of their drinking water), high compliance (65% of children treat 100% of their drinking water), perfect compliance (100% of children treat 100% of their drinking water).

Calibration compliance altered the estimated effect of a perfect placebo (Figure 3.8).

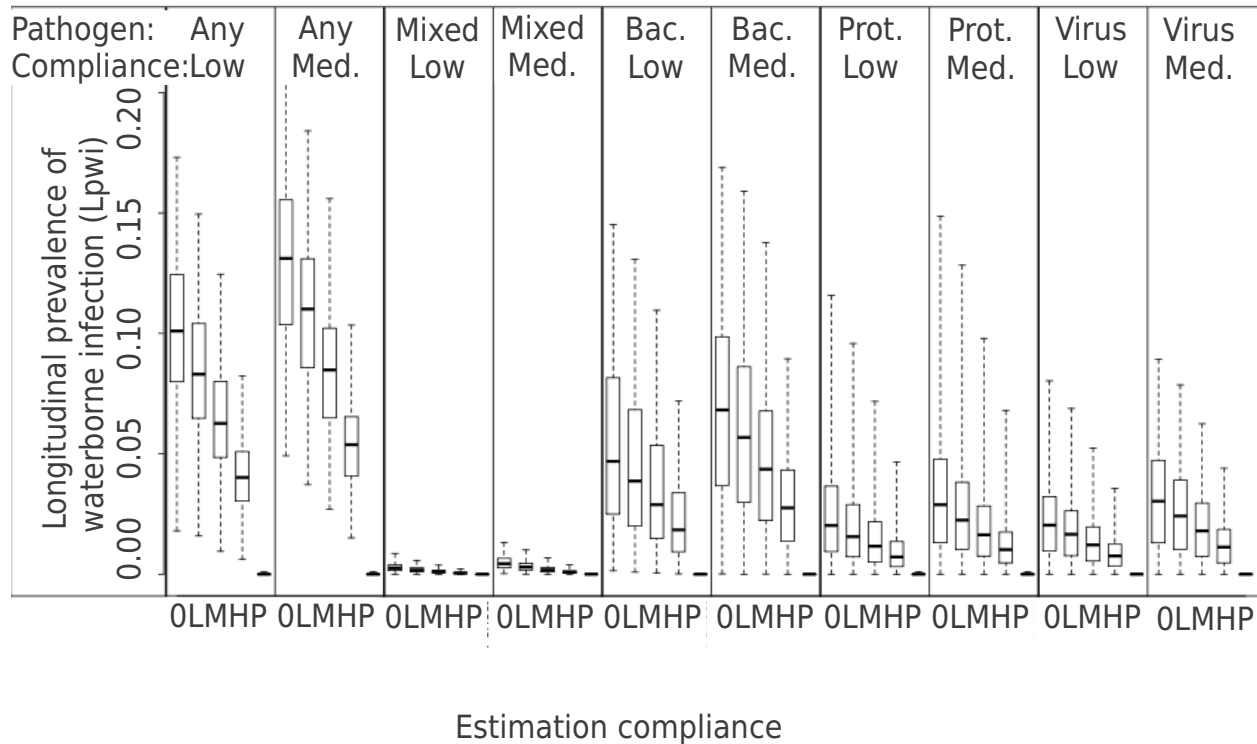
Adjustment for the imperfect placebo increased the estimated preventable fraction of disease by 8 percentage points assuming low calibration compliance (median LPR_{rad} : 0.94 and 0.86 for imperfect and perfect placebo, respectively). Assuming medium calibration compliance, the preventable fraction increased by 22 percentage points (median LPR_{rad} : 0.92 and 0.70 for

imperfect and perfect placebo, respectively).

All three pathogen types contributed substantially to infection and disease (Figure 3.10).

Mixed infections accounted for about 2% of infections.

Figure 3.10. Longitudinal prevalence of waterborne infection in the estimation step



Longitudinal prevalence of waterborne infection (LP_{wi}) during the estimation step with a perfect placebo, by pathogen type. LP_{wi} is higher for medium calibration compliance, compared with low calibration compliance. Compliance levels: None (0; 0% of children treating 0% of their drinking water, i.e., perfect placebo), low compliance (L; 65% of children treat 1/3 of their drinking water), medium compliance (M; 65% of children treat 2/3 of their drinking water), high compliance (H; 65% of children treat 100% of their drinking water), perfect compliance (P; 100% of children treat 100% of their drinking water).

3.5. Discussion

Results from an epidemiological study may only be relevant to the ecological and social conditions of the communities studied. However, quantitative microbial risk assessment (QMRA) models that are calibrated to epidemiologic data can predict risk under scenarios that

were not actually studied, known as counterfactual scenarios (J. J. Kim et al., 2007; Tuite et al., 2010). A calibration process to epidemiological data can facilitate the use of QMRA models where the environmental contamination data required by those models are difficult to measure. There are many situations where epidemiological data can be used to provide parameters to QMRA models, such as in a developing country context where direct measures of disease risk are frequently measured but environmental contamination data are rare. In this context, the modeling framework described herein was used to generalize results from the Lifestraw RCT (Boisson et al., 2010).

Generalizing across different compliance scenarios, this model quantified the relationship between compliance and HWT effectiveness. Our analysis suggests that perfect compliance in the Lifestraw RCT communities would yield an LPR_{rad} of 0.13, suggesting that 87% of reported diarrhea could be prevented by consumption of treated water. This result, suggesting that only 13% of diarrhea in the Lifestraw RCT community was caused by non-waterborne transmission, is consistent with a HWT trial in a refugee camp in which there was 95% compliance and an 83% reduction in diarrhea prevalence (Doocy & Burnham, 2006), as well as numerous field trials of ceramic filters indicating risk ratios < 0.5 (T. Clasen, I. G. Roberts, et al., 2009; Hunter, 2009). Noncompliance will result in an underestimate of protective effect compared to a situation in which there is perfect compliance. Additionally, these trials and our risk assessment model do not account for the interdependency of other transmission routes. Thus, it is unclear whether this model accurately assesses the proportion of diarrhea associated with drinking water.

The reduction in effectiveness due to the imperfect placebo also depended on the assumed calibration compliance level during the Lifestraw RCT. Assuming low calibration compliance, the imperfect placebo decreased effectiveness (preventable fraction) by 8 percentage points.

Effectiveness decreased 22 percentage points assuming medium calibration compliance (Figure 3.8, page 111).

Compliance has several components. This analysis used a simple formulation in which each child had a daily probability of using the device; if the device was used, a fixed proportion of water was treated. In reality, some people may be highly consistent users or nonusers, while others might use the device occasionally (e.g., drinking untreated water while working outside the home). Effectiveness might differ given perfect use by 50% of a population, compared to an entire population treating 50% of their water. These distinctions are further explored in chapter 4.

These results are consistent with the reasonable expectation that increasing compliance should increase effectiveness. Recent systematic reviews suggest a positive but not statistically significant relationship between compliance and effectiveness, perhaps due to difficulty in measuring compliance (B. F. Arnold & Colford, 2007; T. Clasen, I. G. Roberts, et al., 2009; Waddington et al., 2009). Modeling has shown how even occasional treatment failures by a water treatment plant could cause high levels of diarrheal disease in populations (Hunter et al., 2009). Collectively, these results show the importance of consistent treatment of drinking water, by well-managed municipal plants or by high compliance with HWT. Future studies should examine the joint effects of compliance and \log_{10} removals by a HWT device. QMRA models can be used to extend the results shown in Figure 3.8 (page 111) by providing estimates for risk reductions as a function of both compliance and device efficacy. For example, for a given compliance level, what are the expected risk reductions when using a device that provides 5 \log_{10} removal, versus a device that provides 3 \log_{10} removal (see section 4.4.4, page 150)? Such analyses could provide performance metrics and standards that address not only microbiological efficacy but also the correct, consistent and sustained use of a device by the target population.

3.5.1. Predicting pathogen concentrations in drinking water sources

Few data exist on pathogen concentrations that individuals ingest via drinking water in developing countries. Calibration of this model used the epidemiological data from the Lifestraw RCT to predict concentrations of pathogens in untreated water (Figure 3.7, page 109). The predicted *Giardia* concentrations are consistent with measurements from southeastern Brazilian raw water sources (< 0.1 to 3.4 cysts/L) (Razzolini et al., 2010), but lower than other Brazilian (2.5 to 120 cysts/L) (Neto et al., 2010) or Honduran source waters (2.4 to 21 cysts/L) (Solo-Gabriele et al., 1998). The predicted diarrheagenic *E. coli* concentrations are much lower than the $2.5 \cdot 10^5$ -- $1.6 \cdot 10^7$ CFU of enterotoxigenic *E. coli* detected in sewage-impacted Indian rivers (Singh et al., 2010). Furthermore, the predicted rotavirus concentrations are substantially lower than measurements from polluted creeks in Sao Paulo, Brazil (geometric mean, ~2.7 focus-forming units/L) (Mehnert & Stewien, 1993). It is reasonable that the clear source water in the Lifestraw RCT site would have lower pathogen concentrations than the above water sources, many of which were from polluted urbanized environments.

Measuring viable pathogen concentrations in drinking water (and other exposure routes) in several epidemiological studies would increase the accuracy and precision of risk estimates, as indicators of fecal contamination (e.g., coliform bacteria or *E. coli*) correlate poorly with presence or concentrations of actual pathogens (American Water Works Association, 1999; Leclerc et al., 2002; Toranzos et al., 1988).

3.5.2. Calibration of microbial risk assessment models

Microbial risk models, like many environmental models, can never be fully validated (Oreskes et al., 1994). However, they can be confirmed through a calibration process using epidemiological data. We present a framework that first calibrates using epidemiological data, and second estimates risks under differing counterfactual scenarios. The calibration process

transforms uninformed priors to informed posteriors by constraining the model using the epidemiological outcome data from the Lifestraw RCT. The low percentage of calibration runs that were consistent (<0.3%) indicates that the trial data imparted substantial information to the model by constraining the acceptable parameter space. Source water pathogen data would be useful to further calibrate the model for the Congo field site.

Risk models informed by epidemiological data are powerful tools to generalize beyond the context in which epidemiological studies are conducted. Models can be used to inform study design and intervention strategies; epidemiological studies can calibrate these models. Few published examples take this approach, but see: (J. J. Kim et al., 2007; Tuite et al., 2010). The framework provided here can facilitate such future research activities. Results from our examination of HWT field trials indicate that compliance and pathogen concentrations in source water are particularly important processes to characterize. Data from these processes would enhance the calibration step, providing the opportunity to describe other unobserved aspects of the system. Additionally, pathogen measurements from other environmental sources (e.g., hands, food, feces) would facilitate the extension of our model system to consider transmission by multiple environmental pathways (Li et al., 2009); such transmission models would allow investigation of interdependency of multiple transmission routes, and ultimately multiple interventions.

Although additional data on pathogen concentration and compliance and the extension to considering transmission are important steps forward in informing intervention strategies, our analysis provides the important and robust conclusion that effectiveness is highly sensitive to compliance, suggesting that trials of household level interventions should measure compliance as carefully and as effectively as possible. Compliance guidelines should be developed for HWT interventions, in addition to the microbial reduction guidelines for HWT devices recently

published by WHO (Sobsey & Joe Brown, 2011).

4. THE JOINT EFFECTS OF EFFICACY AND COMPLIANCE IN HOUSEHOLD WATER TREATMENT EFFECTIVENESS

This chapter consists of a manuscript and its supplementary material that has been peer-reviewed and accepted for publication. It has been reformatted and reorganized:

Enger, K.S., Nelson, K.L., Rose, J.B., Eisenberg, J.N.S. The joint effects of efficacy and compliance: a study of household water treatment effectiveness against childhood diarrhea. Publication forthcoming in the journal *Water Research*.

4.1. Abstract

The effectiveness of household water treatment (HWT) at reducing diarrheal disease is related to the intrinsic efficacy of the HWT method at removing pathogens, how people comply with HWT, and the relative contributions of other pathogen exposure routes. Although many HWT methods are efficacious at removing or inactivating pathogens, their effectiveness within actual communities is decreased by imperfect compliance. However, the quantitative relationship between compliance and effectiveness is poorly understood. To assess the effectiveness of HWT on childhood diarrhea incidence via drinking water for three pathogen types (bacterial, viral, and protozoan), a quantitative microbial risk assessment (QMRA) model was developed. The model allowed examination of the relationship between \log_{10} removal values (LRVs) and compliance with HWT for scenarios varying by: baseline incidence of diarrhea; etiologic fraction of diarrhea by pathogen type; pattern of compliance; and size of contamination spikes in source water. Benefits from increasing LRVs strongly depend on compliance. For perfect compliance, diarrheal incidence decreases as LRVs increase. However, if compliance is incomplete, there are diminishing returns from increasing LRVs in most of the scenarios considered. Higher LRVs are more beneficial if: contamination spikes are large; contamination levels are generally high; or

some people comply perfectly. The effectiveness of a HWT intervention at the community level may be limited by low compliance, such that the benefits of high LRVs are not realized. Patterns of compliance with HWT should therefore be measured during HWT field studies and HWT dissemination programs. Studies of pathogen concentrations in a variety of developing-country source waters are also needed. Guidelines are needed for measuring and promoting compliance with HWT, in addition to recently published WHO HWT efficacy guidelines.

4.2. Introduction

An effective intervention can be defined as one that reduces disease (i.e., is efficacious) and one that people use (i.e., they comply). For example, a drug or vaccine must be protective, and people must take the drug or receive the vaccine; contaminated water must be correctly treated, and people must drink the treated water. Both efficacy and compliance must be evaluated when assessing the ability of an intervention to reduce illness; both are dynamic factors that can vary over time. Household water treatment (HWT) interventions are an interesting example that illustrates these two factors, where pathogen removal characterizes efficacy and behavior characterizes compliance. This chapter examines the joint effects of 1) pathogen removal by a HWT device, and 2) the degree to which communities use the device. It focuses on the protective effects of HWT against diarrhea in developing countries, a leading cause of morbidity and mortality (Kosek et al., 2003).

Household water treatment (HWT) is a common strategy for reducing diarrhea in developing countries. HWT technologies include chlorination, filtration, solar disinfection (SODIS), and boiling. Systematic reviews of field trials suggest that HWT is effective in preventing diarrhea (B. F. Arnold & Colford, 2007; T. Clasen, I. G. Roberts, et al., 2009). However, lack of blinding and publication bias are important issues in the HWT literature that may exaggerate effectiveness (Schmidt & Cairncross, 2009; Waddington et al., 2009; Hunter,

2009); see also pages 17 and 65.

Antimicrobial effectiveness of HWT is commonly measured by \log_{10} reduction values (LRVs) from laboratory testing. Such tests use indicator organisms to represent the three main classes of waterborne pathogens: viruses, bacteria, and protozoan cysts. LRVs are a common metric for assessing different HWT methods (Sobsey & Joe Brown, 2011; Sobsey et al., 2008). The United States standard for HWT “microbiological water purifiers” is LRVs of 6 for bacteria (99.9999% inactivation), 4 for viruses, and 3 for protozoa (USEPA, 1987). The World Health Organization (WHO) recommends that “highly protective” devices have LRVs of 4 for bacteria, 5 for viruses, and 4 for protozoa (Sobsey & Joe Brown, 2011); see also Table 2.2, page 65. The WHO recommendations use a quantitative microbial risk assessment (QMRA) assuming perfect compliance and an acceptable risk level of 10^{-6} disability-adjusted life-years (DALYs) for diarrheal disease from each pathogen type (Sobsey & Joe Brown, 2011).

In contrast, compliance, the extent to which persons (or a population) use a HWT method, is often poorly defined and poorly measured. Compliance (sometimes referred to as adherence) has many dimensions. Individuals might reject a HWT method because of cost, difficulty using HWT, or taste of treated water. Well-established theory regarding adoption of new technologies indicates that 10%-20% of a community will not use a new technology, even after acceptance by most of the community (E. M. Rogers, 2003). Furthermore, preventive practices (such as HWT) that require consistent individual effort to reduce the probability of an adverse effect have difficulty spreading. This is because the benefit (e.g., bouts of diarrhea averted) is a ‘non-event’ that is distant in time from adopting the practice; therefore, the benefit gained is not obvious to the user (E. M. Rogers, 2003). HWT devices might simultaneously be used frequently and inconsistently. For example, someone might drink treated water at home, but untreated water

while working. During a HWT field trial in rural Congo, nearly all households sometimes drank untreated water (Boisson et al., 2010).

Although the variable and incomplete nature of compliance is widely recognized, it is often unmeasured or incompletely measured by field trials. A review of 30 relatively well-conducted field trials of water quality interventions found that 7 did not report compliance, and 9 measured compliance by “occasional observation” only (T. Clasen, I. G. Roberts, et al., 2009). Furthermore, consumption of treated water was never directly measured (T. Clasen, I. G. Roberts, et al., 2009). Studies that report compliance find that communities rarely use HWT methods 100% of the time. For example, a meta-analysis of HWT chlorination studies indicated a median of 78% of samples having detectable free chlorine (range 36-100% over 12 studies) (B. F. Arnold & Colford, 2007).

Compliance is difficult to measure and is subject to Hawthorne effects (where people's behavior changes because they know that they are being observed) and other biases. Participants in a trial might report that they use the intervention more frequently than they actually do. Compliance might increase during a trial because study personnel remind people to use HWT (deliberately or not). Field trials over longer periods show lower HWT effectiveness against diarrhea; decreasing compliance over time is one explanation (Hunter, 2009). It is particularly difficult to determine the amount of untreated water that HWT users consume outside the home.

Despite not being well measured, compliance clearly influences HWT effectiveness, because HWT can only prevent diarrhea if people use it (Duflo et al., 2007). Field measurements of LRVs tend to be lower than laboratory-measured LRVs for many reasons, such as differing water quality or suboptimal maintenance of HWT devices (Sobsey et al., 2008). Nonetheless, the benefits from HWT might be eroded by slight noncompliance. For example, a risk assessment of diarrheal infection from intermittent treatment by a Ugandan water treatment plant estimated that

water treatment failure for one day per year increased the annual probability of enterotoxigenic *Escherichia coli* (ETEC) infection via drinking water from 0.001 to 0.1 (Hunter et al., 2009).

The relationship between compliance and LRVs (which measure efficacy) can be illustrated mathematically:

$$d = u(1 - c) + uc10^{-L} \quad (4.1)$$

where d is the dose of pathogens consumed, u is pathogens per liter of untreated water, c is compliance (the proportion of drinking water treated), L is the LRV of the HWT method, and one liter of water is consumed. Using equation 4.1 assuming that source water contains 10,000 pathogens per liter, 5 LRVs of pathogens are inactivated, and 1% of drinking water is untreated, then 100 pathogens are ingested. For LRVs of 4, 3, 2, and 1, the numbers of pathogens consumed are, respectively: 101, 110, 199, and 1090. The dose (and therefore the infection risk) is very similar for LRVs of 5, 4, and 3 (100 to 110 pathogens), which leads to the hypothesis tested in this chapter: incomplete compliance results in marginal reductions in diarrheal disease as LRVs increase.

4.3. Materials and methods

To test the hypothesis, a QMRA model was used to simulate waterborne transmission of diarrheal infection (bacteria, protozoa, and viruses) in children aged less than five years. This model was based on the model in chapter 3 (Enger et al., 2012) that simulated a randomized controlled trial of the LifeStraw® Family filter (LFF; a HWT device) in rural Congo (Boisson et al., 2010). The model was programmed in MATLAB 7.12 and Octave 3.2; results were analyzed with R 2.11. The model only considered diarrhea transmitted by drinking water, omitting other transmission routes (e.g., contaminated food, objects, or hands). Parameter values for the model are summarized in chapter 7 and Table 7.1 (page 230).

Four important concepts in the model are: compliance, baseline incidence, etiologic

fractions, and short-term contamination spikes. They are described in the following paragraphs.

4.3.1. Compliance

Compliance with HWT within a community was modeled considering three groups of children: 1) children who exclusively consumed treated water (“perfect compliance”); 2) children who never consumed treated water (“no compliance”); and 3) children who consumed fixed proportions of treated and untreated drinking water (“partial compliance”). Overall compliance (c) was calculated as follows:

$$c = (1 - (a + n))p + a \tag{4.2}$$

where a is the proportion of children who always use HWT, n is the proportion of children who never use HWT, and p is the proportion of water treated by partial compliers. For a given value of c , three types of compliance at the community level were defined: α) c children with perfect compliance and the remainder with no compliance; β) $c/2$ children with perfect compliance, $(1 - c)/2$ children with no compliance, and the remainder partially comply, treating a fraction c of their daily water intake (Table 4.1); γ) all children partially comply by treating a fraction c of their water. If $c = 1$ or 0 , only compliance type α is possible.

Table 4.1. Compliance among individuals in each model run, given compliance type β

Overall compliance (a proportion)	Proportion of simulated children who are:			Proportion of water treated by partial compliers
	Perfect compliers	Noncompliers	Partial compliers	
1	1	0	0	Nonapplicable
0.99	0.495	0.005	0.5	0.99
0.95	0.475	0.025	0.5	0.95
0.8	0.4	0.1	0.5	0.8
0	0	1	0	Nonapplicable

4.3.2. Baseline incidence and etiologic fraction

To further generalize the results, differences in the baseline incidence of diarrhea and the relative contributions of viruses, bacteria, and protozoa to diarrheal incidence (etiologic fractions) were considered. The average incidence categories were: 0 – 2 (low); 2 – 6 (medium); and 6 – 12 (high) episodes per child-year (Kosek et al., 2003). Three sets of etiologic fractions were used (Table 4.2), based on reviews of etiologic studies of childhood diarrhea (Lanata & W. Mendoza, 2002; Ramani & Gagandeep Kang, 2009); the determination of these etiologic fractions is discussed in detail on page 129.

Table 4.2. Criteria for the calibration step of the QMRA model

Description	Midpoint and range of etiologic fractions for childhood diarrhea		
	Bacteria	Protozoa	Viruses
Etiologic fractions A. High bacteria, medium protozoa, low viruses	55% 47.5 to 62.5%	30% 22.5 to 37.5%	15% 7.5 to 22.5%
Etiologic fractions B. High bacteria, medium viruses, low protozoa	55% 47.5 to 62.5%	15% 7.5 to 22.5%	30% 22.5 to 37.5%
Etiologic fractions C. Bacteria slightly predominating over protozoa and viruses	40% 32.5 to 47.5%	30% 22.5 to 37.5%	30% 22.5 to 37.5%

The incidence ranges were: low, 0-2 episodes per child-year; medium, >2-6 episodes per child-year; and high, >6-12 episodes per child-year.

4.3.3. Short-term contamination spikes

Measurements from surface waters indicate that concentrations of indicator organisms are highly variable (Boehm, 2007; K. Levy, A. E. Hubbard, K. L. Nelson, et al., 2009). The variability of pathogen concentrations is expected to be similar or greater, and abnormally large spikes of contamination might occur occasionally. Spikes of pathogen concentrations were simulated on random days, assuming that each spike lasted exactly one day, there were n spikes

per year, and the spike height was x fold higher than the mean baseline concentration on days lacking a spike. For this analysis, $x = 1$ (no spikes), 10, 10^3 , or 10^5 , and $n = 5$. To aid comparison between spike scenarios, the mean number of pathogens in t daily 1-liter samples of source water was held constant regardless of spike height: b_0 is the mean baseline concentration in a scenario without spikes, and b_s is the mean baseline concentration in a scenario with spikes.

Solving for b_s gave the appropriate mean baseline concentration during spike scenarios:

$$b_0 t = b_s(t - n) + n x b_s \quad \rightarrow \quad b_s = b_0 t / (n x + t - n) \quad (4.3)$$

4.3.4. Calibration step

The simulation was implemented in two steps: calibration and estimation. The calibration step simulated transmission of diarrheal infection by drinking water in the absence of HWT, and is described in detail below. It estimated concentrations of bacteria, viruses, and protozoa that were consistent with: 1) assumptions of low, medium, or high incidence of diarrhea; and 2) assumptions about the relative importance of these pathogen types to diarrheal etiology (Table 4.2, page 125). The estimation step used these pathogen concentrations to estimate the risk of diarrhea under various HWT scenarios, defined by different LRVs and different levels of compliance; estimation is described in more detail on page 130, in section 4.3.5

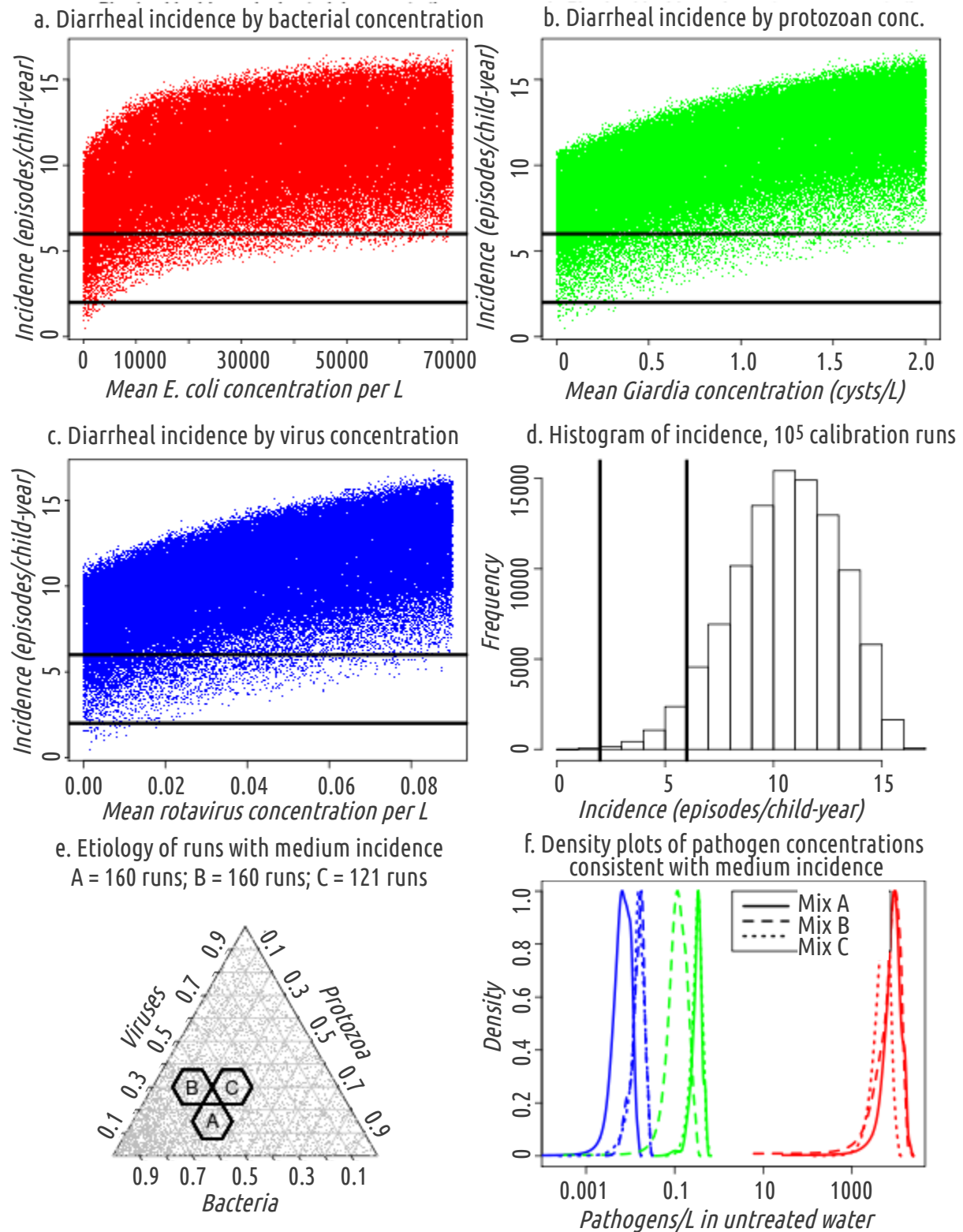
The calibration step modeled waterborne diarrheal infection and disease in a simulated community prior to the introduction of an HWT intervention. The calibration step was run 12 times (3 incidence levels \times 4 spike heights) and each of these yielded 3 sets of pathogen concentrations (since there were 3 sets of etiologic fractions), for a total of 36 calibration scenarios. Each calibration step consisted of 100,000 model runs. Each model run began by randomly selecting a mean pathogen concentration for each of the three pathogen types independently. The pathogen concentrations in simulated drinking water were randomly drawn

daily from a gamma distribution, with the scale parameter determined by the mean pathogen concentration, and the shape parameter determined by the distribution of thermotolerant coliforms measured in source water from a rural area of the Congo (Boisson et al., 2010; Enger et al., 2012). The central 95% of that distribution spanned $1.4 \log_{10}$. Therefore, concentrations could vary 25 fold from day to day, even in the absence of spikes.

Each calibration run followed 100 simulated children over 1 year with no HWT use. The output of each run yielded a community incidence of diarrheal disease and etiologic fractions for the three pathogen types. The mean pathogen concentrations used by a calibration run were retained for use in the estimation step if: 1) the incidence of diarrhea estimated from the model run fell into the appropriate range (Figure 4.1d); and 2) the proportions of diarrhea episodes attributable to bacteria, protozoa, or viruses fell into one of the three sets of etiologic fractions (Table 4.2, page 125; Figure 4.1e). Thus the calibration process yielded sets of pathogen concentrations consistent with 9 distinct calibration scenarios for each of the 4 spike scenarios, for a total of 36 separate calibration scenarios.

Figure 4.1. Calibration results assuming medium incidence of diarrhea

Calibration results, medium incidence. Black lines denote acceptable ranges for incidence and etiology.



See text immediately before & after this panel of charts for further explanation.

For this analysis, calibration of the model required consideration of the largest possible ranges of three joint pathogen concentrations that could conceivably yield an incidence value in the desired range. Consequently, all calibration steps used ranges of pathogen concentrations that began at 0 and ended at an empirically determined concentration where the minimum incidence of diarrhea rose above the highest acceptable incidence value. The resulting distributions of pathogen concentrations that were consistent with the calibration process had negative skew (Figure 4.1f) because a scatterplot of randomly selected pathogen concentrations by incidence forms a gradually increasing band of points (Figure 4.1a, b, & c), and pathogen concentrations that fall within a narrow horizontal band are accepted.

Determination of etiologic fractions

The etiologic fractions were obtained from a review (Lanata & W. Mendoza, 2002) of 266 etiology studies of inpatients, outpatients, and communities, grouped by WHO subregion. Only community studies were considered because the inpatient and outpatient cases are a more severe subset of the full range of diarrhea cases in a given community. Furthermore, only studies from developing WHO subregions AfroD, AfroE, AmroB, and SearoD were considered, because those subregions had 5 or more community-based studies covering most major diarrheal pathogens (*Salmonella*, *Shigella*, *Campylobacter*, enterotoxigenic *Escherichia coli* [ETEC], enteropathogenic *E. coli* [EPEC], *Giardia*, *Cryptosporidium*, and rotavirus). When the pathogens were collapsed into the categories of bacteria, protozoa, and viruses, the four regions broadly agreed on the proportions of diarrhea cases attributed to these three categories. Among cases of diarrhea for which an etiologic agent was identified, bacteria predominated (60%), followed by protozoa (15-30%) and viruses (10-20%). However, the contribution of viruses was probably underestimated by Lanata and Mendoza (2002), since rotavirus was the only virus examined by the studies they reviewed. A more recent review (Ramani & Gagandeep Kang, 2009) of hospital-

based studies of viral gastroenteritis in children in developing countries indicated that about 64% of cases were attributable to rotavirus (other viral gastroenteritides were attributed to caliciviruses, adenoviruses, and astroviruses, or had no etiology determined). Using this information to adjust the conclusions drawn from Lanata and Mendoza's (2002) work, it is suggested that bacteria accounted for approximately 55% of episodes and protozoa and viruses each accounted for 15-30% of episodes, providing a basis for the ranges of etiologic fractions used for calibration (Table 4.2, page 125).

4.3.5. Estimation step

The estimation step modeled waterborne diarrheal infection and disease in a simulated community where a HWT intervention is being used. Each estimation scenario used marker pathogen concentrations from one of the 36 calibration scenarios, for 5000 simulated children over 50 years. Estimation scenarios were defined by the treatment efficacy of the device and the level of compliance. Specifically, combinations of three factors were used: 1) LRVs of the HWT device against all three pathogen types (1, 2, 3, 4, or 5); 2) overall compliance by the community ($c = 1, 0.99, 0.95, 0.80, \text{ or } 0$); 3) type of compliance by the community ($\alpha, \beta, \text{ or } \gamma$; described in detail on page 124). An estimation step was run once for every possible combination of these three factors, for each of the 36 sets of pathogen concentrations from calibration. Each estimation step for a given scenario had 70 to 150 model runs. If a particular calibration step supplied more than 150 sets of pathogen concentrations, 150 sets were randomly sampled for use in the corresponding estimation scenarios. Incidences and incidence ratios (IRs) were determined for various combinations of compliance and device effectiveness; IRs were relative to scenarios in which no HWT was used.

4.3.6. Replication of the WHO model

For comparison, the QMRA model from the WHO HWT recommendations (Sobsey & Joe

Brown, 2011) was replicated in R 2.11. The model was slightly modified to output incidence instead of DALYs, and its assumption of 94% immunity to rotavirus was eliminated, since the model developed in this chapter only considers young children. To obtain total diarrhea incidence from the WHO model, the bacterial, protozoan, and viral diarrhea incidences from its output were summed.

4.4. Results

4.4.1. Calibration step

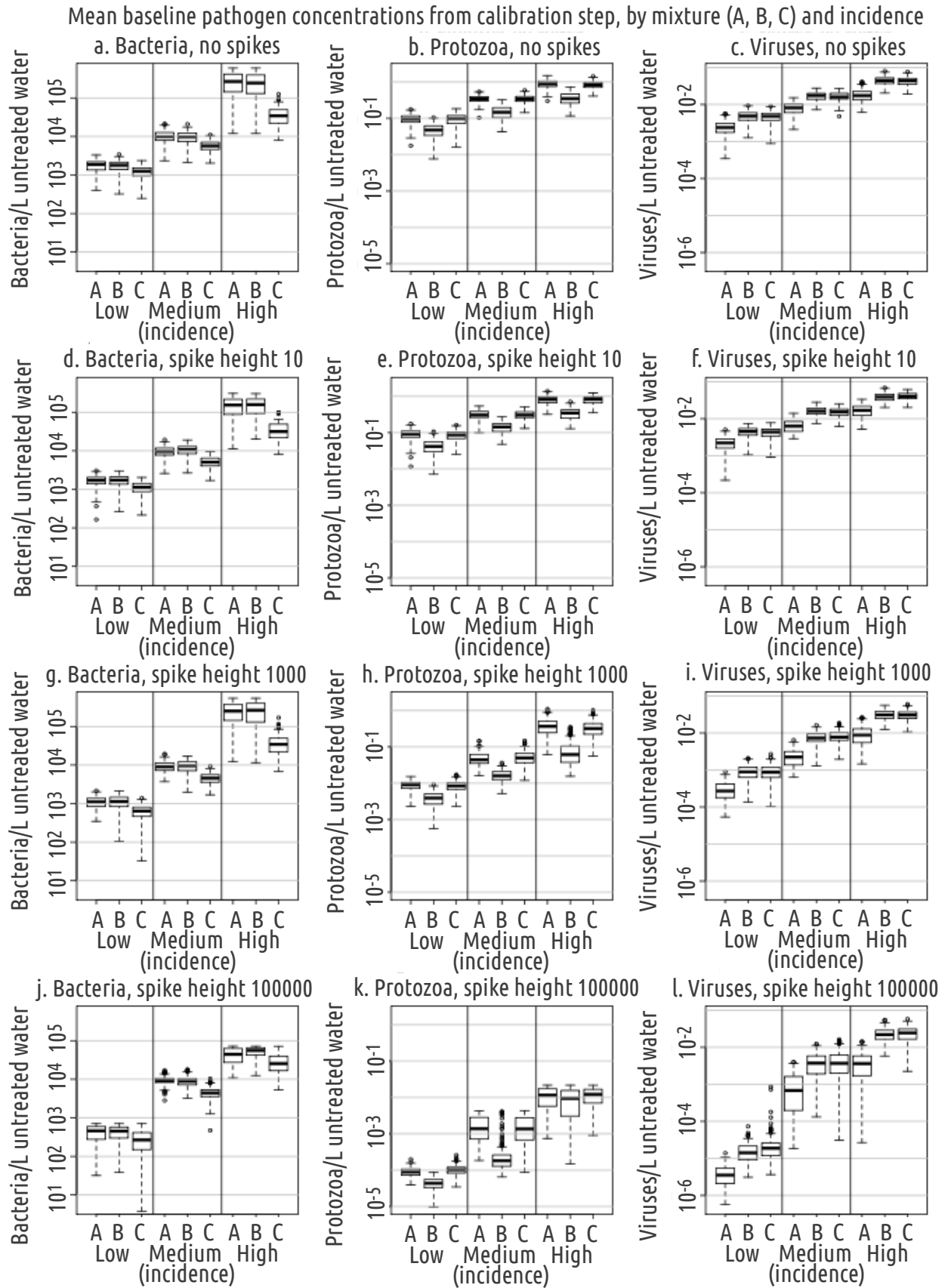
Each calibration step (100,000 runs) yielded 70 to 1164 runs consistent with each of the 36 calibration scenarios. An example of typical calibration output is shown in Figure 4.1 (page 128); if a run was consistent with the incidence criterion (Figure 4.1d), it was tested for consistency with pathogen mixtures A, B, or C (Figure 4.1e; Table 4.2, page 125).

The resulting pathogen concentrations in untreated drinking water, as suggested by the model, are shown in Figure 4.2; without spikes, median estimates ranged from 1100 to 120,000 bacteria/L, 0.06 to 1 protozoa/L, and 0.003 to 0.04 viruses/L (Figure 4.2a, b, and c). It is not clear whether these concentrations are reasonable because pathogen concentrations in source waters have seldom been measured in developing countries. These estimated concentrations (Figure 4.2) are generally lower than the few published measurements available from developing countries (Enger et al., 2012); see also chapter 3, page 116.

Baseline bacterial concentrations during spike scenarios were relatively insensitive to the size of the spike (Figure 4.2, a, d, g, and j). This was because the dose response functions for the three pathogen types are such that the probability of *E. coli* infection increased relatively slowly with \log_{10} increases in dose, compared to *Giardia* or rotavirus (Figures 3.3 and 4.6, pages 99 and 138). In addition, the morbidity ratio (probability of diarrhea, given infection) was much lower

(0.214) for bacterial infection, in contrast to protozoan (0.59) or viral (0.40) infection. Therefore, the maximum contribution to diarrheal morbidity was reached proportionally sooner for bacteria than for protozoa or viruses (note plateauing of the scatterplot in Figure 4.1a on page 128, compared with 4.1b and 4.1c). Since all three pathogen types were increased proportionally during a spike, the impact on overall diarrheal morbidity from protozoa and viruses was therefore larger during a spike.

Figure 4.2. Mean baseline pathogen concentrations from calibration



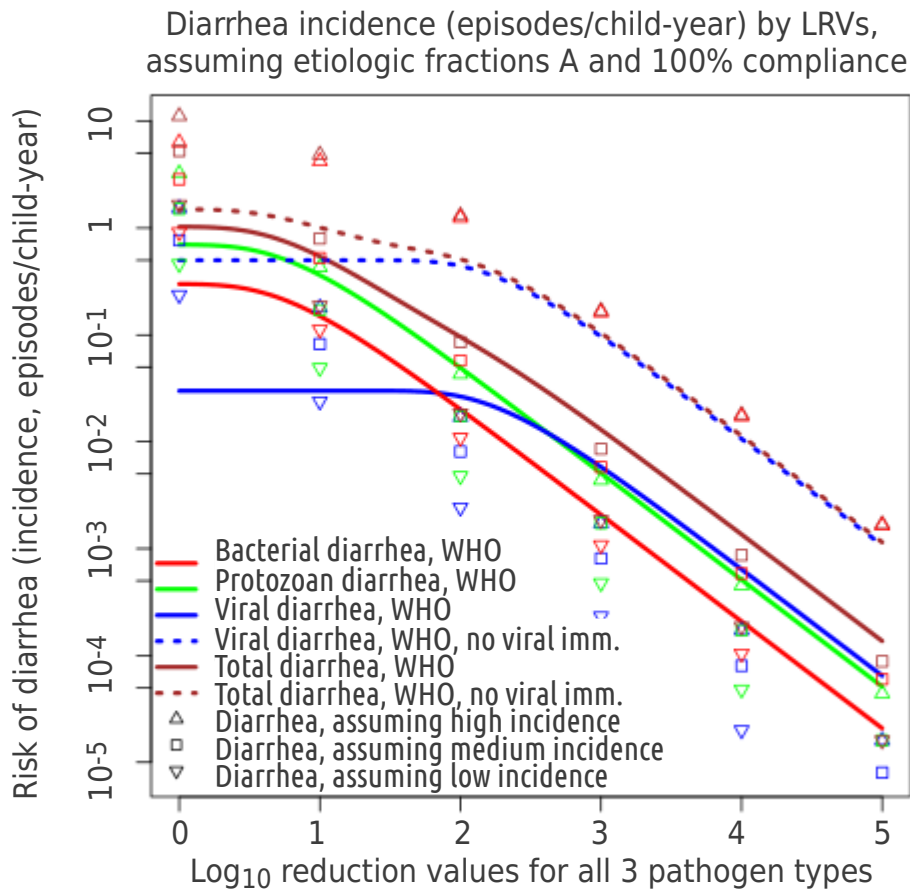
4.4.2. Estimation step

Each estimation step consisted of 70 to 150 model runs, representing distributions of incidences and incidence ratios (IRs) of diarrheal disease given a particular scenario. The medians of all incidence distributions are shown in Figures 4.10-4.13, page 145. Figures 4.4-4.5 and Figures 4.7-4.9 present subsets of those results for clarity, expressed as IRs.

Comparison with the WHO QMRA model

This model and the WHO model (Sobsey & Joe Brown, 2011) produced reasonably consistent estimates of diarrhea risk reduction, assuming 100% compliance (Figure 4.3). This occurred despite differing pathogens and parameter values in each model, as well as substantial differences in model structure (this model considers community-level risk; the WHO model considers individual risk). The WHO model (modified to assume no viral immunity) best resembled this model assuming high incidence. However, the WHO model did not account for repeated episodes of diarrhea in one year, hence its incidence could not be greater than one episode/child-year for each pathogen type. The WHO model indicated that bacteria contribute less childhood diarrhea than protozoa or viruses. This model assumed the opposite, based on reviews of childhood diarrheal etiology (Lanata & W. Mendoza, 2002; Ramani & Gagandeep Kang, 2009). Consequently, although the WHO model results were somewhat similar to the results presented here with respect to total incidence of diarrhea, they differed greatly in the contribution of particular types of pathogens. Furthermore, the WHO model does not account for repeated episodes of diarrhea. Therefore it cannot return incidences higher than 1 for each marker pathogen, and differs substantially from this model when the LRV is 0.

Figure 4.3. Comparison with WHO model

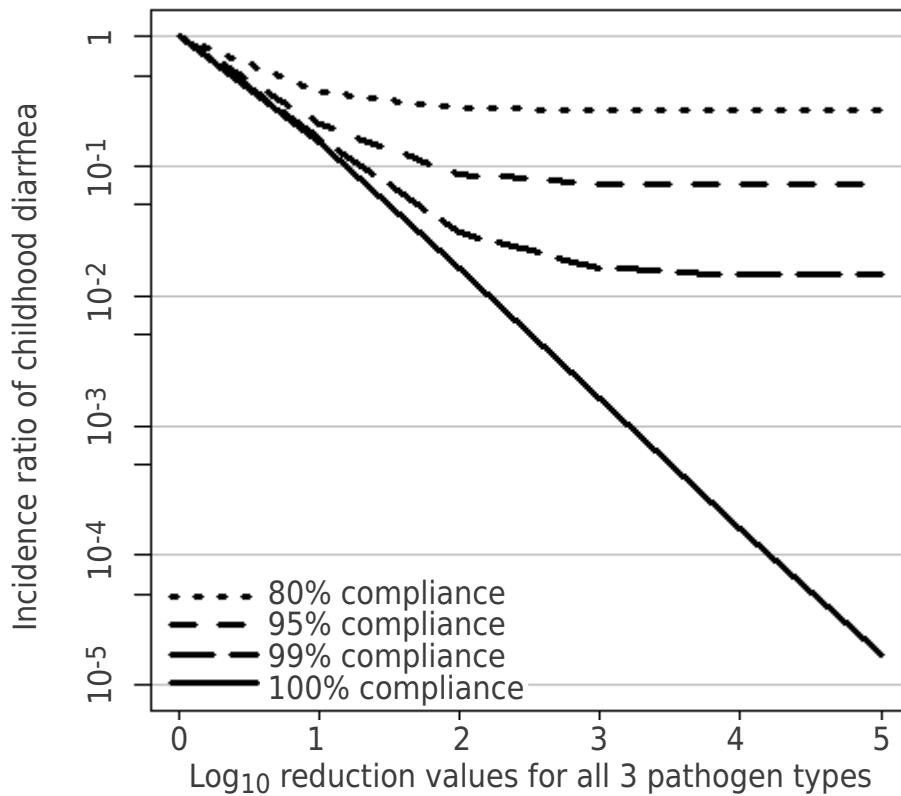


Colors apply to symbols in the same way as lines; for example, bacterial diarrhea in the two models can be compared by comparing the red lines to the red symbols.

Effect of LRVs given imperfect compliance

If compliance slightly decreased to 99% and there were no pathogen spikes, this model predicted little or no additional benefit from LRVs above 3 in many scenarios (e.g., Figure 4.4). If compliance is 80%, there was little benefit from increasing LRVs beyond 2. This behavior was similar regardless of compliance type (α , β , γ), pathogen mixture (A, B, C), or incidence level (low, medium, high) (Figures 4.10-4.13, page 145).

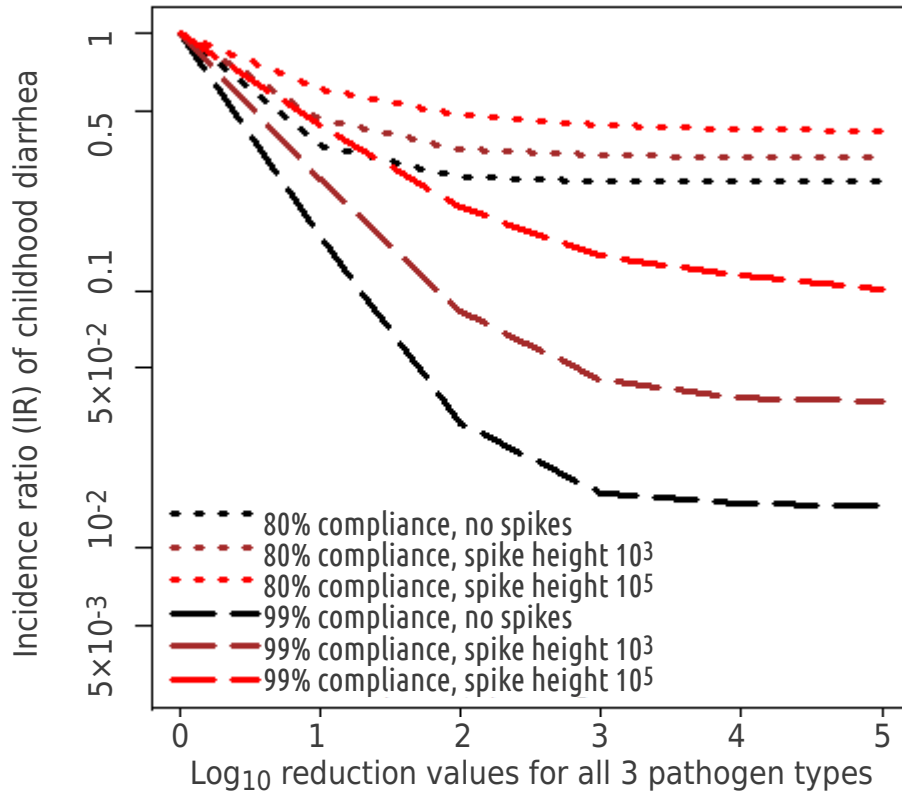
Figure 4.4. Effect of compliance with HWT on the incidence ratio of diarrhea, by LRV



Assuming medium incidence, compliance type β , no spikes, and pathogen mixture A. More scenarios are in section 4.4.3.

If pathogen spikes were included, the incidence ratio increased as spike height increases, for all LRVs (Figure 4.5). This effect was not due to an overall increase in incidence, because the model was calibrated to maintain the same incidence (baseline pathogen concentrations were also reduced to compensate for spike height; Equation 4.3). Rather, the increase in IR was due to the nonlinearity of the dose-response functions at high doses; during a large spike, reducing dose x -fold might only reduce risk by a factor less than x (Figure 4.6). Diminishing returns from LRV increases were still seen when spikes were introduced. Spikes 10 times above baseline gave similar results as no spikes (Figures 4.10-4.12, page 145).

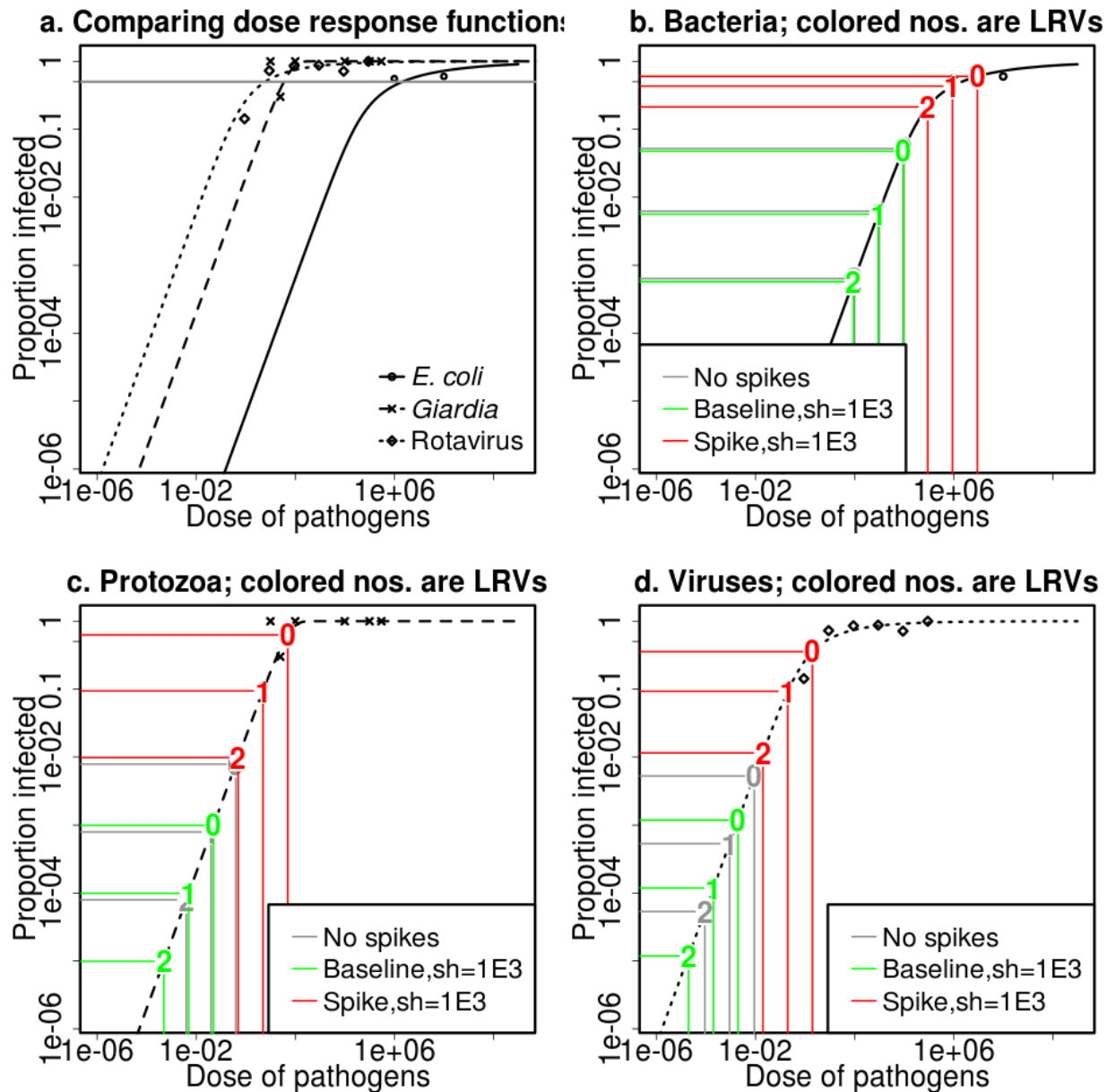
Figure 4.5. Effect of compliance and spikes on the IR of childhood diarrhea, by LRVs



Assuming medium incidence, compliance type β , and pathogen mixture A. More scenarios are in section 4.4.3.

Figure 4.6. Effect of dose response function nonlinearity at high doses

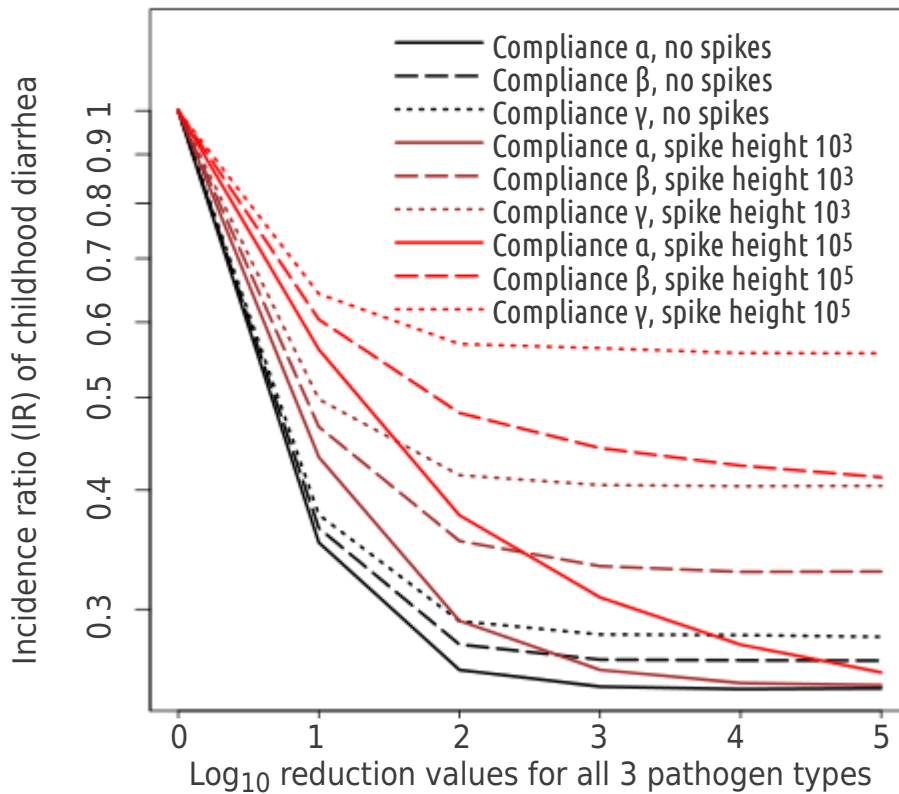
Example with etiologic fraction A, medium incidence, and spike height 1000



Since dose response functions are linear when doses are relatively low, reducing dose by a factor x also reduces the probability of infection by x . If spikes were absent, the pathogen concentrations used in the model were generally in the linear region. However, at high doses (such as during spikes), reducing the dose by x might only reduce the probability of infection by some factor y , where $y < x$. For example, assuming medium incidence and spike height (sh) of 1000, the daily dose of bacteria ingested during a spike was $\sim 10^7$ (chart b). Reducing the dose by 1 LRV, from 10^7 to 10^6 , reduced the probability of infection by $\sim 1/4$. This effect was less marked for the other two pathogen types (charts c & d), but it is still present. See Figure 3.3 (page 99) for a semilog plot of chart a.

Compliance type changed the relationship between LRVs, IR, and spikes (Figure 4.7). When there were no spikes, the results for compliance types α , β , and γ were similar. However, as spike height increased, α had the lowest IRs and γ had the highest IRs. Figure 4.7 also shows additional benefit from LRVs 4 and 5 for the highest spike scenario (10^5 -fold baseline). The benefits were greatest for compliance type α , in which children either complied perfectly or not at all. The benefits were smaller but still evident for β , in which children complied perfectly, partially, or not at all (see also Table 4.1, page 124). In contrast, under γ every child complied partially, always consuming some untreated water.

Figure 4.7 Incidence ratio of diarrhea by compliance level & type, spikes, and LRVs

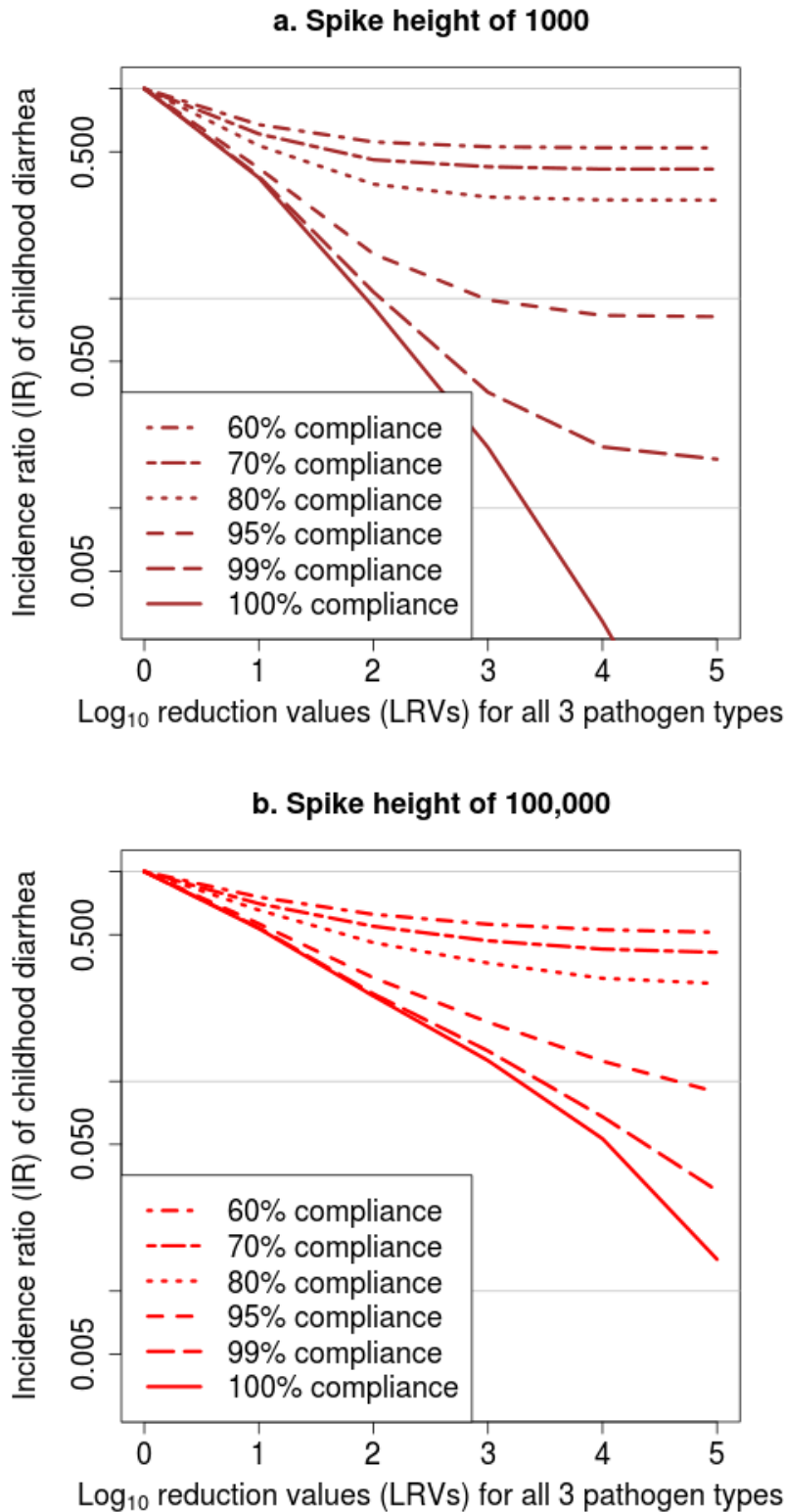


Assuming medium incidence, compliance of 0.8, and pathogen mixture A. See page 124 for a detailed description of compliance types.

Benefits from higher LRVs was more pronounced under conditions of high incidence, large

spikes, compliance type α , and high compliance (Figure 4.8). The benefits decreased as compliance decreased.

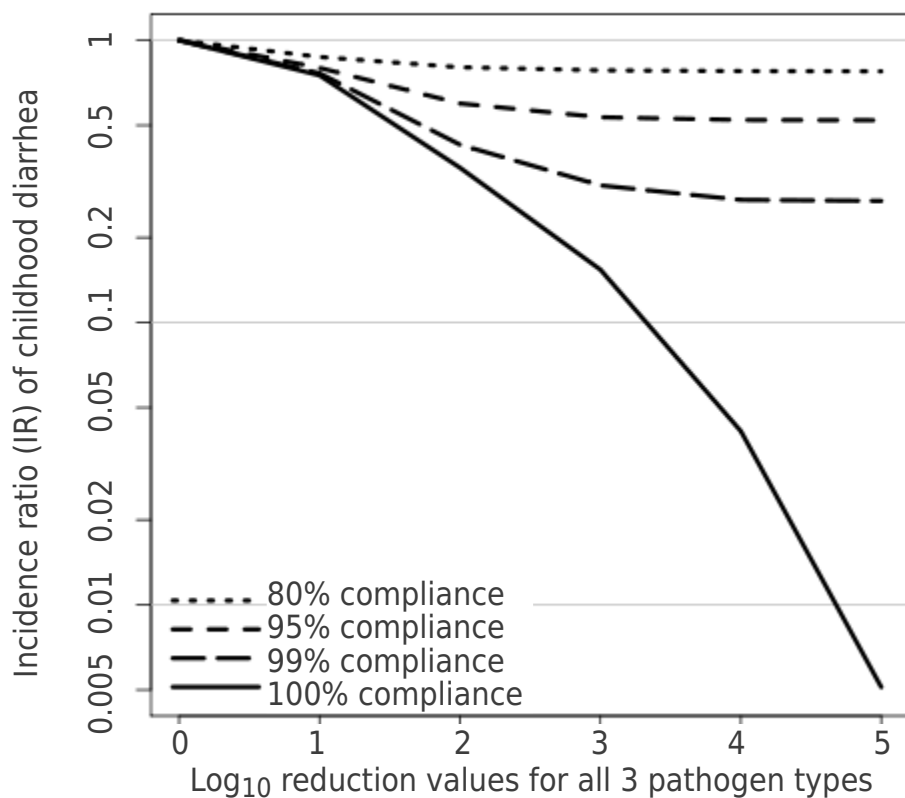
Figure 4.8. Effect of compliance on IR if large contamination spikes occur



Assuming medium incidence, compliance type α , and pathogen mixture C.

There is little information available regarding pathogen concentrations in source waters in developing countries; this is why the model was calibrated to obtain these pathogen concentrations in the first place. If the pathogen concentrations obtained from calibration are too low, that might make high-LRV HWT appear less effective than it actually was. This was evaluated with additional estimation runs using pathogen concentrations obtained by calibrating to high incidence (the three rightmost panels of Figure 4.2, page 133), and multiplying them by 10 or 100. An example is shown in Figure 4.9, and all such runs are shown in Figure 4.14 (page 150). Diminishing returns from increasing LRVs remained apparent. However, LRVs of 4 sometimes represented an improvement over LRVs of 3, particularly if incidence was high (Figure 4.14b) or if compliance was 99%.

Figure 4.9. Effect of compliance on IR with extreme pathogen concentrations



Estimation results using mean pathogen concentrations 100-fold the calibrated values for high incidence. Pathogen mixture A, calibration type β .

Pathogen mixture C tended to give lower IRs than A or B if incidence was high or spike height was high (Figures 4.10-4.12). As incidence was increased from low to high, the effects of pathogen mixture and compliance type increased.

4.4.3. Charts of median incidences from all estimation scenarios

Median incidence values from all estimation scenarios are shown below in Figures 4.10-4.14. Figures 4.4, 4.5, 4.7, and 4.8 above are subsets of Figures 4.10-4.12, expressed as incidence ratios (IRs) and charted as lines. Figure 4.13 is similar, but uses various combinations of LRVs for bacteria, viruses, and protozoa that are consistent with recent WHO guidelines (Sobsey & Joe

Brown, 2011) and USEPA guidelines for HWT (USEPA, 1987). Figures 4.10-4.14 display all combinations of overall compliance (0, 0.8, 0.95, 0.99, 1), compliance type (α , β , γ), and pathogen mixture (A, B, C) for a particular combination of incidence levels (low, medium, high) and spike scenarios (5 spikes per year that are 1 \times , 10 \times , 1000 \times , or 100,000 \times baseline mean pathogen concentration). Each symbol represents the median value of a distribution of estimation results from a scenario defined by a particular combination of these variables. Note that if compliance equals 0 (complete noncompliance, equivalent to LRV=0) or 1 (perfect compliance), then only compliance type α is possible.

Figure 4.10. Detailed estimation results (low incidence)

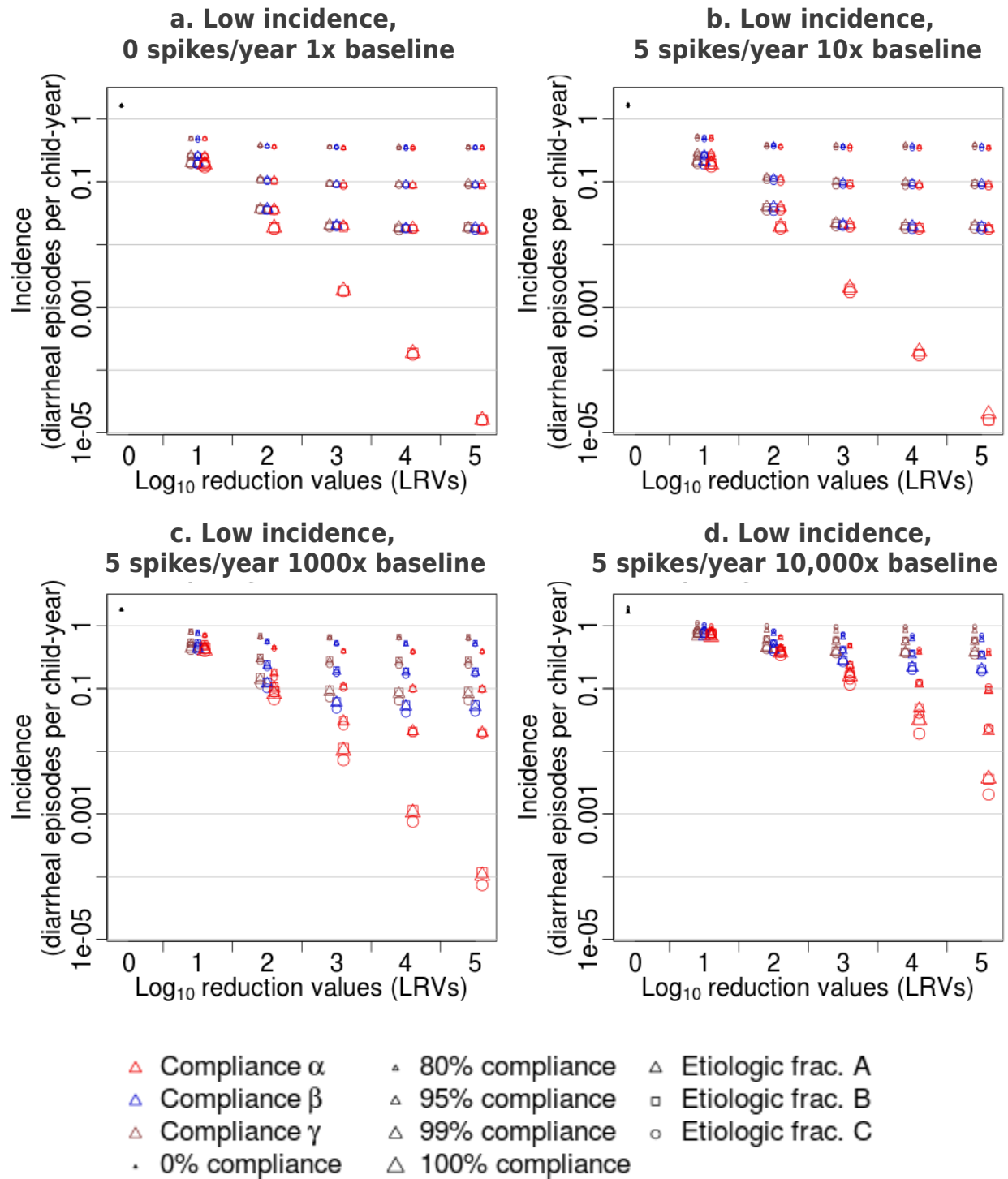


Figure 4.11. Detailed estimation results (medium incidence)

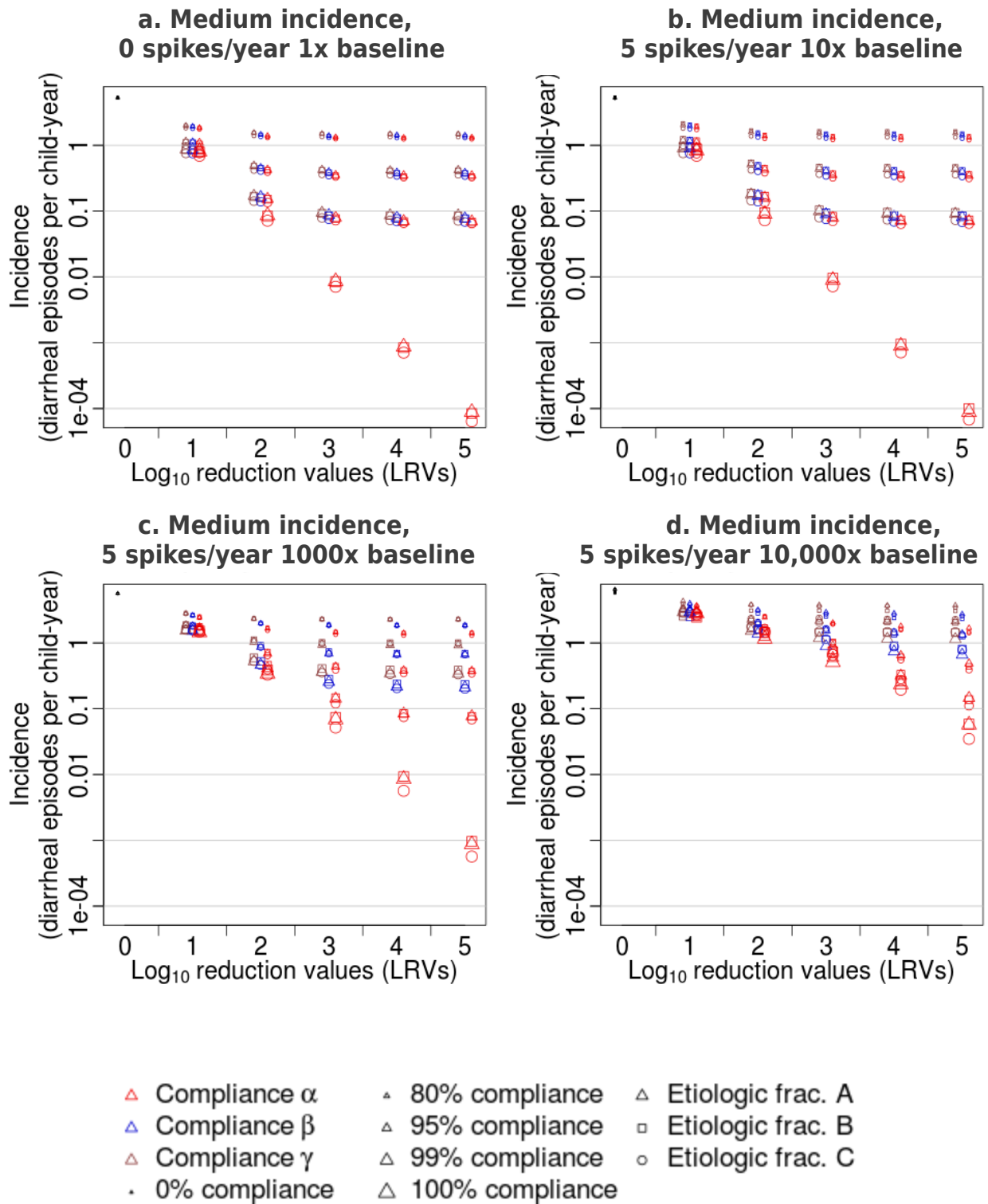
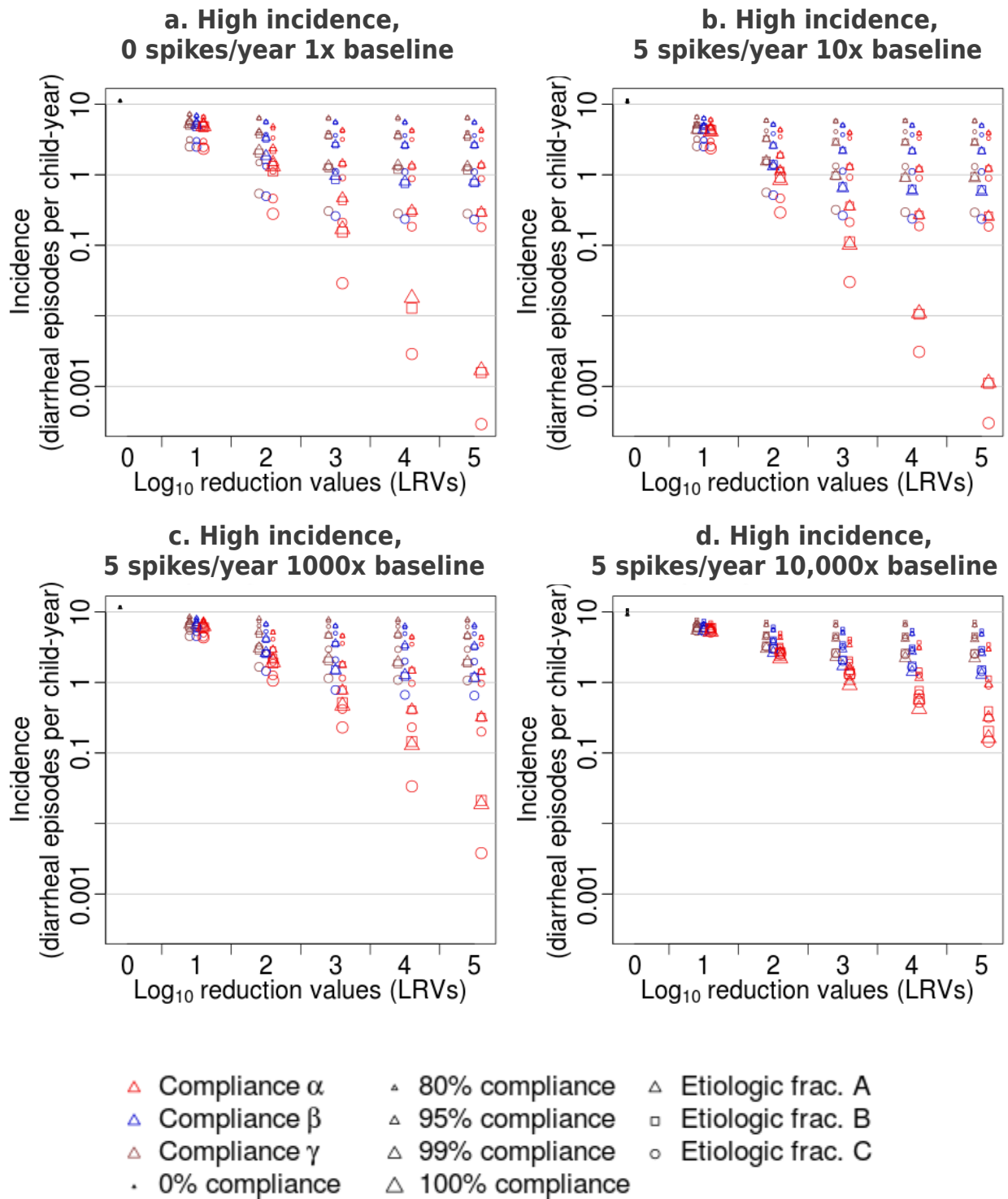


Figure 4.12. Detailed estimation results (high incidence)



Estimation steps were also run for LRV values that were consistent with existing recommendations from the WHO and the USEPA (Sobsey & Joe Brown, 2011; USEPA, 1987) (Figure 4.13). The recommendations consist of sets of three LRVs: bacterial, viral, and protozoan. These runs did not include spikes. They were divided into 3 groups: I) complete noncompliance, i.e., 0:0:0; II) LRVs consistent with the WHO interim target, where two of the three marker pathogens met the WHO protective type but the third had an LRV of 0; III) as II, but the third marker pathogen had an LRV of 1; IV) a series of LRVs including the WHO protective target of 2:3:2 and the WHO highly protective target of 4:5:4; and V) the USEPA standard of 6:4:3. The WHO protective target was generally more protective than the WHO interim targets, but there was little difference between the WHO protective target (2:3:2), the WHO highly protective target (4:5:4), and the USEPA target (6:4:3) unless compliance was 95% or higher.

Figure 4.13. Detailed estimation results, WHO/EPA recommended LRVs

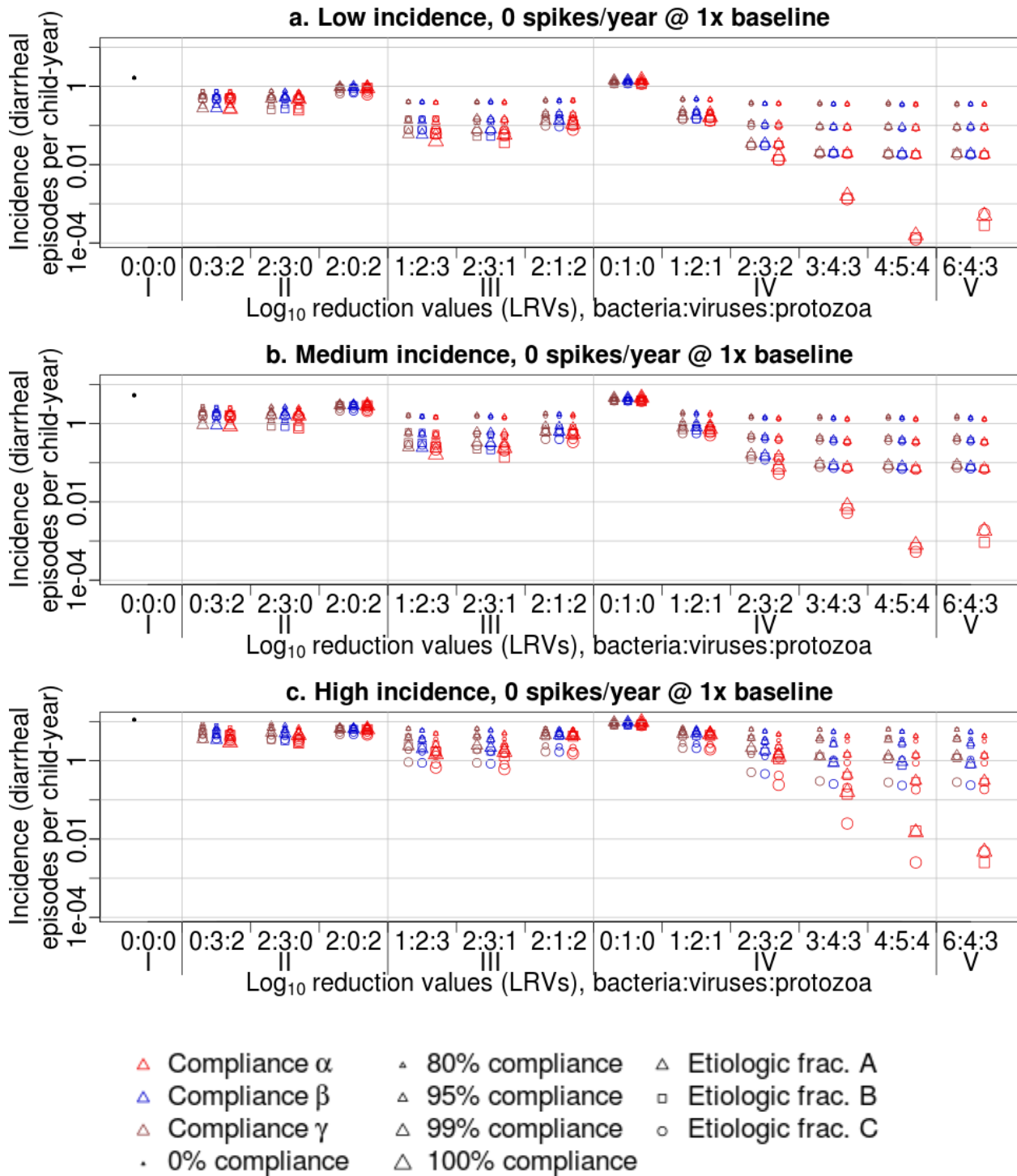
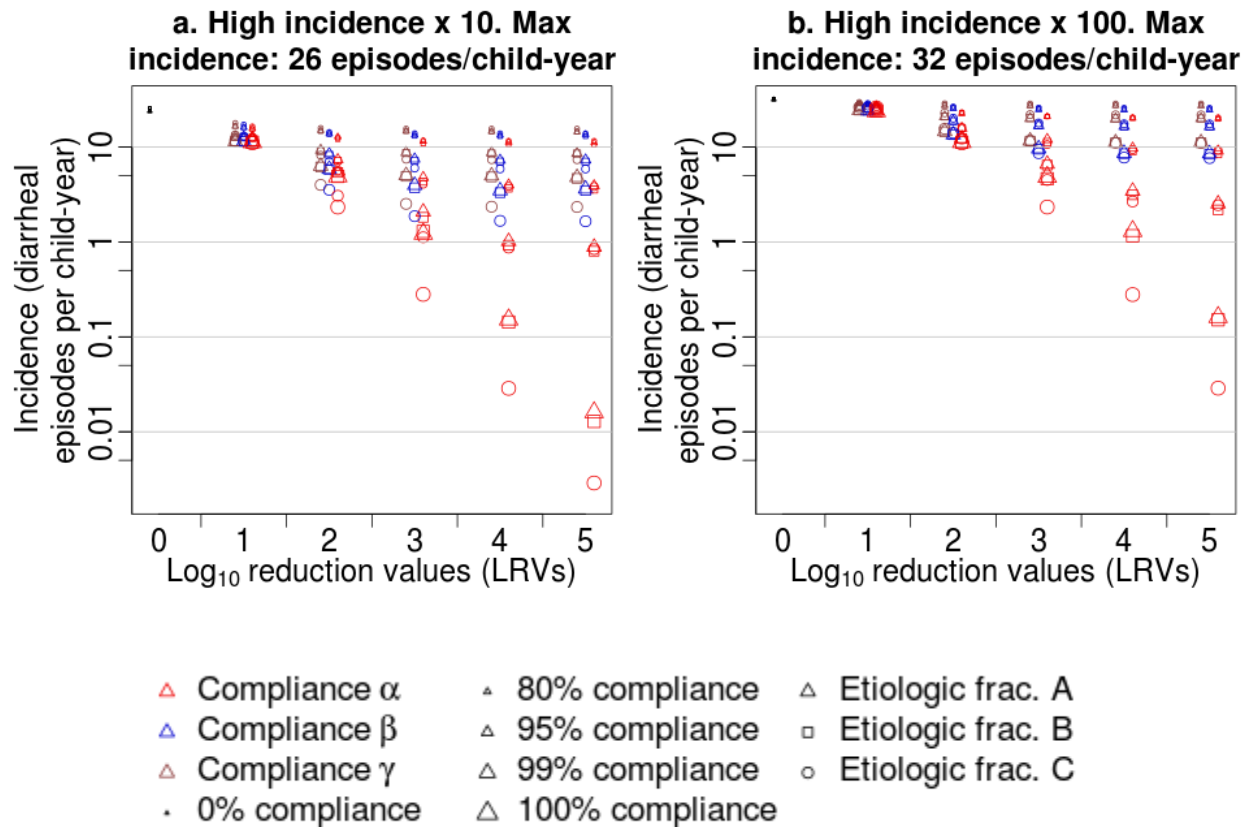


Figure 4.14. Detailed estimation results, extremely high baseline pathogen concentrations



4.4.4. Assessing impact of high LRVs: significance testing & classification trees

The Wilcoxon rank sum test was used to assess whether incidence significantly differed depending on LRVs. This was done by considering all scenarios in Figures 4.10-4.12 except those with overall compliance of 0 or 1, and comparing pairs of scenarios that had identical compliance levels, compliance type, etiologic fractions, and spike height, but had differing LRVs (3 vs. 5 for all pathogen types). The Wilcoxon test detected a difference between the two incidence distributions at $p < 0.05$ in 153 of the 324 comparisons (47.2%). However, because a statistically significant difference does not necessarily mean an important difference, two measures of importance (which were chosen *a priori*) were further considered: an incidence difference (ID) > 0.2 diarrheal episodes per child-year, and an incidence ratio (IR) < 0.9 .

Classification trees were then constructed (rpart package version 3.1. for R) to describe which scenarios showed improvement in diarrheal incidence if LRVs were increased from 3 to 5.

The ID criterion was more restrictive (fewer scenarios met it than the IR criterion) and therefore the tree was simpler (Figure 4.15). Most of the scenarios meeting the criterion (32/43) had a spike height of 100,000, and those also had the following 2 characteristics: 1) medium or high incidence; and 2) some perfect compliers in the community (i.e., compliance types α or β). Of the remaining 11 scenarios that met the criterion but had a spike height $< 10^5$, all of these had high incidence and a spike height of 1000; further, 10 of them were compliance types α or β . In addition, 18 of the 43 scenarios that met the criterion had an overall compliance of 99%.

The IR tree was more permissive (94/324 scenarios met the criterion) and the tree was structured somewhat differently (Figure 4.16). However, 71/94 had spike heights of 1000 or 10^5 , and 61/94 had overall compliance of 99%. As with the ID tree, compliance types α and β tended to meet the criterion more frequently than γ , and higher incidence also met the criterion more frequently than medium or low incidence.

Figure 4.15. Classification tree for incidence difference (ID) criterion

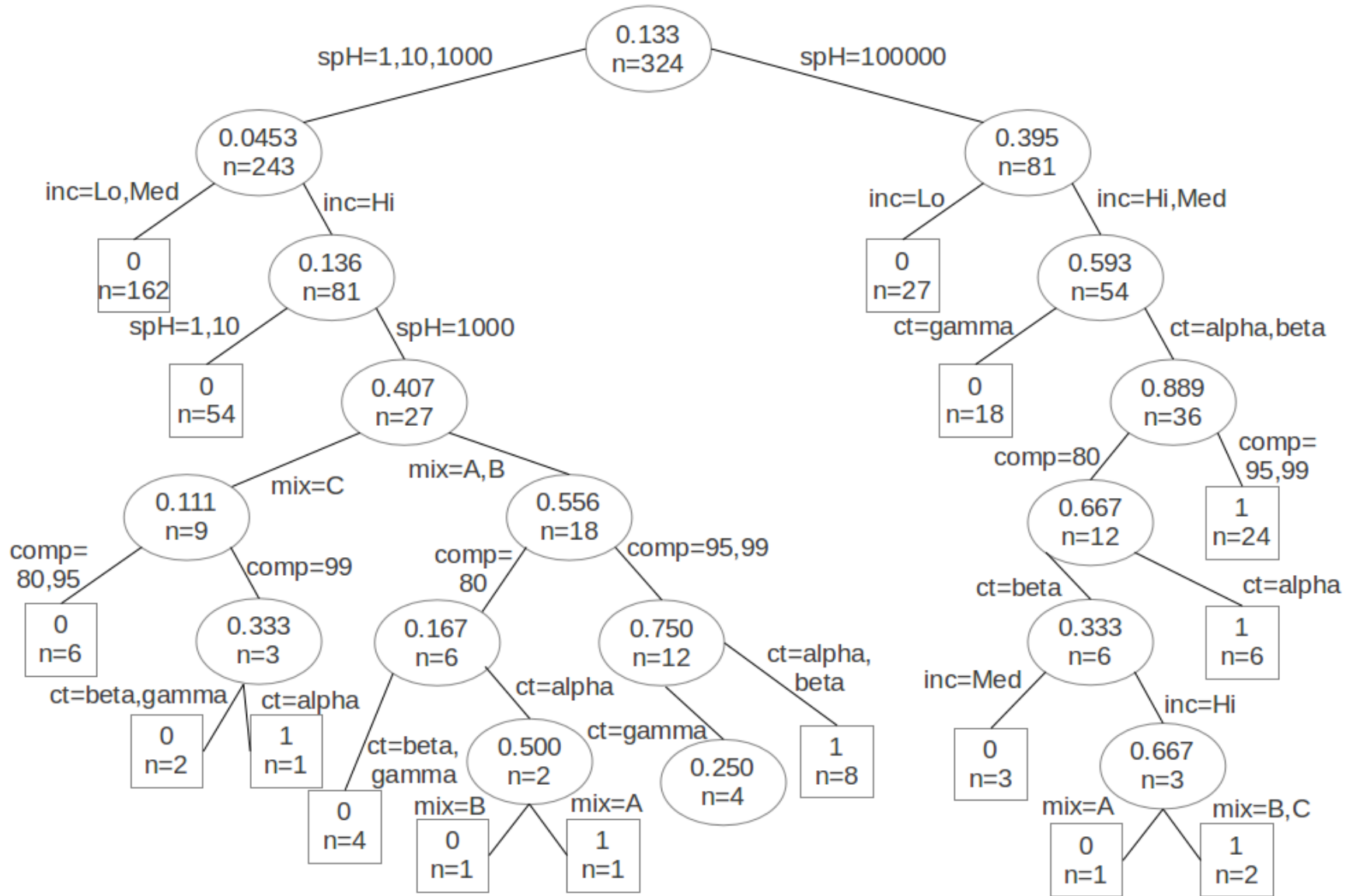
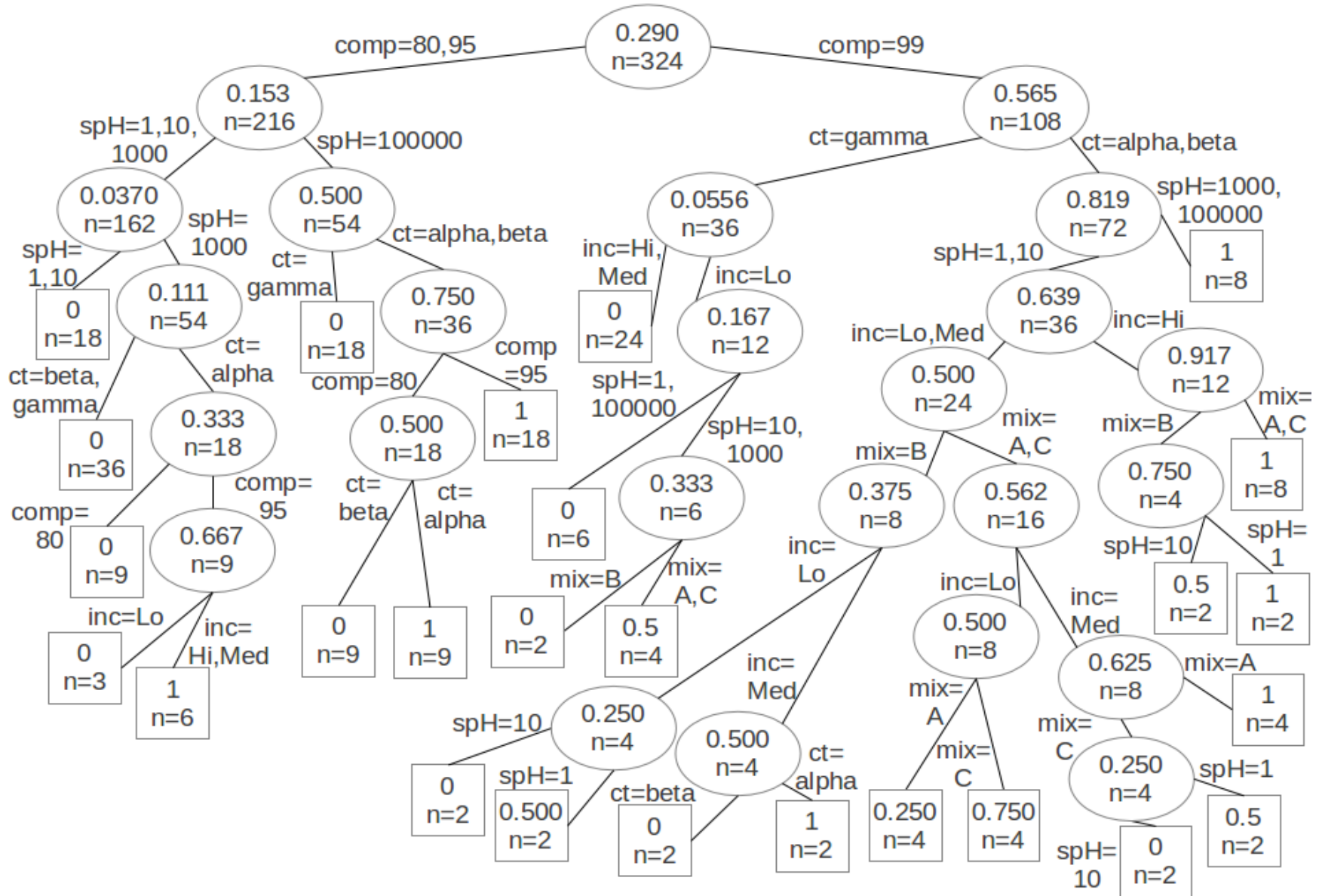


Figure 4.16. Classification tree for incidence ratio (IR) criterion



4.5. Discussion

The model developed in this chapter indicated that the risk of diarrhea decreases linearly (on a log-log scale) with pathogen removal by HWT under perfect compliance conditions. This is a direct consequence of the fact that the dose response relationships are linear in the range of pathogen doses that individuals usually receive (Figure 4.6, page 138). These results are somewhat consistent with those reported in the WHO guidelines on health-based targets for HWT devices (Sobsey & Joe Brown, 2011). Although the model used in this analysis has a similar QMRA approach as the WHO model, there are some important distinctions. For example, the three indicator pathogens used here were pathogenic *E. coli*, rotavirus, and *Giardia*, whereas the WHO guidelines used *Campylobacter*, rotavirus, and *Cryptosporidium*. Time-varying pathogen concentrations were calibrated to reflect realistic diarrhea incidence levels, as well as realistic etiologic fractions of pathogens. The WHO guidelines assumed constant concentrations of the pathogens in sewage, and that drinking water was contaminated with 0.01% sewage. Most importantly, this model relaxed the assumption of perfect compliance, examining the joint effects of varying compliance and LRVs. Assuming imperfect compliance, its results differed greatly from the WHO model, particularly for higher LRVs.

For many of the scenarios with imperfect compliance, diminishing health improvements from increasing LRVs were observed; similar conclusions from a differently structured QMRA model were published recently (Joe Brown & T. Clasen, 2012). Specifically, when the variation in pathogen concentration was limited to 25 fold (i.e., no spikes) and compliance was 99%, little additional diarrhea was prevented for LRVs above 3 (Figure 4.4, page 136). Assuming 80% compliance, LRVs above 2 prevented little additional diarrhea. If spikes occurred and some of the population complied perfectly (compliance types α or β), LRVs above 3 sometimes prevented additional episodes of diarrhea.

These results indicate the importance of including compliance in risk estimations and in policy development, and also emphasize the importance of understanding the different dimensions of compliance. For example, some people may never comply, others may comply when they are home but not when they are away from home, and yet others may comply only during periods of perceived high risk. These simulations suggest that, given a particular overall compliance level within a community (i.e., a proportion of person-time spent complying), HWT scenarios that include more perfectly complying individuals prevent more diarrhea. Although the implications of these different dimensions of compliance are not well understood, it is clearly difficult to obtain long-term, high compliance with household interventions in developing countries (Makutsa et al., 2001; B. Arnold et al., 2009; Luby et al., 2009).

Difficulty in achieving high compliance with intervention strategies also extends to sanitation and hygiene interventions. Handwashing compliance is incomplete in both industrialized (Bischoff et al., 2000) and developing countries, especially with soap (V. A. Curtis & Cairncross, 2003). Despite the obvious importance of sanitation in removing pathogens from the environment and breaking the fecal-oral cycle of transmission, approximately half the population of southern Asia and sub-Saharan Africa openly defecates or has an unimproved latrine (World Health Organization & UNICEF, 2010). Even if latrines are available, they might not be used consistently (B. F. Arnold et al., 2010; Banda et al., 2007; Montgomery et al., 2010).

4.5.1. Information needed to inform models of diarrheal infection transmission

Although many of the modeled scenarios had diminishing health improvements beyond 3 LRVs (and sometimes beyond 2 LRVs), scenarios were identified where LRVs above 3 were beneficial (e.g., Figures 4.7 and 4.8, page 139). Understanding which scenarios are most realistic requires a better characterization of the variability in pathogen concentrations in source waters, the relative proportions of pathogens in contaminated water, and the extent to which these

pathogens are also transmitted through other environmental pathways. These issues are further discussed below.

Pathogen concentrations in source waters

Little information is available on the variability of pathogen concentrations in source waters. Even point measurements of pathogen concentrations are scarce (Enger et al., 2012), and it is unclear what reasonable spike concentrations would be. However, contamination spikes are plausible due to various mechanisms, including stormwater runoff, defecation directly into source waters, or washing of contaminated items like diapers. In the absence of spikes, we assumed that the daily concentration of pathogens varied over a 25-fold range, consistent with measurements of thermotolerant coliforms in source water in rural Congo (Boisson et al., 2010) and *E. coli* in a rural Ecuadorian stream (K. Levy, A. E. Hubbard, K. L. Nelson, et al., 2009).

Etiology of diarrheal disease

The contribution of different pathogens to diarrheal disease is also uncertain and depends upon ecology, sociology, and infrastructure. Published etiologic fractions include diarrhea from all transmission routes, not only drinking water; pathogen profiles for different routes, for example food versus water, will differ. In addition, the true distribution of etiologies may differ from the distribution of reported etiologies. For example, certain bacteria may be more frequently identified because they are easier to culture. For this analysis, three broad mixtures of etiologic fractions were chosen, based on the most comprehensive information available. Future research, particularly from the Global Enterics Multicenter Study (University of Maryland, 2012), will further clarify diarrheal etiology.

Routes of transmission other than drinking water

Finally, this model only accounts for infection via drinking water. Additional routes of transmission (e.g., contaminated hands, objects, or food) operate in underdeveloped

communities. Considering these routes would decrease the apparent effectiveness of a HWT device, since these routes would affect users and nonusers of HWT alike. This model also does not account for infection transmission between individuals. Effective HWT would reduce the number of infected people, thus reducing pathogen shedding, thus indirectly preventing infection in people not using HWT. This would increase the apparent effectiveness of HWT, assuming imperfect compliance (Halloran et al., 1991). Although examining effectiveness in the context of multiple transmission pathways is important, it probably would not affect our general conclusions about the joint effects of compliance and LRVs on the effectiveness of HWT interventions.

4.5.2. Conclusions

Recent WHO guidelines (Sobsey & Joe Brown, 2011) provide an important framework for evaluating the health benefits of HWT devices resulting from their LRVs. However, the simulation results presented here indicate that prevention of diarrhea by HWT is limited by compliance. Thus, the classification system in the WHO guidelines incompletely informs HWT users and promoters regarding effectiveness of devices if < 100% of drinking water is treated. For promoters of HWT, these simulation results emphasize that facilitating consistent, sustained use is extremely important when deciding which devices to use for a program, in addition to the antimicrobial efficacy of the device.

HWT cannot greatly reduce transmission of diarrheal disease by drinking water unless compliance is high and sustained. More research is necessary to understand the full complexities of compliance, to explicitly measure compliance in intervention trials, and to incorporate compliance in development policy. This chapter provides a modeling framework that examines the impact of compliance on the effectiveness of interventions, as an initial step toward more complete consideration of compliance by researchers, policymakers, and development workers.

5. TRANSMISSION MODEL OF DIARRHEAL INFECTION

5.1. Abstract

Although quantitative microbial risk assessment (QMRA) models are useful tools for describing infectious disease risks in many different circumstances, they ignore two important characteristics of particular importance for diarrheal infections in developing countries: 1) secondary transmission effects, and 2) multiple transmission routes.

To further investigate whether: 1) diminishing returns are seen with increasing HWT \log_{10} reduction values (LRVs) under incomplete compliance; and 2) whether the pattern of compliance within a community affects diarrhea prevention, an environmental infection transmission system (EITS) model was constructed. The model incorporated household structure, multiple routes of transmission (including drinking water, environmental exposure, and contacts between households), and pathogen shedding by infected individuals. It also included simultaneous transmission of bacterial, viral, and protozoan pathogens. Although the model can simulate the effects of additional interventions, including sanitation, handwashing, and safe storage of treated drinking water, this work concentrates upon HWT only.

A calibration step determined values for seven parameters in the EITS model that were consistent with high diarrheal incidence in developing countries. Subsequently, those parameter values were reused in estimation steps that applied HWT interventions with varying levels of antimicrobial effectiveness and varying patterns of compliance to simulated communities. The results resembled previous conclusions from QMRA models: 1) LRVs above 3 seldom prevent additional diarrhea, and 2) the pattern of compliance alters HWT effectiveness, even if overall compliance was held constant. In contrast to the QMRA models, the EITS model indicated that LRVs above 3 prevented little additional diarrhea, even when compliance was perfect.

5.2. Introduction

Infection transmission models have been often been used to better understand infectious diseases, with the goal of improving the health of the population by controlling or stopping transmission (Keeling & Rohani, 2008). However, transmission models have seldom been applied to diarrheal disease in developing countries, a particularly severe and widespread public health problem. Although it is clear that clean drinking water, effective sanitation, and hygienic behavior can effectively control diarrheal infections, the best ways to achieve these goals are far from clear, given the serious resource constraints in developing countries.

Modeling systems of transmission of diarrheal infections would allow simulation of various interventions, providing insight on which interventions, or combinations of interventions, might yield the largest reductions in diarrheal disease. Environmental infection transmission system (EITS) models (Li et al., 2009) are particularly suited to modeling diarrheal infections because they can explicitly describe the diverse pathways that pathogens can take through the environment (e.g. Figure 2.7, page 77). Furthermore, substantial information is available regarding pathogen inactivation or removal by various interventions (Table 2.1, page 64), which can be simulated in an EITS model by removing the appropriate proportion of pathogens from particular compartments at appropriate times.

The QMRA model in chapter 4 described diminishing returns from increasing household water treatment (HWT) \log_{10} reduction values (LRVs) when compliance with HWT is imperfect. However, that model did not account for any secondary transmission effects of HWT. If an intervention is used by some households in a community, those households will benefit because some infections will be prevented; however, because complying households have fewer infections, they release fewer pathogens into the environment, and consequently non-complying

households benefit also. These secondary transmission effects would prevent additional disease beyond what was measured in the QMRA model in chapter 4. They might also affect the point at which diminishing returns are seen from increasing LRVs when compliance is imperfect.

The primary goal of the EITS model described below was to determine whether increasing LRVs from HWT still leads to diminishing returns given imperfect compliance, in the context of a model with secondary transmission and multiple transmission routes. In addition, the EITS model is well-suited to exploring additional questions (which time considerations did not permit including in this chapter), such as:

- The nature of interaction between two joint interventions: under what circumstances is it positive, negative, or absent? Previously published research (see page 46 for a summary) suggests two hypotheses:
 - Sanitation usually interacts positively with other interventions.
 - Two non-sanitation interventions applied jointly usually interact negatively.
- How much diarrhea prevented by an intervention is due to immediate effects on those who use it, and how much is due to indirect effects, since healthier people shed fewer pathogens, reducing risk to the community as a whole?

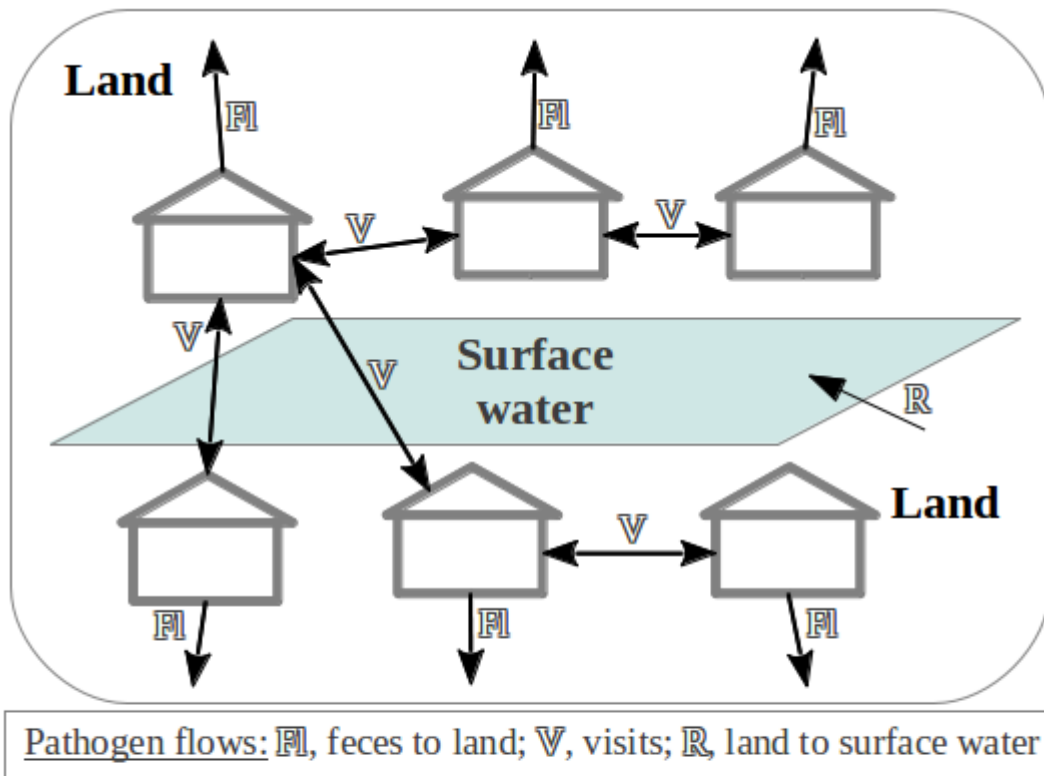
5.3. Materials and methods

5.3.1. General description of the model

An EITS model was programmed using MATLAB software (version 7.13, R2011b) that simulated a small isolated community in a developing country (parameter values for the model are summarized in chapter 7 and Table 7.1, page 230). The community consisted of 200 households with an average of 5 people per household. The community gathered their drinking water from a surface water source (Figure 5.1). The system had four types of compartments

containing pathogens: 1) the land where the community resided; 2) the surface water from which the community obtained drinking water; 3) stored drinking water within each household; and 4) the household environment, which represented the hands of household members as well as fomites and surfaces within the household. Since each household (Figure 5.2) contained a stored drinking water compartment and a household environment compartment, and there was only one land compartment and one surface water compartment in the community, each model run contained $2h + 2$ distinct compartments, where h is the number of households in the community. Each compartment contained pathogens of three distinct types: bacteria, viruses, and protozoa. Each household contained variable numbers of people, who could be either young children less than five years old (18% of the population), or children/adults aged 5 years or more (82% of the population). These two types of people are called 'children' and 'adults' for brevity. The only source of pathogens was infected people, who contaminated the community's land and their household environment through defecation. Pathogens gradually moved from land into surface water; randomly scheduled rainfall events increased the rate of this transfer. Households were randomly connected with one another, allowing exchange of pathogens between households (Figure 5.1). All pathogens were attenuated exponentially over time in all compartments, representing a variety of processes that can remove pathogens completely from the system (e.g., inactivation; sedimentation; transport in flowing water; percolation below the soil surface; etc.).

Figure 5.1. Simplified overview of the simulated community

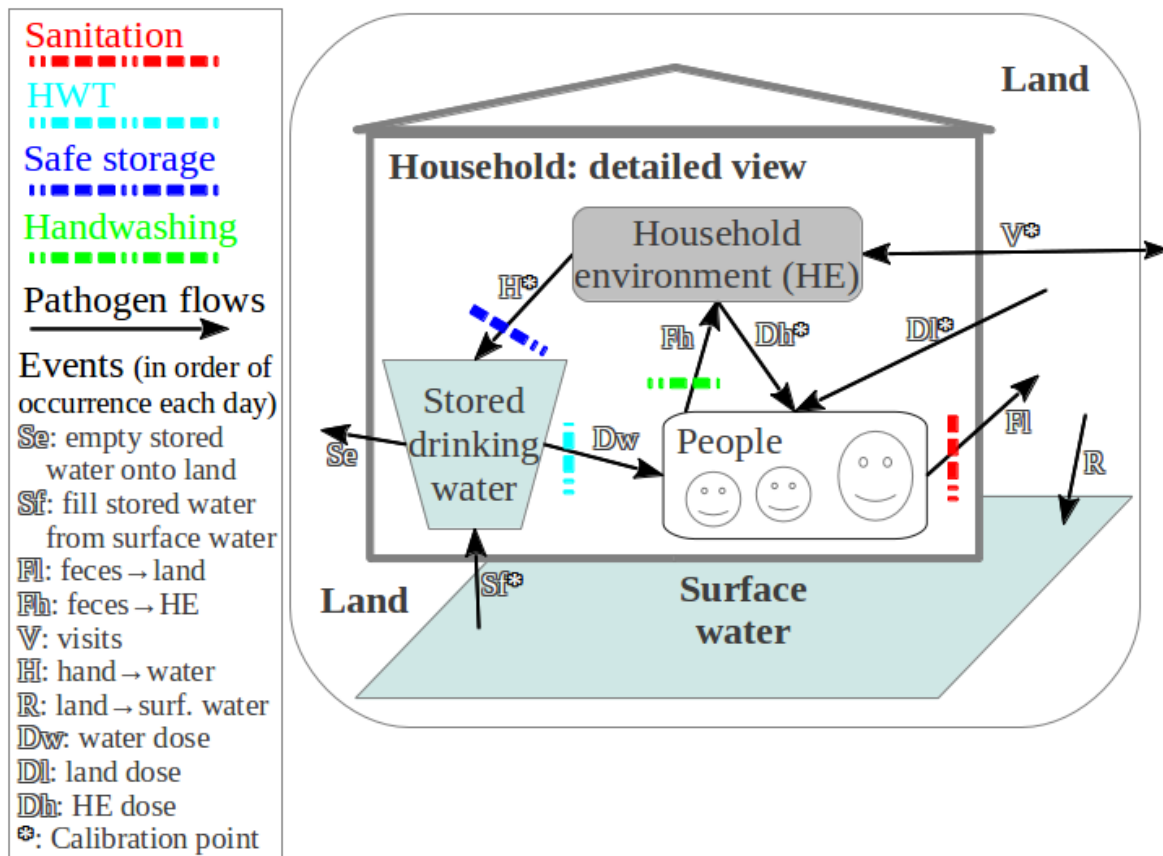


Each simulated community contained about 200 households.

Each household contained a group of people and two compartments for pathogens: 1) a container for stored water, and 2) a more abstract household environment compartment representing pathogens that were available for ingestion on hands, objects, and surfaces (Figure 5.2). Stored water was collected at the beginning of each day from the community's surface water; the stored water could subsequently be contaminated by pathogens in the household environment. Four types of interventions could operate: household water treatment (HWT) destroyed pathogens in the stored water immediately after collection, safe storage prevented recontamination of stored water by hands, handwashing prevented contamination of hands after defecation, and sanitation prevented contamination of land by defecation. Households could

comply perfectly, incompletely, or not at all with these interventions. Each day, each person ingested variable doses of three pathogen types (bacteria, viruses, and protozoa, represented by diarrheagenic *Escherichia coli*, rotavirus, and *Giardia duodenalis*) from their household's stored drinking water, their hands, and the land outside the household.

Figure 5.2. Structure of each household within the simulated community

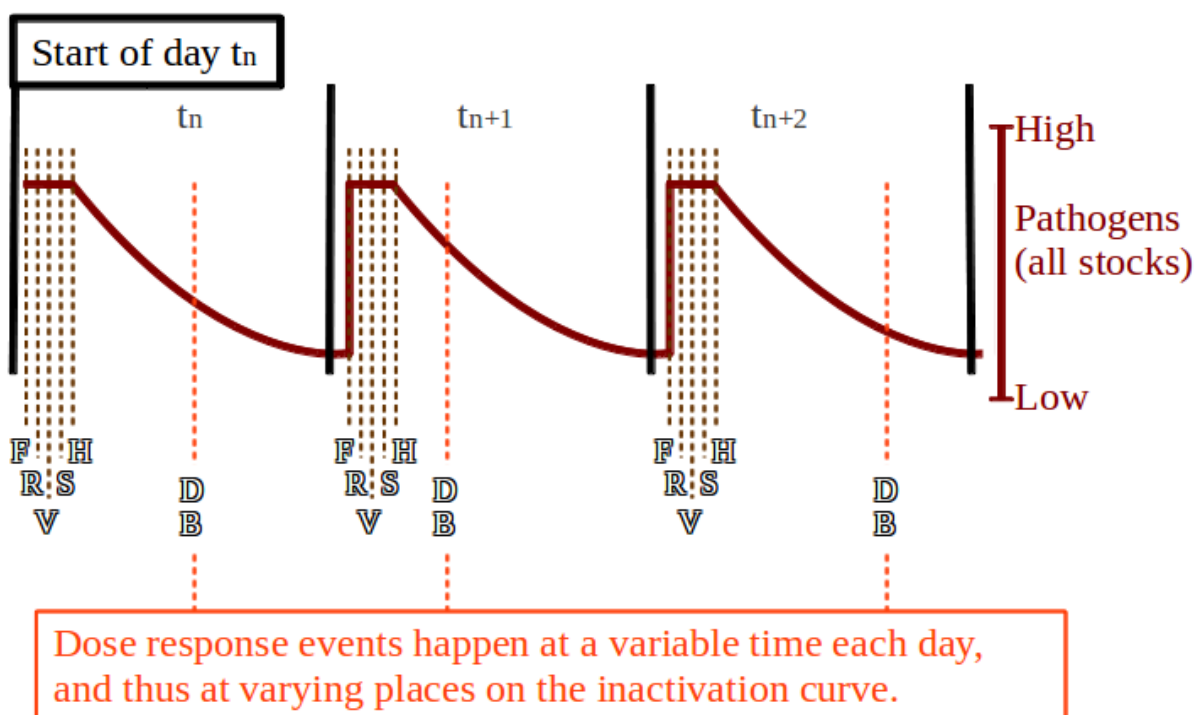


The four transfer calibration parameters marked by '*' influence movement of pathogens between compartments. There are also three attenuation calibration parameters describing attenuation of the three pathogen types in all compartments; they are not shown in this figure, but see Table 5.1 (page 169).

The model used discrete timesteps of one day. Each day included four types of events: 1) contamination events, which represented defecation and were the only source of new pathogens in the system; 2) pathogen transfers, which described movement of pathogens between

compartments; 3) exposure events, which described ingestion of pathogens by people and the subsequent possibility of new infections (i.e., transfers of pathogens from compartments to people); and 4) pathogen attenuation, which described removal of infectious pathogens from the system in each compartment over time (corresponding to inactivation, sequestration, or any other means that could render pathogens unable to ever contact a host). Since many of the parameter values describing these events are highly uncertain, seven abstract 'calibration parameters' (see Figure 5.2, page 163, and Tables 5.1 and 5.2, page 169) were varied during each model run in a calibration step consisting of many thousands of model runs; values of these parameters that yielded acceptable results were retained for use in subsequent estimation steps. The events occurred in a particular sequence, as shown in Figure 5.3. To avoid arbitrarily biasing the doses of pathogens ingested by people, exposure events happened at a variable time each day. This meant that pathogens introduced into the system at the beginning of each day would be attenuated for some random fraction of a day before they were ingested, introducing additional variability into the doses that people received.

Figure 5.3. Daily progression of the EITS model of diarrheal infections



Events, in order of occurrence each day

- F**: Defecation adds pathogens to land (F_l) & household environment (F_h), reduced by sanitation & handwashing.
- R**: 'Runoff': pathogen transfer, land to surface water. Temporarily increased by random rain events.
- V**: Visits transfer pathogens between households' environments.
- S**: Each household empties their stored drinking water (S_e), stores fresh surface water (S_f), and might apply household water treatment (HWT).
- H**: Household environment pathogens contaminate stored water; safe storage prevents this.
- D**: People ingest pathogen doses: stored water (D_w), land (D_l), & household environment (D_h).
- B**: Baseline exposures are randomly assigned.

Summary of the steps in each model run

The events marked in Figures 5.1-5.3 are explained in more detail below; events that include calibration parameters are marked with a *.

1. There were two contamination events representing defecation. The total number of pathogens excreted per infected person is determined by the grams of feces excreted per

person per day multiplied by the number of pathogens per gram of feces. Adults excrete more pathogens than children due to higher fecal output, and symptomatic people likewise excrete more pathogens than asymptomatic people due to higher fecal output.

The pathogens excreted by a person are distributed as follows:

a) Contamination of the land (Fl)

- Most of the pathogens excreted by infected individuals follow this route.

b) Contamination of the person's household environment (Fh); summarizes hand, surface, and fomite contamination related to anal cleansing

- A small proportion of the pathogens excreted by infected individuals follow this route.

2. There are four main pathogen transfers:

a) Land to surface water (R); summarizes runoff due to rainfall events, soil erosion, people washing or bathing in the water, etc.

- This event is defined by daily rates of pathogen transfer from land to water. There are two rates: a lower one for non-rainy days and a higher one for rainy days (rain events occur randomly, 14 days apart on average).

b) * Surface water to stored drinking water (Se and Sf), representing daily resupply of stored water by each household

- All pathogens remaining in each stored drinking water compartment at the beginning of each day are transferred onto the land (Se); this is a very small transfer, but is included for completeness.
- Each household's stored drinking water is refilled directly from surface water at the beginning of each day (Sf). The amount of pathogens transferred is

determined by multiplying the pathogens in surface water by a calibration parameter (CP_{Sf}); this is analogous to dilution.

c) * Household environment to stored water (H)

- The transfer of pathogens from each household's environment to its stored drinking water was determined through multiplication by a calibration parameter ($CP_{H\&Dh}$); event 3b (below) also used $CP_{H\&Dh}$.

d) * Transfer of pathogens by inter-household visits (V), representing pathogens carried between households on hands, fomites, food, etc.

- A set of two-way links between households was randomly created at the beginning of each model run. Each day, a subset of these links was randomly selected, signifying visits occurring that day. Each visit represented a two-way transfer of pathogens between the households' environment compartments. Transfer of pathogens from household A to household B was: (number of pathogens in household A's environment) / (number of people in household A) * (calibration parameter CP_V denoting proportion of pathogens transferred). Household B simultaneously transferred pathogens to household A in the same manner.

3. There were four exposure events, where susceptible people may become exposed and subsequently develop infection or disease.

a) Drinking water ingestion (Dw)

- Determined by: (liters of water consumed daily) \times (number of pathogens in stored drinking water) / (volume of stored drinking water container).

b) * Ingestion of pathogens from the contaminated household environment (Dh)

- The number of pathogens in the household environment multiplied by a calibration parameter ($CP_{H\&Dh}$; see also event 2c above).
- c) * Ingestion of pathogens from land (DI); summarizing people playing or working in soil, consumption of locally grown food, etc.
- The number of hand-mouth events per day multiplied by a calibration parameter (CP_{DI}). Each person's dose was subtracted from the outside environment.
- d) Baseline exposure (B)
- Each day, people were randomly chosen to become exposed to each of the three pathogen types, independently from the doses they had received. This simulated importation of infection.
4. Exponential attenuation of all pathogens in all compartments:
- a) * Decay of pathogens in all compartments was based on published rates in raw water, multiplied by 3 independent calibration parameters (one for each pathogen type; CP_{atten}).

Table 5.1. Summary of calibration parameters

Transfer calibration parameters	
CP_{Sf}	Transfer of pathogens from surface water to stored drinking water (analogous to dilution of pathogens in surface water)
$CP_{H\&Dh}$	Transfer of pathogens from the household environment to: 1) stored water; or 2) people, who then ingest them
CP_V	Transfer of pathogens between the household environment compartments of a pair of households
CP_{DI}	Transfer of pathogens from land to people (who then ingest them)
Attenuation calibration parameters	
CP_{atten} (bacteria)	Daily exponential inactivation/removal rate in all compartments, bacteria.
CP_{atten} (viruses)	Daily exponential inactivation/removal rate in all compartments, viruses.
CP_{atten} (protozoa)	Daily exponential inactivation/removal rate in all compartments, protozoa.

Different calibration parameter values were sampled from uniform distributions (Table 5.2) for each calibration run of the model.

Table 5.2. Ranges over which calibration parameters were sampled

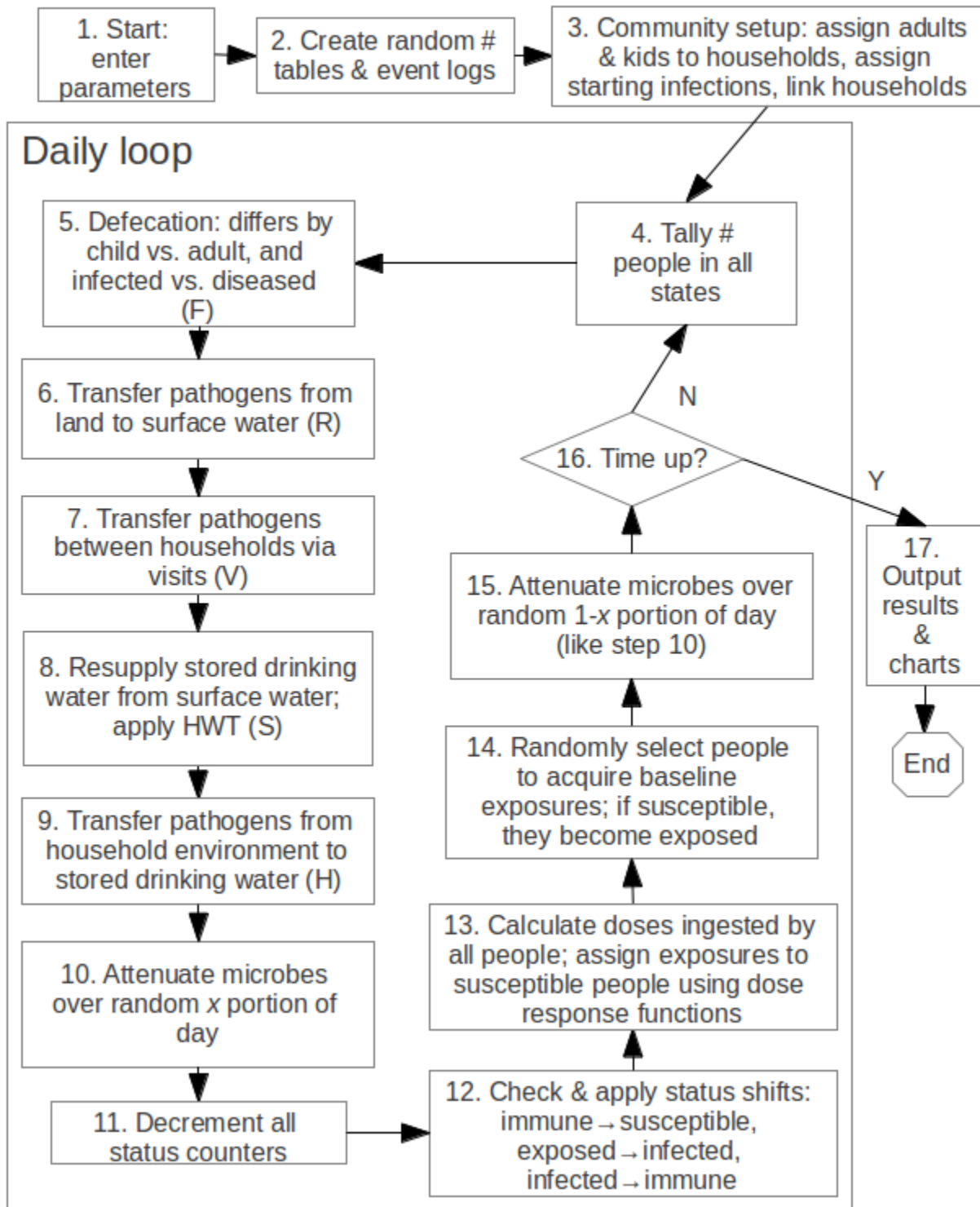
Calibration parameter	Lower limit	Upper limit
Transfer calibration parameters (proportions of pathogens transferred)		
CP_{Sf}	10^{-6}	$10^{-2.4} \approx 4.0 \times 10^{-3}$
$CP_{H\&Dh}$	10^{-5}	$10^{-1.5} \approx 0.032$
CP_V	10^{-5}	$10^{-0.5} \approx 0.32$
CP_{DI}	$10^{-11.5} \approx 3.2 \times 10^{-12}$	10^{-6}
Attenuation calibration parameters (daily rates of inactivation, removal, etc.)		
CP_{atten} (bacteria)	$10^{-0.5} \approx 0.32$	10^2
CP_{atten} (viruses)	10	10^3
CP_{atten} (protozoa)	$10^{0.6} \approx 4.0$	10^3

Calibration parameters were sampled from uniform distributions on the \log_{10} scale. These ranges were determined empirically before calibration; see page 182 for further explanation.

5.3.2. Technical description of the model

The mechanics of the model are described in detail in the flowchart below (Figure 5.4).

Figure 5.4. Flowchart of the operations of the EITS model



Each numbered step in this flowchart is described in detail below.

Step 1: Enter parameters

Each simulation run began by reading in 53 parameters, 7 of which were calibration parameters that were randomly varied during the calibration step (Tables 5.1 and 5.2, page 169), and 37 of which were obtained from published scientific literature. The remaining parameters were based on expert opinion. Further discussion regarding choice of particular (non-calibration) parameters is in chapter 7, and the parameter values are summarized in Table 7.1, page 230.

Step 2: Create random number tables and output logs

All random numbers needed for the model run were pre-generated when the model run began (which reduced processing time). Matrices were also generated to store output from the model.

Step 3: Set up the simulated community

Assignment of children and adults to households

Simulated communities contained n households ($n \approx 200$), each containing a certain number of people randomly drawn from a Poisson distribution whose mean was 5 people/household. If a household was assigned 0 people, it was discarded; thus, the number of households in the simulated community varied slightly between runs. Children and adults were randomly assigned to households as follows:

1. The total number of adults in the community was determined by multiplying the number of people in the community (~ 1000) by the proportion of the community consisting of people aged five years or older (0.82) (Ayad et al., 1994); see page 228 for further discussion.
2. One person in each household was designated an adult.
3. Each remaining adult was randomly assigned to a household that was not already filled with adults.

4. Once all adults had been assigned to households, the remaining people were considered children under five years of age.

Assignment of compliance to households

Compliance with HWT, handwashing, and sanitation was considered to be a characteristic of the household, not a characteristic of each person. The community had a particular compliance level and compliance type that was designated at the beginning of each model run. Compliance was described by overall compliance (proportion of person-time spent complying) as well as compliance type (α , households complied perfectly or not at all; β , households complied perfectly, partially, or not at all; γ , all households complied partially; for a more detailed description, see section 4.3.1, page 124). Using these compliance parameters for the community, each household was randomly assigned a compliance level for each intervention, describing the proportion of time that household used it. In some simulations that included HWT, safe storage was applied to all households complying perfectly or partially with HWT, since safe storage is a particular characteristic of some HWT interventions.

Data structures for tracking household and person characteristics

Household composition and compliance were stored using a household matrix with one row per household, and a people matrix with one row per person (Table 5.3). Children were at the top of the people matrix, and adults were at the bottom. These matrices also explicitly tracked the amounts of pathogens in household compartments, as well as infection status of individual people. The household matrix had one row per household, and contained information about household-level compartments of microbes. The people matrix had one row per person, and contained information about infection state (-1 = susceptible; 0 = immune; 1 = exposed; 2 = infected; 3 = diseased), the number of days remaining in the infection state (if susceptible, these are negative), and the row in the household matrix that each person belonged to. Community-

level compartments of microbes were tracked separately (by 3-element vectors); 'nMw' is the number of microbes in the water reservoir, and 'nMI' is the number of microbes on the land.

Table 5.3. Description of matrices tracking households and people

Column number	Household matrix ('HHs')	People matrix ('People')
1	Number of people in the household	Infection state (bacteria)
2	Number of adults (aged ≥ 5 years)	Infection state (viruses)
3	Number of children (aged < 5 years)	Infection state (protozoa)
4	Number of bacteria in household environment*	Time counter† (bacteria).
5	Number of viruses in household environment*	Time counter† (viruses).
6	Number of protozoa in household environment*	Time counter† (protozoa).
7	Number of bacteria in household's water	Household that the person belongs to (row number in 'HHs')
8	Number of viruses in household's water	Number of people in the person's household
9	Number of protozoa in household's water	Unused‡
10	Unused‡	Unused‡
11	Unused‡	Unused‡
12	Unused‡	Unused‡
13	Compliance with sanitation (proportion)	Unused‡
14	Compliance with HWT (proportion)	Unused‡
15	Compliance with handwashing (proportion)	Unused‡
16	Unused‡	Unused‡
17	Whether household has safe storage (binary)	Unused‡

Infection states: -1 = susceptible; 0 = immune; 1 = exposed; 2 = infected; 3 = diseased

* The only sources of microbes on hands were from a) defecation; or b) inter-household visits.

† Each time counter was an integer denoting the number of days remaining in a particular person's infection state. Each day, 1 was subtracted from all time counters. When the time counter reached 0, the infection state changes. If the infection state was 'susceptible', the time counter was negative and denoted the number of days a person has spent susceptible to a particular pathogen. See page 178 for more detail.

‡ Unused rows were reserved for future expansion of the model, or were emptied as the design of the model changed during its development.

Generating connections between households

Households were connected by a randomly generated network to allow pathogens to be directly exchanged between households. An Erdős–Rényi random network was generated using the 'erdrey' function from the CONTEST toolbox for MATLAB (A. Taylor & Higham, 2008), using an estimate of mean network degree (number of connections per household) observed in rural Ecuadorian villages (Zelner et al., 2012).

Initial infection status

Initially, all persons in the village were classified as exposed; this exposure lasted for a randomly determined period from 0-18 days for each person, after which they developed infection or disease. This prevented the development of large oscillations in infection prevalence, in the same fashion as in the QMRA models in the two preceding chapters (see page 96).

Step 4: Start daily loop and tally people in all states

Once the community had been generated, the first simulated day in the community could begin. The numbers of people in each infection state for each pathogen (exposed, infected, diseased, immune, or susceptible) were tallied and stored. If the tallies did not agree with the number of people in the simulated community, an error would occur and the simulation would end.

Step 5: Defecation

People who were infected with a pathogen excreted pathogens into the environment. For each pathogen type and each infected person:

$$P_e = fdn \tag{5.1}$$

where P_e is the number of pathogens excreted, f is grams of feces excreted daily by a person without diarrhea, d is the number of defecation events per day (which depends on whether the person is symptomatic [3 events] or asymptomatic [1 event]), and n is the number of

pathogens per gram of feces.

Most of these pathogens were deposited on the land, but a small proportion entered the hands compartment for that person's household:

$$P_h = P_e h \quad (5.2)$$

where P_h is the number of pathogens added to hands by a particular person, and h is the proportion of feces that remain on the hands (1/1000). The number of pathogens deposited on land by a particular person is then $P_l = P_e - P_h$. If sanitation or handwashing interventions were included in the model run, LRVs were then applied to P_l and P_h according to sanitation and handwashing compliance, respectively, in the same way as HWT interventions (see step 7 below, Equation 5.4).

Step 6: Transfer pathogens from land to surface water

Each day, a small proportion (0.001) of pathogens was transferred from land to surface water, representing movement of pathogens by processes such as soil erosion, groundwater movement, laundering or defecating directly into water, etc. If a rain event occurred on a given day (probability of 1/14), a higher proportion (0.05) of pathogens were transferred from land to surface water. These proportions probably vary greatly depending on climate and hydrogeology, and were chosen by expert opinion, due to the absence of data; see page 228 for further discussion.

Step 7: Transfer pathogens between households via visits

Conceptually, each visit constitutes a contact between 2 people from different households, each of which gives a proportion of their pathogens to the other. A subset of the possible connections between households (as determined in step 3, page 171) was randomly chosen, based on a daily probability (2/7, i.e., twice weekly on average) of a contact occurring for each

connection. It was possible for a household to visit more than 1 other households per day. For each visit and each pathogen type, the number of pathogens transferred from one particular household to one other household (P_t) was:

$$P_t = CP_V \times P_b / N \quad (5.3)$$

where CP_V is a calibration parameter representing the proportion of pathogens transferred between households during a visit, P_b is the number of pathogens in the household environment before the contact, and N is the number of people in the household.

Step 8: Resupply stored water & apply household water treatment (HWT)

The first event to occur each day was resupplying the stored water within each household. For simplicity, any pathogens remaining in the stored water compartment were transferred to the land, and then a proportion of pathogens in the surface water (calibration parameter CP_{Sf}) was transferred into each household's stored water compartment. Log₁₀ reduction values (LRVs) were then applied to the pathogens in the stored water of each household, depending on the household's compliance with HWT:

$$P_t = P_b(1 - c) + P_b c 10^{-L}$$

(5.4)

where P_t is the number of pathogens remaining after treatment, P_b is the number of pathogens before treatment, c is the proportion of pathogens treated by the household, and L is the LRV of the treatment method. The value of c for each household was determined at the beginning of each model run, based on the overall compliance level and the compliance type within the simulated community (see step 3, page 171). Equation 5.4 was also used to calculate pathogen removal or inactivation from handwashing or sanitation (described in step 5, page 174).

Step 9: Pathogen transfer from household environment to drinking water

Pathogens within each household environment compartment could be transferred to its stored drinking water compartment, simulating recontamination of water when people remove it from the storage container.

$$P_t = CP_{H\&Dh} \times P_b S \quad (5.5)$$

where P_t is the number of pathogens transferred, $CP_{H\&Dh}$ is a calibration parameter describing the proportion of pathogens transferred from hands to water, P_b is the number of pathogens in the household environment before the transfer, and S equals 0 if the household uses safe storage, and 1 otherwise. Thus safe storage completely blocked the transfer of pathogens from the household environment to stored drinking water.

Step 10: First attenuation of pathogens

Pathogens in all compartments were exponentially attenuated (i.e., completely removed from the system) according to a particular rate r for each pathogen type. Since attenuation affected the doses of pathogens received by people (Figure 5.3, page 165), it was initially applied over a random proportion x of the day, after which people ingest doses of pathogens:

$$P_r = P_b e^{-rx} \quad (5.6)$$

where P_r is the number of pathogens remaining in a particular compartment after decay, and P_b is the number of those pathogens in that compartment before decay. The daily attenuation rate r was given by three values of a calibration parameter (CP_{atten}), one value for each pathogen type. CP_{atten} was the only calibration parameter that took different values for each pathogen type.

Steps 11 & 12: Decrement all status counters and apply status shifts

At all times, each person had a particular infection status for each pathogen type: susceptible, exposed, infected, diseased, or immune. A status counter measured the number of days remaining for that status. Every day, 1 was subtracted from all status counters. When a status counter reached 0, that person transitioned to the next state: exposed → infected (possibly diseased) → immune → susceptible (Table 5.3). When a person transitioned from exposed to infected or diseased, a duration of infectiousness was randomly chosen from a distribution. The durations of the immune and exposed states were fixed (for parameter values, see Table 7.1, page 230; but recall that the initial exposure duration varied; step 3, page 171).

Step 13: Calculation of daily pathogen doses and dose response

All people ingested daily doses of pathogens, removing those pathogens from the system. If a person was susceptible to a pathogen, they might become exposed to that pathogen as a result of that day's dose. The dose of each pathogen type was the sum of three component doses: pathogens from stored drinking water, pathogens from land, and pathogens from the household environment.

Pathogens from stored water

Each person's dose of each pathogen type from stored water (D_w) was determined as follows:

$$D_w = P_w I / V \quad (5.7)$$

where P_w is the number of pathogens in the household's stored water, I is the amount of stored water ingested in liters per day (which is higher for adults and lower for children), and V is the volume (in liters) of the household's stored drinking water container.

Pathogens from land

Each day, people ingest a dose D_l of pathogens from land, with children ingesting more pathogens than adults:

$$D_l = CP_{Dl} \times P_l(Z_c/Z_a) \text{ for children, and}$$

$$D_l = CP_{Dl} \times P_l \text{ for adults,} \tag{5.8}$$

where CP_{Dl} is a calibration parameter representing the proportion of pathogens on the land ingested daily; P_l is the number of pathogens on the land; and Z_c and Z_a are the numbers of daily hand-mouth contacts for children and adults (respectively 328/day and 130/day; USEPA, 2011). D_l could be considered to include pathogens from the environment outside the household, including soil and locally grown food.

Pathogens from the household environment

A proportion of the pathogens in a household's environment at the time of dose calculation was ingested by its occupants. Children ingested more of those pathogens than adults, as described immediately above. The dose D_h received in this manner by a particular occupant was:

$$D_h = CP_{H\&Dh} \times P_h(Z_c/Z_a) \text{ for children, and}$$

$$D_h = CP_{H\&Dh} \times P_h \text{ for adults,} \tag{5.9}$$

where $CP_{H\&Dh}$ is a calibration parameter representing the proportion of pathogens in the household environment ingested daily; and P_h is the number of pathogens in the household environment.

Dose response: converting pathogen doses to exposure and infection

Each of the three doses D_w , D_l , and D_h were then removed from the appropriate compartments (stored drinking water, land, and household environment), and were summed for each person and each pathogen to obtain the total dose for each pathogen received by each

person. Dose response functions were then used to convert the doses into probabilities of infection, in the same manner as in chapters 3 and 4 (Equations 3.2 and 3.3, page 98). Based on the resulting probabilities of infection, it was randomly determined whether each person who was susceptible to a particular pathogen type became exposed to that pathogen. Any person who became exposed would eventually develop infection (see steps 11 and 12, page 178), provided the model run did not end first.

Step 14: Assignment of baseline exposures

In order to force the model to have a minimum non-zero incidence of diarrhea, baseline exposures were assigned. All people had an equal daily probability per pathogen type S_p of being selected for a baseline exposure. If a person was selected and was also susceptible, their status was changed to exposed. If a non-susceptible person was selected, nothing occurred.

$$S_p = (T_p B_c) / (365 M_p) \quad (5.10)$$

where T_p is the proportion of the baseline incidence in children attributable to that pathogen type (0.5 for bacteria, 0.25 each for viruses and protozoa, based on etiologic fractions discussed in chapter 4, page 129), M_p is the morbidity ratio in children for that pathogen type, and B_c is the baseline incidence of diarrhea in children from all pathogen types. The value for B_c was assumed to be 0.5 episodes per child-year, since measurements of diarrheal incidence in developed countries are often roughly 0.5 to 1.5 episodes per person-year (M. E. Wilson, 2005), and the best possible outcome of various interventions in a developing country community would be to reduce diarrheal incidence to levels found in developed countries.

Step 15: Second inactivation of pathogens

Pathogens in all compartments were then inactivated over the remainder of the day $(1 - x)$ in the same manner as in step 11, Equation 5.6, page 177.

Step 16: Continue to next day, or end simulation

Each model run was scheduled to terminate after a certain number of days. Once pathogens had been inactivated over the final portion of the day, the simulation terminated if the required number of days had elapsed, or continued to the next day (see step 4, page 174) if not.

5.3.3. Calibration and estimation

All of the parameters used in this model are substantially uncertain, and communities vary across many characteristics that are difficult to measure, but could nonetheless influence the transmission of diarrheal infections. In the calibration step, the model was run many times while varying the values of the calibration parameters (Table 5.2, page 169). If a model run's output was consistent with certain calibration criteria chosen *a priori* (Table 5.4), its calibration parameter values were retained and used for estimation later.

The calibration criteria (Table 5.4) considered incidence and etiologic fractions in children in developing countries aged less than five years. Observational studies of diarrhea in developing countries show that the incidence of diarrhea has a wide range, depending on the country and the community (Kosek et al., 2003); a relatively high incidence of childhood diarrheal disease (6-12 episodes per child-year) was chosen because the calibration step is meant to simulate conditions in a very poorly developed community with essentially no hygienic or sanitary infrastructure. Studies of diarrheal etiology also indicate that bacterial diarrhea is generally more common than viral or protozoan diarrhea; see page 19 for further discussion (Lanata & W. Mendoza, 2002). However, the importance of differing transmission routes is unknown; therefore, model runs consistent with the first two criteria were subdivided by the relative importance of transmission via stored drinking water, so that estimation could be conducted for varying scenarios of high, medium, and low waterborne transmission.

The sets of calibration parameter values that fit the calibration criteria (Table 5.4) can be

considered to represent a variety of differing communities. In the estimation step, differing interventions can be applied to these communities at a variety of compliance levels to determine the likely effect of these interventions on diarrheal disease.

Table 5.4. Criteria for calibrating the transmission model

Description	Units	Criteria		
Incidence of childhood* diarrheal disease	episodes per child-year*	6 to 12		
Etiologic fractions	proportion of childhood* diarrheal episodes by pathogen type	32.5 to 62.5% bacterial 7.5 to 37.5% viral 7.5 to 37.5% protozoan		
Route importance	proportion of childhood* diarrheal episodes from drinking water	Mostly water: >2/3 to 100%	Water & non-water similar: 1/3 to 2/3	Mostly non-water: 0 to <1/3

See Table 5.1 (page 169) for descriptions of the 7 calibration parameters that were varied to obtain these values during the calibration process.

*Children aged less than 5 years.

Determination of calibration parameter ranges

The calibration process used a single set of ranges for the calibration parameters (Table 5.2, page 169). The ranges were determined by running a series of mock calibration processes consisting of about 30,000 runs each. The first mock calibration process used extremely wide ranges of calibration parameter values (7 to 9 orders of magnitude). For subsequent processes, these ranges were gradually narrowed, one calibration parameter at a time, guided by the distributions of calibration runs within each calibration parameter range that were consistent with the calibration criteria (Table 5.4). For example, if CP_{Sf} was varied from 10^{-9} to 10^{-2} , and no model runs whose results were consistent with the calibration criteria were seen from 10^{-9} to 10^{-8} , CP_{Sf} might be varied from 10^{-8} to 10^{-2} next. Random sampling of all calibration parameters

was carried out as if they were uniformly distributed on the \log_{10} scale. The calibration process was done once, producing three mutually exclusive sets of parameters: 1) a set consistent with predominantly ($> 2/3$) waterborne transmission; 2) a set with similar waterborne and non-waterborne transmission; and 3) a set consistent with predominantly ($> 2/3$) non-waterborne transmission. Those sets of parameters were then used as the basis for estimation scenarios in which HWT interventions were applied to the simulated community.

All runs of the EITS model were on MATLAB 7.13 (R2011b; 64 bit). The program code for the EITS model was originally developed using Octave 3.2.3, but it runs on either Octave or MATLAB without modifying the code. Each run of the model took about 16 seconds in MATLAB; for comparison, each run required about 30 seconds in Octave. Output from the model was analyzed with R 2.15.1.

5.4. Results

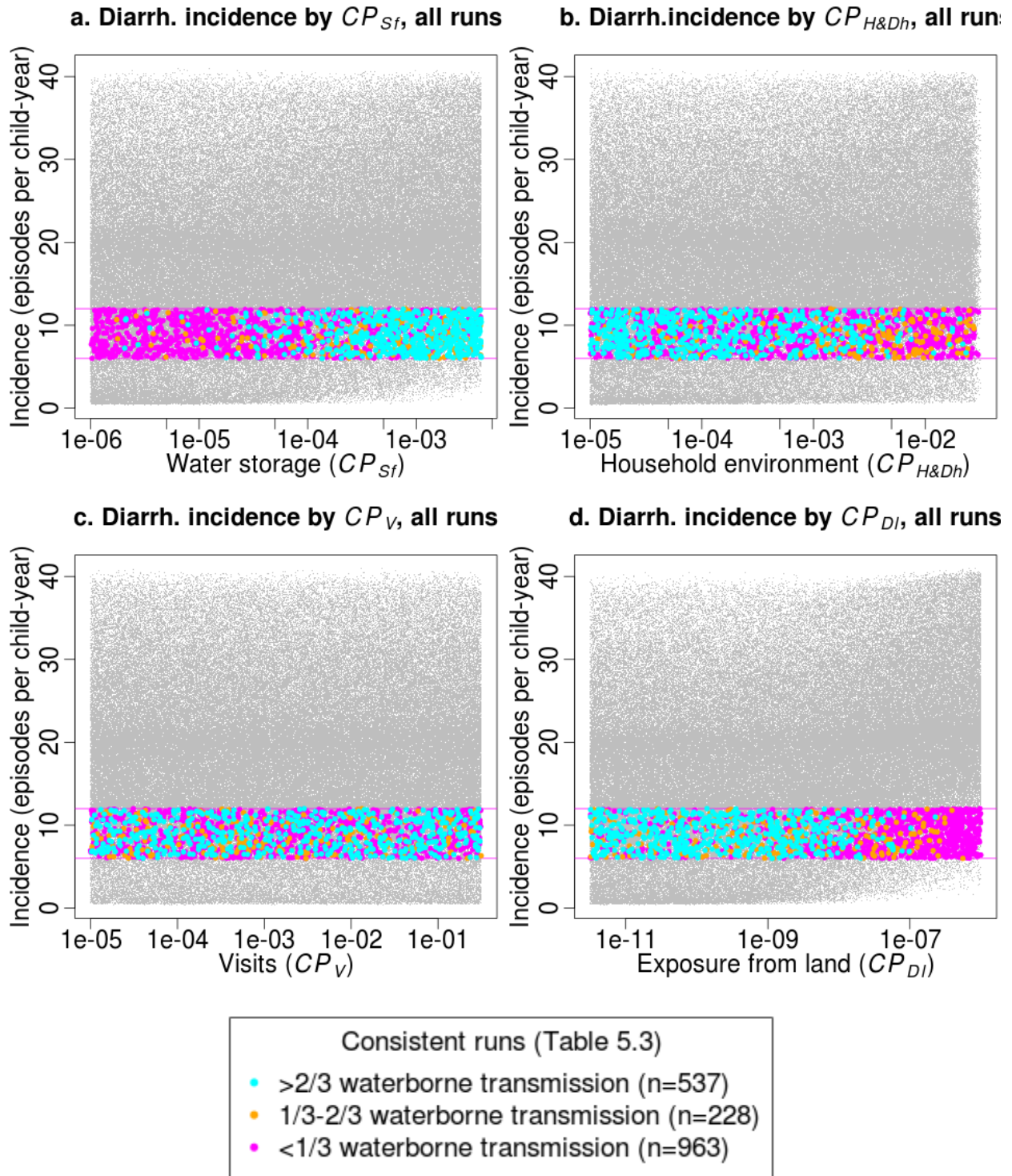
5.4.1. Calibration

The calibration step included 228,480 model runs. Calibration runs whose output was consistent with the criteria in Table 5.4 were considered to represent distinct communities with differing characteristics that were summarized by their calibration parameter values. There were 1728 (0.756%) consistent runs; of these, 963 (55.7%) had $< 1/3$ waterborne transmission, 129 (13.2%) had $1/3$ to $2/3$ waterborne transmission, and 537 (31.1%) had $> 2/3$ waterborne transmission.

Charts of calibration output are presented in Figures 5.5 and 5.6. Incidence of diarrhea in children was highly variable for all transfer calibration parameter values, ranging from roughly 0.5 to 40 episodes per child-year (Figure 5.5). The proportion of cases of diarrhea attributable to waterborne transmission increased as CP_{Sf} increased or as CP_{DI} decreased (Figure 5.5a and d,

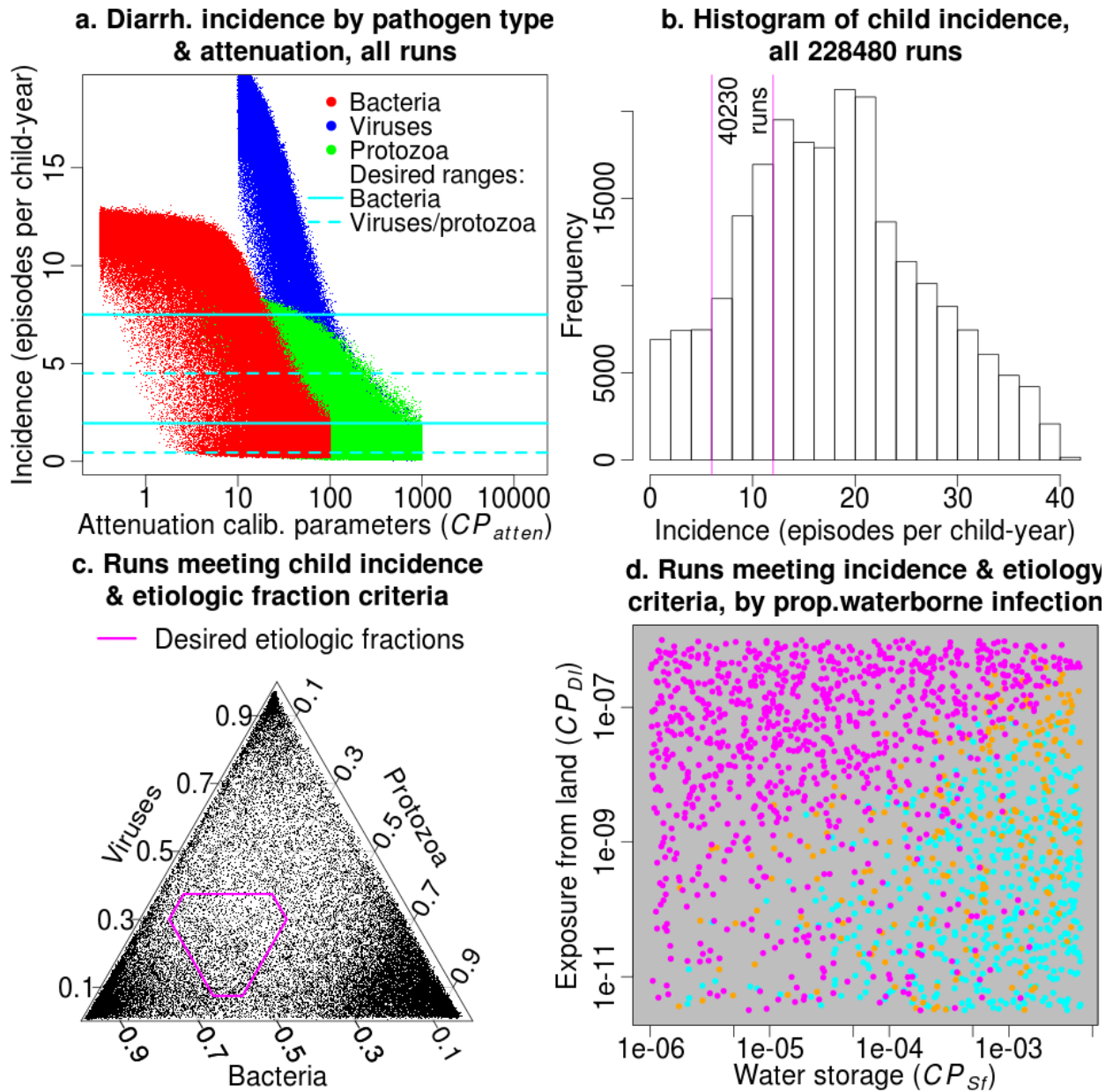
and Figure 5.6d). Bacterial pathogens generally had lower attenuation rates (CP_{atten}) than viral or protozoan pathogens, though this was not universal (Figure 5.6a). The distribution of diarrheal incidence in children over all the calibration runs was multimodal (Figure 5.6b) with peaks near 0, 9, and 20 episodes per child-year; the last two peaks roughly correspond with the maximum incidence levels attainable by each pathogen type, visible on Figure 5.6a where the point clouds approach their maxima as CP_{atten} decreases. Of the 40,230 (17.6%) calibration runs that met the incidence criterion, it was common for a single pathogen type to predominate (Figure 5.6c, note densely clustered dots near each corner of the triangle).

Figure 5.5. Calibration output from EITS model, transfer parameters



For further explanation, see text immediately before & after this figure.

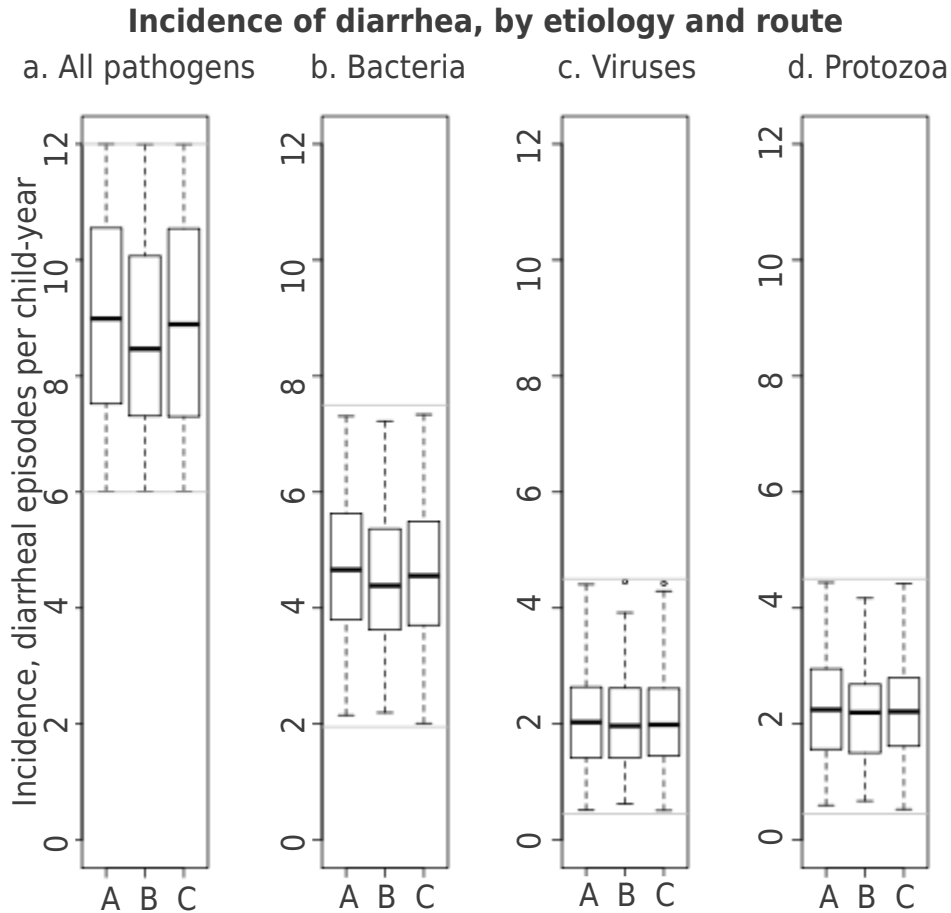
Figure 5.6. Calibration output from EITS model, scatterplots & histograms



Incidence of diarrhea in children was unrelated to the proportion of diarrhea in children that was waterborne (Figure 5.7). Although incidence of diarrhea appeared slightly lower if 1/3 to 2/3 of diarrhea was waterborne for all pathogens and for bacteria (Figure 5.7, a and b), the distributions were not significantly different after adjusting for multiple comparisons (Holm

method, n=12).

Figure 5.7. Simulated diarrhea incidence in children (calibration step)



A, > 2/3 of incidence from waterborne route; *B*, 1/3-2/3 of incidence from waterborne route; *C*, < 1/3 of incidence from waterborne route.

In the calibration runs that were consistent with the calibration criteria (Table 5.4, page 182), six of the seven calibration parameters were associated with diarrheal incidence among children by multiple linear regression (Table 5.5; for descriptions of the calibration parameters, see Table 5.1, page 169). In contrast with all other calibration parameters, CP_V (describing transfer of pathogens between households) was not significantly associated with diarrheal incidence, or the proportion of incidence that was waterborne (Table 5.5). Substantial portions of

the variation in incidence was explained by the calibration parameters, with adjusted R^2 values from 0.3 to 0.8 depending on the dependent variable that was used.

Table 5.5. Association of calibration parameters with incidence

Dependent variable	n	Linear regression parameter estimates* and p-values								R^2 ‡
		Intercept	CP_{Sf}	$CP_{H\&DH}$	CP_V	CP_{DI}	CP_{atten}			
							Bacteria	Viruses	Protozoa	
Child incidence, all runs	228,480	67.8 2×10^{-16}	1.67 2×10^{-16}	1.17 2×10^{-16}	0.0303 2×10^{-6}	1.52 2×10^{-16}	-5.51 2×10^{-16}	-9.88 2×10^{-16}	-0.010 2×10^{-16}	0.806
Child incidence, consistent runs, A†	537	41.3 2×10^{-16}	3.88 2×10^{-16}	0.190 0.005	-0.012 0.7	0.291 5×10^{-10}	-9.00 2×10^{-16}	-1.31 2×10^{-9}	-1.32 4×10^{-9}	0.551
Child incidence, consistent runs, B†	228	32.5 3×10^{-4}	2.65 2×10^{-16}	0.121 0.4	0.0135 0.8	0.252 0.004	-6.47 2×10^{-16}	-1.06 0.006	-0.901 0.01	0.411
Child incidence, consistent runs, C†	963	28.9 2×10^{-16}	0.373 2×10^{-10}	0.400 1×10^{-12}	0.0458 0.2	0.890 2×10^{-16}	-3.81 2×10^{-16}	-1.54 5×10^{-13}	-0.687 0.002	0.284
Proportion of incidence that is waterborne, consistent runs	1728	0.209 0.1	-0.107 2×10^{-16}	-0.107 2×10^{-16}	-0.001 0.7	-0.110 2×10^{-16}	-0.146 4×10^{-7}	0.066 0.0006	-0.043 0.02	0.762

* The \log_{10} of the calibration parameters were the independent variables in the linear models.

† A, > 2/3 of incidence was waterborne; B, 1/3-2/3 of incidence was waterborne; C, < 1/3 of incidence was waterborne.

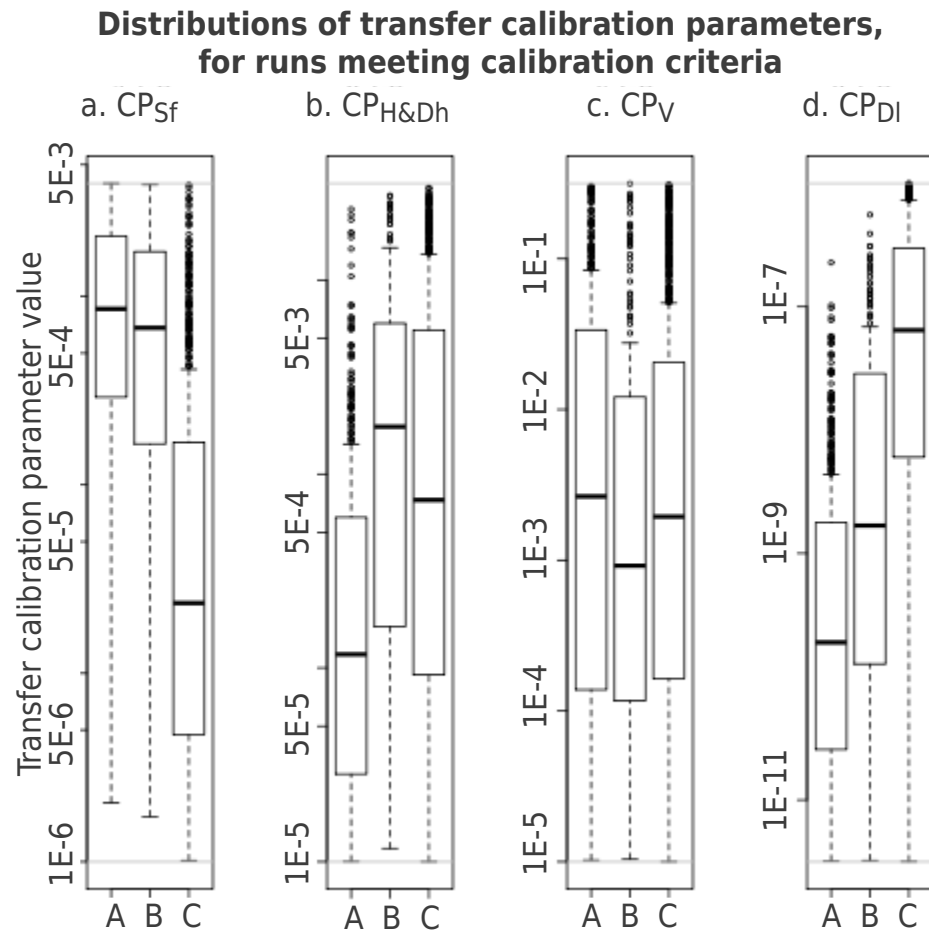
‡ The adjusted R^2 was calculated by the `lm()` function in R 2.15.1.

For descriptions of the calibration parameters, see Table 5.1, page 169.

The influence of the calibration parameter values on the proportion of childhood diarrhea that was waterborne was particularly apparent for the transfer calibration parameters, which were proportions describing daily movement of pathogens between compartments. Lower values of

CP_{Sf} (related to dilution of pathogens in surface water) were associated with less waterborne transmission, and similarly, lower values of CP_{DI} (related to dispersion of pathogens on land) were associated with less non-waterborne transmission (Figure 5.8, a and d). Lower values of $CP_{H\&Dh}$ (describing transfer of pathogens out of the household environment) also favored more waterborne transmission (Figure 5.8b), though there was a significant tendency ($p = 0.009$, Wilcoxon rank sum test, Figure 5.8b) for $CP_{H\&Dh}$ values to be higher if 1/3 to 2/3 of transmission was waterborne (category B), compared with $< 1/3$ of incidence being waterborne (category C). There was also some weaker statistical evidence that CP_V (describing transfer of pathogens between households) values were lower if 1/3 to 2/3 of transmission was waterborne (category B, Figure 5.8c), compared with $> 2/3$ of incidence being waterborne (category A; $p = 0.03$) or $< 1/3$ of incidence being waterborne (category C; $p = 0.08$).

Figure 5.8. Distributions of transfer calibration parameters

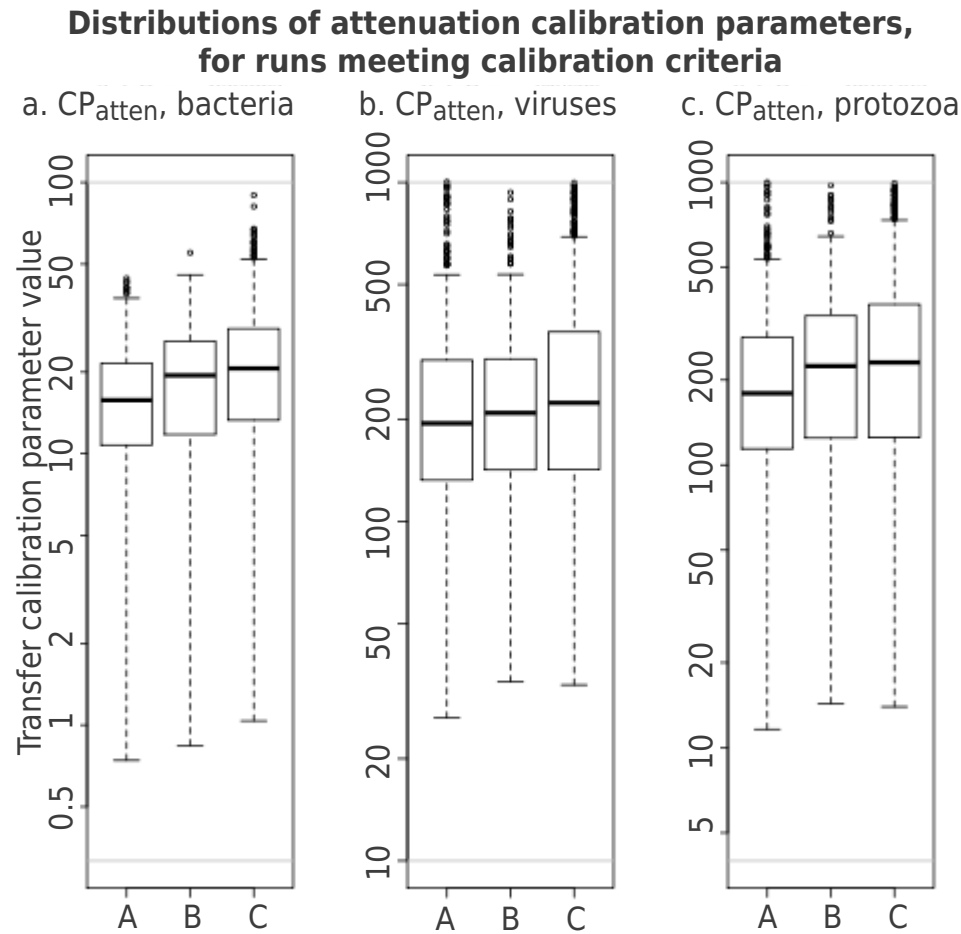


A, > 2/3 of incidence was waterborne; B, 1/3-2/3 of incidence was waterborne; C, < 1/3 of incidence was waterborne. Grey horizontal lines represent the ranges over which the calibration parameters were sampled.

The values of the attenuation calibration factors varied little in relation to the proportion of incidence that was waterborne, although there was a slight trend for their values to increase as the proportion of waterborne transmission decreased (Figure 5.9). In general, the values of the calibration parameters that were consistent with the calibration criteria spanned the ranges over which they were sampled during the calibration step (Table 5.2, page 169), although the attenuation parameter values did not quite meet the lower bound. The value of CP_{atten} for

bacteria was approximately 20, compared with approximately 200 for viruses and protozoa. The model therefore suggested that protozoa and viruses are attenuated roughly 10 times faster than bacteria.

Figure 5.9. Distributions of attenuation calibration parameters



A, > 2/3 of incidence was waterborne; B, 1/3-2/3 of incidence was waterborne; C, < 1/3 of incidence was waterborne. Grey horizontal lines represent the ranges over which the calibration parameters were sampled.

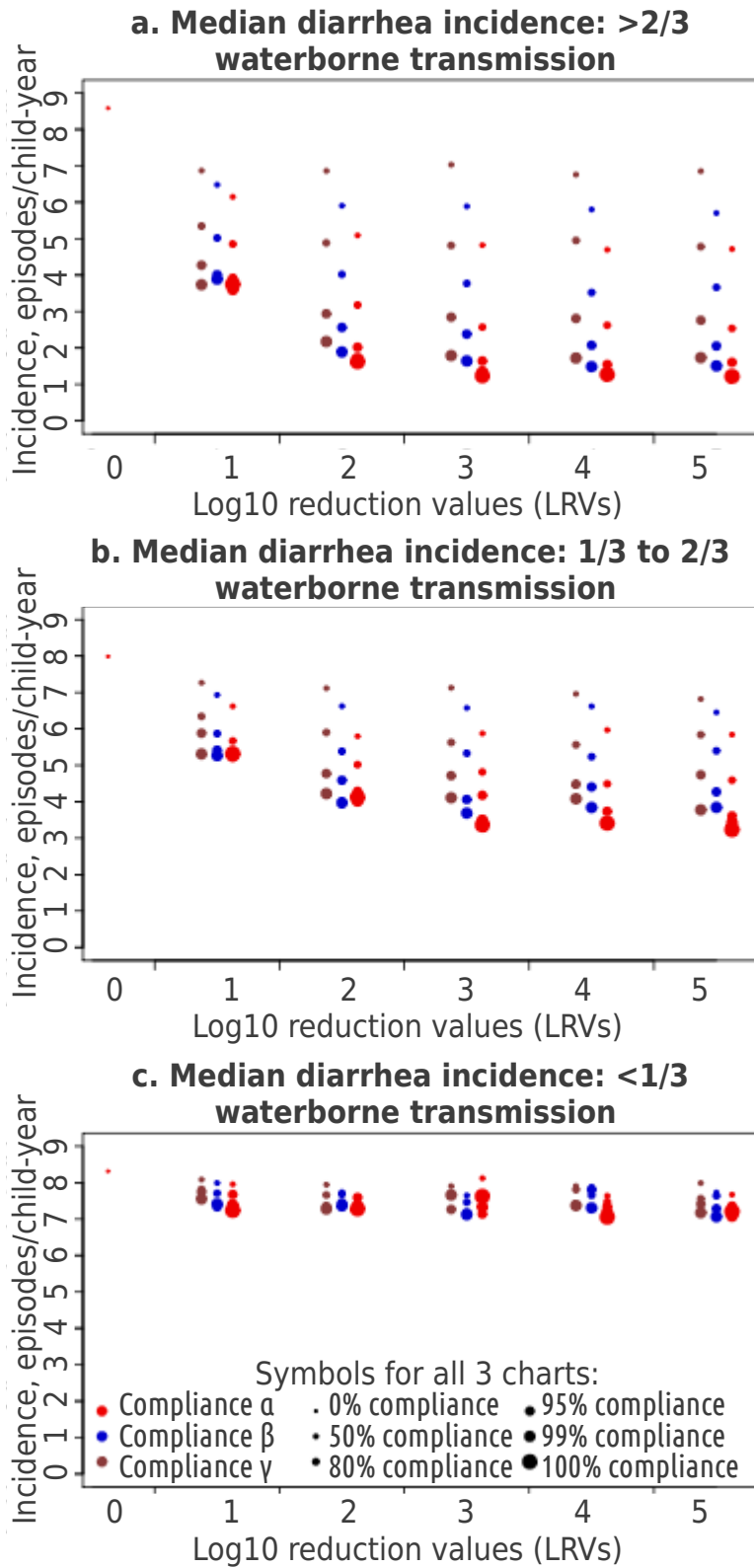
5.4.2. Estimation

A random sample of 100 parameter sets was taken from each of the three groups of calibration runs consistent with the calibration criteria. Using these parameter sets, a HWT

intervention with LRVs of 1, 2, 3, 4, or 5 against all pathogen types was applied with community-wide compliance of 50%, 80%, 95%, 99%, or 100%, and compliance types of α , β , or γ (see page 124 for further discussion of compliance types). The HWT intervention prevented little or no diarrhea where $< 1/3$ of childhood diarrhea was waterborne (Figure 5.10c), even if compliance was perfect. If $1/3$ - $2/3$ of childhood diarrhea was waterborne (Figure 5.10b), the median incidence decreased from 8 episodes per child-year to 3 episodes per child-year if compliance was perfect and LRVs were 3 or higher; however, if compliance was 50%, median incidence was approximately 7 episodes per child-year for compliance type γ and 6 episodes per child-year for compliance type α , regardless of LRV. If $> 2/3$ of childhood diarrhea was waterborne (Figure 5.10a), median incidence was further decreased, particularly for higher compliance levels; there were approximately 1.3 episodes per child-year if compliance was perfect. As found in chapter 4, LRVs higher than 3 generally did not prevent additional diarrhea.

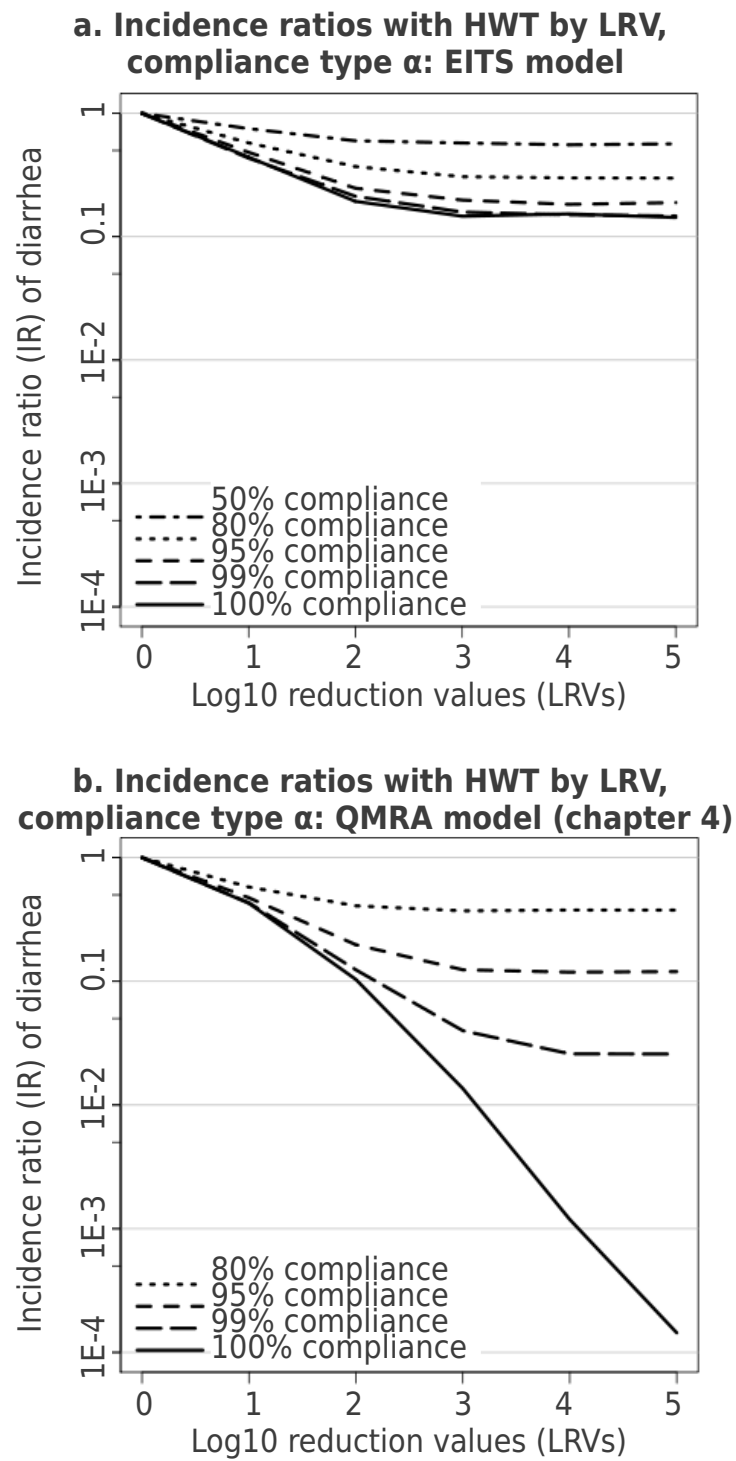
There was substantial variation around each of the median estimates displayed in Figure 5.10; interquartile ranges generally spanned about 4 episodes per child-year (data not shown). At the extremes, incidence levels ranging from 1 to 20 were obtained.

Figure 5.10. Estimation step, incidence by LRV of HWT



Although the EITS model is not directly comparable with the QMRA model described in chapter 4, it is possible to make a rough comparison (Figure 5.11). Both models indicate decreasing incidence ratios as compliance decreases. However, increasing compliance in the QMRA model leads to greater risk reductions than in the EITS model. Furthermore, in the EITS model, there is little or no improvement in IR from LRVs above 3 even if compliance is perfect. This differs from the QMRA model, where risk continues to decrease as LRVs increase if compliance is perfect.

Figure 5.11. Comparison of EITS and QMRA results



5.5. Discussion

5.5.1. Calibration step

All calibration parameters were varied over wide ranges; the transfer calibration factors (describing the proportion of pathogens transferred daily between compartments) varied over 4 or more orders of magnitude (Figure 5.8), and the attenuation calibration factors (describing inactivation or removal of the three pathogen types) varied over about 2 orders of magnitude (Figure 5.9). Although 1728 sets of calibration parameters were found that were consistent with childhood diarrheal incidence values and etiologic fractions that appear common in developing countries (Table 5.4, page 182), it is not certain that these calibration parameter sets reflect realistic conditions in developing country communities, because of uncertainty in parameter values that necessitated calibration in the first place.

All of the calibration parameters strongly influenced the outcome of the model, except for the calibration parameter describing pathogen exchanges during inter-household visits (CP_V). The routes influenced by CP_V and CP_{DI} (which describes ingestion of pathogens from land) are parallel transmission routes; they both allow pathogens to move from one household to another. However, the inter-household route influenced by CP_{DI} is more direct (feces from household A → land → new host in household B) than the inter-household route influenced by CP_V (feces from household A → household environment A → household environment B → new host in household B). Furthermore, the CP_{DI} route links all households with all other households every day, while the CP_V route links only a subset of households daily. Although removing inter-household visits might simplify the model without reducing its utility, household structure remains a useful component of the model (shown by the effect of $CP_{H\&Dh}$ on diarrheal incidence

and the proportion of incidence that is waterborne, Figure 5.8b, page 190).

The values of the daily attenuation rates (CP_{atten}) of the three pathogen types were very high, with median estimates of $\sim 20/\text{day}$ for bacteria, and $\sim 200/\text{day}$ for viruses and protozoa. These correspond to half-lives of 50 minutes and 5 minutes, respectively, and are much faster than decay rates of $\sim 0.6/\text{day}$ (half life of 1700 minutes, or 1.2 days) for *E. coli* (Flint, 1987), norovirus (Pancorbo et al., 1987), and *Giardia* (Wickramanayake et al., 1985; deRegnier et al., 1989) measured in unfiltered natural waters at $\sim 20^\circ\text{C}$ (see page 227 for further discussion). However, attenuation includes many processes in addition to decay that prevent pathogens from contacting hosts, such as: percolation or burial in soil; sedimentation or settling in water; ingestion by animals; and removal from the community by flowing water.

5.5.2. Estimation step

The EITS model corroborates the QMRA model in chapter 4 in several respects. Both models agree that increasing HWT LRVs beyond 3 is unlikely to prevent much additional diarrhea, although it might provide some benefit for intermediate levels of compliance when waterborne transmission predominates. Furthermore, in both models, compliance type α (where people either comply perfectly with HWT or don't comply at all) tends to prevent more diarrhea in both models than compliance type γ (where everyone complies partially with HWT); compliance type β is intermediate between α and γ .

The EITS model indicates that HWT can prevent very little diarrhea if waterborne transmission is low in relation to other transmission routes, even if LRVs are high and compliance is perfect (Figure 5.10c, page 193).

5.5.3. Limitations of the EITS model

Wherever possible, published measurements from peer-reviewed publications were used for parameter values (see chapter 7 for detailed discussion). However, reliable information from

developing countries was not available for all parameter values. Furthermore, it is possible that appropriate values for some parameters were not found during the literature review, despite being documented in the literature. Ideally, a rigorous meta-analysis would be conducted for each parameter in this model, but this was not practical given the resources available for the research and the large number of parameters. Nonetheless, multiple documented measurements were sought for each parameter; discussions of the decision process for various parameter values are in chapter 7.

Most parameters used in the model had fixed values. Because few measurements have been reported for many parameters, it is possible that some of them are improperly specified, possibly affecting the model's results. However, the seven calibration parameters introduced variation along all possible routes that pathogens could take within the model. This potentially allowed the model to self-adjust for missing or misspecified parameter values in the calibration step, but would also confound sensitivity analysis of the fixed model parameters. The model could be refined in the future by incorporating updated information regarding its parameter values, which would mean that the aspects of each route that are covered by each calibration parameter would decrease, yielding a more completely specified model requiring calibration over fewer variables or narrower ranges of values.

Although the model suggests that visits between households are a relatively unimportant route of transmission, this conclusion might depend upon the way visits are simulated. The model describes visits by an exchange of pathogens between two household environment compartments (Equation 5.3, page 176), which might represent a brief visit or an exchange of food or items. However, children are frequently cared for by friends or relatives in different households, and the entry of an infected child into an otherwise uninfected household could be a particularly important exposure route. The entry of an infected person into an uninfected

household for a long period, particularly a small child who defecates in the new household, could represent a much more intense exposure. In principle, the model could be modified to better represent such a situation, which might increase the importance of visits to infection transmission.

This EITS model does not include zoonotic transmission of diarrheal pathogens. Domestic animals live in close proximity with humans (particularly in rural areas), and can be major reservoirs of pathogenic *E. coli*, *Campylobacter* (C. R. Young et al., 1999), and *Cryptosporidium parvum* (Xiao, 2010); however, giardiasis is probably not a zoonosis under most circumstances (Cacciò et al., 2005). The lack of animal sources of transmission in this model may inappropriately penalize transmission of bacterial and protozoan infections relative to viral infections. This may partially explain the relatively low attenuation rates ($CP_{atten} \approx 20$) of for bacteria, compared to viruses and protozoa ($CP_{atten} \approx 200$), because selecting lower values of CP_{atten} for bacteria than for viruses or protozoa is the only way for the calibration step to favor bacterial transmission. If the model omits a route that bacteria use preferentially, the calibration process must select lower values of CP_{atten} for bacteria than for viruses or protozoa in order to meet the calibration criteria (Table 5.4, page 182), essentially compensating for the absence of the bacterial route.

The model also does not explicitly include exposure to pathogens through contaminated food, although doses of pathogens ingested from the household environment compartments and the land compartment could be considered to include such exposures. Many bacterial pathogens can grow in food, and this is not explicitly incorporated in the model either, although the lower attenuation rates (CP_{atten}) of bacteria compared to viruses or protozoa (Figure 5.5e) could also be considered to reflect this.

Transmission of rotavirus in this model might also be inappropriately high because the model considers adults and children to have the same concentration of virus in their feces, regardless of whether they are infected or diseased. However, it has been reported that rotavirus particles are roughly 10 times as concentrated in child feces as in adult feces (Vollet et al., 1979). Furthermore, rotavirus infections in adults tend to be shorter and milder than in children (Hrdy, 1987); thus adults might also have a shorter period of communicability than children.

5.5.4. Insights gained during the model construction process

The initial formulation of this model used a Gillespie algorithm framework (Keeling & Rohani, 2008) to model many discrete events occurring at particular rates over time using short timesteps of variable length separating each event; this resulted in a series of randomly selected events, separated by time periods of random (but very short) length. This framework was chosen in an attempt to build upon the previously published model described on page 81, which used similar methodology (J. N. S. Eisenberg et al., 2007). Although the Gillespie algorithm version of the model had similar compartments and flows (resembling Figures 5.1 and 5.2, page 162) to the EITS model described here, there were many thousands of events occurring daily, even within a very small community (~20 households). In order for the model to run in a reasonable amount of time, it was necessary to simplify it into a framework with discrete daily time steps.

An original ambitious goal of this work was to construct a highly mechanistic model that was thoroughly grounded in published theory and measurements. However, in the course of reviewing the literature, it became clear that many important parameter values were highly uncertain or completely unknown. Furthermore, there is likely to be large variation in many aspects of real diarrheal infection transmission networks; for example, hydrogeologic characteristics affecting contamination of groundwater, or differing cultural practices in areas of hygiene, food preparation, food cultivation, etc. (for further discussion of uncertainty regarding

particular parameter values, see chapter 7). In order to incorporate this variation, the calibration parameters were added to key routes and varied during the calibration process to determine sets of calibration parameters leading to realistic childhood diarrheal incidence levels (page 181). It is not clear how to interpret these calibration parameter values, since they encompass many different processes. For example, the calibration parameter CP_{DI} is a daily proportion of pathogens from the land that are ingested by each person. It is a crude way of summarizing many different processes that are poorly understood, for example: dispersal of pathogens over an area; the likelihood of contacting feces that are clustered in space in unknown ways; ingestion of pathogens along with soil; transfer of pathogens from soil to skin to mouth; etc. All models represent a compromise between realism and practicality, since highly detailed models are difficult to construct and analyze. In the case of this EITS model, some simplifications were necessary due to lack of information.

The model was originally intended to be a closed system, in which people could only be infected by ingesting pathogens excreted by other people in the community. However, this led to frequent stochastic pathogen extinctions during the calibration step, particularly of bacteria. The probability of extinction was reduced by increasing the size of the community. However, extinctions remained frequent even if the community had 4000 households, and the model ran too slowly when simulating such a large community. Since extinction of pathogen types in an underdeveloped community is unlikely, random exposures of people to all pathogen types were added to the model, in such a way that a mean baseline incidence rate of 0.5 diarrheal episodes per child year was obtained (step 14, page 180). The baseline incidence represents importation of infection from outside the community, or infections arising from pathogen sources that are not explicitly modeled (e.g., animal feces or imported food).

5.5.5. Future applications of the model

Readily achievable applications

At this writing, it has not yet been possible to thoroughly explore many aspects of this complicated EITS model. This section briefly describes additional aspects of the transmission and prevention of diarrheal infections that could be explored without modifying the model code.

The model is not limited to HWT interventions; it can also simulate the action of sanitation and hygiene interventions, and assess scenarios in which different interventions are applied together. Furthermore, it can simulate the effect of safe storage of drinking water, by making the assumption that each household owning a safe storage container completely blocks transmission of pathogens from the household environment to stored drinking water. Although many HWT interventions include a safe storage container, this is not always the case.

The model could be used to assess different interventions simultaneously, or in series. Field trials report widely varying efficacies for particular interventions; one explanation might be that a community's response to an intervention might depend upon other interventions that are already in place. For example, if a community with predominantly non-waterborne diarrhea transmission (e.g., Figure 5.10c) acquires latrines, diarrheal incidence would decrease, but the importance of the waterborne route would be larger for the remaining disease transmission; this would be expected to increase the effectiveness of HWT interventions (Figure 5.10a). Furthermore, two interventions may interact positively or negatively when they are applied simultaneously (see page 46 for further discussion). Such scenarios can be simulated using this model.

The time required for the full benefits of an intervention to be realized may be of interest. Because the model is dynamic, it can be run for a period of time without an intervention; once an intervention is applied, diarrheal incidence could be tracked over time to determine how rapidly

a lower equilibrium is reached. Since giardiasis in particular has an incubation period and disease duration of roughly two weeks (Rendtorff, 1954; A. M. Jokipii & L. Jokipii, 1977), it might take months for the full preventive effect of an intervention to be attained. A long time period between distribution of an intervention and the appearance of its benefits might lead to reduced compliance by the community because it initially appears ineffective, even if it would have been effective in the long term.

Although the burden of diarrheal infections and disease in people aged five years or more is poorly understood, particularly in developing countries, this model explicitly includes them. Although it is impossible to say whether the model properly accounts for infections in such people, examination of infection and disease status in the simulated population could generate hypotheses that future studies might test.

A sensitivity analysis would provide further information about the impact of the 46 fixed parameter values on the results from the model. Such an analysis would take the form of repeated estimation steps, modifying each fixed parameter in turn to determine its effect on the incidence of diarrhea in children, as well as the relative contribution of the waterborne route of transmission. Although an extremely thorough sensitivity analysis would need to consider alterations to the calibration parameter sets resulting from changing the fixed parameter values, it is not practical to recalibrate the model repeatedly in order to explore this effect.

Applications requiring substantial modifications to the model

Additional aspects of diarrheal infection transmission and prevention could be explored through further modifications to the program code of the model, and subsequent analysis of its behavior.

Households who are not using a particular intervention might nonetheless benefit directly because other households in the community are using them; participating households thus

prevent diarrhea for themselves directly, and prevent diarrhea in the rest of the community indirectly since fewer people are infected and excreting pathogens. Incidence in noncompliers could be measured before and after a simulated intervention to determine the magnitude of indirect protective effects of interventions, in a manner similar to Halloran et al. (2002; see page 83).

Bouts of diarrheal disease, or asymptomatic infections with gastrointestinal pathogens, are believed to degrade resistance to further bouts of diarrhea, probably through malnutrition and consequent impairment of immunity (see page 14 for further discussion). Although it is unclear precisely how this occurs, information is available regarding correlation of diarrheal episodes within individuals (Schmidt et al., 2009). Calibration parameters could be added that modify the dose response relationship(s), signifying reduced resistance to infection or disease following a previous infection, and select values of these parameters that yield distributions of cases within individuals similar to those observed in the field. Since repeated diarrhea episodes are an important factor in mortality, this could facilitate estimation of the mortality benefit attributable to particular interventions.

The model might further be used to examine the emergence of persistent diarrhea. Since coinfections with diarrheal pathogens are common, a single diarrheal episode might consist of multiple overlapping infections. Processes by which diarrheal episodes are reported and measured in the field could be incorporated into the model, in a similar fashion to the incorporation of imperfect recall in the QMRA model in chapter 3. By calibrating the model such that the distributions of the durations of modeled diarrheal episodes match observed distributions of diarrheal episode durations, insight could be gained regarding the development of persistent diarrhea in developing country communities.

5.5.6. Conclusions

The results from the EITS model corroborate the conclusions from the QMRA model regarding compliance (chapter 4): increasing LRVs beyond (approximately) 3 is unlikely to prevent additional diarrhea, and more diarrhea is prevented if more households comply perfectly (i.e., compliance type α is superior to compliance types β or γ). By incorporating multiple transmission routes, the EITS model also showed that improving compliance with HWT can greatly improve diarrhea prevention if waterborne transmission accounts for at least 1/3 of diarrheal incidence. This finding holds even for relatively low HWT LRVs of 1 or 2. More attention should be paid to improving compliance with HWT, rather than improving the antimicrobial efficacy of HWT.

The relative importance of the various transmission routes for diarrheal infections are unknown; in the EITS model, it depends upon the values of the calibration parameters. The calibration parameters themselves represent a variety of processes which are poorly understood, such as: transfer of pathogens from soil or objects to hands or mouth; decay or attenuation of pathogens in many different media, temperatures, or humidities; and many others. However, transfers of pathogens via households visiting each other (affected by the calibration parameter CP_{γ}) did not greatly affect diarrheal incidence or the proportion of incidence attributable to the waterborne route, in contrast to the other 6 calibration parameters.

The EITS model is a simplification of an extremely complicated infection transmission system incorporating multiple pathogens using multiple transmission routes. In addition, many aspects of diarrheal infection transmission through the environment are poorly understood. The EITS model attempts to account for this uncertainty through the calibration process; calibration produces sets of calibration parameter values that could be considered to represent distinct

communities with differing environmental characteristics. However, it is unknown which sets of parameter values best represent actual communities. Important aspects of transmission may also be improperly specified or missing (e.g., conceptualizing visits to represent shared child care might increase the importance of visits; incorporating domestic animals or foodborne transmission might favor bacterial infections). Nonetheless, the results regarding compliance are robust across widely varying sets of calibration parameter values, increasing confidence that they should also apply in real communities.

Future descriptive studies and field trials results will allow refinement of EITS models for diarrhea by clarifying appropriate values for parameters and reducing the need for calibration. Future models will provide guidance for public health and development professionals regarding effective ways to prevent diarrhea in developing countries.

6. CONCLUSIONS

6.1. Summary of research

Despite the large body of literature regarding the transmission of diarrheal infections and the various interventions used to prevent those infections, much remains to be learned. Constructing models simulating the transmission and control of these infections can yield conclusions useful for designing diarrhea prevention programs, and also highlights aspects of diarrheal infections that require further research. This dissertation describes three such models: a quantitative microbial risk assessment (QMRA) model simulating waterborne transmission during a published household water treatment field trial (chapter 3); a generalization of that QMRA model to examine the effect on diarrheal incidence by varying HWT compliance and antimicrobial effectiveness (measured by \log_{10} reduction values [LRVs] under different scenarios (chapter 4); and a more complex environmental infection transmission system (EITS) model that further examines compliance and LRVs, as well as multiple routes of transmission (chapter 5).

By simulating a published field trial (Boisson et al., 2010) of a HWT device (chapter 3), the model estimated that similar trials with perfect compliance would have found a longitudinal prevalence ratio for childhood diarrhea of about 0.1, in contrast to an LPR of 0.9 with low compliance (an LPR of 0.8 was estimated by the actual trial, but was not statistically significant). Although a goal of the model was to adjust for bias caused by an imperfect placebo HWT device, uncertainty about the level of compliance greatly influenced the effect of the imperfect placebo. The model also predicted concentrations of diarrheagenic bacteria, viruses, and protozoa in the source water of the field study communities that were consistent with the limited published measurements of pathogen concentrations in other developing country source waters.

Since the importance of compliance was highlighted by the first QMRA model, it was modified to determine how childhood diarrhea incidence changed if HWT compliance and LRVs were varied under differing scenarios (chapter 4). These scenarios were defined by various combinations of: 1) baseline incidence of childhood diarrhea; 2) the pattern of compliance within the community (put simply, the proportion of people who complied perfectly was varied while holding compliance constant at the community level); 3) the size of randomly scheduled spikes of pathogens in untreated drinking water; and 4) etiologic fractions of childhood diarrhea cases attributable to bacteria, viruses, and protozoa. In general, LRVs of 5 prevented little or no additional diarrhea compared to LRVs of 3. However, LRVs of 5 sometimes prevented additional diarrhea if incidence was high, there were large contamination spikes, or many people complied perfectly with HWT.

Although the two QMRA models were informative, they included only the waterborne route of transmission. Although that route is important, diarrheal pathogens can also be transmitted by other routes, e.g., contaminated hands, soil, or objects. The relative importance of these routes is unknown. Furthermore, it was unclear whether findings from the QMRA models would hold in a more complicated and realistic model. An EITS model was programmed (chapter 5), using many functions and parameters from the QMRA models, but incorporating additional functionality such as pathogen shedding by infected people, household structure, multiple routes of transmission, and attenuation (e.g., decay or sequestration) of pathogens in the environment.

The EITS model suggested that viruses and protozoa are attenuated roughly 10 times faster than bacteria in the environment. This may be partially explained by the absence of certain aspects of pathogen transmission that favor bacteria, such as the ability of bacteria to multiply in food or within nonhuman animals. The calibration process may have selected lower attenuation

rates for bacteria in order to meet the calibration criteria, essentially compensating for these missing pathways. It also suggested that visits between households are a relatively minor route for transmission of pathogens, though this might differ if the intensity of exposure were greater during visits (e.g., an infected child being cared for in a different household while the parents were working).

Further analysis of the EITS model will yield additional results, in particular the assessment of interaction between two differing interventions applied to the same community.

6.2. Implications for future diarrheal research and prevention efforts

6.2.1. Conduct and description of field trials

Field trials of public health interventions in developing countries are expensive and difficult to perform, and subject to biases that are difficult to avoid. Models simulating existing field trials can be used to draw inferences regarding what the trial might have measured under different conditions. If biases present during a field trial are well understood, the biases themselves can be simulated within the model, allowing inference of what the trial might have measured in the absence of bias. Although it is impossible to be sure that a model perfectly represents what happened, they can nonetheless be useful for generalizing additional knowledge from field trials.

It is often unclear how to interpret results from a single field trial of a public health intervention, and modeling field trials requires detailed information about the conduct of the trial and the study site. Human communities are extremely diverse, and the health impact from an intervention in one community might differ from the health impact of the identical intervention in a different community. However, detailed characteristics of study communities are seldom published, though there are exceptions (Mata, 1978). Table 6.1 lists information that should be published (likely in supplemental material) about any community that participates in a field trial.

Although it may not always be possible to gather detailed information about each of these aspects, qualitative or semi-quantitative assessments of many of them could be quickly obtained from focus group discussions with several community members. Establishment of multi-year research relationships between study communities and diverse teams of researchers (e.g., epidemiologists, anthropologists, microbiologists, and others) would facilitate collection, analysis, and publication of this information.

Table 6.1. Important community characteristics for measurement in field trials

Sanitation & hygiene	Water source & treatment	Micro-biology‡	Social & community	Nutritional	Epidemi-ology	Climate & geography
Open defecation	Water source(s)	Untreated water†	Community cohesion	Stunting	Diarrheal incidence†	Temperature†
Type and usage of latrines*	Distance to water source	Treated water†	Child care & schooling	Wasting†	Diarrheal longitudinal prevalence†	Humidity†
Hand-washing practices*	Drinking of boiled water*	Human feces†	Household size	Pattern of breast-feeding	Diarrheal mortality†	Precipitation frequency†
Domestic animals	Other HWT methods*	Animal feces†	Cultural practices	Staple foods	All-cause mortality†	Precipitation amount†
Anal cleansing	Water consumption†	Food†	Government or NGO programs*	Weaning foods	Diarrhea in adults†	Solar irradiation†
Attitudes toward child feces	Water usage†	Hands†	Population density	Dietary fiber	HIV/AIDS	Roads & rivers
			Socio-economic status	Protein source(s)	Locally important diseases	Urban/rural

* Should include repeated measurements of compliance over at least 1 year.

† Should be measured repeatedly over at least 1 year.

‡ Ideally, pathogens should be directly quantified, though this is difficult.

HWT: household water treatment. NGO: non-governmental organization. HIV: human immunodeficiency virus.

6.2.2. Compliance with interventions that prevent diarrheal infections

HWT methods are typically assessed based on their antimicrobial efficacy, assuming perfect compliance. However, perfect compliance is unattainable. HWT methods would be more holistically assessed, and public health would be better protected, if HWT guidelines were based on compliance as well as LRVs. Results from these models also indicate that the currently recommended LRVs for HWT (particularly 6 for bacteria, 4 for viruses, and 3 for protozoa (USEPA, 1987), which are sought after in order to advertise that a device 'meets US guidelines') are higher than necessary, since LRVs of 5 prevented little or no additional diarrhea in most scenarios studied, compared with LRVs of 3. Nonetheless, LRVs above 3 might provide additional benefit in certain situations, such as when transmission is primarily waterborne, compliance is high, and large spikes of drinking water contamination are known to occur.

The pattern of compliance within communities is also important. The QMRA and EITS models agree that, for a given value of compliance at the community level, more diarrhea is prevented if more people within the community comply perfectly. For example, it is better for 80% of the population to comply perfectly and 20% to not comply at all, than if the entire population treated 80% of their daily water intake.

Despite its importance, compliance is difficult to measure. At a minimum, the female head-of-household can be asked if they comply; however, this will probably overestimate compliance because some noncompliers will claim they comply out of politeness or a desire to give the 'right' answer. Less biased methods include structured observation (which can still give biased results due to Hawthorne effects) or directly measuring drinking water to determine if it has been treated (this is relatively easy for HWT chlorination methods). Electronic devices have also been attached to HWT devices to record usage data. However, it is not sufficient to know whether households are using a HWT method; it is also important to know how much untreated water

people drink. Even small amounts of untreated water can greatly increase the risk of diarrhea, but carefully quantifying intake of untreated water would require following people throughout their community to observe when, where, and how much water they drank. However, it might be possible to roughly estimate a community's untreated water intake with focus group discussions about water consumption behavior, perhaps combined with a quantitative estimate of total water intake (e.g., Akpata 2004).

Every field trial should measure compliance as carefully as possible, since compliance greatly affects the measured effectiveness of interventions.

APPENDICES

7. APPENDIX A: DISCUSSION OF PARAMETER VALUES USED IN THE MODELS

The models described in this dissertation use a variety of parameter values, most of which were obtained from peer-reviewed scientific publications. Many of the same parameter values are used in all three models. Only children aged less than five years were considered in the two QMRA models described in chapters 3 and 4; both children and 'adults' (defined as people aged 5 years or older) were considered in the environmental infection transmission system (EITS) model described in chapter 5.

Wherever possible, parameter values were chosen to reflect conditions in developing countries with warm climates. However, some parameter values only appear to have been measured in industrialized countries.

Parameter values are discussed roughly in order of progression from exposure to infection to development of disease to shedding of pathogens.

7.1. Water ingestion rate

The daily water ingestion rate for young children was obtained from a study (Akpatha, 2004) which provided bottled water to mothers of 50 rural Nigerian children aged one to three years, and measured the amount remaining in the bottles at the end of the day. This study successfully cross-validated itself by obtaining similar estimates when asking mothers to estimate water intake based on common household measures. Relative humidity during the study was approximately 87% and mean maximum ambient temperature was approximately 31°C.

For adults, a study of Kenyan distance runners reported a mean daily water intake of 2.3 L/day (Fudge et al., 2008). This value also agrees with U.S. Army planning parameters of 2.3 L/day for the sum of urine and sweat losses during physical work in hot climates (USEPA, 2011), as well as the water intake of 2.0 L among children aged 7 to 9 years in the Nigerian study described above (Akpatha, 2004).

The ingestion rates for drinking water described above are much higher than other standard values measured in industrialized countries, e.g., 317 mL/day for two year olds and 1043 mL/day for people over 20 years of age (USEPA, 2011). However, people in industrialized countries are more likely to drink bottled beverages, reducing their exposure to possibly contaminated drinking water; they are also more likely to reside in cooler (possibly air-conditioned) environments, which reduces their demand for fluids.

7.2. Hand-mouth contacts per day

Young children mouth hands and objects more frequently than older children or adults, which indicates that they should have more exposure to pathogens in the environment. Children aged 0 to 5 years contact their mouths with their hands or an object about 15.8 times per hour, while children aged 6-10 years have about 8.1 such contacts per hour (USEPA, 2011). Assuming that this behavior in 6-10 year olds is similar to adults, and further assuming that young children are awake 12 hours per day while adults are awake 16 hours per day, young children have 330 daily contacts with their mouth while adults have 130. The ratio of these (2.54) was used to weight environmental exposure to pathogens in the EITS model more heavily for children.

7.3. Log₁₀ reduction values (LRVs) attributable to interventions

LRVs are particularly relevant to HWT interventions, but handwashing can also be considered to apply LRVs to pathogens adhering to hands. Such LRVs, as well as USEPA and WHO recommendations for antimicrobial effectiveness of HWT devices, are summarized in Tables 2.1 and 2.2, page 64.

7.3.1. LRVs attributable to sanitation

Although sanitation clearly removes pathogens from the environment, it is not straightforward to assign a simple measure of antimicrobial efficacy to it. The effectiveness of

sanitation depends on the manner in which it is constructed and located; pathogens may be carried out of latrines by groundwater or runoff, subsequently contaminating water wells or surface water sources. Removal of feces from latrines for use as fertilizer is another possible way for people to be exposed. Nonetheless, to simulate sanitation in the EITS model, it was necessary to assign some level of antimicrobial efficacy to it. Since anal cleansing materials are often disposed of in a wastebasket or on the ground rather than in the toilet in developing countries, it could be considered that a well-constructed and well-situated latrine would remove all feces from the community, except for those involved in anal cleansing. The amount of feces involved in anal cleansing was determined as follows: a daily average of 4.4 mg of fecal nitrogen has been measured on toilet paper (Calloway et al., 1971); feces are approximately 1.9% nitrogen (Rivero-Marcotegui et al., 1998); thus approximately $0.0044 / 0.019 = 0.23$ g of feces would remain on anal cleansing materials from a single defecation event, which occurs approximately once per person per day (Weaver, 1988). This figure is roughly corroborated by a report of an average of 0.1 g of feces per set of underwear in university students (Gerba, 2001). Considering that an adult on a high-fiber diet excretes approximately 225 g of feces daily (Davies et al., 1986), approximately one thousandth ($0.23/225$) of daily fecal output would not enter the latrine; this corresponds to an LRV of 3.

7.3.2. LRVs attributable to intervention and placebo filters in the Lifestraw RCT

The first QMRA model (chapter 3) used \log_{10} reduction values (LRVs) for the LifeStraw Family Filter (LFF) from a laboratory study of the device (T. Clasen, Naranjo, et al., 2009). In contrast to the LRV of 6.9 for *E. coli* reported in the laboratory study, the mean LRV from functioning LFFs for thermotolerant coliforms (TTC) reported by the Lifestraw RCT was only 2.98; removal of other organisms was not determined during the RCT (Boisson et al., 2010). However, the mean LRV of 2.98 for TTC was determined by assuming a count of 1 CFU per 100

mL where TTC were not detected in the filtered water (64% of samples); the median LRV was therefore technically > 3.1 , and the maximum LRV measurable in any sample taken during the RCT was 4.5. Therefore, the average LRV for TTC in the RCT was greater than the value of 2.98 that was reported; it is impossible to know how much greater.

In addition, given the size exclusion treatment mechanism by the ultrafilter membrane in the LFF (20 nm pore size), it seems unlikely that the treatment performance in the field was much lower than what was observed in the laboratory. Although TTC might have passed through some devices due to non-visible leaks in the gaskets or defects in the membrane material, the presence of TTC in outlet samples was not correlated with specific devices, nor were there trends in time over the one-year field trial. In addition, intervention and placebo devices were monitored and repaired as necessary by the study team (Boisson et al., 2010). Thus, a more likely explanation for the lower LRVs measured in some intervention LFFs is that contamination of the outlet tube with TTC occurred. Based on this reasoning as well as the detection limit issues described above, it was determined that the most reasonable LRVs for the intervention LFF against the three pathogen types were the values determined in the laboratory study (T. Clasen, Naranjo, et al., 2009).

For the placebo device, the Lifestraw RCT reported a mean LRV of 1.05 (i.e., 91% removed) for thermotolerant coliforms (TTC) (Boisson et al., 2010). Further analysis of the data revealed that the LRV varied greatly between -1 (indicating a tenfold higher concentration in filtered water) and 3 (99.9% removed). There was no evidence that removal by the placebo increased significantly over time, or was restricted to particular defective devices. It is likely that some of the observed variability in removal was due to the inherent variability in water quality and measurement associated with indicator organisms in water sources (K. Levy, A. E. Hubbard, K. L. Nelson, et al., 2009). Based on the observed performance and the design, mechanisms of

removal may have included: (1) straining by the prefilter of the LFF, especially if TTC were attached to larger particles; and (2) adhesion to the biofilm that likely developed on surfaces, in particular the rubber tubing that replaced the filter membrane cartridge. These mechanisms might not have occurred during laboratory testing because of the short testing period (3 weeks) and simpler challenge water quality (no removal of the three test organisms by the placebo device was measured in the laboratory). Since the mechanism of this reduction is unclear, it is difficult to predict removal of viruses and protozoan cysts by the placebo device in the field, but it seems likely that if some bacteria were removed, some viruses and protozoa were also removed. Thus, the simplest assumption was chosen: application of the same LRV of 1.05 to all three marker pathogens.

7.4. Dose response functions

Dose response functions are critical for translating estimated dose into a probability of infection. Such functions can be obtained from studies in which volunteers are fed widely varying doses of pathogens, and monitored for development of the response (e.g., infection or disease). These studies have been conducted on healthy adults because it would be unethical to conduct such studies on children. It is likely that dose response relationships in developing-country children differ in unknown ways; ill or malnourished children could have decreased resistance to infection, or they could have increased immunity due to more frequent exposure. The use of infection as the response, combined with morbidity ratios from developing countries to determine disease, constitutes a crude means of adjustment for this uncertainty. The dose response functions used in this model are graphed in Figure 3.3 (page 99, semilog plot) and Figure 4.6 (page 138, log-log plot), and the equations are given on page 98. The same dose response functions and parameters were used in all three models.

7.4.1. Dose response for *E. coli* infection

Although there are many dose response models describing *E. coli* disease (Haas et al., 1999; P F M Teunis et al., 1996), few data are available regarding dose response of *E. coli* infection. In published *E. coli* feeding studies, data regarding *E. coli* dose and infection are concentrated in the top half of the dose response curve (i.e., > 50% of the participants become infected), meaning that the response at low dose levels is very uncertain (Anon, 2012; P F M Teunis et al., 1996). Only one feeding study (H L DuPont et al., 1971) appears to exist that uses infection as the response and also includes data in the lower half of the dose response curve; it had 3 dose levels and 19 participants, who were fed enteroinvasive *E. coli* (EIEC). The data were used to fit a beta-Poisson dose response function in R version 2.12, using maximum likelihood methods (Anon, 2012). Although dose response models of infection are available for enterohemorrhagic *E. coli* and *Shigella* species., they are not appropriate for these models because they are much more infectious than diarrheagenic *E. coli* (J P Nataro & J B Kaper, 1998; Anon, 2012).

7.4.2. Dose response for rotavirus and *Giardia*

We used previously published analyses to parameterize the dose response functions (J B Rose et al., 1991; Haas et al., 1993). The feeding data for these fits have also been published (Rendtorff, 1954; Ward et al., 1986).

7.5. **Incubation periods**

Incubation periods were only used in the EITS model (chapter 5), not the two QMRA models (chapters 3 and 4). The incubation period (time from exposure to development of disease) was assumed to be the same as the prepatent period (time from exposure to detection of the pathogen in the host, i.e., shedding of the pathogen), although this is not always the case; *Giardia* disease tends to precede excretion of cysts by roughly one week (A. M. Jokipii & L.

Jokipii, 1977).

The median incubation period across nine ETEC outbreaks in the USA was 1.75 days (Dalton et al., 1999). The mean incubation period for rotavirus disease was approximately 3.2 days in 5 adults who were voluntarily exposed (Kapikian et al., 1983). For *Giardia* infection, the prepatent period was used; a study where *Giardia* cysts were fed to imprisoned adult men gave a mean of 13.5 days (Rendtorff, 1954), while a study of Finnish giardiasis patients who had recently visited an area of Russia with endemic giardiasis reported a median prepatent period of 14 days (A. M. Jokipii & L. Jokipii, 1977).

7.6. Morbidity ratios

The morbidity ratio for a pathogen is the number of symptomatic infections in a community divided by the total number of infections by that pathogen. They can be obtained by systematically examining stools from a population of children, regardless of their diarrhea status. Asymptomatic infection with diarrheal pathogens is common, particularly in developing countries. Morbidity ratios for young children were used in all three models; morbidity ratios for adults were only used in the EITS model.

7.6.1. Morbidity ratios in young children

The morbidity ratio for *E. coli* (0.21) was taken from a study (Vergara et al., 1996) of children under 5 years of age in a subtropical, rural region of Argentina, in which ETEC and EPEC were measured in feces quarterly over two years.

The morbidity ratio for rotavirus (0.36) was provided by a cohort study (Fischer et al., 2002) from birth to age two years in periurban children in Guinea-Bissau, where stool specimens were collected weekly and 116 rotavirus infections were identified in 94 children.

The morbidity ratio for *Giardia* (0.59) was obtained from 210 children aged 1 month to 9 years in Peruvian periurban districts with poor sanitation; each child contributed one stool

sample over the course of a year (Pérez Córdón et al., 2008). Another study (M S Prado et al., 2005) suggests a much lower morbidity ratio (~0.03) for *Giardia* in Brazilian children aged 6-45 months, but was not used for these reasons: 1) stool samples were collected repeatedly for each child over four months, so multiple samples may have been taken during the same infection episode; 2) information about individual episodes of diarrhea was not provided; 3) its low morbidity ratio is inconsistent with higher morbidities for *Cryptosporidium* (Kirkpatrick et al., 2008; Bushen et al., 2007; Pérez Córdón et al., 2008), which *Giardia* also represents by proxy in these models.

7.6.2. Morbidity ratios in adults

Suitable morbidity ratios for diarrheagenic *E. coli* infection in adults appear to be unavailable. Although two studies of diarrheagenic *E. coli* infection in adult Kenyan food handlers have been published (Oundo et al., 2008; Onyango et al., 2009), they considered diarrhea only in terms of loose stools, without regard for stool frequency. Since three or more stools per day is commonly used as a criterion for diarrhea (USAID et al., 2005), these studies probably overestimated the morbidity ratio by counting 1-2 loose stools per day as diarrhea (they reported 0.34 and 0.62, respectively). Therefore, the morbidity ratio for *E. coli* in children described above (Vergara et al., 1996) was also used for adults.

Morbidity ratios for rotavirus infections in adults are also difficult to find, but they should be substantially less than morbidity ratios in children due to acquired immunity. In developing countries, frequent rotavirus exposure throughout life probably contributes to maintenance of immunity in adults (Bishop, 1996). Morbidity ratios of 5/16 among adults and 6/28 among children were measured in a prospective study of rotavirus infection in middle-class families served by a particular pediatric practice in the USA (Rodriguez et al., 1987); although the morbidity ratio among children in Guinea-Bissau (Fischer et al., 2002) of 0.36 is similar to the

above ratio of 5/16 in adults, it is rather higher than the morbidity ratio among children in that pediatric practice. The morbidity ratios from Rodriguez et al. (1987) were not used because they did not appear representative of rotavirus disease in developing countries, where children are more likely to acquire serious rotavirus disease and adults may have stronger immunity due to more frequent exposure. In the absence of better information, a value of 4/18 was used in the EITS model for the rotavirus morbidity ratio in adults, based a study of 18 adult volunteers ingesting rotavirus from 0.2g of stool of an ill child, four of whom became ill (Kapikian et al., 1983). However, only twelve volunteers showed a serologic response, and only five volunteers shed the virus, four of whom had diarrhea (Kapikian et al., 1983). The volunteers were chosen partly based on their low antibody titers against rotavirus.

The morbidity ratio for *Giardia* was obtained from a study of Pakistani men and their children aged two to twelve years, who lived in an area where untreated wastewater was used for agricultural irrigation. Although 67.2% of the study population was infected, only 2.8% of the infected people reported diarrhea during the week before a stool sample was collected (Ensink et al., 2006), similar to the Brazilian study of young children described above (M S Prado et al., 2005). Although the number of study participants who were children aged under five years was not reported, the population is likely to be dominated by persons aged five years or older.

7.7. Durations of illness and infection

Determination of the duration of infection in underdeveloped settings is difficult, since immunity is incomplete for most diarrheal pathogens, and reinfection is common. What appears to be a long or intermittent period of infection (or illness) may in fact be multiple reinfections. Also, when childhood diarrheal disease is studied in developing countries, it is usually done in a hospital or clinic setting; measurements of the duration of infection or disease are biased upward in these situations since more severely ill children are more likely to seek care. Also, the duration

of infection as measured by detection of a pathogen in stool does not necessarily coincide with the duration of illness. Since information on asymptomatic infection is difficult to obtain, the duration of illness was assumed to be the same as the duration of infectiousness. The same durations of illness and infectiousness were used for adults and children.

7.7.1. Duration of diarrheagenic *E. coli* infection and illness

The duration of *E. coli* diarrhea (mean of 3.0 days) was obtained from a study (Estrada-Garcia et al., 2009) of Mexican children aged less than two years, who were monitored prospectively for the development of infection or disease from diarrheagenic *E. coli*. A gamma distribution was fit to these data for use in the three models. A feeding study of ETEC and EPEC in healthy United States adults found a similar mean duration of illness of 3.4 days (R E Black et al., 1982). However, a study of 10 ETEC outbreaks among adults in the USA reported median durations of illness from 4 to 6 days (Dalton et al., 1999).

7.7.2. Duration of rotavirus infection and illness

A mean rotavirus diarrhea duration of approximately 2.5 days was reported by a study (Kapikian et al., 1983) of four experimentally infected adult volunteers, with durations of 1, 2, 3, and 4 days. Accordingly, a uniform distribution from 1 to 4 was used for viral infection duration in all three models.

7.7.3. Duration of *Giardia* infection and illness

An outbreak investigation of giardiasis from a contaminated water supply system in Massachusetts yielded a mean duration of *Giardia* disease of 11.3 days (Kent et al., 1988); a gamma distribution was fit to these data, and used in all three models. Although a higher mean duration of 18.3 days was reported in a feeding study of healthy imprisoned men (Rendtorff, 1954), the value from the outbreak was used in the models because it was from a more diverse population. *Cryptosporidium* diarrhea in developing countries can be similarly long-lasting;

mean durations of diarrhea of 21 days for *C. hominis* and 13 days for *C. parvum* were reported in a cohort study of children in a Brazilian shantytown during their first 5 years of life (Bushen et al., 2007).

7.8. Duration of immunity

Full treatment of immunity to diarrheal infections is extremely complicated, and it is beyond the scope of this dissertation. For *E. coli*, *Campylobacter*, and *Giardia*, immunity protects against disease, but does not necessarily prevent reinfection (R H Gilman et al., 1988; Cravioto et al., 1990; Valentiner-Branth et al., 2003; A. H. Havelaar et al., 2009). For rotavirus, immunity is long-lasting, but is strain-specific; many differing strains exist, and adults can still develop rotaviral infection and disease (Wenman et al., 1979; Kapikian et al., 1983; Ward et al., 1986). Further considering that each pathogen type also represents other similar pathogens (e.g., *Campylobacter* for bacteria, norovirus for viruses, and *Cryptosporidium* for protozoa), repeated infections therefore occur for all three pathogen types. Therefore, some information regarding immunity is already included in these models in the form of the morbidity ratios discussed above (page 220).

Infection with diarrheagenic *E. coli*, rotavirus, or *Giardia* tends to persist for several days, as long as a week, after symptoms resolve (R E Black et al., 1982; Rendtorff, 1954; Kapikian et al., 1983). Therefore the QMRA models assumed that immunity from each pathogen lasted 7 days after infection resolved. However, this assumption was reconsidered in the course of constructing the EITS model, in which each pathogen type had a period of immunity lasting one day, principally to separate episodes of disease.

7.9. Incomplete recall of diarrheal disease

The probability of remembering and reporting an episode of diarrheal illness is lower if the illness had resolved longer ago. Pakistani mothers were asked to recall diarrhea in their children

aged < 5 years (Zafar et al., 2010); the proportion who were reported ill was high for the previous two days, but dropped sharply for the third day. The amount of diarrhea reported for the third through the sixth day was similar. Incomplete recall was only considered in the first QMRA model (chapter 3); the other models measured diarrheal incidence as if it was reported perfectly.

7.10. Fecal excretion of pathogens

7.10.1. Concentrations of pathogens in feces

The numbers of pathogens per gram of feces in infected people have seldom been measured. Measurements of EIEC and ETEC in six adult male volunteers with dysentery or diarrhea yielded estimates from 10^8 to 10^9 CFU/g (H L DuPont et al., 1971).

The available studies of *Giardia* cyst concentrations in stool are somewhat problematic. Although the EITS model uses a mean measurement (5.7×10^5 cysts/g) from 15 Colombian children aged three to seven years (Danciger & M. Lopez, 1975), these children were divided into three groups of five children each: high excretors, low excretors, and mixed excretors. It was unclear which type predominated, or how the children were chosen for the study, although cyst concentrations were similar in formed and diarrheic stools (Danciger & M. Lopez, 1975). A study that quantitated *Giardia* cysts in the feces of seven ill soldiers (Porter, 1916) presented detailed time series of cyst concentrations; however, all of these soldiers were ill for several months, had been ill for at least 1.5 months before examination of their feces, had received differing chemotherapies, and were probably more severely affected than their peers. Cyst concentrations varied widely from day to day, ranging from 0 to 2×10^7 cysts/g.

Although rotavirus particles can reach extremely high concentrations ($\sim 10^{11}$ /g) in stool when measured by electron microscopy, only a tiny fraction appears to be viable; an average of 2×10^6 FFU/g has been reported from 5 fecal specimens from rotavirus diarrhea patients (Ward et

al., 1984).

7.10.2. Amount of feces excreted

To determine the amount of pathogens excreted given a concentration of pathogens in feces, the volume of feces must also be known. The EITS model used fecal output measurements from individuals consuming high-fiber diets: Nigerian children aged 6-60 months defecated 109 g on average (SD 54) (Akinbami et al., 1995), while adult British vegans defecated 225g on average (SD 91) (Davies et al., 1986). Although people living in developing countries might be expected to consume more fiber than people in industrialized countries, this is not necessarily the case; a dietary survey of urban adult Nigerian hospital patients and medical students indicated similarly low daily dietary fiber intake (~8g) and daily stool mass (~140g) as people in industrialized countries (Ogunbiyi, 1978).

Although the effect of fiber intake on pathogen concentrations in feces is unclear, gastrointestinal pathogens are associated with the intestinal wall, and therefore the amount of feces shed might be more related to the internal surface area of the gut that has been colonized, rather than the mass of feces passing through the gut. Diets higher in fiber yield larger and softer stools, and promote healthy microbial communities in the gut, leading to improved nutrient absorption; fiber may also reduce risk of diarrheal disease in some circumstances (Brownawell et al., 2012). Therefore communities where high-fiber diets are common might experience less diarrhea and also excrete fewer pathogens into their environment.

The number of daily defecation events would be expected to affect the amount of pathogens present on hands and in the household environment, because each defecation event represents an episode of hand contamination with feces. In industrialized countries, healthy adults and young children generally have between 1 and 2 bowel movements per day (Weaver, 1988); the Nigerian medical students mentioned above had a mean of 0.89 bowel movements per

day (Ogunbiyi, 1978). Two studies of young children hospitalized for acute diarrhea in Peru and Bangladesh indicated similar mean daily fecal outputs of 365g and 310g (Lembcke et al., 1989; S. K. Roy et al., 1997), and a study of Tunisian adults with acute diarrhea reported mean daily fecal output of 499g (Hamza et al., 1999). Although the Peruvian study reported 9.5 (SD 5.0) stools during the 24 hours before admission, and the Tunisian study reported 6.3 (SD 2.5) stools the day of admission, people seeking care are likely to have more severe diarrhea. As a simplifying assumption, the EITS model used a value of 3 stools/day for all persons with diarrhea; when multiplied by the fecal output in persons without diarrhea, it roughly matches the fecal output described in the Peruvian, Bangladeshi, and Tunisian studies above.

7.11. Inactivation, removal, or attenuation of pathogens in the environment

Inactivation of pathogens is highly variable, depending on the type of pathogen, temperature, humidity, and the microenvironment where the pathogen resides (e.g., water, skin, feces, soil, etc.). Furthermore, some pathogenic bacteria may multiply in contaminated food. An exponential decay rate of 0.6 per day was initially considered as a basis for all pathogens in all EITS model compartments; it was based on daily rates of: 0.64 for *E. coli* in unfiltered river water that was bottled and stored at 15°C (Flint, 1987); 0.58 for rotavirus in water from a creek that received discharges of domestic sewage, which was then bottled and stored at 20°C (Pancorbo et al., 1987); and 0.55 for *Giardia* fitted to data from inactivation studies at 15-20°C in distilled water, tap water, and *in situ* river water (Wickramanayake et al., 1985; deRegnier et al., 1989). However, pathogens may be removed from a system by many means beyond simple decay, such as being transported in water or soil particles, ingestion by animals, sedimentation in water, sequestration in underwater sediments or underground, etc. Furthermore, pathogen decay rates can vary greatly depending on temperature, solar radiation, humidity, and other factors. Since the actions of these various attenuation mechanisms are unclear and are likely to vary

greatly in different communities, the EITS model assigned varying values to three attenuation calibration parameters without reference to particular values from the literature.

7.12. Pathogen movement from land to surface water

Modeling the movement and inactivation of pathogens within land and surface water is a large and complicated area of research, and is beyond the scope of this dissertation. The EITS model used a few simple parameter values to describe pathogen movement from land to water; 0.1% on dry days, 5% on rainy days, with rain occurring once every 14 days on average. In chapter 5, the calibration parameters CP_{DI} , CP_{Sf} , and CP_{atten} can be considered to include variability in scenarios due to hydrogeological considerations.

7.13. Demographic parameters

Infection transmission could be affected by demography and community structure, since children are more susceptible to diarrheal infections than adults, and people share an environment with other residents of their household. A household size of roughly five people is common in many developing countries, although it can range from four to eight (Ayad et al., 1994). Household size across a wide variety of industrialized and developing countries can be adequately represented by a Poisson distribution truncated at zero (Jennings et al., 1999). In many sub-Saharan African countries, on average 18% of each household are children aged less than five years; however, it is lower in other developing regions, e.g., 12-17% among several Latin American countries (Ayad et al., 1994).

7.14. Summary table of all parameter values

Table 7.1 summarizes parameter values used in all three models described in this dissertation. The two QMRA models (chapters 3 and 4) used very similar parameter sets. The EITS model (chapter 5) also used many of the parameters from the QMRA models, as well as

some additional parameters. Parameters that were varied during the estimation steps of the three models are not included here (but see chapters 3, 4, and 5, pages 105, 130, and 183). The variable names used in the model code are given in the table; the QMRA models are designated 'Q', while the EITS model is designated 'E'. Where necessary, the first and second QMRA models (chapters 3 and 4) are designated 'Q1' and 'Q2'. Some variables were stored in vectors; where necessary, the element of the vector is denoted in parentheses, e.g., 'VariableName(1)'.

Table 7.1. Summary table of all parameter values

Description of parameter values, and pages where discussed	Value	Variable name(s)	Reference
Water ingestion, young children (page 214)	1.178 L/day	Q: drinkKids E: Wdd(1)	(Akpata, 2004)
Water ingestion, adults (page 214)	2.3 L/day	E: Wdd(2)	(Fudge et al., 2008)
Size of stored drinking water container in each household	25 L	E: Ws	
Daily probability of a rainfall event (page 228)	1/14	E: rRain	
Proportion of pathogens moving from land to surface water daily (page 228)			Kara L. Nelson, personal communication
Without a rainfall event	0.001	E: xRunoff(1)	..
With a rainfall event	0.05	E: xRunoff(2)	..
Log ₁₀ reduction values (LRVs), intervention group, LifeStraw RCT (pages 87 & 216)			(T. Clasen, Naranjo, et al., 2009)
<i>Escherichia coli</i>	6.9	Q1: LRInt(1)	..
Rotavirus	4.7	Q1: LRInt(3)	..
<i>Giardia</i>	3.6	Q1: LRInt(2)	..
LRVs, placebo group, LifeStraw RCT, all 3 pathogen types (page 216)		Q1: LRPla	
Calibration step	1.05	..	(Boisson et al., 2010)
Estimation step	0	..	
LRV for handwashing, all pathogen types (page 64)	0.46	E: lHand	(Luby et al., 2001)
LRV for sanitation, all pathogen types (page 215)	3	E: lSan	(Calloway et al., 1971; Rivero-Marcotegui et al., 1998)
Additional LRVs, particularly for HWT: see Tables 2.1 and 2.2, page 64.			
Dose response function parameters (pages 98 & 218)			(Anon, 2012)

Table 7.1 (cont'd)

Description of parameter values, and pages where discussed	Value	Variable name(s)	Reference
<i>E. coli</i> (enteroinvasive); beta-Poisson parameters	$\alpha = 0.155$ $N_{50} = 2.11 \times 10^6$	Q&E: alpha(1) Q&E: KorN50(1)	(H L DuPont et al., 1971)
Rotavirus; beta-Poisson parameters	$\alpha = 0.2531$ $N_{50} = 6.171$	Q: alpha(3) E: alpha(2) Q: KorN50(3) E: KorN50(2)	(Ward et al., 1986; Haas et al., 1993)
<i>Giardia</i> ; exponential k parameter	0.0198	Q: KorN50(2)	(Rendtorff, 1954; J B Rose et al., 1991)
Baseline incidence (infections/person-year) with the three pathogen types (page 180)			(Lanata & W. Mendoza, 2002; M. E. Wilson, 2005)
<i>Escherichia coli</i>	1.17	E: BaseInf(1)	..
Rotavirus	0.347	E: BaseInf(1)	..
<i>Giardia</i>	0.212	E: BaseInf(1)	..
Incubation periods (page 219)			
<i>Escherichia coli</i> (ETEC)	2 days	E: latent(1)	(Dalton et al., 1999)
Rotavirus	3 days	E: latent(2)	(Kapikian et al., 1983)
<i>Giardia</i>	14 days	E: latent(3)	(Rendtorff, 1954; A. M. Jokipii & L. Jokipii, 1977)
Morbidity ratios for children (proportion of infected who are symptomatic; page 220)			
<i>Escherichia coli</i>	0.214	Q: MorbidityK(1) E: MRk(1)	(Vergara et al., 1996)
Rotavirus	0.397	Q: MorbidityK(3) E: MRk(2)	(Fischer et al., 2002)

Table 7.1 (cont'd)

Description of parameter values, and pages where discussed	Value	Variable name(s)	Reference
<i>Giardia</i>	0.590	Q: MorbidityK(2) E: MRk(3)	(Pérez Córdón et al., 2008)
Morbidity ratios for adults (proportion of infected who are symptomatic; page 221)			
<i>Escherichia coli</i>	0.214	E: MRa(1)	(Vergara et al., 1996)
Rotavirus	0.222	E: MRa(2)	(Kapikian et al., 1983)
<i>Giardia</i>	0.03	E: MRa(3)	(Ensink et al., 2006)
Distributions of duration of infection and infectiousness (page 222)			
<i>Escherichia coli</i> (gamma distribution, mean 3 days)	shape = 1.775 scale = 1.690	Q&E: durEc.m	(Estrada-Garcia et al., 2009)
Rotavirus (uniform distribution; mean 2.5 days)	Range 1-4 days	Q&E: durRo.m	(Kapikian et al., 1983)
<i>Giardia</i> (gamma distribution, mean 11 days)	shape = 3.206 scale = 3.431	Q&E: durGi.m	(Kent et al., 1988)
Period of immunity for all pathogen types (QMRA models; page 224)	7 days	Q: ImmuneTimes	
Period of immunity for all pathogen types (EITS model; page 224)	1 day	E: immune	
Chance of remembering diarrhea >2 days in the past (page 224)	0.54	Q1: remembrance	(Zafar et al., 2010)
Number of pathogens per gram of feces (page 225)			
<i>Escherichia coli</i>	5×10 ⁸	E: Mpgf(1)	(H L DuPont et al., 1971)
Rotavirus	2×10 ⁶	E: Mpgf(2)	(Ward et al., 1984)
<i>Giardia</i>	5.7×10 ⁵	E: Mpgf(3)	(Danciger & M. Lopez, 1975)
Fecal output, children aged < 5 years, high fiber diet (page 226)	109 g/day	E: fpp(1)	(Akinbami et al., 1995)
Fecal output, people aged five years or more, high fiber diet (page 226)	225 g/day	E: fpp(2)	(Davies et al., 1986)

Table 7.1 (cont'd)

Description of parameter values, and pages where discussed	Value	Variable name(s)	Reference
Feces entering household environment per defecation event (page 226)	0.23 g	E: fHands	(Calloway et al., 1971; Rivero-Marcotegui et al., 1998)
Defecation events per day, people without diarrhea (page 226)	1	E: rPoopInf	
Defecation events per day, people with diarrhea (page 226)	3	E: rPoopIll	
Number of mouth contacts per day, children (page 215)	330/day		(USEPA, 2011)
Number of mouth contacts per day, adults (page 215)	130/day		..
Ratio of the above mouth contacts per day, children/adults	2.54	E: wHandMouth	..
No. children <5 years, LifeStraw RCT intervention group	85	Q1: nKidsInt	(Boisson et al., 2010)
No. children <5 years, LifeStraw RCT placebo group	105	Q1: nKidsPla	..
Demographic and community information (page 228)			
Number of households in the simulated community	200	E: nHH	
Proportion of the community that was children aged < 5 years	0.18	E: pKids	(Ayad et al., 1994)
Mean persons per household (Poisson distribution)	5	E: mPHH	(Ayad et al., 1994; Jennings et al., 1999)
Mean household network degree (connections per household)	5.3	E: meanDeg	(Zelner et al., 2012)
Daily probability of a visit, per household connection	2/7	E: rVisit	
Shape parameter for all gamma distributions of pathogen type concentrations (the two QMRA models only) ^a	1.85	Q: Shapes	(Boisson et al., 2010)
Longitudinal prevalence of reported diarrhea for each LifeStraw RCT group			
Intervention (LP_{Irad})	0.0749	Q1: LongPrevs(2)	..
Placebo (LP_{Prad})	0.0896	Q1: LongPrevs(1)	..

Table 7.1 (cont'd)

Description of parameter values, and pages where discussed	Value	Variable name(s)	Reference
Longitudinal prevalence ratio of reported diarrhea (LPR_{rad}) measured by the LifeStraw RCT	0.872	Q1: longPrevKids-DesiredRatio	..
Compliance with device use, first QMRA model: chance of using device on a given day			
Calibration step	0.65		..
Estimation step	0, 0.65, or 1.00		
Compliance with device use: If using device on a given day, proportion of water treated			
Calibration step	2/3 or 1/3		

^a The scale parameters for the gamma distributions of pathogen concentrations in the QMRA models were calculated from the mean concentrations that were obtained during calibration of the model in chapter 3 (Table 3.2, page 96).

8. APPENDIX B: ADDITIONAL INTERVENTIONS FOR MITIGATING DIARRHEA

Although this dissertation gives particular attention to household water treatment, sanitation, and handwashing, there are other important interventions that prevent or mitigate diarrhea. They are briefly reviewed here.

8.1. Nutritional interventions

Diarrhea and malnutrition form a vicious cycle in which diarrhea leads to malnutrition, which reduces resistance to disease, which leads to more diarrhea (Motarjemi et al., 1993). Diarrheal illnesses contribute more strongly to malnutrition than other infections, such as respiratory infections (Motarjemi et al., 1993). Children with diarrhea often refuse food, contributing to malnutrition, although they may still accept breast milk eagerly (de Zoysa et al., 1991). Mothers sometimes believe that food should be withheld from children with diarrhea, which can further aggravate the diarrhea-malnutrition cycle (C. E. Taylor & Greenough, 1989). In addition, young children with diarrhea commonly refuse food, leading to (or exacerbating) malnutrition, but it has been shown in several countries that they do not usually lose their appetite for breastmilk (Huffman & Combest, 1990).

8.1.1. Breastfeeding

Breast milk can provide all necessary nutrients for infants aged 6 months or less; even in very hot conditions, breast milk can supply all necessary fluid intake (Huffman & Combest, 1990). In addition, malnourished mothers can still produce sufficient good-quality milk to support exclusive breastfeeding (Huffman & Combest, 1990). Although the World Health Organization recommends exclusive breastfeeding for the first 6 months of life, and continuing to breastfeed until age two and a half years or longer, exclusive breastfeeding is uncommon (about 35% of infants aged 0 to 4 months) (Hill et al., 2004). Breast milk transfers maternal antibodies from mother to child, providing passive immunity. Although breastfeeding infants can

still acquire diarrheal infections, they seldom lead to malnutrition while the child is exclusively breastfeeding (Motarjemi et al., 1993).

Both exclusive and partial breastfeeding are strongly protective against diarrhea, particularly in infants. In a review (Feachem & Koblinsky, 1984) of this topic, comparing any breastfeeding against no breastfeeding, the relative risk for diarrhea was 0.33 in 0-3 month olds, 0.42 for 3-6 month olds, and 0.71 for 6-11 month olds. However, protection from diarrhea does not appear to continue after breastfeeding has stopped (Feachem & Koblinsky, 1984).

Exclusive breastfeeding appears to be effective enough to neutralize the effects of poor sanitation on diarrheal risk (VanDerslice et al., 1994; Feachem & Koblinsky, 1984). However, supplementing breastmilk with small amounts of contaminated water can double diarrheal risk, and after adding additional foods or weaning, the situation is even worse (VanDerslice et al., 1994). Dose-response relationships have been observed, with additional non-breastmilk feeds incrementally increasing diarrheal mortality risk, and breastmilk feeds similarly reducing mortality risk (Huffman & Combest, 1990).

8.1.2. Zinc supplementation

Zinc deficiency has broad effects, particularly on immunity, but its symptoms are not obvious (Bhutta et al., 1999). Zinc is not stored in the body and frequent intake is therefore required (Bhutta et al., 1999). Treating diarrhea with 20 mg/day (10 mg/day for <6 month olds) oral zinc supplementation for 10-14 days can make disease less severe (USAID et al., 2005).

Several trials of zinc supplementation have been reviewed (Bhutta et al., 1999), finding that it was effective in reducing both the incidence of diarrhea by 18% and prevalence by 25% in children under 3 years of age. It was even more effective (41% reduction) on pneumonia incidence.

Cereal flours can be fortified with zinc, which would make sustained zinc supplementation

attainable without any particular action by the family, provided that common brands of flour are supplemented. In the absence of a fortified staple food, regular supplementation seems unlikely to be sustainable. Zinc fortification of wheat flour is mandatory in Indonesia, Jordan, Mexico, and South Africa; the latter two countries also require zinc fortification of maize flour (Kenneth H Brown et al., 2010).

8.1.3. Other nutrients

Vitamin A also appears to be beneficial in reducing mortality and severe diarrhea, but may not impact diarrheal incidence overall (Long et al., 2007). Iron supplementation may actually exacerbate diarrhea in some cases (Long et al., 2007). There has been relatively little study of other micronutrients (Long et al., 2007).

8.2. Treatment of diarrheal illness

There are 5 broad rules for home treatment of diarrhea (USAID et al., 2005):

1. Give more fluids than usual, particularly breastmilk and hygienically prepared oral rehydration solution (ORS; see below)
2. Supplement the child's diet with zinc for 10-14 days
3. Continue to feed the child, particularly well-cooked, energy-rich, non-sugary foods that are easy to digest
4. Return to the clinic if the child is dehydrated or fails to improve after 3 days
5. Do not give antibiotics unless there is blood in the stool, signifying dysentery.

8.2.1. Oral rehydration salts/solution/therapy (ORS)

Use of ORS drastically reduces dehydration, and consequently death, due to diarrhea. The original WHO-recommended ORS formulation of ORS from the 1960s reduced dehydration and death, but did not otherwise impact diarrhea severity or duration (Atia & Buchman, 2009). An improved lower-osmolarity formula began to be promoted in 2002 (Hill et al., 2004). Compared

with the original formulation, it has been shown to decrease stool volume in at least 5 trials and to decrease diarrheal duration in at least 2 trials (Atia & Buchman, 2009). It is unclear to what extent this might reduce secondary transmission of diarrhea.

Zinc supplementation can also be used in conjunction with ORS. Although it does not greatly decrease disease duration, it decreases stool output in comparison with ORS alone (Bahl et al., 2002; Bhatnagar et al., 2004).

8.2.2. Access to care

Improved access to medical care is probably most useful for preventing mortality from severe or complicated diarrhea. However, it could also impact diarrhea through decreasing the duration of disease via effective treatment. It might also indirectly prevent disease by promoting other behaviors, such as handwashing or good nutrition.

8.3. Vaccination

Among the diarrheal pathogens, only rotavirus has a highly effective vaccine; it is 85-95% effective against severe rotavirus gastroenteritis in children (CDC, 2012). However, a cholera vaccine that is about 50% effective is available, and vaccines are available for other pathogens that are transmitted by the fecal-oral route, such as hepatitis A virus, poliovirus, and *Salmonella enterica* Typhi (which causes typhoid fever) (CDC, 2012; CDC, 2011). Rotavirus vaccine is not generally available in developing countries, although rotavirus vaccination programs are being investigated. The rotavirus vaccine is less effective in children in developing countries, compared with developed countries, but it remains effective enough to have a large public health impact (Cherian et al., 2012). However, vaccination against enterotoxigenic *E. coli* (ETEC) is considered feasible, and vaccine candidates are being researched (R. I. Walker et al., 2007).

9. APPENDIX C: SOURCE CODE FOR THE MODELS

9.1. Overview of the source code for the models

All models were written in Octave, which is a freely available open source programming language for Linux, Mac, or Windows (<http://www.gnu.org/software/octave/>). Octave and MATLAB have nearly identical syntax, and the code runs on either platform without modifications (though it runs faster with MATLAB). There is no formal documentation for these models, aside from this dissertation; however, the code is extensively commented.

The source code for all three models consists of several text files that can be viewed in any text editor: one main file and several functions/subroutines that it calls. They have the extension “.m” and are called 'm-files'. To do a test run of any of the models, copy all m-files to a single directory, start Octave or MATLAB, set the working directory to that directory (if you don't know how, type 'help cd'), and type the name of the appropriate main file (omitting the '.m'). If copies of the source code cannot be found online, the source code can be recreated simply by copying it from this dissertation and pasting it into a text editor.

All three models should run without errors on Octave 3.2 or 3.3, as well as MATLAB 7.11-7.13. It is very likely that they will also run on later versions of Octave or MATLAB.

All three models are licensed under the GNU General Public License (see also the copyright notice on page 4), with the exception of the file “erdrey.m” used in the EITS model, which is from the CONTEST toolbox (A. Taylor & Higham, 2008) and is reproduced with the permission of the authors (Des Higham, personal communication, 24 August 2012).

9.2. GNU General Public License

The following license applies to all computer code in this dissertation, except otherwise noted; see 'erdrey.m' in section 9.5 The license has been copied directly, without modification,

from <http://www.gnu.org/licenses/gpl.txt>.

GNU GENERAL PUBLIC LICENSE
Version 3, 29 June 2007

Copyright (C) 2007 Free Software Foundation, Inc. <<http://fsf.org/>>
Everyone is permitted to copy and distribute verbatim copies
of this license document, but changing it is not allowed.

Preamble

The GNU General Public License is a free, copyleft license for software and other kinds of works.

The licenses for most software and other practical works are designed to take away your freedom to share and change the works. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change all versions of a program--to make sure it remains free software for all its users. We, the Free Software Foundation, use the GNU General Public License for most of our software; it applies also to any other work released this way by its authors. You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for them if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs, and that you know you can do these things.

To protect your rights, we need to prevent others from denying you these rights or asking you to surrender the rights. Therefore, you have certain responsibilities if you distribute copies of the software, or if you modify it: responsibilities to respect the freedom of others.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must pass on to the recipients the same freedoms that you received. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

Developers that use the GNU GPL protect your rights with two steps: (1) assert copyright on the software, and (2) offer you this License giving you legal permission to copy, distribute and/or modify it.

For the developers' and authors' protection, the GPL clearly explains that there is no warranty for this free software. For both users' and

authors' sake, the GPL requires that modified versions be marked as changed, so that their problems will not be attributed erroneously to authors of previous versions.

Some devices are designed to deny users access to install or run modified versions of the software inside them, although the manufacturer can do so. This is fundamentally incompatible with the aim of protecting users' freedom to change the software. The systematic pattern of such abuse occurs in the area of products for individuals to use, which is precisely where it is most unacceptable. Therefore, we have designed this version of the GPL to prohibit the practice for those products. If such problems arise substantially in other domains, we stand ready to extend this provision to those domains in future versions of the GPL, as needed to protect the freedom of users.

Finally, every program is threatened constantly by software patents. States should not allow patents to restrict development and use of software on general-purpose computers, but in those that do, we wish to avoid the special danger that patents applied to a free program could make it effectively proprietary. To prevent this, the GPL assures that patents cannot be used to render the program non-free.

The precise terms and conditions for copying, distribution and modification follow.

TERMS AND CONDITIONS

0. Definitions.

"This License" refers to version 3 of the GNU General Public License.

"Copyright" also means copyright-like laws that apply to other kinds of works, such as semiconductor masks.

"The Program" refers to any copyrightable work licensed under this License. Each licensee is addressed as "you". "Licensees" and "recipients" may be individuals or organizations.

To "modify" a work means to copy from or adapt all or part of the work in a fashion requiring copyright permission, other than the making of an exact copy. The resulting work is called a "modified version" of the earlier work or a work "based on" the earlier work.

A "covered work" means either the unmodified Program or a work based on the Program.

To "propagate" a work means to do anything with it that, without

permission, would make you directly or secondarily liable for infringement under applicable copyright law, except executing it on a computer or modifying a private copy. Propagation includes copying, distribution (with or without modification), making available to the public, and in some countries other activities as well.

To "convey" a work means any kind of propagation that enables other parties to make or receive copies. Mere interaction with a user through a computer network, with no transfer of a copy, is not conveying.

An interactive user interface displays "Appropriate Legal Notices" to the extent that it includes a convenient and prominently visible feature that (1) displays an appropriate copyright notice, and (2) tells the user that there is no warranty for the work (except to the extent that warranties are provided), that licensees may convey the work under this License, and how to view a copy of this License. If the interface presents a list of user commands or options, such as a menu, a prominent item in the list meets this criterion.

1. Source Code.

The "source code" for a work means the preferred form of the work for making modifications to it. "Object code" means any non-source form of a work.

A "Standard Interface" means an interface that either is an official standard defined by a recognized standards body, or, in the case of interfaces specified for a particular programming language, one that is widely used among developers working in that language.

The "System Libraries" of an executable work include anything, other than the work as a whole, that (a) is included in the normal form of packaging a Major Component, but which is not part of that Major Component, and (b) serves only to enable use of the work with that Major Component, or to implement a Standard Interface for which an implementation is available to the public in source code form. A "Major Component", in this context, means a major essential component (kernel, window system, and so on) of the specific operating system (if any) on which the executable work runs, or a compiler used to produce the work, or an object code interpreter used to run it.

The "Corresponding Source" for a work in object code form means all the source code needed to generate, install, and (for an executable work) run the object code and to modify the work, including scripts to control those activities. However, it does not include the work's System Libraries, or general-purpose tools or generally available free programs which are used unmodified in performing those activities but

which are not part of the work. For example, Corresponding Source includes interface definition files associated with source files for the work, and the source code for shared libraries and dynamically linked subprograms that the work is specifically designed to require, such as by intimate data communication or control flow between those subprograms and other parts of the work.

The Corresponding Source need not include anything that users can regenerate automatically from other parts of the Corresponding Source.

The Corresponding Source for a work in source code form is that same work.

2. Basic Permissions.

All rights granted under this License are granted for the term of copyright on the Program, and are irrevocable provided the stated conditions are met. This License explicitly affirms your unlimited permission to run the unmodified Program. The output from running a covered work is covered by this License only if the output, given its content, constitutes a covered work. This License acknowledges your rights of fair use or other equivalent, as provided by copyright law.

You may make, run and propagate covered works that you do not convey, without conditions so long as your license otherwise remains in force. You may convey covered works to others for the sole purpose of having them make modifications exclusively for you, or provide you with facilities for running those works, provided that you comply with the terms of this License in conveying all material for which you do not control copyright. Those thus making or running the covered works for you must do so exclusively on your behalf, under your direction and control, on terms that prohibit them from making any copies of your copyrighted material outside their relationship with you.

Conveying under any other circumstances is permitted solely under the conditions stated below. Sublicensing is not allowed; section 10 makes it unnecessary.

3. Protecting Users' Legal Rights From Anti-Circumvention Law.

No covered work shall be deemed part of an effective technological measure under any applicable law fulfilling obligations under article 11 of the WIPO copyright treaty adopted on 20 December 1996, or similar laws prohibiting or restricting circumvention of such measures.

When you convey a covered work, you waive any legal power to forbid circumvention of technological measures to the extent such circumvention is effected by exercising rights under this License with respect to the covered work, and you disclaim any intention to limit operation or modification of the work as a means of enforcing, against the work's users, your or third parties' legal rights to forbid circumvention of technological measures.

4. Conveying Verbatim Copies.

You may convey verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice; keep intact all notices stating that this License and any non-permissive terms added in accord with section 7 apply to the code; keep intact all notices of the absence of any warranty; and give all recipients a copy of this License along with the Program.

You may charge any price or no price for each copy that you convey, and you may offer support or warranty protection for a fee.

5. Conveying Modified Source Versions.

You may convey a work based on the Program, or the modifications to produce it from the Program, in the form of source code under the terms of section 4, provided that you also meet all of these conditions:

- a) The work must carry prominent notices stating that you modified it, and giving a relevant date.
- b) The work must carry prominent notices stating that it is released under this License and any conditions added under section 7. This requirement modifies the requirement in section 4 to "keep intact all notices".
- c) You must license the entire work, as a whole, under this License to anyone who comes into possession of a copy. This License will therefore apply, along with any applicable section 7 additional terms, to the whole of the work, and all its parts, regardless of how they are packaged. This License gives no permission to license the work in any other way, but it does not invalidate such permission if you have separately received it.
- d) If the work has interactive user interfaces, each must display Appropriate Legal Notices; however, if the Program has interactive interfaces that do not display Appropriate Legal Notices, your work need not make them do so.

A compilation of a covered work with other separate and independent works, which are not by their nature extensions of the covered work, and which are not combined with it such as to form a larger program, in or on a volume of a storage or distribution medium, is called an "aggregate" if the compilation and its resulting copyright are not used to limit the access or legal rights of the compilation's users beyond what the individual works permit. Inclusion of a covered work in an aggregate does not cause this License to apply to the other parts of the aggregate.

6. Conveying Non-Source Forms.

You may convey a covered work in object code form under the terms of sections 4 and 5, provided that you also convey the machine-readable Corresponding Source under the terms of this License, in one of these ways:

- a) Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by the Corresponding Source fixed on a durable physical medium customarily used for software interchange.
- b) Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by a written offer, valid for at least three years and valid for as long as you offer spare parts or customer support for that product model, to give anyone who possesses the object code either (1) a copy of the Corresponding Source for all the software in the product that is covered by this License, on a durable physical medium customarily used for software interchange, for a price no more than your reasonable cost of physically performing this conveying of source, or (2) access to copy the Corresponding Source from a network server at no charge.
- c) Convey individual copies of the object code with a copy of the written offer to provide the Corresponding Source. This alternative is allowed only occasionally and noncommercially, and only if you received the object code with such an offer, in accord with subsection 6b.
- d) Convey the object code by offering access from a designated place (gratis or for a charge), and offer equivalent access to the Corresponding Source in the same way through the same place at no further charge. You need not require recipients to copy the Corresponding Source along with the object code. If the place to copy the object code is a network server, the Corresponding Source

may be on a different server (operated by you or a third party) that supports equivalent copying facilities, provided you maintain clear directions next to the object code saying where to find the Corresponding Source. Regardless of what server hosts the Corresponding Source, you remain obligated to ensure that it is available for as long as needed to satisfy these requirements.

e) Convey the object code using peer-to-peer transmission, provided you inform other peers where the object code and Corresponding Source of the work are being offered to the general public at no charge under subsection 6d.

A separable portion of the object code, whose source code is excluded from the Corresponding Source as a System Library, need not be included in conveying the object code work.

A "User Product" is either (1) a "consumer product", which means any tangible personal property which is normally used for personal, family, or household purposes, or (2) anything designed or sold for incorporation into a dwelling. In determining whether a product is a consumer product, doubtful cases shall be resolved in favor of coverage. For a particular product received by a particular user, "normally used" refers to a typical or common use of that class of product, regardless of the status of the particular user or of the way in which the particular user actually uses, or expects or is expected to use, the product. A product is a consumer product regardless of whether the product has substantial commercial, industrial or non-consumer uses, unless such uses represent the only significant mode of use of the product.

"Installation Information" for a User Product means any methods, procedures, authorization keys, or other information required to install and execute modified versions of a covered work in that User Product from a modified version of its Corresponding Source. The information must suffice to ensure that the continued functioning of the modified object code is in no case prevented or interfered with solely because modification has been made.

If you convey an object code work under this section in, or with, or specifically for use in, a User Product, and the conveying occurs as part of a transaction in which the right of possession and use of the User Product is transferred to the recipient in perpetuity or for a fixed term (regardless of how the transaction is characterized), the Corresponding Source conveyed under this section must be accompanied by the Installation Information. But this requirement does not apply if neither you nor any third party retains the ability to install modified object code on the User Product (for example, the work has been installed in ROM).

The requirement to provide Installation Information does not include a requirement to continue to provide support service, warranty, or updates for a work that has been modified or installed by the recipient, or for the User Product in which it has been modified or installed. Access to a network may be denied when the modification itself materially and adversely affects the operation of the network or violates the rules and protocols for communication across the network.

Corresponding Source conveyed, and Installation Information provided, in accord with this section must be in a format that is publicly documented (and with an implementation available to the public in source code form), and must require no special password or key for unpacking, reading or copying.

7. Additional Terms.

"Additional permissions" are terms that supplement the terms of this License by making exceptions from one or more of its conditions. Additional permissions that are applicable to the entire Program shall be treated as though they were included in this License, to the extent that they are valid under applicable law. If additional permissions apply only to part of the Program, that part may be used separately under those permissions, but the entire Program remains governed by this License without regard to the additional permissions.

When you convey a copy of a covered work, you may at your option remove any additional permissions from that copy, or from any part of it. (Additional permissions may be written to require their own removal in certain cases when you modify the work.) You may place additional permissions on material, added by you to a covered work, for which you have or can give appropriate copyright permission.

Notwithstanding any other provision of this License, for material you add to a covered work, you may (if authorized by the copyright holders of that material) supplement the terms of this License with terms:

- a) Disclaiming warranty or limiting liability differently from the terms of sections 15 and 16 of this License; or
- b) Requiring preservation of specified reasonable legal notices or author attributions in that material or in the Appropriate Legal Notices displayed by works containing it; or
- c) Prohibiting misrepresentation of the origin of that material, or requiring that modified versions of such material be marked in reasonable ways as different from the original version; or

- d) Limiting the use for publicity purposes of names of licensors or authors of the material; or
- e) Declining to grant rights under trademark law for use of some trade names, trademarks, or service marks; or
- f) Requiring indemnification of licensors and authors of that material by anyone who conveys the material (or modified versions of it) with contractual assumptions of liability to the recipient, for any liability that these contractual assumptions directly impose on those licensors and authors.

All other non-permissive additional terms are considered "further restrictions" within the meaning of section 10. If the Program as you received it, or any part of it, contains a notice stating that it is governed by this License along with a term that is a further restriction, you may remove that term. If a license document contains a further restriction but permits relicensing or conveying under this License, you may add to a covered work material governed by the terms of that license document, provided that the further restriction does not survive such relicensing or conveying.

If you add terms to a covered work in accord with this section, you must place, in the relevant source files, a statement of the additional terms that apply to those files, or a notice indicating where to find the applicable terms.

Additional terms, permissive or non-permissive, may be stated in the form of a separately written license, or stated as exceptions; the above requirements apply either way.

8. Termination.

You may not propagate or modify a covered work except as expressly provided under this License. Any attempt otherwise to propagate or modify it is void, and will automatically terminate your rights under this License (including any patent licenses granted under the third paragraph of section 11).

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, you do not qualify to receive new licenses for the same material under section 10.

9. Acceptance Not Required for Having Copies.

You are not required to accept this License in order to receive or run a copy of the Program. Ancillary propagation of a covered work occurring solely as a consequence of using peer-to-peer transmission to receive a copy likewise does not require acceptance. However, nothing other than this License grants you permission to propagate or modify any covered work. These actions infringe copyright if you do not accept this License. Therefore, by modifying or propagating a covered work, you indicate your acceptance of this License to do so.

10. Automatic Licensing of Downstream Recipients.

Each time you convey a covered work, the recipient automatically receives a license from the original licensors, to run, modify and propagate that work, subject to this License. You are not responsible for enforcing compliance by third parties with this License.

An "entity transaction" is a transaction transferring control of an organization, or substantially all assets of one, or subdividing an organization, or merging organizations. If propagation of a covered work results from an entity transaction, each party to that transaction who receives a copy of the work also receives whatever licenses to the work the party's predecessor in interest had or could give under the previous paragraph, plus a right to possession of the Corresponding Source of the work from the predecessor in interest, if the predecessor has it or can get it with reasonable efforts.

You may not impose any further restrictions on the exercise of the rights granted or affirmed under this License. For example, you may not impose a license fee, royalty, or other charge for exercise of rights granted under this License, and you may not initiate litigation (including a cross-claim or counterclaim in a lawsuit) alleging that any patent claim is infringed by making, using, selling, offering for

sale, or importing the Program or any portion of it.

11. Patents.

A "contributor" is a copyright holder who authorizes use under this License of the Program or a work on which the Program is based. The work thus licensed is called the contributor's "contributor version".

A contributor's "essential patent claims" are all patent claims owned or controlled by the contributor, whether already acquired or hereafter acquired, that would be infringed by some manner, permitted by this License, of making, using, or selling its contributor version, but do not include claims that would be infringed only as a consequence of further modification of the contributor version. For purposes of this definition, "control" includes the right to grant patent sublicenses in a manner consistent with the requirements of this License.

Each contributor grants you a non-exclusive, worldwide, royalty-free patent license under the contributor's essential patent claims, to make, use, sell, offer for sale, import and otherwise run, modify and propagate the contents of its contributor version.

In the following three paragraphs, a "patent license" is any express agreement or commitment, however denominated, not to enforce a patent (such as an express permission to practice a patent or covenant not to sue for patent infringement). To "grant" such a patent license to a party means to make such an agreement or commitment not to enforce a patent against the party.

If you convey a covered work, knowingly relying on a patent license, and the Corresponding Source of the work is not available for anyone to copy, free of charge and under the terms of this License, through a publicly available network server or other readily accessible means, then you must either (1) cause the Corresponding Source to be so available, or (2) arrange to deprive yourself of the benefit of the patent license for this particular work, or (3) arrange, in a manner consistent with the requirements of this License, to extend the patent license to downstream recipients. "Knowingly relying" means you have actual knowledge that, but for the patent license, your conveying the covered work in a country, or your recipient's use of the covered work in a country, would infringe one or more identifiable patents in that country that you have reason to believe are valid.

If, pursuant to or in connection with a single transaction or arrangement, you convey, or propagate by procuring conveyance of, a covered work, and grant a patent license to some of the parties

receiving the covered work authorizing them to use, propagate, modify or convey a specific copy of the covered work, then the patent license you grant is automatically extended to all recipients of the covered work and works based on it.

A patent license is "discriminatory" if it does not include within the scope of its coverage, prohibits the exercise of, or is conditioned on the non-exercise of one or more of the rights that are specifically granted under this License. You may not convey a covered work if you are a party to an arrangement with a third party that is in the business of distributing software, under which you make payment to the third party based on the extent of your activity of conveying the work, and under which the third party grants, to any of the parties who would receive the covered work from you, a discriminatory patent license (a) in connection with copies of the covered work conveyed by you (or copies made from those copies), or (b) primarily for and in connection with specific products or compilations that contain the covered work, unless you entered into that arrangement, or that patent license was granted, prior to 28 March 2007.

Nothing in this License shall be construed as excluding or limiting any implied license or other defenses to infringement that may otherwise be available to you under applicable patent law.

12. No Surrender of Others' Freedom.

If conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot convey a covered work so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not convey it at all. For example, if you agree to terms that obligate you to collect a royalty for further conveying from those to whom you convey the Program, the only way you could satisfy both those terms and this License would be to refrain entirely from conveying the Program.

13. Use with the GNU Affero General Public License.

Notwithstanding any other provision of this License, you have permission to link or combine any covered work with a work licensed under version 3 of the GNU Affero General Public License into a single combined work, and to convey the resulting work. The terms of this License will continue to apply to the part which is the covered work, but the special requirements of the GNU Affero General Public License, section 13, concerning interaction through a network will apply to the combination as such.

14. Revised Versions of this License.

The Free Software Foundation may publish revised and/or new versions of the GNU General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies that a certain numbered version of the GNU General Public License "or any later version" applies to it, you have the option of following the terms and conditions either of that numbered version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of the GNU General Public License, you may choose any version ever published by the Free Software Foundation.

If the Program specifies that a proxy can decide which future versions of the GNU General Public License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Program.

Later license versions may give you additional or different permissions. However, no additional obligations are imposed on any author or copyright holder as a result of your choosing to follow a later version.

15. Disclaimer of Warranty.

THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

16. Limitation of Liability.

IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN

WRITING
WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MODIFIES AND/OR
CONVEYS
THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES,
INCLUDING ANY
GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT
OF THE
USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO
LOSS OF
DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU
OR THIRD
PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER
PROGRAMS),
EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE
POSSIBILITY OF
SUCH DAMAGES.

17. Interpretation of Sections 15 and 16.

If the disclaimer of warranty and limitation of liability provided
above cannot be given local legal effect according to their terms,
reviewing courts shall apply local law that most closely approximates
an absolute waiver of all civil liability in connection with the
Program, unless a warranty or assumption of liability accompanies a
copy of the Program in return for a fee.

END OF TERMS AND CONDITIONS

9.3. The QMRA model simulating the Lifestraw field trial (chapter 3)

The model consists of several text files containing necessary functions and subroutines; the core program is 'QMRAv13_20110414.m'. Simulation options are set by the choice of several values at the top of the files 'QMRAv13_20110414.m' and 'GetTrialParams.m'. These options default to values that generate a single test run of the simulation.

The source code is found below. The filename of each of the source code files is found in the copyright information near the top of each file. Each file starts on a new page.

%QMRA for Lifestraw Family RCT in DRC (Boisson 2010), coded by Kyle S. Enger (engerkyl@msu.edu) with suggestions from Joe Eisenberg & Bryan Mayer.
%It was used to produce the manuscript, Linking quantitative microbial risk assessment and epidemiological data: Informing safe drinking water trials in developing countries, by Kyle S. Enger, Kara L. Nelson, Thomas Clasen, Joan B. Rose, and Joseph N. S. Eisenberg, published in Environmental Science and Technology in 2012.

%{
COPYRIGHT INFORMATION

Copyright 2011 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (QMRAv13_20110414.m) is part of QMRAv13_20110414.

QMRAv13_20110414 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRAv13_20110414 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRAv13_20110414. If not, see <<http://www.gnu.org/licenses/>>.

%}

%Requires Octave 3.2 or later and the octave-image package.

%Also works rather well on Matlab, running about 4x faster. Requires the statistics toolbox, and possibly others.

%To run, start Octave, change the directory to the location of the program files (using the 'cd' command), and type 'QMRAv13_20110414'.

```

%This code (QMRAv13_20110414.m) is accompanied by several functions/subroutines, which
need to be in the working directory:
%   AssignInf.m:           Stochastically assigns infections to individuals with fixed
durations
%   AssignInfRand.m:      Stochastically assigns infections to individuals with random
durations
%   AssignInfIllRand.m:  Stochastically assigns infections & illnesses to individuals
with random durations
%   CalcDiarrhWeeks.m:   Determines whether a week with 1+ days of diarrhea is actually
reported as a 'diarrhea week'
%   DRbP.m:              Beta-Poisson dose response model
%   DRchoose.m:         Executes the appropriate dose response model and determines
illness
%   DRexp.m:            Exponential dose response model
%   durEc.m:           Randomly pick a duration for E. coli infection
%   durGi.m:           Randomly pick a duration for Giardia infection
%   durRo.m:           Randomly pick a duration for rotavirus infection
%   Examine1Run.m:      Allows inspection of the complete simulated data from a single
run of this code.
%   GetTrialParams.m:   Generates a series of trial parameters instead of determining
them stochastically.
%   OutQMRAmerge.m:     Allows conglomeration of output from multiple model
executions (e.g., if parallel processing).

%Kids are defined as those < 5 yrs of age. All others are considered adults.
%====Initial housekeeping====
clear all;
Octave = size(ver('Octave'),1); %Indicator of whether the code is running under Octave
(1) or Matlab (0).
if Octave == 0; format compact; format shortg; end;
StartTime = clock;           %Recording start time of simulation. Is used to timestamp output
file.
%ignore_function_time_stamp('all'); %Hopefully saves time by preventing Octave from

```

checking if functions have changed during the run. However, makes debugging difficult (altered functions aren't updated).

```
%====Setting simulation options====More simulation options are in GetTrialParams.m====
```

```
badPlacebo = 1; %Whether the simulation is run with an imperfect placebo (1) or a perfect placebo (0).
```

```
loops = 1; %No. Monte Carlo iterations of QMRA, if parameters are chosen stochastically. If 1, & noStochParams != 1, stores all run info.
```

```
noStochParams = 1; %If 1, use GetTrialParams.m to iterate over a series of parameter choices (not choose them stochastically). Resets 'loops'.
```

```
testing = 1; %Equals 1 if testing (production of plots of infection prevalence) is desired. Only works if there are <= 25 runs.
```

```
compliance100 = 0; %Equals 1 if device is used perfectly, always, by everyone.
```

```
dailyVariation = 1; %Equals 1 if pathogen concentrations are allowed to vary by person by day, instead of taking a single fixed value.
```

```
randomDurations = 1; %Equals 1 if illness durations are randomized instead of taking a single fixed value. Renders Duration vector mostly moot (still used to choose start time of surveys).
```

```
storeStatus = 1; %If 1, store infection & disease status, to determine 'actual' (i.e., reported & unreported) burden of infection.
```

```
%====Housekeeping based on simulation options====
```

```
%cd ~/Dropbox/Octave/Lifestraw; %Selecting working directory. Assumes Linux.
```

```
%cd QMRAv13; %TODO: Update whenever a new version of this code is produced.
```

```
%if randomDurations == 1; IllDurations; end %Reading in necessary functions to randomly choose illness durations.
```

```
if noStochParams == 1; GetTrialParams; end %GetTrialParams.m reads in pathogens/L & background measure to be tried.
```

```
if testing == 1 && loops > 5 && noStochParams == 0;
```

```
    error('Do not use so many loops with testing code enabled, or you will flood the screen with graphs');
```

```
end
```

```
%====Ending housekeeping. Starting parameter values:====
```

```
%Water concentration and disease parameter values
```

```
% These 4 vectors have 3 elements corresponding to our 3 pathogens of interest:
```

```

%      [ETEC/EPEC, Giardia, rotavirus]
PDiseaseK = [0.107, 0.064 , 0.098]; PDiseaseK = PDiseaseK / sum(PDiseaseK);
%Proportion of kid diarrhea episodes due to various pathogens, adjusted to sum to 1
(assume unknown episodes are distributed just like known episodes)
PathogensLMax = [2e5, 1.35, 0.18]; %Maximum pathogens/L; minimum is zero for
all. 2fold empirically observed levels that led to LP exceeding the 95% CI for the
placebo group, each pathogen taken individually.
pathogensLcv = sqrt(3407044) / 2509.329; %Coeff. of var. calc. from variance & mean
of 'cfbef' (Boisson 2010 water qual. data), high outliers (>= 30000 CFU) removed.
MorbidityK = [0.214 , 0.59 , 0.397]; %Proportion of infected 'kids' (<5y) with
diarrhea.
Duration = [82.1/24, 18.3, 2.5]; Duration = round(Duration); %Duration of diarrhea
(days)
prevDiarrhBaseKidsMax = 0.0972;
LongPrevs = [0.103, 0.0896]; %Vector of raw long. prev. values from Boisson dataset,
kids w. placebo, then kids w. intervention).
longPrevRangeMult = 0.4; %Deprecated; multiplier (between 0 & 1) to determine acceptable
range for longitudinal prevalence & its ratio.
longPrevKidsDesiredRatio = 0.872; %Ratio from study of long. prev. in intervention kids
to placebo kids.
prevDiarrhBaseKids = LongPrevs(2); %Baseline non-waterborne reported prevalence of
disease. Can be no greater than that observed in the RCT.
ImmuneTimes = [7 7 7]; %Length of immune period for all pathogens.
%Parameters: population & exposure information. Model by household later.
nKidsInt = 85; nKidsPla = 105; %Boisson 2010.
drinkKids = 1.178; %drinkKidsSD = 0.186; %daily water intake, L/d, kids
pUse = mean([0.685, 0.757, 0.483, 0.670]); %Mean of people using the device the
previous day, int. & placebo, at 8 & 14 months.
pTreat = 2/3; %Proportion of water treated, if device is being used.
Try 1, 2/3, & 1/3.
if compliancel00 == 1; pUse = 1; pTreat = 1; end; %If perfect compliance is desired,
override above 2 lines.
%Parameters: device effectiveness information

```

```

LRsInt = [6.9, 4.7, 3.6];           %Log reductions [bacteria, viruses, protozoa] by the
intervention device, with upper & lower ranges
LRsPla = [1.05, 1.05, 1.05];       %As above, for 'placebo' device
if badPlacebo == 0; LRsPla = [0 0 0]; end; %If a perfect placebo is being modeled,
override above line.
%Parameters: dose response, order as above [ETEC/EPEC, Giardia, rotavirus]
KorN50 = [2111912, 0.01982, 6.171]; %Exponential k parameter or beta-Poisson N50
parameter
alpha = [0.1549, NaN, 0.2531];     %Presence/absence of alpha value determines
beta-Poisson or exponential dose resp.
%Bias parameters
remembrance = 0.54; %Proportion of diarrhea episodes remembered (and reported) if they
ended >2d before being surveyed; assume perfect recall if episode is on day 0, 1, or 2
%Study parameters - relating to how the study was conducted
recallPeriod = 7; %Number of days in the past over which people were asked to remember
diarrheal episodes
interval = 30; %Interval between beginnings of recall periods
nRecallPeriods = 12; %Number of recall periods (i.e., number of simulated diarrhea
surveys)
daysBurnIn = ceil(max(ImmuneTimes) + max(Duration) + recallPeriod) * 4; %Days required
for prevalence to reach equilibrium (simulation starts with nobody infected). Allows
ample margin for reaching equilibrium.
%====Ending parameter values====

maxTime = daysBurnIn + (nRecallPeriods-1)*interval; %Time over which to run each
simulation.
%Creating output structure for storing results from main QMRA loop
OutQMRA =
struct('StartTime',StartTime,'Fit',NaN(1,loops),'KILP',NaN(1,loops),'KPLP',NaN(1,loops),'
LPR',NaN(1,loops),...
'EcL',NaN(1,loops),'GiL',NaN(1,loops),'RoL',NaN(1,loops),'PrevBase',NaN(1,loops));
%'Fit' is no longer used.
if storeStatus == 1; %Optionally creating cell array for more detailed output. Not

```


needed for calibration step.

```
DailyStatus = cell(2, loops); %Each cell in the array needs to contain a matrix  
(rows are days, columns are variables).
```

```
DailyStatus(:) = {NaN(maxTime,22)}; %1st row for intervention, 2nd row for  
placebo.
```

```
%Columns are: 1:3 = new inf. Ec,Gi,&Ro; 4:6 = new cases; 7:14 = num. inf.  
(0, Ec, Gi, Ro, EcGi, EcRo, GiRo, EcGiRo); 15:22 = as inf., but ill.
```

```
end;
```

```
tic %Starts timer
```

```
if noStochParams == 1; loops = size(TrialParams); loops = loops(1); end %Resets loops  
if a series of trial pathogens/L values is being used.
```

```
for i = 1:loops; %====Starting main QMRA loop.==== Loops once for each QMRA run. i  
indexes each loop.
```

```
%====Randomly generating parameters for this iteration====
```

```
switch(noStochParams);
```

```
case 0;
```

```
PathogensLmeans = rand(1,length(MorbidityK)) .* PathogensLMax; %Uniform  
sampling of the mean value for each pathogen.
```

```
prevDiarrhBaseKids = rand(1,1)*prevDiarrhBaseKidsMax;%For Matlab.
```

```
case 1; %Pulling parameters from previous runs consistent with RCT.
```

```
PathogensLmeans = TrialParams(i,1:3);
```

```
prevDiarrhBaseKids = TrialParams(i,4);
```

```
otherwise
```

```
error('noStochParams must be 0 or 1');
```

```
end
```

```
OutQMRA.EcL(i)=PathogensLmeans(1); OutQMRA.GiL(i)=PathogensLmeans(2);
```

```
OutQMRA.RoL(i)=PathogensLmeans(3);
```

```
OutQMRA.PrevBase(i) = prevDiarrhBaseKids; %This line & previous store the varying  
parameter values.
```

```
%====End random parameter generation - start setup of values/vectors/matrices used  
throughout simulation====
```

```
%Computing daily doses of pathogens ingested in drinking water, using water drunk  
per day and log reduction values
```

```

    %Computing parameter values for gamma distribution of pathogens in water
    Scales = (pathogensLcv * PathogensLmeans).^2 ./ PathogensLmeans;
    Shapes = PathogensLmeans ./ Scales;
    %Assigning infections randomly based on responses, assuming infections with
different pathogens are independent.
    % A person can have only 1 infection per pathogen.
    OutPrevs = struct('KidsInt',NaN(1,nRecallPeriods),'KidsPla',NaN(1,nRecallPeriods));
    %Creating struct to hold prevalences from surveys.
    %Creating person-pathogen matrices; everybody starts infected for a random amount of
time, to reduce periodicity from constant disease duration.
    KidsInt = rand(nKidsInt,length(Duration)) * 2*(max(Duration)+max(ImmuneTimes));
    KidsPla = rand(nKidsPla,length(Duration)) * 2*(max(Duration)+max(ImmuneTimes));
    if storeStatus == 1; %Optionally, making corresponding matrices to store disease
info.
        KidsIntD = ones(size(KidsInt)); %NaN means never infected, 0 means
uninfected, 1 means infected, 2 means diseased.
        KidsPlaD = ones(size(KidsPla)); %Note that everyone starts off infected with
everything, just as with KidsInt & KidsPla above.
    end
    OutputFields = fieldnames(OutPrevs);
    if Octave == 1; fflush(stdout); end; %Forces a write to screen.
    %=====Code for testing purposes only=====
    if testing == 1; %Only runs when testing code. These vectors needed for charting
infection prevalence over the entire simulation.
        KidsIntInfPrev = NaN(nKidsInt,1); %Creating a vector to hold infection
prevalence info, intervention group.
        KidsPlaInfPrev = NaN(nKidsPla,1); %As above, placebo group.
    end;
    %=====End code for testing purposes=====
    %Now storing all person-pathogen matrices, but only if exactly 1 loop is requested.
This repeats at the end of each day.
    if loops == 1;
        PPmatrices(1).KidsInt = KidsInt; PPmatrices(1).KidsPla = KidsPla;

```

```

        PPmatrices(1).KidsIntD = KidsIntD; PPmatrices(1).KidsPlaD = KidsPlaD;
    end;
    %=====Begin daily loop, t indexes the days=====
    for t = 1:maxTime;
        KidsInt = KidsInt - 1;    %Note: a value of 0 signifies the first day of
immunity.
        KidsPla = KidsPla - 1;
        if storeStatus == 1;    %Optionally, tracking recovery from infection/disease.
            KidsIntD(find(KidsInt <= 0)) = 0;
            KidsPlaD(find(KidsPla <= 0)) = 0;
        end
        %Computing doses in untreated water, varying for each child, each day.
        Dose.KidsInt = NaN(nKidsInt,length(Duration));    %Matrix of doses per child
(intervention). Columns are pathogens.
        Dose.KidsPla = NaN(nKidsPla,length(Duration));    %Matrix of doses per child
(placebo). Columns are pathogens.
        RandComp = rand(max(nKidsInt,nKidsPla),2);    %Random numbers for determining
compliance (i.e., use of device).
        for j = 1:length(Duration);    %Looping over pathogens to determine
daily doses for ea. person.
            if isnan(Shapes(j)) == 1;    %If mean pathogen conc. is set to 0,
pathogen conc. is always 0.
                Dose.KidsInt(:,j) = 0;
                Dose.KidsPla(:,j) = 0;
            else
                if dailyVariation == 1;
                    Dose.KidsInt(:,j) = gamrnd(Shapes(j),Scales(j),[nKidsInt,1]) *
drinkKids;    %Initial untreated dose
                    Dose.KidsPla(:,j) = gamrnd(Shapes(j),Scales(j),[nKidsPla,1]) *
drinkKids;
                else
                    Dose.KidsInt(:,j) = PathogensLmeans(j) * drinkKids;    %Dose
becomes the mean dose if daily variation is turned off.

```

```

        Dose.KidsPla(:,j) = PathogensLmeans(j) * drinkKids;
    end
    end %Next 2 lines: Determining who complies and has a nonzero dose
    (because log reduction would fail on zero dose).
    Comp.KidsInt = intersect(find(RandComp(1:nKidsInt,1) < pUse),
    find(Dose.KidsInt(:,j) > 0));
    Comp.KidsPla = intersect(find(RandComp(2:nKidsPla,1) < pUse),
    find(Dose.KidsPla(:,j) > 0));
    %Now applying LRs, if using device. Includes adjustment for partial
    treatment of water (pTreat).
    Dose.KidsInt(Comp.KidsInt,j) = 10.^(log10(Dose.KidsInt(Comp.KidsInt,j) *
    pTreat) - LRsInt(j)) + Dose.KidsInt(Comp.KidsInt,j) * (1 - pTreat);
    Dose.KidsPla(Comp.KidsPla,j) = 10.^(log10(Dose.KidsPla(Comp.KidsPla,j) *
    pTreat) - LRsPla(j)) + Dose.KidsPla(Comp.KidsPla,j) * (1 - pTreat);
    end
    %Computing responses (diarrheal illness) using custom functions DRexp() and
    DRbP().
    Responses.KidsInt(:,1) = DRbP(KorN50(1),alpha(1),Dose.KidsInt(:,1)); %Note:
    response matrices correspond to the person-path. matrices.
    Responses.KidsInt(:,2) = DRexp(KorN50(2),Dose.KidsInt(:,2));
    Responses.KidsInt(:,3) = DRbP(KorN50(3),alpha(3),Dose.KidsInt(:,3));
    Responses.KidsPla(:,1) = DRbP(KorN50(1),alpha(1),Dose.KidsPla(:,1));
    Responses.KidsPla(:,2) = DRexp(KorN50(2),Dose.KidsPla(:,2));
    Responses.KidsPla(:,3) = DRbP(KorN50(3),alpha(3),Dose.KidsPla(:,3));
    switch(randomDurations);
    case 1
        switch(storeStatus);
        case 1
            %DailyStatus columns: 1:3 = new inf. Ec,Gi,&Ro; 4:6 = new cases; 7:14 = num. inf.
            (0,Ec,Gi,Ro,EcGi,EcRo,GiRo,EcGiRo); 15:22 = as inf., but ill.
            KidsIntDold = KidsIntD;
            KidsPlaDold = KidsPlaD;
            [KidsInt,KidsIntD] =

```

```

AssignInfIllRand(KidsInt,KidsIntD,Responses.KidsInt,ImmuneTimes,MorbidityK);
    [KidsPla,KidsPlaD] =
AssignInfIllRand(KidsPla,KidsPlaD,Responses.KidsPla,ImmuneTimes,MorbidityK);
    for s = 1:6; %Store counts of new infections & illnesses.
        if s <= 3; %Infections:
            DailyStatus{1,i}(t,s) =
length(intersect(find(KidsIntDold(:,s) == 0), find(KidsIntD(:,s) > 0)));
            DailyStatus{2,i}(t,s) =
length(intersect(find(KidsPlaDold(:,s) == 0), find(KidsPlaD(:,s) > 0)));
        else %Illnesses:
            DailyStatus{1,i}(t,s) =
length(intersect(find(KidsIntDold(:,s-3) == 0), find(KidsIntD(:,s-3) == 2)));
            DailyStatus{2,i}(t,s) =
length(intersect(find(KidsPlaDold(:,s-3) == 0), find(KidsPlaD(:,s-3) == 2)));
        end
    end
    otherwise
        KidsInt =
AssignInfRand(KidsInt,Responses.KidsInt,ImmuneTimes);
        KidsPla =
AssignInfRand(KidsPla,Responses.KidsPla,ImmuneTimes);
    end
    otherwise
        KidsInt = AssignInf(KidsInt,Responses.KidsInt,Duration,ImmuneTimes);
        KidsPla = AssignInf(KidsPla,Responses.KidsPla,Duration,ImmuneTimes);
    end

    if t >= daysBurnIn && mod(t-(daysBurnIn),interval) == 0; %Obtaining results
from diarrhea assessment survey.
%Determining reported diarrhea-weeks. CalcDiarrhWeeks() uses the Person-Pathogen
Matrices, morbidity ratios, and recall of diarrhea episodes to determine if a week was
reported as a week with diarrhea.
        KidsIntRD = CalcDiarrhWeeks(KidsInt,remembrance,recallPeriod,MorbidityK);

```

```

    %Whether diarrhea was reported by each particular person. Note age (last column) is
    removed from the person-pathogen matrix when inputted to the function.
        OutPrevs.KidsInt((t-daysBurnIn)/interval+1) =
sum(KidsIntRD)/length(KidsIntRD); %Getting prevalence for each diarrhea survey
        KidsPlaRD = CalcDiarrhWeeks(KidsPla,remembrance,recallPeriod,MorbidityK);
    %Like above 2 lines, but kid placebo
        OutPrevs.KidsPla((t-daysBurnIn)/interval+1) =
sum(KidsPlaRD)/length(KidsPlaRD);
        fprintf(1,['Day ',num2str(t),'/',num2str(maxTime),' done--']);
        if Octave == 1; fflush(stdout); end; %Forces a write to screen.
    end
    if testing == 1; %=====Testing code=====
        KidsIntInfPrev(t) = sum(max(KidsInt') > 0) / nKidsInt;%If max value over
all pathogens >0, then there is an infection.
        KidsPlaInfPrev(t) = sum(max(KidsPla') > 0) / nKidsPla;%Transposing so that
max is for ea. row instead of ea. column.
        KidsIntInfEc(t) = sum(KidsInt(:,1) > 0) / nKidsInt;
        KidsIntInfGi(t) = sum(KidsInt(:,2) > 0) / nKidsInt;
        KidsIntInfRo(t) = sum(KidsInt(:,3) > 0) / nKidsInt;
        KidsPlaInfEc(t) = sum(KidsPla(:,1) > 0) / nKidsPla;
        KidsPlaInfGi(t) = sum(KidsPla(:,2) > 0) / nKidsPla;
        KidsPlaInfRo(t) = sum(KidsPla(:,3) > 0) / nKidsPla;
    end %=====End testing code=====
    if storeStatus == 1;
        %Columns are: 1:3 = new inf. Ec,Gi,&Ro; 4:6 = new cases; 7:14 = num. inf.
(0, Ec, Gi, Ro, EcGi, EcRo, GiRo, EcGiRo); 15:22 = as inf., but ill.
        for s = 7:22; %All based on 0=uninfected, 1=asymptomatic,
2=ill.
            if s == 7; %Tallying completely uninfected people
                DailyStatus{1,i}(t,s) = length(find(sum(KidsIntD') == 0));
                DailyStatus{2,i}(t,s) = length(find(sum(KidsPlaD') == 0));
            elseif s == 8; %Infected with only Ec
                DailyStatus{1,i}(t,s) = length(intersect(find(KidsIntD(:,1) >

```

```

0), find(sum([KidsIntD(:,2) KidsIntD(:,3)]') == 0)'));
        DailyStatus{2,i}(t,s) = length(intersect(find(KidsPlaD(:,1) >
0), find(sum([KidsPlaD(:,2) KidsPlaD(:,3)]') == 0)')));
        elseif s == 9;           %Infected with only Gi
        DailyStatus{1,i}(t,s) = length(intersect(find(KidsIntD(:,2) >
0), find(sum([KidsIntD(:,1) KidsIntD(:,3)]') == 0)')));
        DailyStatus{2,i}(t,s) = length(intersect(find(KidsPlaD(:,2) >
0), find(sum([KidsPlaD(:,1) KidsPlaD(:,3)]') == 0)')));
        elseif s == 10;          %Infected with only Ro
        DailyStatus{1,i}(t,s) = length(intersect(find(KidsIntD(:,3) >
0), find(sum([KidsIntD(:,1) KidsIntD(:,2)]') == 0)')));
        DailyStatus{2,i}(t,s) = length(intersect(find(KidsPlaD(:,3) >
0), find(sum([KidsPlaD(:,1) KidsPlaD(:,2)]') == 0)')));
        elseif s == 11;          %Tallying infected with Ec & Gi only
        DailyStatus{1,i}(t,s) =
length(intersect(find(prod(KidsIntD(:,1:2)') ~= 0), find(KidsIntD(:,3) == 0)));
        DailyStatus{2,i}(t,s) =
length(intersect(find(prod(KidsPlaD(:,1:2)') ~= 0), find(KidsPlaD(:,3) == 0)));
        elseif s == 12;          %Tallying infected with Ec & Ro only
        DailyStatus{1,i}(t,s) =
length(intersect(find(prod([KidsIntD(:,1) KidsIntD(:,3)]') ~= 0), find(KidsIntD(:,2) ==
0)));
        DailyStatus{2,i}(t,s) =
length(intersect(find(prod([KidsPlaD(:,1) KidsPlaD(:,3)]') ~= 0), find(KidsPlaD(:,2) ==
0)));
        elseif s == 13;          %Tallying infected with Gi & Ro only
        DailyStatus{1,i}(t,s) =
length(intersect(find(prod(KidsIntD(:,2:3)') ~= 0), find(KidsIntD(:,1) == 0)));
        DailyStatus{2,i}(t,s) =
length(intersect(find(prod(KidsPlaD(:,2:3)') ~= 0), find(KidsPlaD(:,1) == 0)));
        elseif s == 14;          %Tallying infected with Ec & Gi & Ro
        DailyStatus{1,i}(t,s) = length(find(prod(KidsIntD(:,1:3)') ~=
0));

```

```

        DailyStatus{2,i}(t,s) = length(find(prod(KidsPlaD(:,1:3)') ~=
0));
    elseif s == 15;        %Tallying completely non-ill people
        DailyStatus{1,i}(t,s) = length(find(sum(KidsIntD' .^2) <= 3));
        DailyStatus{2,i}(t,s) = length(find(sum(KidsPlaD' .^2) <= 3));
    elseif s == 16;        %Ill with only Ec
        DailyStatus{1,i}(t,s) = length(intersect(find(KidsIntD(:,1) ==
2), find(sum(KidsIntD(:,2:3)') .^2) <= 2)'));
        DailyStatus{2,i}(t,s) = length(intersect(find(KidsPlaD(:,1) ==
2), find(sum(KidsPlaD(:,2:3)') .^2) <= 2)'));
    elseif s == 17;        %Ill with only Gi
        DailyStatus{1,i}(t,s) = length(intersect(find(KidsIntD(:,2) ==
2), find(sum([KidsIntD(:,1) KidsIntD(:,3)]' .^2) <= 2)'));
        DailyStatus{2,i}(t,s) = length(intersect(find(KidsPlaD(:,2) ==
2), find(sum([KidsPlaD(:,1) KidsPlaD(:,3)]' .^2) <= 2)'));
    elseif s == 18;        %Ill with only Ro
        DailyStatus{1,i}(t,s) = length(intersect(find(KidsIntD(:,3) ==
2), find(sum(KidsIntD(:,1:2)') .^2) <= 2)'));
        DailyStatus{2,i}(t,s) = length(intersect(find(KidsPlaD(:,3) ==
2), find(sum(KidsPlaD(:,1:2)') .^2) <= 2)'));
    elseif s == 19;        %Tallying ill with Ec & Gi only
        DailyStatus{1,i}(t,s) =
length(intersect(find(prod(KidsIntD(:,1:2)') == 4), find(KidsIntD(:,3) <= 1)));
        DailyStatus{2,i}(t,s) =
length(intersect(find(prod(KidsPlaD(:,1:2)') == 4), find(KidsPlaD(:,3) <= 1)));
    elseif s == 20;        %Tallying ill with Ec & Ro only
        DailyStatus{1,i}(t,s) =
length(intersect(find(prod([KidsIntD(:,1) KidsIntD(:,3)]') == 4), find(KidsIntD(:,2) <=
1)));
        DailyStatus{2,i}(t,s) =
length(intersect(find(prod([KidsPlaD(:,1) KidsPlaD(:,3)]') == 4), find(KidsPlaD(:,2) <=
1)));
    elseif s == 21;        %Tallying ill with Gi & Ro only

```



```

        DailyStatus{1,i}(t,s) =
length(intersect(find(prod(KidsIntD(:,2:3)') == 4), find(KidsIntD(:,1) <= 1)));
        DailyStatus{2,i}(t,s) =
length(intersect(find(prod(KidsPlaD(:,2:3)') == 4), find(KidsPlaD(:,1) <= 1)));
        elseif s == 22;      %Tallying ill with Ec & Gi & Ro
            DailyStatus{1,i}(t,s) = length(find(prod(KidsIntD(:,1:3)') ==
8)));
            DailyStatus{2,i}(t,s) = length(find(prod(KidsPlaD(:,1:3)') ==
8)));
        end
    end
end
%Now storing the person-pathogen matrices at the end of this day, if exactly 1
loop was requested.
if loops == 1;
    PPmatrices(t+1).KidsInt = KidsInt; PPmatrices(t+1).KidsPla = KidsPla;
    PPmatrices(t+1).KidsIntD = KidsIntD; PPmatrices(t+1).KidsPlaD = KidsPlaD;
end;
%sizeof(DailyStatus)      %Debug measure - too big for memory?
end
%disp([' ']) %Adds line feed to separate the progress counters.
%=====End daily loop, begin more testing code=====
if testing == 1;
    maxTimePla = find(KidsPlaInfPrev == max(KidsPlaInfPrev)); %Getting time points
of max. prevalence of infection (1st, if tie).
    disp(['1st time point where max. prevalence (placebo) is seen is
',num2str(maxTimePla(1))]) %Printing first point of max. prevalence.
    figure(i); %Plotting infection prevalence.
    subplot(2,1,1);
    plot([1:maxTime],KidsIntInfPrev,'-k');
    title 'Daily infec. prev. (red=Ec,green=Gi,blue=Ro,black=any), dots = reported
waterborne prev.';
    ylabel 'Proportion affected (intervention)';

```

```

xlim([0 625]);
hold on;
plot([1:maxTime],KidsIntInfEc,'-r');
plot([1:maxTime],KidsIntInfGi,'-g');
plot([1:maxTime],KidsIntInfRo,'-b');
h5 = plot([daysBurnIn:interval:maxTime],OutPrevs.KidsInt,'cx');
set(h5,'linewidth',2);
legend({'\fontsize{10} Any infection (daily)}','{\fontsize{10} E. coli
infection (daily)}','{\fontsize{10} Giardia infection (daily)}','{\fontsize{10} Rotavirus
infection (daily)}','{\fontsize{10} LP_{Irwd} (prior week)}');
hold off;
subplot(2,1,2);
plot([1:maxTime],KidsPlaInfPrev,'-k');
xlim([0 625]);
xlabel 'Time (simulated days)';
ylabel 'Proportion affected (placebo)';
hold on;
plot([1:maxTime],KidsPlaInfEc,'-r');
plot([1:maxTime],KidsPlaInfGi,'-g');
plot([1:maxTime],KidsPlaInfRo,'-b');
h5 = plot([daysBurnIn:interval:maxTime],OutPrevs.KidsPla,'cx');
set(h5,'linewidth',2);
legend({'\fontsize{10} Any infection (daily)}','{\fontsize{10} E. coli
infection (daily)}','{\fontsize{10} Giardia infection (daily)}','{\fontsize{10} Rotavirus
infection (daily)}','{\fontsize{10} LP_{Prwd} (prior week)}');
hold off;
end %=====End testing code=====
%Adding baseline prevalence. Corrected for doublecounting (some infected people
would have been infected anyway by baseline transmission).
OutPrevs.KidsInt = OutPrevs.KidsInt + prevDiarrhBaseKids * (1 - OutPrevs.KidsInt);
OutPrevs.KidsPla = OutPrevs.KidsPla + prevDiarrhBaseKids * (1 - OutPrevs.KidsPla);
i %Printing i as a debug measure
OutQMRA.KILP(i) = mean(OutPrevs.KidsInt); OutQMRA.KPLP(i) = mean(OutPrevs.KidsPla);

```

```

%Output long. prevs.
OutQMRA.LPR(i) = OutQMRA.KILP(i) / OutQMRA.KPLP(i);
if storeStatus == 1;      %Storing incidences for intervention group.
    OutQMRA.IncInfIntEc(i) = sum(DailyStatus{1,i}(maxTime-364:maxTime,1));
%Yearly infection incidence for E. coli.
    OutQMRA.IncIllIntEc(i) = sum(DailyStatus{1,i}(maxTime-364:maxTime,4));
%Yearly illness incidence for E. coli.
    OutQMRA.IncInfIntGi(i) = sum(DailyStatus{1,i}(maxTime-364:maxTime,2));
%Yearly infection incidence for Giardia.
    OutQMRA.IncIllIntGi(i) = sum(DailyStatus{1,i}(maxTime-364:maxTime,5));
%Yearly illness incidence for Giardia.
    OutQMRA.IncInfIntRo(i) = sum(DailyStatus{1,i}(maxTime-364:maxTime,3));
%Yearly infection incidence for rotavirus.
    OutQMRA.IncIllIntRo(i) = sum(DailyStatus{1,i}(maxTime-364:maxTime,6));
%Yearly illness incidence for rotavirus.
    %As above, for placebo group.
    OutQMRA.IncInfPlaEc(i) = sum(DailyStatus{2,i}(maxTime-364:maxTime,1));
%Yearly infection incidence for E. coli.
    OutQMRA.IncIllPlaEc(i) = sum(DailyStatus{2,i}(maxTime-364:maxTime,4));
%Yearly illness incidence for E. coli.
    OutQMRA.IncInfPlaGi(i) = sum(DailyStatus{2,i}(maxTime-364:maxTime,2));
%Yearly infection incidence for Giardia.
    OutQMRA.IncIllPlaGi(i) = sum(DailyStatus{2,i}(maxTime-364:maxTime,5));
%Yearly illness incidence for Giardia.
    OutQMRA.IncInfPlaRo(i) = sum(DailyStatus{2,i}(maxTime-364:maxTime,3));
%Yearly infection incidence for rotavirus.
    OutQMRA.IncIllPlaRo(i) = sum(DailyStatus{2,i}(maxTime-364:maxTime,6));
%Yearly illness incidence for rotavirus.
    %'Actual' longitudinal prevalences (person-days ill or infected, divided by
total person-days observed).
    OutQMRA.LPInfIntEc(i) = sum(sum(DailyStatus{1,i}(maxTime-364:maxTime,[ 8 11 12
14])))/(nKidsInt*365); %Yearly inf. LP, E. coli.
    OutQMRA.LPIllIntEc(i) = sum(sum(DailyStatus{1,i}(maxTime-364:maxTime,[16 19 20

```

```

22]))/ (nKidsInt*365); %Yearly ill LP, E. coli.
    OutQMRA.LPInfIntGi(i) = sum(sum(DailyStatus{1,i}(maxTime-364:maxTime,[ 9 11 13
14]))/ (nKidsInt*365); %Yearly inf. LP, Giardia.
    OutQMRA.LPIllIntGi(i) = sum(sum(DailyStatus{1,i}(maxTime-364:maxTime,[17 19 21
22]))/ (nKidsInt*365); %Yearly ill LP, Giardia.
    OutQMRA.LPInfIntRo(i) = sum(sum(DailyStatus{1,i}(maxTime-364:maxTime,[10 12 13
14]))/ (nKidsInt*365); %Yearly inf. LP, rota.
    OutQMRA.LPIllIntRo(i) = sum(sum(DailyStatus{1,i}(maxTime-364:maxTime,[18 20 21
22]))/ (nKidsInt*365); %Yearly ill LP, rota.
    OutQMRA.LPInfIntMix(i) = sum(sum(DailyStatus{1,i}(maxTime-364:maxTime,11:14)))/
(nKidsInt*365); %Yearly inf. LP, mixed.
    OutQMRA.LPIllIntMix(i) = sum(sum(DailyStatus{1,i}(maxTime-364:maxTime,19:22)))/
(nKidsInt*365); %Yearly ill LP, mixed.
    OutQMRA.LPInfIntAny(i) = sum(sum(DailyStatus{1,i}(maxTime-364:maxTime,8:14)))/
(nKidsInt*365); %Yearly inf. LP, any.
    OutQMRA.LPIllIntAny(i) = sum(sum(DailyStatus{1,i}(maxTime-364:maxTime,16:22)))/
(nKidsInt*365); %Yearly ill LP, any.
    %As above, for placebo group.
    OutQMRA.LPInfPlaEc(i) = sum(sum(DailyStatus{2,i}(maxTime-364:maxTime,[ 8 11 12
14]))/ (nKidsPla*365); %Yearly inf. LP, E. coli.
    OutQMRA.LPIllPlaEc(i) = sum(sum(DailyStatus{2,i}(maxTime-364:maxTime,[16 19 20
22]))/ (nKidsPla*365); %Yearly ill LP, E. coli.
    OutQMRA.LPInfPlaGi(i) = sum(sum(DailyStatus{2,i}(maxTime-364:maxTime,[ 9 11 13
14]))/ (nKidsPla*365); %Yearly inf. LP, Giardia.
    OutQMRA.LPIllPlaGi(i) = sum(sum(DailyStatus{2,i}(maxTime-364:maxTime,[17 19 21
22]))/ (nKidsPla*365); %Yearly ill LP, Giardia.
    OutQMRA.LPInfPlaRo(i) = sum(sum(DailyStatus{2,i}(maxTime-364:maxTime,[10 12 13
14]))/ (nKidsPla*365); %Yearly inf. LP, rota.
    OutQMRA.LPIllPlaRo(i) = sum(sum(DailyStatus{2,i}(maxTime-364:maxTime,[18 20 21
22]))/ (nKidsPla*365); %Yearly ill LP, rota.
    OutQMRA.LPInfPlaMix(i) = sum(sum(DailyStatus{2,i}(maxTime-364:maxTime,11:14)))/
(nKidsPla*365); %Yearly inf. LP, mixed.
    OutQMRA.LPIllPlaMix(i) = sum(sum(DailyStatus{2,i}(maxTime-364:maxTime,19:22)))/

```

```

(nKidsPla*365);          %Yearly ill LP, mixed.
    OutQMRA.LPInfPlaAny(i) = sum(sum(DailyStatus{2,i}(maxTime-364:maxTime,8:14)))/
(nKidsPla*365);          %Yearly inf. LP, any.
    OutQMRA.LPIllPlaAny(i) = sum(sum(DailyStatus{2,i}(maxTime-364:maxTime,16:22)))/
(nKidsPla*365);          %Yearly ill LP, any.
    %Longitudinal prevalence ratios.
    OutQMRA.LPRIInfIntEc(i) = OutQMRA.LPInfIntEc(i) / OutQMRA.LPInfPlaEc(i);
%E. coli, infection.
    OutQMRA.LPRIllIntEc(i) = OutQMRA.LPIllIntEc(i) / OutQMRA.LPIllPlaEc(i);
%E. coli, illness.
    OutQMRA.LPRIInfIntGi(i) = OutQMRA.LPInfIntGi(i) / OutQMRA.LPInfPlaGi(i);
%Giardia, infection.
    OutQMRA.LPRIllIntGi(i) = OutQMRA.LPIllIntGi(i) / OutQMRA.LPIllPlaGi(i);
%Giardia, illness.
    OutQMRA.LPRIInfIntRo(i) = OutQMRA.LPInfIntRo(i) / OutQMRA.LPInfPlaRo(i);
%Rota, infection.
    OutQMRA.LPRIllIntRo(i) = OutQMRA.LPIllIntRo(i) / OutQMRA.LPIllPlaRo(i);
%Rota, illness.
    OutQMRA.LPRIInfIntAny(i) = OutQMRA.LPInfIntAny(i) / OutQMRA.LPInfPlaAny(i);
%Any, infection.
    OutQMRA.LPRIllIntAny(i) = OutQMRA.LPIllIntAny(i) / OutQMRA.LPIllPlaAny(i);
%Any, illness.
end
if mod(i,floor(loops/10)) == 0;    %Progress meter
    disp(['Loop ',num2str(i),' of ',num2str(loops),' complete.'])
    toc
end
end %====Ending main QMRA loop=====
disp(['Program finished.'])
toc %Outputs runtime.
OutQMRA.EndTime = clock;
eval(['save Results/OutQMRA',datestr(StartTime,30),'.mat, OutQMRA;']) %Saving output
file.

```

```

if noStochParams == 1 && loops <= 25; %Displaying results if a series of input
parameters were tested.
    out = [OutQMRA.KILP; OutQMRA.KPLP; OutQMRA.LPR; OutQMRA.EcL; OutQMRA.GiL;
OutQMRA.RoL; OutQMRA.PrevBase]';
    disp('KILP, KPLP, LPR, EcL, GiL, RoL, PrevBase, PrevBase for 0.84 LPR')
    out(:,8) = (0.84 * out(:,2) - out(:,1)) / 0.16 %Calc. necessary PrevBase to get
an LPR of 0.84.
    min(out)
    mean(out)
    max(out)
end
if Octave == 1; %Audio alert when done.
    system('aplay -q /usr/lib/openoffice/basis-link/share/gallery/sounds/ok.wav');
else
    system('start c:\windows\media\tada.wav');
end

```

```
%{  
COPYRIGHT INFORMATION
```

Copyright 2011 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (AssignInf.m) is part of QMRAv13_20110414.

QMRAv13_20110414 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRAv13_20110414 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRAv13_20110414. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
%Just like AssignIll.m, but loops over columns of the person-pathogen matrix instead of rows, & should be faster.
```

```
%Assigns an illness duration to entries of a matrix, where rows are people and columns are pathogens
```

```
%Positive entries mean the person is infected, negative entries (or 0) mean the person has recovered
```

```
%Inputs:
```

```
%   PPmatrix: Matrix with 1 row per person and 1 column per pathogen (Person-Pathogen matrix)
```

```
%   Responses:   Vector of illness responses (probability of illness given dose), 1 entry per pathogen
```

```
%   ResponsesNon: Vector of illness responses for people not using any device
```

```

% pUse:          Probability that a person is not using any device
% durations:    Vector of durations of illnesses, 1 entry per pathogen
% ImmuneTimes:  Vector of durations of immunity, 1 entry per pathogen
%The output (matrixOut) is matrixIn with new illness durations assigned to some of its
entries.

function PPmatrix = AssignInf(PPmatrix,Responses,Durations,ImmuneTimes);
    sizePP = size(PPmatrix);
    Randoms = rand(sizePP); %Random #s for determining infection. One number per person
per pathogen.
    %NewlyInfected = cell(sizePP(2),1); %Initializing cell array to store index
values of people who will be newly infected.
    Immunities = ones(sizePP(1),1) * ImmuneTimes + PPmatrix;
    %Durations = ones(sizePP) * Durations;
    %Immunities = PPmatrix + ImmuneTimes; %Immunities gives days left in (infection +
immune period), or a neg. # if susceptible.
    for i = 1:sizePP(2);
        Immunities(:,i) = PPmatrix(:,i) + ImmuneTimes(i);
        NewlyInfected = intersect(find(Immunities(:,i) <= 0), find(Randoms(:,i) <
Responses(:,i))); %Gets indices of newly infected.
        PPmatrix(NewlyInfected,i) = Durations(i);
    end
end
end

```



```
%{  
COPYRIGHT INFORMATION
```

Copyright 2011 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (AssignInfIllRand.m) is part of QMRAv13_20110414.

QMRAv13_20110414 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRAv13_20110414 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRAv13_20110414. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
%Similar to AssignIll.m, but assigns infection durations randomly.
```

```
%Assigns an illness duration to entries of a matrix, where rows are people and columns are pathogens
```

```
%Positive entries mean the person is ill, negative entries (or 0) mean the person has recovered
```

```
%Inputs:
```

```
% PPmatrix: Matrix with 1 row per person and 1 column per pathogen (Person-Pathogen matrix)
```

```
% Responses: Matrix of illness responses (probability of illness given dose), 1 row per person and 1 column per pathogen
```

```
% ImmuneTimes: Vector of durations of immunity, 1 entry per pathogen
```

```
%The output (matrixOut) is matrixIn with new illness durations assigned to some of its
```

entries.

%It requires the functions durEc(), durGi(), & durRo() in IllDurations.m.

```
function [PPmatrix,PPmatrixD] =
AssignInfIllRand(PPmatrix,PPmatrixD,Responses,ImmuneTimes,MorbidityK);
    %rand('state',28)
    sizePP = size(PPmatrix);
    Randoms = rand(sizePP);          %Random #s for determining infection. One number per
person per pathogen.
    Randoms2 = rand(sizePP); %Random #s for determining disease, as above.
    %NewlyInfected = cell(sizePP(2),1); %Initializing cell array to store index
values of people who will be newly infected.
    Immunities = ones(sizePP(1),1) * ImmuneTimes + PPmatrix; %Adjusts PPmatrix to
account for immunity.
    Durations = NaN(sizePP); %Making & populating a matrix of disease durations (same
size as PPmatrix) to assign to newly infected.
    Durations(:,1) = durEc(sizePP(1));
    Durations(:,2) = durGi(sizePP(1));
    Durations(:,3) = durRo(sizePP(1));
    for i = 1:sizePP(2); %Loop, once for each pathogen.
        NewlyInfected = intersect(find(Immunities(:,i) <= 0), find(Randoms(:,i) <
Responses(:,i))); %Gets indices of newly infected. Note: 0 or less is susc.
        PPmatrix(NewlyInfected,i) = Durations(NewlyInfected,i);
        PPmatrixD(NewlyInfected,i) = 1; %Flags newly infected.
        NewlyIll = intersect(NewlyInfected, find(Randoms2(:,i) < MorbidityK(i)));
        PPmatrixD(NewlyIll,i) = 2;
    end
end
```

```
%{  
COPYRIGHT INFORMATION
```

Copyright 2011 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (AssignInfRand.m) is part of QMRAv13_20110414.

QMRAv13_20110414 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRAv13_20110414 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRAv13_20110414. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
%Similar to AssignIll.m, but assigns infection durations randomly.
```

```
%Assigns an infection duration to entries of a matrix, where rows are people and columns are pathogens
```

```
%Positive entries mean the person is infected, negative entries (or 0) mean the person has recovered
```

```
%Inputs:
```

```
% PPmatrix: Matrix with 1 row per person and 1 column per pathogen (Person-Pathogen matrix)
```

```
% Responses: Matrix of illness responses (probability of infection given dose), 1 row per person and 1 column per pathogen
```

```
% ImmuneTimes: Vector of durations of immunity, 1 entry per pathogen
```

```
%The output (matrixOut) is matrixIn with new illness durations assigned to some of its
```

entries.

%It requires the functions durEc(), durGi(), & durRo() in IllDurations.m.

```
function [PPmatrix] = AssignInfRand(PPmatrix,Responses,ImmuneTimes);
    %rand('state',28)
    sizePP = size(PPmatrix);
    Randoms = rand(sizePP); %Random #s for determining infection. One number per person
per pathogen.
    Immunities = ones(sizePP(1),1) * ImmuneTimes + PPmatrix; %Adjusts PPmatrix to
account for immunity.
    Durations = NaN(sizePP); %Making & populating a matrix of disease durations (same
size as PPmatrix) to assign to newly infected.
    Durations(:,1) = durEc(sizePP(1));
    Durations(:,2) = durGi(sizePP(1));
    Durations(:,3) = durRo(sizePP(1));
    NewlyInfected = intersect(find(Immunities <= 0), find(Randoms < Responses));
    %Gets indices of newly infected. Note: 0 or less is susc.
    PPmatrix(NewlyInfected) = Durations(NewlyInfected);
end
```

```
%{  
COPYRIGHT INFORMATION
```

Copyright 2011 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (CalcDiarrhWeeks.m) is part of QMRAv13_20110414.

QMRAv13_20110414 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRAv13_20110414 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRAv13_20110414. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
%Calculates whether a given week is reported as a week with diarrhea under the reporting scheme in the DRC Lifestraw RCT (Boisson 2010).
```

```
%It considers reduced recall of past diarrheal episodes after 2d ('remembrance') and possible distinct diarrhea episodes in the previous 7d.
```

```
%It operates on a matrix:
```

```
% Rows represent people, columns represent pathogens/illnesses, and entries represent # of days remaining in the illness.
```

```
%It outputs a vector with 1 entry per person, 1 if illness is reported during the week, 0 if not.
```

```
function vec = CalcDiarrhWeeks(InMatrix, remembrance,timeWindow,Morbidity);  
    sizeInMatrix = size(InMatrix);
```

```

Randoms = rand(sizeInMatrix(1),sizeInMatrix(2));
for j = 1:sizeInMatrix(2);
    InMatrix((find(Randoms(:,j) > Morbidity(j))),j) = -9999;
end

    for i = 1:sizeInMatrix(1); %Loop over all people. Only the most recent episode
(largest entry in a row) is used to assign illness.
        %Randoms = rand(1,columns(InMatrix)); %Random numbers for determining
morbidity (these 2 lines moved upward for greater speed)
        %InMatrix(i,find(Randoms > Morbidity)) = -9999; %Apply morbidity ratio: if
asymptomatic, infection is set to -9999, and therefore not reported. Note that this
modification is not passed out of this function.
        if max(InMatrix(i,:)) >= -2;
            vec(i) = 1; %If ill during day 0, 1, or 2, assume illness is always
reported, therefore assign illness.
        elseif (max(InMatrix(i,:)) >= -timeWindow) & rand() < remembrance);
            vec(i) = 1; %Otherwise, if ill during days 3-7, randomly determine if
episode is remembered. If so, assign illness.
        else
            vec(i) = (0); %Otherwise, no illness is remembered or reported. Assign no
illness.
        end
    end
end
end

```

```
%{  
COPYRIGHT INFORMATION
```

Copyright 2011 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (DRbP.m) is part of QMRAv13_20110414.

QMRAv13_20110414 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRAv13_20110414 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRAv13_20110414. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
%Beta-Poisson dose response model, using N50 (default) or beta as a parameter  
%function outvar = DRbP(N50orBeta,alpha,invar,reverse='no',WhichParam='N50') %Ordinarily,  
invar is dose & outvar is response.  
function outvar = DRbP(N50orBeta,alpha,invar,reverse,WhichParam) %Ordinarily, invar is  
dose & outvar is response.  
    if nargin == 3;  
        reverse = 'no'; WhichParam = 'N50';  
    end  
    switch(reverse)  
        case 'no'  
            switch(WhichParam)  
                case 'N50'
```

```

        outvar = 1-(1+(invar/N50orBeta)*(2^(1/alpha)-1)).^-alpha;
    case 'Beta'
        outvar = 1-(1+(invar/N50orBeta)).^-alpha;
    otherwise
        error(['WhichParam must be "N50" or "Beta"'])
    end
case 'yes' %If reverse='yes', invar is response & outvar is dose.
    switch(WhichParam)
        case 'N50'
            outvar = N50orBeta * ( ((1-invar).^(-1/alpha) -1) /
(2^(1/alpha)-1) );
        case 'Beta'
            outvar = N50orBeta * ((1-invar).^(-1/alpha) -1);
        otherwise
            error(['WhichParam must be "N50" or "Beta"'])
        end
    otherwise
        error(['reverse must be "no" or "yes"'])
    end
end
end

```



```
%{  
COPYRIGHT INFORMATION
```

Copyright 2011 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (DRchoose.m) is part of QMRAv13_20110414.

QMRAv13_20110414 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRAv13_20110414 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRAv13_20110414. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
%Returns a vector of response values, determined by dose response models, for several pathogens.
```

```
%Inputs are vectors with 1 entry per pathogen, all in the same order:
```

```
% nonalphas:      k parameter (exponential) or N50 parameter (beta-Poisson)
```

```
% alphas:        alpha parameter (beta-Poisson); NA if exponential model is desired
```

```
% Doses:         Doses of pathogens received per individual (under default behavior;  
see 'reverse' below)
```

```
% morbidities:   Morbidity ratios: proportion of infected who are ill
```

```
% reverse:       Defaults to 'no', determining proportion ill from dose. If 'yes',  
determines dose from proportion ill.
```

```
%The output (outvec) is a vector containing the proportions of exposed who will fall ill.
```

```
%If reverse=='yes', outvec is a vector of doses calculated from invec, the proportions
```

```

ill.
%This code requires several custom functions/subroutines in the working directory:
%   DRexp.m:      Exponential dose response model
%   DRbP.m:      Beta-Poisson dose response model

%function outvec =
%DRchoose(nonalphas,alphas,invec,morbidities=1,reverse='no') %Octave
function outvec = DRchoose(nonalphas,alphas,Doses,morbidities,reverse)
    if nargin == 3;
        morbidities = 1; reverse = 'no';
    end
    if morbidities == 1;
        morbidities = ones(length(nonalphas));
    end
    switch(reverse)
        case 'no'
            for i=1:length(alphas);
                if (isnan(alphas(i))) %if alpha is NA, run exponential dose
response
                    if size(Doses)(1) == 1;
                        outvec(i) = DRexp(nonalphas(i),Doses(i)) * morbidities(i);
                    else
                        outvec(i) = DRexp(nonalphas(i),Doses(:,i)) *
morbidities(i);
                    end
                else %run beta-Poisson dose response
                    outvec(i) = DRbP(nonalphas(i),alphas(i),Doses(i)) *
morbidities(i);
                end
            end
        case 'yes'
            for i=1:length(alphas);
                if (isnan(alphas(i))) %if alpha is NA, run exponential dose

```

```

response
    outvec(i) = DRexp(nonalphas(i),Doses(i) ./
morbidityies(i),'yes','N50') ;
    else
        %run beta-Poisson dose response
        outvec(i) = DRbP(nonalphas(i),alphas(i),Doses(i) ./
morbidityies(i),'yes','N50');
    end
end
    end
otherwise
    error('reverse must be "no" or "yes"')
end
end
end

```

```
%{  
COPYRIGHT INFORMATION
```

Copyright 2011 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (DRexp.m) is part of QMRAv13_20110414.

QMRAv13_20110414 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRAv13_20110414 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRAv13_20110414. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
%Exponential dose response model
```

```
%function outvar = DRexp(k, invar,reverse='no') %Ordinarily, invar is dose & outvar is response.
```

```
function outvar = DRexp(k, invar, reverse) %This works in Matlab.
```

```
    if nargin < 3;
```

```
        reverse = 'no';
```

```
    end
```

```
    switch(reverse);
```

```
        case 'no';
```

```
            outvar = 1-exp(-k * invar);
```

```
        case 'yes'; %If reverse='yes', invar is response & outvar is dose.
```

```
            outvar = log(1-invar)/-k;
```

```
    otherwise
      error(['reverse (last parameter) must be "no" or "yes"'])
    end
  end
```

```
%{  
COPYRIGHT INFORMATION
```

Copyright 2011 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (durEc.m) is part of QMRv13_20110414.

QMRv13_20110414 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRv13_20110414 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRv13_20110414. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
%Functions for calculating vectors of illness durations
```

```
function output = durEc(n);
```

```
    output = round(gamrnd(1.775,1.690,[n,1]));    %Shape, then scale
```

```
    output(find(output == 0)) = 0.1;           %Sets zero durations to 0.1 day instead.
```

```
Will still function as 1 day.
```

```
end
```

```
%{  
COPYRIGHT INFORMATION
```

Copyright 2011 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (durGi.m) is part of QMRAv13_20110414.

QMRAv13_20110414 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRAv13_20110414 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRAv13_20110414. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
%Functions for calculating vectors of illness durations
```

```
function output = durGi(n); %Based on a fit of gamma dist. to limited info from Kent GP 1988.
```

```
    output = round(gamrnd(3.206,3.431,[n,1])); %Shape, then scale
```

```
    output(find(output == 0)) = 0.1; %Sets zero durations to 0.1 day instead.
```

```
Will still function as 1 day.
```

```
end
```

```
%{  
COPYRIGHT INFORMATION
```

Copyright 2011 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (durRo.m) is part of QMRAv13_20110414.

QMRAv13_20110414 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRAv13_20110414 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRAv13_20110414. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
%Functions for calculating vectors of illness durations  
function output = durRo(n); %Based on 4 rotavirus-infected volunteers having durations  
of 1, 2, 3, and 4 days (Kapikian 1983).  
    output = ceil(rand([n,1]) * 4);  
    output(find(output == 0)) = 0.1; %Sets zero durations to 0.1 day instead.  
Will still function as 1 day.  
end
```



```
%{  
COPYRIGHT INFORMATION
```

Copyright 2011 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (Examine1Run.m) is part of QMRAv13_20110414.

QMRAv13_20110414 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRAv13_20110414 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRAv13_20110414. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
%Examining data from a single run of the QMRA LifeStraw model.
```

```
%Converting to susceptible (-9), immune (-1), or diseased (9) for each of the 3 pathogens.
```

```
%====Setting options====
```

```
startTime = 96;      %Time point at which to start looking at the data. Note that 1st matrix corresponds to time 0.
```

```
recode = 1;        %If 1, recode matrix entries to susc./inf./immune.
```

```
%====Finished with options, starting processing.====
```

```
PPMs = PPmatrices; %Making a copy.
```

```
switch(recode);
```

```
    case 1;
```

```
        disp(['Recoding raw numbers to susc./inf./immune.'])
```

```

    for i = startTime:size(PPmatrices)(2); %Starting when the actual simulation starts
(after equilibrium is reached).
        PPMs(i).KidsInt(PPMs(i).KidsInt <= -7) = -9; %7 day immune period; 0 counts as
the 1st immune day, so at -7 they are susc.
        PPMs(i).KidsInt(PPMs(i).KidsInt > 0) = 9; %Susceptible if a positive
integer.
        PPMs(i).KidsInt(abs(PPMs(i).KidsInt) != 9) = -1;%Immune if neither of the above
applies.
        PPMs(i).KidsInt(PPMs(i).KidsInt == 9) = 2; %Infected person-day marked as 2,
so as to more easily distinguish.
    end
    for i = startTime:size(PPmatrices)(2); %Same as above 'for' loop, but placebo.
        PPMs(i).KidsPla(PPMs(i).KidsPla <= -7) = -9;
        PPMs(i).KidsPla(PPMs(i).KidsPla > 0) = 9;
        PPMs(i).KidsPla(abs(PPMs(i).KidsPla) != 9) = -1;
        PPMs(i).KidsPla(PPMs(i).KidsPla == 9) = 2;
    end
    otherwise
        disp(['Not recoding to susc./inf./immune, output will display raw numbers.'])
    end
KidsIntStatusEc = NA(size(PPMs(1).KidsInt)(1), size(PPmatrices)(2)-(startTime-1));
KidsIntStatusGi = KidsIntStatusEc;
KidsIntStatusRo = KidsIntStatusEc;
KidsPlaStatusEc = NA(size(PPMs(1).KidsPla)(1), size(PPmatrices)(2)-(startTime-1));
KidsPlaStatusGi = KidsPlaStatusEc;
KidsPlaStatusRo = KidsPlaStatusEc;
for i = startTime:size(PPmatrices)(2); %Starting when the actual simulation starts
(after equilibrium is reached).
    for j = 1:size(PPMs(1).KidsInt)(1);
        KidsIntStatusEc(j,i-startTime+1) = PPMs(i).KidsInt(j,1);
        KidsIntStatusGi(j,i-startTime+1) = PPMs(i).KidsInt(j,2);
        KidsIntStatusRo(j,i-startTime+1) = PPMs(i).KidsInt(j,3);
    end
end

```

```

    for j = 1:size(PPMs(1).KidsPla)(1);
        KidsPlaStatusEc(j,i-startTime+1) = PPMs(i).KidsPla(j,1);
        KidsPlaStatusGi(j,i-startTime+1) = PPMs(i).KidsPla(j,2);
        KidsPlaStatusRo(j,i-startTime+1) = PPMs(i).KidsPla(j,3);
    end
end
%Now can visually inspect the 6 status matrices that have been output.
%Should be a way to collapse them also (run-length encoding?), but not yet implemented.

function outmatrix = coll(inmatrix,nMaxRuns) %Inefficient but hopefully works.
    sizeM = size(inmatrix);
    maxk = 1; %Initializing counter to determine the maximum number of runs ever seen
    during the function call.
    for i = 1:sizeM(1); %Loop over all rows
        for j = 2:sizeM(2); %Loop over each entry per row
            if j == 2; %Special procedure for first iteration, since there could
                be a transition (or not) between the 1st 2 entries.
                    k = 1; %Initiating run counter;
                    if inmatrix(i,j) != inmatrix(i,j-1);
                        if inmatrix(i,j-1) == -9; dur(k) = -1;
                        elseif inmatrix(i,j-1) == -1; dur(k) = 0.001;
                        elseif inmatrix(i,j-1) == 2; dur(k) = 1;
                        else stop('Unexpected value in matrix (not -9, -1, or 2).')
                        end
                        k = k + 1; %Increment run counter
                        if inmatrix(i,j) == -9; dur(k) = -1;
                        elseif inmatrix(i,j) == -1; dur(k) = 0.001;
                        elseif inmatrix(i,j) == 2; dur(k) = 1;
                        else stop('Unexpected value in matrix (not -9, -1, or 2).')
                        end
                    else %If type of run does not change in 1st 2 entries
                        if inmatrix(i,j) == -9; dur(k) = -2;
                        elseif inmatrix(i,j) == -1; dur(k) = 0.002;
                    end
                end
            end
        end
    end
end

```

```

        elseif inmatrix(i,j) == 2; dur(k) = 2;
        else stop('Unexpected value in matrix (not -9, -1, or 2).')
        end
    end
    elseif j == sizeM(2);    %Special procedure for last iteration - need to
drop it since it is probably incomplete.
        dur(k) = 0;
    else %If j (the column) is anything greater than 2, but not the last
column:
        if inmatrix(i,j) != inmatrix(i,j-1);    %If there is a transition,
reset the run:
            k = k + 1;
            if inmatrix(i,j) == -9; dur(k) = -1;
            elseif inmatrix(i,j) == -1; dur(k) = 0.001;
            elseif inmatrix(i,j) == 2; dur(k) = 1;
            else stop('Unexpected value in matrix (not -9, -1, or 2).')
            end
        else %If there is no transition, extend the run
            if inmatrix(i,j) == -9; dur(k) = dur(k) - 1;
            elseif inmatrix(i,j) == -1; dur(k) = dur(k) + 0.001;
            elseif inmatrix(i,j) == 2; dur(k) = dur(k) + 1;
            else stop('Unexpected value in matrix (not -9, -1, or 2).')
            end
        end
    end
end
if k > maxk;
    maxk = k %Updates & displays maxk (largest no. runs seen so far). Use as
guide for entering nMaxRuns.
end
outmatrix(i,:) = padarray(dur,[0, nMaxRuns - length(dur)],0,'post');
end
outmatrix(:,1) = 0; %Sets all 1st runs to 0 (they are likely to be incomplete).

```

```

    mean(outmatrix(find(outmatrix >= 1)))    %Print mean duration of illness
end

function z = GraphColl(outmatrix) %Graphs output of above function.
    bins = unique(outmatrix(find(outmatrix < 0)));
    if length(bins) == 1; binsT=NaN(3,1); binsT(2)=bins; binsT(1)=binsT(2)-1;
binsT(3)=binsT(2)+1; bins=binsT; end
    subplot(2,2,1);
    hist(outmatrix(find(outmatrix < 0)),bins)    %Durations of susceptibility
    bins = unique(outmatrix(find(outmatrix > 0 & outmatrix < 1)));
    if length(bins) == 1; binsT=NaN(3,1); binsT(2)=bins; binsT(1)=binsT(2)-0.001;
binsT(3)=binsT(2)+0.001; bins=binsT; end
    subplot(2,2,2);
    hist(outmatrix(find(outmatrix > 0 & outmatrix < 1)),bins)    %Durations of immunity
    bins = unique(outmatrix(find(outmatrix >= 1)));
    if length(bins) == 1; binsT=NaN(3,1); binsT(2)=bins; binsT(1)=binsT(2)-1;
binsT(3)=binsT(2)+1; bins=binsT; end
    subplot(2,2,3);
    hist(outmatrix(find(outmatrix >= 1)),bins)    %Durations of illness
end

%histc(PlaRo(find(PlaRo < 1 & PlaRo > 0)),[0:0.001:max(PlaRo(find(PlaRo < 1 & PlaRo >
0)))]))    %Awful (but functional) way to get counts of possible values for immunity
length.

```

```
%{  
COPYRIGHT INFORMATION
```

Copyright 2011 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (GetTrialParams.m) is part of QMRAv13_20110414.

QMRAv13_20110414 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRAv13_20110414 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRAv13_20110414. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
%This script generates a series of parameters for input into the main code, rather than  
stochastically generating them.
```

```
reps = 1;          %# of times to run each parameter set.  
readParams = 0;   %To read parameters generated by a previous model run (1) or not  
(0).  
paramSets = 10;   %# of parameter sets to be run, if generating them  
systematically.  
%loops = paramSets * reps;   %Deliberately overwrites 'loops' in the main code.  
switch(readParams);  
    case 1;  
        TrialParams = dlmread('Results/RunsThatFit.csv',' ',1,1);   %Access a file  
returned by ReadOutput.r.
```

```

    paramSets = size(TrialParams, 1);    %Overwrites 'paramSets' above.
    disp(['Reading parameter values from ',num2str(paramSets),' trials.'])
case 0;
    MinDoses = [2e4, 0, 0]; %Minimum non-zero dose. A good choice is dose that
infects 1% of population (ID1).
    ID1s = [7.5697E3, 5.0708E-1, 1.7280E-2];    %ID1 for ETEC, Giardia, & rota.
    MinDoses = ID1s * .1;    %Uncomment if ID1s are desired.
    TrialParams = zeros(paramSets,size(MinDoses, 2));
    for i = 1:length(MinDoses); %Populating all cells except the 1st row with the
minimum nonzero dose.
        TrialParams(2:paramSets,i) = MinDoses(i);
    end
    TrialParams(2,:) = MinDoses;    %1st run is 0 pathogens; 2nd run is the
minimum nonzero dose.
    for i = 3:paramSets;    %Uncomment the particular line desired. Comment all to
check multiple replicates of the same dose.
        %TrialParams(i,:) = MinDoses * (i-1);    %Linearly increases the dose
on each model run.
        TrialParams(i,:) = 2 * TrialParams(i-1,:); %Doubles the dose on each
model run.
        %TrialParams(i,2) = 10 * TrialParams(i-1,2); %Doubles the dose for only 1
pathogen, leaving others constant.
    end
    TrialParams(:,4) = 0;    %Sets a single value for non-waterborne diarrhea
prevalence.
    %disp('Will cycle through these parameters, 1 run per set.')
    %trialPrevDiarrhBaseKids = 0 %Sets a single value for non-waterborne diarrhea
prevalence.
    %TrialParams    %Print to screen, so we see that this file was
executed & to view the trial params.
    otherwise;
        error('readParams must be 0 or 1');
end

```

```

%Replicating the parameter sets.
TP = TrialParams; %Making a copy, for use in loop below.
if reps > 1;
    for i = 1:reps-1;
        TrialParams = [TrialParams; TP]; %Appending copies of the parameter sets.
    end
end
TrialParams = sortrows(TrialParams); %Sorting so that identical parameter values are
next to each other.
loops = paramSets; %Overwriting 'loops' variable in main QMRA code.
disp(['Using ',num2str(paramSets),' parameter sets, ',num2str(reps),' times each,
totaling ',num2str(loops),' runs.'])
if paramSets <= 25;
    disp(['Parameter sets are as follows:'])
    TP
end

```



```
%{  
COPYRIGHT INFORMATION
```

Copyright 2011 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (OutQMRAmerge.m) is part of QMRAv13_20110414.

QMRAv13_20110414 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRAv13_20110414 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRAv13_20110414. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
%Pulls together QMRA output files from multiple Octave threads. Runs surprisingly fast!  
clear all;
```

```
%First, enter the desired name of the .CSV:
```

```
filename = {'Results/MergedOutQMRA.csv'};
```

```
%Now enter as many files as necessary, each one containing the 'workspaces' from a  
thread.
```

```
Files = {'Results/OutQMRA20110315T130458.mat'};
```

```
rows = 0; %Initializing variable to count up total number of rows.
```

```
for i = 1:length(Files);
```

```
    eval(dispatch(['load ',char(Files(i)), ' OutQMRA;'])) %Loads the 'OutQMRA' struct stored  
in .mat file, overwriting that object if it exists.
```

```
    %disp(['File ',num2str(i),' took ',num2str(),' to run
```

```

',num2str(length(OutQMRA.CaL)),...
    %' loops (' ,num2str(),' per loop.')]
    eval(dispatch(['OutQMRA',num2str(i),'=OutQMRA;'])) %Copies it and adds a numeric
suffix to the name.
    rows = rows + length(OutQMRA.EcL);
end
clear OutQMRA; %Removes the initial copy of the last file loaded.

%CSVmatrix = NA(rows,length(fieldnames(OutQMRA1))-2; %Creating the output matrix. Each
row is a QMRA iteration.
for i = 1:length(Files);
    %i %For debugging
    eval(dispatch(['OutQMRA = OutQMRA',num2str(i),';'])) %Taking 'OutQMRAx' and creating a
copy called 'OutQMRA' to work from.
    OutQMRA = rmfield(OutQMRA, 'StartTime'); OutQMRA = rmfield(OutQMRA, 'EndTime');
    CSVmatrix = OutQMRA.Fit'; %Initializing a matrix that will become a .CSV by
transposing the first structure field into it.
    for [val,key] = OutQMRA; %This special syntax allows looping over all elements of
the structure.
        %key %For debugging
        if strcmp(char(key),'Fit') == 0; %Don't do anything for the 'Fit' element
because we took care of that 2 lines before.
            CSVmatrix = [CSVmatrix, val']; %Transpose fields into columns & bind
into the matrix.
        end
    end
    eval(dispatch(['CSVmatrix',num2str(i),' = CSVmatrix;']))
    clear CSVmatrix;
end

CSVmatrix = CSVmatrix1; %Initializing output matrix.
for i = 2:length(Files);
    eval(dispatch(['CSVmatrix = [CSVmatrix; CSVmatrix',num2str(i),';']))

```

end

```
fn=fieldnames(OutQMRA);
nFields = numel(fn);      %http://stackoverflow.com/questions/5292437/how-to-concat-cell-
array-of-strings-in-matlab
fn(1:nFields-1) = strcat(fn(1:nFields-1),{' ',''});
file = fopen(filename,'w+');
fprintf(file,'%s',disp([fn{:}]));
fclose(file);
eval(["dlmwrite('" ,char([filename]),"', 'CSVmatrix',' -append');"])
disp(['Done; .mat files have been merged and output to ',char(filename),' in
',char(pwd),' /Results/'])
```

9.4. The QMRA model investigating compliance and LRVs (chapter 4)

The model (referred to as “QMRA2v5” for short) consists of several text files containing necessary functions and subroutines; the core program is 'Main.m'. Simulation options are set by the choice of several values at the top of the files 'Main.m' and 'GetTrialParams.m'. These options default to values that generate a single test run of the simulation. Other options are set when calling the function 'Main.m' and are described within that file; for example, the following can be submitted at the Octave (or MATLAB) prompt to do a test calibration run:

```
Main(0,1,0,[0 0 0],[0 0 0],[1e5 1 .1],1,'test.csv','NA.csv',1,0,1,0)
```

The source code is found below. The filename of each of the source code files is found in the copyright information at the top of each file. Although QMRA2v5 uses some filenames that are identical to those in QMRAv13_20110414, the content of its files differs.

```
%Main function for running the model QMRA2v5, used for chapter 4 of Kyle S. Enger's Ph.D.
dissertation,
% as well as the manuscript "The joint effects of efficacy and compliance: a study of
household water treatment effectiveness against childhood diarrhea".
%Allows multiple runs (e.g., on computing cluster) using different parameters.
%Implements a QMRA model originally based on Boisson 2010 (PLoS One) RCT of Lifestraw
Family filtration device in the Dem. Rep. of the Congo.
%It was produced by modifying QMRAv13_20110414.
```

```
%{
COPYRIGHT INFORMATION
```

Copyright 2012 Kyle S. Enger.

This file (Main.m) is part of QMRA2v5.

QMRA2v5 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRA2v5 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRA2v5. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
%Requires Octave 3.2 or later.
```

```
%Also works well on MATLAB; the results in chapter 4 of the accompanying dissertation
were all produced with MATLAB.
```

```

%If running on MATLAB, requires the statistics toolbox, and possibly others.
%This code (Main.m, the core component of QMRA2v5) is accompanied by several
functions/subroutines, which need to be in the working directory:
% AnalyzeInfectionFromSpikes.m: Optional code to determine the proportion of all
infections that are caused by contamination spikes
% AssignInf.m: Stochastically assigns infections to individuals with fixed
durations
% AssignInfRand.m: Stochastically assigns infections to individuals with random
durations
% AssignInfIllRand.m: Stochastically assigns infections & illnesses to individuals
with random durations
% CalcDiarrhWeeks.m: Determines whether a week with 1+ days of diarrhea is actually
reported as a 'diarrhea week'; not actually used for ch. 4 analysis
% CalibrationLoopFuncCompile.m: Code that calls Main() in order to facilitate parallel
processing of many differently parameterized calibration runs
% DRbP.m: Beta-Poisson dose response model
% DRchoose.m: Executes the appropriate dose response model and determines
illness
% DRexp.m: Exponential dose response model
% durEc.m: Randomly pick a duration for E. coli infection
% durGi.m: Randomly pick a duration for Giardia infection
% durRo.m: Randomly pick a duration for rotavirus infection
% EstimationLoopFuncCompileV2.m: Code that calls Main() in order to facilitate
parallel processing of many differently parameterized estimation runs
% EstimationLoopFuncCompileV2PC.m: As above, but for estimation runs with perfect
compliance
% EstimationLoopFuncCompileV2Untreated.m: As above, but for estimation runs with
complete noncompliance
% Examine1Run.m: Allows inspection of the complete simulated data from a single
run of this code
% GetTrialParams.m: Generates a series of trial parameters instead of determining
them stochastically
% OutQMRAmerge.m: Allows conglomeration of output from multiple model

```

executions (e.g., if parallel processing)

%Function arguments to Main():

%pPerfUse: Probability of using the device perfectly

%pNoUse: Probability of not using the device at all

%overallCompliance: Overall proportion of person-time complying. $\text{overallCompliance} = \text{pPerfUse} + \text{pTreat} * (1 - \text{pPerfUse} - \text{pNoUse})$

%pTreat: Proportion of water treated if using the device imperfectly. Calculated from pPerfUse, pNoUse, & OverallCompliance.

%LRs: Log10 reductions attributable to the device: vector (bac., protozoa, viruses)

%PathogensLMin: During calibration, the minimum mean concentration of the 3 marker pathogens (bac., protozoa, viruses); usually [0 0 0].

%PathogensLMax: During calibration, the maximum mean concentration of the 3 marker pathogens (bac., protozoa, viruses); [0 0 0] for estimation.

%calibRuns: Number of calibration runs; 0 for estimation runs

%outFilename: Filename for storing output as a .CSV.

%inFilename: File containing parameter values from calibration (pathogen concentrations). Only matters for estimation phase.

%multConc: Multiplier for concentrations obtained from calibration. Used for estimation step to torture-test extreme concentrations.

%nSpikesY: Number of 1-day pathogen spikes (all 3 pathogens spike at once) per year. Set to 0 to turn them off.

%multSpikes: Defines size of spikes. Multiplier above mean baseline level.

%useAllParamSets: If 1, and if estimation runs are being done, use all available parameter sets from calibration (instead of a subsample of parameter sets).

%Note that more options are available to be set in the first few lines of this function; % using them as function arguments would have been cumbersome.

```
function [OutQMRAmatrix OutQMRA] = Main(pPerfUse, pNoUse, overallCompliance, LRs, PathogensLMin, PathogensLMax, calibRuns, outFilename, inFilename, multConc, nSpikesY, multSpikes, useAllParamSets); %TODO: Designate necessary output (overall measures & distributions, plus vals of 1st 3 params).
```

```

Octave = size(ver('Octave'),1); %Indicator of whether the code is running under Octave
(1) or Matlab (0).
StartTime = clock;
%====Setting simulation options====More simulation options are in GetTrialParams.m====
dailyVariation = 1; %Equals 1 if pathogen concentrations are allowed to vary by person by
day, instead of taking a single fixed value.
randomDurations = 1; %Equals 1 if illness durations are randomized instead of taking
a single fixed value. Renders Duration vector (below) mostly moot (it is still used to
choose start time of surveys).
storeStatus = 1; %If 1, store infection & disease status, to determine 'actual' (i.e.,
reported & unreported) burden of infection.
StoreAllDailyStatuses = 1; %If 1, store all daily statuses for all runs in a struct.
%====Housekeeping based on simulation options====
if calibRuns > 0;
    loops = calibRuns;
    noStochParams = 0;
else
    noStochParams = 1;
    GetTrialParams; %GetTrialParams.m reads in pathogens/L & background measure to
be tried. It contains its own options - check before running.
end
if loops <= 25; testing = 1; else testing = 0; end
if overallCompliance == 0 | pNoUse == 1; pTreat = 0; pPerfUse = 0; overallCompliance = 0;
pNoUse = 1; %Avoids x/0 error below
elseif overallCompliance == 1 | pPerfUse == 1; pTreat = 1; pPerfUse = 1;
overallCompliance = 1; pNoUse = 0; %Avoids x/0 error below
else pTreat = (overallCompliance - pPerfUse) / (1 - pPerfUse - pNoUse);
%Calculating pTreat so as to be able to hold overallCompliance constant over mult.
runs.
end
if pTreat < 0 - eps | pTreat > 1 + eps; error('pTreat of ',num2str(pTreat),' is
impossibly > 1 or < 0: check pNoUse, pPerfUse, & overallCompliance!'); end

```



```

%====Ending housekeeping. Starting parameter values:====
%Water concentration and disease parameter values
% Several vectors have 3 elements corresponding to our 3 pathogens of interest:
% [ETEC/EPEC, Giardia, rotavirus], i.e., bacteria, protozoa, viruses
%PathogensLMax = [2e5, 1.35, 0.18]; %Maximum pathogens/L; minimum is zero for
all. 2fold empirically observed levels that led to LP exceeding the 95% CI for the
placebo group, each pathogen taken individually.
pathogensLcv = sqrt(3407044) / 2509.329; %Coeff. of var. calc. from variance & mean
of 'cfbef' (Boisson 2010 water qual. data), high outliers (>= 30000 CFU) removed. For
calc. of scaled gamma dists.
MorbidityK = [0.214 , 0.59 , 0.397]; %Proportion of infected 'kids' (<5y) with
diarrhea.
Duration = [82.1/24, 18.3, 2.5]; Duration = round(Duration); %Duration of infection
(days). Although length & max of this vector are still used, actual infection duration
is determined by durEc.m, durGi.m, & durRo.m.
prevDiarrhBaseKidsMax = 0.0972; %Upper limit of non-waterborne reported diarrhea
prevalence.
LongPrevs = [0.103, 0.0896]; %Vector of raw long. prev. values from Boisson dataset,
kids w. placebo, then kids w. intervention).
prevDiarrhBaseKids = LongPrevs(2); %Baseline non-waterborne reported diarrhea prevalence.
No greater than observed in the RCT. Only used if no random variation (overwritten
otherwise).
ImmuneTimes = [7 7 7]; %Length of immune period for all pathogens.
%Parameters: population & exposure information. Model by household later.
%nKidsInt = 85; nKidsPla = 105; %Boisson 2010.
nKids = 100; %No longer simulating an intervention trial - simply a
set of counterfactuals for comparison.
drinkKids = 1.178; %drinkKidsSD = 0.186; %daily water intake, L/d, kids
%pTreat = 2/3; %Proportion of water treated, if device is being used.
Try 1, 2/3, & 1/3.
%if compliancel00 == 1; pUse = 1; pTreat = 1; end; %If perfect compliance is desired,
override above 2 lines.
%Parameters: device effectiveness information

```

```

%LRs = [6.9, 3.6, 4.7];          %Log reductions [bacteria, protozoa, viruses] by the
intervention device, with upper & lower ranges
%LRsPla = [1.05, 1.05, 1.05];    %As above, for 'placebo' device
%if badPlacebo == 0; LRsPla = [0 0 0]; end; %If a perfect placebo is being modeled,
override above line.
%Parameters: dose response, order as above [ETEC/EPEC, Giardia, rotavirus]
KorN50 = [2111912, 0.01982, 6.171]; %Exponential k parameter or beta-Poisson N50
parameter
alpha = [0.1549, NaN, 0.2531];    %Presence/absence of alpha value determines
beta-Poisson or exponential dose resp.
%Bias parameters
remembrance = 0.54; %Proportion of diarrhea episodes remembered (and reported) if they
ended >2d before being surveyed; assume perfect recall if episode is on day 0, 1, or 2
%Study parameters - relating to how the study was conducted
recallPeriod = 7; %Number of days in the past over which people were asked to remember
diarrheal episodes
interval = 31; %Interval between beginnings of recall periods. Must be 31 to avoid
undercounting a year as 360d.
nYears = 1; %For easy adjustment of the length of the simulation. Used later to
properly calculate incidence & LP.
nRecallPeriods = 12 * nYears; %Number of recall periods (i.e., number of simulated
diarrhea surveys)
daysBurnIn = ceil(max(ImmuneTimes) + max(Duration) + recallPeriod) * 4; %Days required
for prevalence to reach equilibrium (simulation starts with nobody infected). Allows
ample margin for reaching equilibrium.
%====Ending parameter values====

maxTime = daysBurnIn + (nRecallPeriods-1)*interval; %Time over which to run each
simulation.
%Creating output structure for storing results from main QMRA loop
OutQMRA =
struct('StartTime',StartTime,'Fit',NaN(1,loops),'KILP',NaN(1,loops),'KPLP',NaN(1,loops),'
LPR',NaN(1,loops),...

```

```

'EcL',NaN(1,loops),'GiL',NaN(1,loops),'RoL',NaN(1,loops),'PrevBase',NaN(1,loops));
%'Fit' is no longer used.
if storeStatus == 1; %Optionally creating cell array for more detailed output. Not
needed for calibration step.
    DailyStatus = cell(1, loops); %Each cell in the array needs to contain a matrix
(rows are days, columns are variables).
    DailyStatus(:) = {NaN(maxTime,22)}; %1st row for intervention, 2nd row for
placebo.
    %Columns are: 1:3 = new inf. Ec,Gi,&Ro; 4:6 = new cases; 7:14 = num. inf.
(0,Ec,Gi,Ro,EcGi,EcRo,GiRo,EcGiRo); 15:22 = as inf., but ill.
end;
tic %Starts timer
if noStochParams == 1; loops = size(TrialParams,1); end %Resets loops if a series of
trial pathogens/L values is being used.
MeanDoses = zeros(loops,size(MorbidityK,2)); %Prepopulating a matrix for storing mean
doses (helps in checking whether spikes are working properly).
for i = 1:loops; %====Starting main QMRA loop.==== Loops once for each QMRA run. i
indexes each loop.
    %====Randomly generating parameters for this iteration====
    switch(noStochParams);
        case 0;
            PathogensLmeans = rand(1,length(MorbidityK)) .* (PathogensLMax - PathogensLMin)
+ PathogensLMin; %Uniform sampling of the mean value for each pathogen.
            prevDiarrhBaseKids = rand(1,1)*prevDiarrhBaseKidsMax;
        case 1; %Pulling parameters from previous runs consistent with RCT.
            PathogensLmeans = TrialParams(i,1:3) * multConc; %Allows scaling up of dose
values obtained through calibration.
            if size(TrialParams,2) <= 3;
                prevDiarrhBaseKids = 0;
            else
                prevDiarrhBaseKids = TrialParams(i,4);
            end
        otherwise

```

```

        error('noStochParams must be 0 or 1');
    end
    if nSpikesY > 0;    %Adjusting the baseline mean downward if spikes are used, so
that overall mean concentrations remain the same.
        nSpikes = nSpikesY * nYears;
        simTime = maxTime - daysBurnIn;
        PathogensLmeansBase = PathogensLmeans * simTime/(nSpikes*multSpikes+simTime);
        SpikeTimes = randperm(simTime);
        SpikeTimes = SpikeTimes(1:nSpikes);
        SpikeTimes = sort(SpikeTimes + daysBurnIn); %Preallocating the times for the
spikes.
        SpikeTimes(end + 1) = 0; %Needed to avoid an error when spike counter advances
past the last spike.
        nextSpike = 1; %Counter for working through the list of spikes.
    else PathogensLmeansBase = PathogensLmeans;
    end
    OutQMRA.EcL(i)=PathogensLmeans(1); OutQMRA.GiL(i)=PathogensLmeans(2);
    OutQMRA.RoL(i)=PathogensLmeans(3);
    OutQMRA.PrevBase(i) = prevDiarrhBaseKids;    %This line & previous store the varying
parameter values.
    %====End random parameter generation - start setup of values/vectors/matrices used
throughout simulation====
    %Computing daily doses of pathogens ingested in drinking water, using water drunk
per day and log reduction values
    %Computing parameter values for gamma distribution of pathogens in water
    %Scales = (pathogensLcv * PathogensLmeansBase).^2 ./ PathogensLmeans; %This seems
the same as next line.
    Scales = pathogensLcv ^2 * PathogensLmeansBase;
    %Shapes = PathogensLmeansBase ./ Scales;    %This seems the same as next line.
    Shapes = 1 / pathogensLcv ^ 2; Shapes = [Shapes Shapes Shapes]; %Shape parameter is
identical for all 3 pathogen types.
    %Assigning infections randomly based on responses, assuming infections with
different pathogens are independent.

```

```

%   A person can have only 1 infection per pathogen.
OutPrevs = struct('Kids',NaN(1,nRecallPeriods)); %Creating struct to hold
prevalences from surveys.
%Creating person-pathogen matrices; everybody starts infected for a random amount of
time, to reduce periodicity from constant disease duration.
KidsPPM = rand(nKids,length(Duration)) * 2*(max(Duration)+max(ImmuneTimes));
%KidsPla = rand(nKidsPla,length(Duration)) * 2*(max(Duration)+max(ImmuneTimes));
if storeStatus == 1; %Optionally, making corresponding matrices to store disease
info.
    KidsPPMD = ones(size(KidsPPM)); %NaN means never infected, 0 means
uninfected, 1 means infected, 2 means diseased.
    %KidsPlaD = ones(size(KidsPla)); %Note that everyone starts off infected with
everything, just as with KidsInt & KidsPla above.
end
OutputFields = fieldnames(OutPrevs);
if Octave == 1; fflush(stdout); end; %Forces a write to screen so that
sim progress can be seen.
%=====Code for testing purposes only=====
if testing == 1 && loops <= 15; %Only runs when testing code. These vectors
needed for charting infection prevalence over the entire simulation.
    KidsInfPrev = NaN(maxTime,1); %Creating a vector to hold infection prevalence
info, intervention group.
    %KidsPlaInfPrev = NaN(nKidsPla,1); %As above, placebo group.
end;
%=====End code for testing purposes=====
%Now storing all person-pathogen matrices, but only if exactly 1 loop is requested.
This repeats at the end of each day.
if loops == 1;
    PPmatrices(1).KidsPPM = KidsPPM;
    PPmatrices(1).KidsPPMD = KidsPPMD;
end;
%=====Begin daily loop, t indexes the days=====
for t = 1:maxTime;

```

```

KidsPPM = KidsPPM - 1; %Note: a value of 0 signifies the first day of
immunity.
%KidsPla = KidsPla - 1;
if storeStatus == 1; %Optionally, tracking recovery from infection/disease.
    KidsPPMD(KidsPPM <= 0) = 0;
    %KidsPlaD(find(KidsPla <= 0)) = 0;
end
%Computing doses in untreated water, varying for each child, each day.
Dose.Kids = NaN(nKids,length(Duration)); %Matrix of doses per child
(intervention). Columns are pathogens.
%Dose.KidsPla = NaN(nKidsPla,length(Duration)); %Matrix of doses per child
(placebo). Columns are pathogens.
RandComp = rand(nKids,1); %Random numbers for determining compliance (i.e.,
use of device).

    for j = 1:length(Duration); %Looping over pathogens to determine
daily doses for ea. person.
        if isnan(Shapes(j)) == 1; %If mean pathogen conc. is set to 0,
pathogen conc. is always 0.
            Dose.Kids(:,j) = 0;
            %Dose.KidsPla(:,j) = 0;
        else
            if dailyVariation == 1;
                Dose.Kids(:,j) = gamrnd(Shapes(j),Scales(j),[nKids,1]) *
drinkKids; %Initial untreated dose
                %Dose.KidsPla(:,j) = gamrnd(Shapes(j),Scales(j),[nKidsPla,1]) *
drinkKids;
            else
                Dose.Kids(:,j) = PathogensLmeansBase(j) * drinkKids; %Dose
becomes the mean dose if daily variation is turned off.
                %Dose.KidsPla(:,j) = PathogensLmeansBase(j) * drinkKids;
            end
        end
    end
end

```

```

        %Next 2 lines: Determining who complies and has a nonzero dose (because
log reduction would fail on zero dose).
        Comp.Kids.Perf = find((RandComp(1:nKids,1) < pPerfUse) &
(Dose.Kids(:,j) > 0));
        Comp.Kids.Imperf = find((RandComp(1:nKids,1) > (pPerfUse+pNoUse)) &
(Dose.Kids(:,j) > 0));
        %Now applying LRs, if using device. Includes adjustment for partial
treatment of water (pTreat) for imperfect compliers.
        Dose.Kids(Comp.Kids.Perf,j) = 10.^(log10(Dose.Kids(Comp.Kids.Perf,j)) -
LRs(j));
        Dose.Kids(Comp.Kids.Imperf,j) = 10.^(log10(Dose.Kids(Comp.Kids.Imperf,j) *
pTreat) - LRs(j)) + Dose.Kids(Comp.Kids.Imperf,j) * (1 - pTreat);
    end

    if nSpikesY > 0 && t == SpikeTimes(nextSpike);
        Dose.Kids = Dose.Kids * multSpikes;
        %disp(['Time ',num2str(t),' , means = ',num2str(mean(Dose.Kids))])
%This line for testing only
        nextSpike = nextSpike + 1;
    end
    MeanDoses(t,:) = mean(Dose.Kids);

    %Computing responses (diarrheal illness) using custom functions DRexp() and
DRbP().
    Responses.Kids(:,1) = DRbP(KorN50(1),alpha(1),Dose.Kids(:,1)); %Note:
response matrices correspond to the person-path. matrices.
    Responses.Kids(:,2) = DRexp(KorN50(2),Dose.Kids(:,2));
    Responses.Kids(:,3) = DRbP(KorN50(3),alpha(3),Dose.Kids(:,3));
    %Responses.KidsPla(:,1) = DRbP(KorN50(1),alpha(1),Dose.KidsPla(:,1));
    %Responses.KidsPla(:,2) = DRexp(KorN50(2),Dose.KidsPla(:,2));
    %Responses.KidsPla(:,3) = DRbP(KorN50(3),alpha(3),Dose.KidsPla(:,3));
    switch(randomDurations);
        case 1

```

```

        switch(storeStatus);
            case 1
                %DailyStatus columns: 1:3 = new inf. Ec,Gi,&Ro; 4:6 = new cases; 7:14 = num. inf.
                (0, Ec, Gi, Ro, EcGi, EcRo, GiRo, EcGiRo); 15:22 = as inf., but ill.
                KidsPPMDold = KidsPPMD;
                %KidsPlaDold = KidsPlaD;
                [KidsPPM, KidsPPMD] =
AssignInfIllRand(KidsPPM, KidsPPMD, Responses.Kids, ImmuneTimes, MorbidityK);
                %[KidsPla, KidsPlaD] =
AssignInfIllRand(KidsPla, KidsPlaD, Responses.KidsPla, ImmuneTimes, MorbidityK);
                for s = 1:6; %Store counts of new infections & illnesses.
                    if s <= 3; %Infections:
                        %DailyStatus{1,i}(t,s) =
length(intersect(find(KidsPPMDold(:,s) == 0), find(KidsPPMD(:,s) > 0)));
                        DailyStatus{1,i}(t,s) = sum((KidsPPMDold(:,s) == 0) &
(KidsPPMD(:,s) > 0)); %Should run much faster than above.
                        %DailyStatus{2,i}(t,s) =
length(intersect(find(KidsPlaDold(:,s) == 0), find(KidsPlaD(:,s) > 0)));
                    else %Illnesses:
                        %DailyStatus{1,i}(t,s) =
length(intersect(find(KidsPPMDold(:,s-3) == 0), find(KidsPPMD(:,s-3) == 2)));
                        DailyStatus{1,i}(t,s) = sum((KidsPPMDold(:,s-3) == 0) &
(KidsPPMD(:,s-3) == 2)); %Should run much faster than above.
                        %DailyStatus{2,i}(t,s) =
length(intersect(find(KidsPlaDold(:,s-3) == 0), find(KidsPlaD(:,s-3) == 2)));
                    end
                end
            otherwise
                KidsPPM =
AssignInfRand(KidsPPM, Responses.Kids, ImmuneTimes);
                %KidsPla =
AssignInfRand(KidsPla, Responses.KidsPla, ImmuneTimes);
            end
    end

```



```

        otherwise
            KidsPPM = AssignInf(KidsPPM,Responses.Kids,Duration,ImmuneTimes);
            %KidsPla = AssignInf(KidsPla,Responses.KidsPla,Duration,ImmuneTimes);
        end

        if t >= daysBurnIn && mod(t-(daysBurnIn),interval) == 0; %Obtaining results
from diarrhea assessment survey.
%Determining reported diarrhea-weeks. CalcDiarrhWeeks() uses the Person-Pathogen
Matrices, morbidity ratios, and recall of diarrhea episodes to determine if a week was
reported as a week with diarrhea.
            KidsRD = CalcDiarrhWeeks(KidsPPM,remembrance,recallPeriod,MorbidityK);
            %Whether diarrhea was reported by each particular person. Note age (last column) is
removed from the person-pathogen matrix when inputted to the function.
            OutPrevs.Kids((t-daysBurnIn)/interval+1) = sum(KidsRD)/length(KidsRD);
            %Getting prevalence for each diarrhea survey
            %KidsPlaRD = CalcDiarrhWeeks(KidsPla,remembrance,recallPeriod,MorbidityK);
            %Like above 2 lines, but kid placebo
            %OutPrevs.KidsPla((t-daysBurnIn)/interval+1) =
sum(KidsPlaRD)/length(KidsPlaRD);
            %fprintf(1,['d',num2str(t),'/',num2str(maxTime),'|']); %Progress counter.
            %if Octave == 1; fflush(stdout); end; %Forces a write to screen.
        end
        if testing == 1 && loops <= 15; %====Testing code====
            KidsInfPrev(t) = sum(max(KidsPPM') > 0) / nKids;%If max value over all
pathogens >0, then there is an infection.
            %KidsPlaInfPrev(t) = sum(max(KidsPla') > 0) / nKidsPla;%Transposing so
that max is for ea. row instead of ea. column.
            KidsInfEc(t) = sum(KidsPPM(:,1) > 0) / nKids;
            KidsInfGi(t) = sum(KidsPPM(:,2) > 0) / nKids;
            KidsInfRo(t) = sum(KidsPPM(:,3) > 0) / nKids;
            %KidsPlaInfEc(t) = sum(KidsPla(:,1) > 0) / nKidsPla;
            %KidsPlaInfGi(t) = sum(KidsPla(:,2) > 0) / nKidsPla;
            %KidsPlaInfRo(t) = sum(KidsPla(:,3) > 0) / nKidsPla;

```

```

end          %====End testing code====
if storeStatus == 1;      %Storing lots of daily information about the model
run.
    %Columns are: 1:3 = new inf. Ec,Gi,&Ro; 4:6 = new cases; 7:14 = num. inf.
    (0,Ec,Gi,Ro,EcGi,EcRo,GiRo,EcGiRo); 15:22 = as inf., but ill.
        for s = 7:22;          %All based on 0=uninfected, 1=asymptomatic,
2=ill.
            if s == 7;          %Tallying completely uninfected people
                %DailyStatus{1,i}(t,s) = length(find(sum(KidsPPMD') == 0));
                DailyStatus{1,i}(t,s) = sum(sum(KidsPPMD,2) == 0);    %Should be
faster than above.
                    %DailyStatus{2,i}(t,s) = length(find(sum(KidsPlaD') == 0));
                elseif s == 8;      %Infected with only Ec
                    %DailyStatus{1,i}(t,s) = length(intersect(find(KidsPPMD(:,1) >
0), find(sum([KidsPPMD(:,2) KidsPPMD(:,3)]') == 0)')));
                    DailyStatus{1,i}(t,s) = sum((KidsPPMD(:,1) > 0) & (sum([KidsPPMD(:,2)
KidsPPMD(:,3)],2) == 0));    %Should be faster than above.
                    %DailyStatus{2,i}(t,s) = length(intersect(find(KidsPlaD(:,1) >
0), find(sum([KidsPlaD(:,2) KidsPlaD(:,3)]') == 0)')));
                elseif s == 9;      %Infected with only Gi
                    %DailyStatus{1,i}(t,s) = length(intersect(find(KidsPPMD(:,2) >
0), find(sum([KidsPPMD(:,1) KidsPPMD(:,3)]') == 0)')));
                    DailyStatus{1,i}(t,s) = sum((KidsPPMD(:,2) > 0) & (sum([KidsPPMD(:,1)
KidsPPMD(:,3)],2) == 0));    %Should be faster than above.
                    %DailyStatus{2,i}(t,s) = length(intersect(find(KidsPlaD(:,2) >
0), find(sum([KidsPlaD(:,1) KidsPlaD(:,3)]') == 0)')));
                elseif s == 10;     %Infected with only Ro
                    %DailyStatus{1,i}(t,s) = length(intersect(find(KidsPPMD(:,3) >
0), find(sum([KidsPPMD(:,1) KidsPPMD(:,2)]') == 0)')));
                    DailyStatus{1,i}(t,s) = sum((KidsPPMD(:,3) > 0) & (sum([KidsPPMD(:,1)
KidsPPMD(:,2)],2) == 0));    %Should be faster than above.
                    %DailyStatus{2,i}(t,s) = length(intersect(find(KidsPlaD(:,3) >
0), find(sum([KidsPlaD(:,1) KidsPlaD(:,2)]') == 0)')));

```

```

elseif s == 11;    %Tallying infected with Ec & Gi only
    %DailyStatus{1,i}(t,s) =
length(intersect(find(prod(KidsPPMD(:,1:2)') ~= 0), find(KidsPPMD(:,3) == 0)));
    DailyStatus{1,i}(t,s) = sum((prod(KidsPPMD(:,1:2),2) ~= 0) &
(KidsPPMD(:,3) == 0));    %Should be faster than above.
    %DailyStatus{2,i}(t,s) =
length(intersect(find(prod(KidsPlaD(:,1:2)') ~= 0), find(KidsPlaD(:,3) == 0)));
elseif s == 12;    %Tallying infected with Ec & Ro only
    %DailyStatus{1,i}(t,s) =
length(intersect(find(prod([KidsPPMD(:,1) KidsPPMD(:,3)]') ~= 0), find(KidsPPMD(:,2) ==
0)));
    DailyStatus{1,i}(t,s) = sum((prod([KidsPPMD(:,1) KidsPPMD(:,3)],2) ~=
0) & (KidsPPMD(:,2) == 0));    %Should be faster than above.
    %DailyStatus{2,i}(t,s) =
length(intersect(find(prod([KidsPlaD(:,1) KidsPlaD(:,3)]') ~= 0), find(KidsPlaD(:,2) ==
0)));
elseif s == 13;    %Tallying infected with Gi & Ro only
    %DailyStatus{1,i}(t,s) =
length(intersect(find(prod(KidsPPMD(:,2:3)') ~= 0), find(KidsPPMD(:,1) == 0)));
    DailyStatus{1,i}(t,s) = sum((prod(KidsPPMD(:,2:3),2) ~= 0) &
(KidsPPMD(:,1) == 0));    %Should be faster than above.
    %DailyStatus{2,i}(t,s) =
length(intersect(find(prod(KidsPlaD(:,2:3)') ~= 0), find(KidsPlaD(:,1) == 0)));
elseif s == 14;    %Tallying infected with Ec & Gi & Ro
    %DailyStatus{1,i}(t,s) = length(find(prod(KidsPPMD(:,1:3)') ~=
0));
    DailyStatus{1,i}(t,s) = sum(prod(KidsPPMD(:,1:3),2) ~= 0);    %Should
be faster than above.
    %DailyStatus{2,i}(t,s) = length(find(prod(KidsPlaD(:,1:3)') ~=
0));
elseif s == 15;    %Tallying completely non-ill people
    %DailyStatus{1,i}(t,s) = length(find(sum(KidsPPMD' .^2) <= 3));
    DailyStatus{1,i}(t,s) = sum(sum(KidsPPMD' .^2) <= 3);    %Should be

```

faster than above.

```
        %DailyStatus{2,i}(t,s) = length(find(sum(KidsPlaD' .^2) <= 3));
elseif s == 16;           %Ill with only Ec
        %DailyStatus{1,i}(t,s) = length(intersect(find(KidsPPMD(:,1) ==
2), find(sum(KidsPPMD(:,2:3)' .^2) <= 2)'));
        DailyStatus{1,i}(t,s) = sum((KidsPPMD(:,1) == 2) &
(sum(KidsPPMD(:,2:3)' .^2)' <= 2)); %Should be faster than above.
        %DailyStatus{2,i}(t,s) = length(intersect(find(KidsPlaD(:,1) ==
2), find(sum(KidsPlaD(:,2:3)' .^2) <= 2)'));
elseif s == 17;           %Ill with only Gi
        %DailyStatus{1,i}(t,s) = length(intersect(find(KidsPPMD(:,2) ==
2), find(sum([KidsPPMD(:,1) KidsPPMD(:,3)]' .^2) <= 2)'));
        DailyStatus{1,i}(t,s) = sum((KidsPPMD(:,2) == 2) &
(sum([KidsPPMD(:,1) KidsPPMD(:,3)]' .^2)' <= 2)); %Should be faster than above.
        %DailyStatus{2,i}(t,s) = length(intersect(find(KidsPlaD(:,2) ==
2), find(sum([KidsPlaD(:,1) KidsPlaD(:,3)]' .^2) <= 2)'));
elseif s == 18;           %Ill with only Ro
        %DailyStatus{1,i}(t,s) = length(intersect(find(KidsPPMD(:,3) ==
2), find(sum(KidsPPMD(:,1:2)' .^2) <= 2)'));
        DailyStatus{1,i}(t,s) = sum((KidsPPMD(:,3) == 2) &
(sum(KidsPPMD(:,1:2)' .^2)' <= 2)); %Should be faster than above.
        %DailyStatus{2,i}(t,s) = length(intersect(find(KidsPlaD(:,3) ==
2), find(sum(KidsPlaD(:,1:2)' .^2) <= 2)'));
elseif s == 19;           %Tallying ill with Ec & Gi only
        %DailyStatus{1,i}(t,s) =
length(intersect(find(prod(KidsPPMD(:,1:2)') == 4), find(KidsPPMD(:,3) <= 1)));
        DailyStatus{1,i}(t,s) = sum((prod(KidsPPMD(:,1:2),2) == 4) &
(KidsPPMD(:,3) <= 1)); %Should be faster than above.
        %DailyStatus{2,i}(t,s) =
length(intersect(find(prod(KidsPlaD(:,1:2)') == 4), find(KidsPlaD(:,3) <= 1)));
elseif s == 20;           %Tallying ill with Ec & Ro only
        %DailyStatus{1,i}(t,s) =
length(intersect(find(prod([KidsPPMD(:,1) KidsPPMD(:,3)]') == 4), find(KidsPPMD(:,2) <=
```

```

1)));
        DailyStatus{1,i}(t,s) = sum((prod([KidsPPMD(:,1) KidsPPMD(:,3)],2) ==
4) & (KidsPPMD(:,2) <= 1)); %Should be faster than above.
        %DailyStatus{2,i}(t,s) =
length(intersect(find(prod([KidsPlaD(:,1) KidsPlaD(:,3)]) == 4), find(KidsPlaD(:,2) <=
1)));
        elseif s == 21; %Tallying ill with Gi & Ro only
        %DailyStatus{1,i}(t,s) =
length(intersect(find(prod(KidsPPMD(:,2:3)') == 4), find(KidsPPMD(:,1) <= 1)));
        DailyStatus{1,i}(t,s) = sum((prod(KidsPPMD(:,2:3),2) == 4) &
(KidsPPMD(:,1) <= 1)); %Should be faster than above.
        %DailyStatus{2,i}(t,s) =
length(intersect(find(prod(KidsPlaD(:,2:3)') == 4), find(KidsPlaD(:,1) <= 1)));
        elseif s == 22; %Tallying ill with Ec & Gi & Ro
        %DailyStatus{1,i}(t,s) = length(find(prod(KidsPPMD(:,1:3)') ==
8));
        DailyStatus{1,i}(t,s) = sum(prod(KidsPPMD(:,1:3),2) == 8); %Should
be faster than above.
        %DailyStatus{2,i}(t,s) = length(find(prod(KidsPlaD(:,1:3)') ==
8));
        end
    end
end
%Now storing the person-pathogen matrices at the end of this day, if exactly 1
loop was requested.
if loops == 1;
    PPmatrices(t+1).KidsPPM = KidsPPM;
    PPmatrices(t+1).KidsPPMD = KidsPPMD;
end
%sizeof(DailyStatus) %Debug measure - too big for memory?
if nSpikesY > 0;
    DailyStatus{2,i} = SpikeTimes; %Storing the spike times
    DailyStatus{3,i} = daysBurnIn; %Storing the equilibration duration

```

```

(for debugging)
    end
end
end
%disp([' ']) %Adds line feed to separate the progress counters.
%=====End daily loop, begin more testing code=====
if testing == 1 && loops <= 15;
    maxPrevTime = find(KidsInfPrev == max(KidsInfPrev)); %Getting time points of
max. prevalence of infection (1st, if tie).
    disp(['1st time point where max. prevalence is seen is
',num2str(maxPrevTime(1))]) %Printing first point of max. prevalence.
    figure(i); %Plotting infection prevalence.
    %set(f1, 'Position', [5 5 1024 768]);
    subplot(2,2,1);
    plot([1:maxTime]',KidsInfPrev,'-k');
    title 'Daily infection prevalence; Xs = reported waterborne prevalence';
    ylabel 'Proportion affected'; xlabel 'Time';
    %xlim([0 625]);
    hold on;
    plot([1:maxTime],KidsInfEc,'-r');
    plot([1:maxTime],KidsInfGi,'-g');
    plot([1:maxTime],KidsInfRo,'-b');
    h5 = plot([daysBurnIn:interval:maxTime],OutPrevs.Kids,'cx');
    set(h5,'linewidth',2);
    legend({'\fontsize{10} Any infection}','\fontsize{10} E. coli
infection}','\fontsize{10} Giardia infection}','\fontsize{10} Rotavirus
infection}','\fontsize{10} LP_{Irwd} (prior week)');
    hold off;
    subplot(2,2,2);
    semilogy([1:maxTime],MeanDoses(:,1),'-r');
    title 'Daily mean dose of pathogens in drinking water (untreated)';
    ylabel 'Daily mean dose'; xlabel 'Time';
    %xlim([0 625]);
    hold on;

```

```

semilogy([1:maxTime],MeanDoses(:,2),'-g');
semilogy([1:maxTime],MeanDoses(:,3),'-b');
legend('\fontsize{10} E. coli','\fontsize{10} Giardia','\fontsize{10}
Rotavirus');
hold off;
subplot(2,2,3);
plot([1:maxTime],sum(DailyStatus{1,i}(:,1:3),2),'-k');
title 'Daily infection incidence (new infections each day)';
ylabel 'New infections'; xlabel 'Time';
%xlim([0 625]);
%ylim([0 nKids]);
hold on;
plot([1:maxTime],DailyStatus{1,i}(:,1),'-r');
plot([1:maxTime],DailyStatus{1,i}(:,2),'-g');
plot([1:maxTime],DailyStatus{1,i}(:,3),'-b');
legend('\fontsize{10} Total new infections','\fontsize{10} E. coli
infections','\fontsize{10} Giardia infections','\fontsize{10} Rotavirus
infections');
hold off;
subplot(2,2,4);
plot([1:maxTime],sum(DailyStatus{1,i}(:,4:6),2),'-k');
title 'Daily illness incidence (new illnesses each day)';
ylabel 'New illnesses'; xlabel 'Time';
%xlim([0 625]);
%ylim([0 nKids]);
hold on;
plot([1:maxTime],DailyStatus{1,i}(:,4),'-r');
plot([1:maxTime],DailyStatus{1,i}(:,5),'-g');
plot([1:maxTime],DailyStatus{1,i}(:,6),'-b');
legend('\fontsize{10} Total new illnesses','\fontsize{10} E. coli
illnesses','\fontsize{10} Giardia illnesses','\fontsize{10} Rotavirus illnesses');
hold off;
if nSpikesY > 0;

```

```

        disp([num2str(sum(sum(DailyStatus{1,i}(SpikeTimes(1:end-
1),1:3)))/sum(sum(DailyStatus{1,i}(daysBurnIn:end,1:3)))),' of infections due to spikes
(not including equilibration).']);
    end
end %=====End testing code=====
%Adding baseline prevalence. Corrected for doublecounting (some infected people
would have been infected anyway by baseline transmission).
OutPrevs.Kids = OutPrevs.Kids + prevDiarrhBaseKids * (1 - OutPrevs.Kids);
%OutPrevs.KidsPla = OutPrevs.KidsPla + prevDiarrhBaseKids * (1 - OutPrevs.KidsPla);
%i %Printing i as a debug measure
OutQMRA.rLP(i) = mean(OutPrevs.Kids); %Output long. prev. of reported diarrhea.
%OutQMRA.LPR(i) = OutQMRA.KILP(i) / OutQMRA.KPLP(i);
if storeStatus == 1; %Storing incidences.
    OutQMRA.IncInfEc(i) = sum(DailyStatus{1,i}(maxTime-(nYears*365)+1:maxTime,1));
%Count of infections with E. coli.
    OutQMRA.IncIllEc(i) = sum(DailyStatus{1,i}(maxTime-(nYears*365)+1:maxTime,4));
%Count of illnesses with E. coli.
    OutQMRA.IncInfGi(i) = sum(DailyStatus{1,i}(maxTime-(nYears*365)+1:maxTime,2));
%Count of infections with for Giardia.
    OutQMRA.IncIllGi(i) = sum(DailyStatus{1,i}(maxTime-(nYears*365)+1:maxTime,5));
%Count of illnesses with Giardia.
    OutQMRA.IncInfRo(i) = sum(DailyStatus{1,i}(maxTime-(nYears*365)+1:maxTime,3));
%Count of infections with for rotavirus.
    OutQMRA.IncIllRo(i) = sum(DailyStatus{1,i}(maxTime-(nYears*365)+1:maxTime,6));
%Count of illnesses with rotavirus.
    %'Actual' longitudinal prevalences (person-days ill or infected, divided by
total person-days observed).
    OutQMRA.LPInfEc(i) = sum(sum(DailyStatus{1,i}(maxTime-(nYears*365)+1:maxTime,
[ 8 11 12 14])))/(nKids*nYears*365); %Yearly inf. LP, E. coli.
    OutQMRA.LPILLec(i) = sum(sum(DailyStatus{1,i}(maxTime-(nYears*365)+1:maxTime,
[16 19 20 22])))/(nKids*nYears*365); %Yearly ill LP, E. coli.
    OutQMRA.LPInfGi(i) = sum(sum(DailyStatus{1,i}(maxTime-(nYears*365)+1:maxTime,
[ 9 11 13 14])))/(nKids*nYears*365); %Yearly inf. LP, Giardia.

```



```

        OutQMRA.LPIllGi(i) = sum(sum(DailyStatus{1,i}(maxTime-(nYears*365)+1:maxTime,
[17 19 21 22])))/(nKids*nYears*365);    %Yearly ill LP, Giardia.
        OutQMRA.LPInfRo(i) = sum(sum(DailyStatus{1,i}(maxTime-(nYears*365)+1:maxTime,
[10 12 13 14])))/(nKids*nYears*365);    %Yearly inf. LP, rota.
        OutQMRA.LPIllRo(i) = sum(sum(DailyStatus{1,i}(maxTime-(nYears*365)+1:maxTime,
[18 20 21 22])))/(nKids*nYears*365);    %Yearly ill LP, rota.
        OutQMRA.LPInfMix(i) = sum(sum(DailyStatus{1,i}(maxTime-
(nYears*365)+1:maxTime,11:14)))/(nKids*nYears*365);    %Yearly inf. LP, mixed.
        OutQMRA.LPIllMix(i) = sum(sum(DailyStatus{1,i}(maxTime-
(nYears*365)+1:maxTime,19:22)))/(nKids*nYears*365);    %Yearly ill LP, mixed.
        OutQMRA.LPInfAny(i) = sum(sum(DailyStatus{1,i}(maxTime-
(nYears*365)+1:maxTime,8:14)))/(nKids*nYears*365);    %Yearly inf. LP, any.
        OutQMRA.LPIllAny(i) = sum(sum(DailyStatus{1,i}(maxTime-
(nYears*365)+1:maxTime,16:22)))/(nKids*nYears*365);    %Yearly ill LP, any.
        %Mean daily pathogens per liter.
        OutQMRA.PathLMeanEc(i) = PathogensLmeans(1);
        OutQMRA.PathLMeanGi(i) = PathogensLmeans(2);
        OutQMRA.PathLMeanRo(i) = PathogensLmeans(3);
    end

        %fprintf(1,['d',num2str(t),'/',num2str(maxTime),'|']); %Progress counter.

        if i == 1; fprintf(1,['Finished loops: ',num2str(i)]); else; fprintf(1,
['-',num2str(i)]); end;
        if mod(i,floor(loops/10)) == 0;    %Progress meter
            disp(['Loop ',num2str(i),'/',num2str(loops),'. Done in ',num2str((toc/i) *
(loops-i) / 60 / 60),'h.'])
            toc
        end
        if i == 5; disp([' ']); disp(['=== Will finish ',num2str(loops),' loops in ~
',num2str(toc*loops/5/60/60),' h ===']); end;
end %=====Ending main QMRA loop=====
%Converting to a matrix so as to concatenate output from multiple calls, and subsequent

```

```

easy output to .CSV for analysis in R.
OutQMRAMatrix = NaN(loops,14);
OutQMRAMatrix(:,1) = 1:1:loops; %ID designation.
OutQMRAMatrix(:,2) = pPerfUse;
OutQMRAMatrix(:,3) = pNoUse;
OutQMRAMatrix(:,4) = pTreat;
OutQMRAMatrix(:,5) = overallCompliance;
OutQMRAMatrix(:,6) = LRs(1);
OutQMRAMatrix(:,7) = LRs(2);
OutQMRAMatrix(:,8) = LRs(3);
OutQMRAMatrix(:,9) = OutQMRA.PathLMeanEc';
OutQMRAMatrix(:,10) = OutQMRA.PathLMeanGi';
OutQMRAMatrix(:,11) = OutQMRA.PathLMeanRo';
OutQMRAMatrix(:,12) = OutQMRA.LPInfAny';
OutQMRAMatrix(:,13) = OutQMRA.LPIllAny';
OutQMRAMatrix(:,14) = OutQMRA.IncIllEc';
OutQMRAMatrix(:,15) = OutQMRA.IncIllGi';
OutQMRAMatrix(:,16) = OutQMRA.IncIllRo';
OutQMRAMatrix(:,17) = sum(OutQMRAMatrix(:,14:16)')'; %Creating the total
incidence column, IncIllTot.
OutQMRAMatrix(:,18) = OutQMRAMatrix(:,14) ./ OutQMRAMatrix(:,17); %pIncIllEc,
proportion of illness from E. coli.
OutQMRAMatrix(:,19) = OutQMRAMatrix(:,15) ./ OutQMRAMatrix(:,17); %pIncIllGi,
proportion of illness from Giardia.
OutQMRAMatrix(:,20) = OutQMRAMatrix(:,16) ./ OutQMRAMatrix(:,17); %pIncIllRo,
proportion of illness from rotavirus.
OutQMRAMatrix(:,21) = OutQMRAMatrix(:,17) / (nKids * nYears); %IncIllECY,
incidence in episodes per child per year.

%Write the results of a calibration to a file.
dlmwrite(outFilename, OutQMRAMatrix, '-append');
disp([' Results written to ',outFilename, '.']);

```

```

disp(['Mean path. concs. (Ec, Gi,
Ro): ', num2str(mean(MeanDoses(daysBurnIn+1:maxTime,:)))]); %Checking pathogen
concentrations.

%eval(['save Results/OutQMRA', datestr(StartTime,30), '.mat, OutQMRA;']) %Saving output
file.
if noStochParams == 1 && loops <= 15; %Displaying results if a series of input
parameters were tested.
    out = [OutQMRA.rLP; OutQMRA.IncIllEc; OutQMRA.IncIllGi; OutQMRA.IncIllRo];
    disp('rLP, IncIllEc, IncIllGi, IncIllRo, IncIllAll, pIncEc, pIncGi, pIncRo')
    out(5,:) = sum(out(2:4,:));
    out(6,:) = out(2,:) ./ out(5,:);
    out(7,:) = out(3,:) ./ out(5,:);
    out(8,:) = out(4,:) ./ out(5,:);
    %out = out'
    %min(out)
    %mean(out)
    %max(out)
end

if StoreAllDailyStatuses == 1; %If storing everything, overwrite OutQMRA with it, so
it can be used within Octave.
    OutQMRA = DailyStatus;
end

end %End function Main().

```

```
%Octave script to obtain proportion of infections that are from spikes.
%Must set StoreAllDailyStatuses == 1 in Main.m. This analyzes the OutQMRA cell array
that's output by Main.m.
```

```
%{
COPYRIGHT INFORMATION
```

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (AnalyzeInfectionFromSpikes.m) is part of QMRA2v5.

QMRA2v5 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRA2v5 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRA2v5. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
1;
function Output = pS(mix,inc,sH);
    [OutQMRAmatrix OutQMRA] = Main(0,1,0,[0 0 0],[0 0 0],[0 0
0],0,'TestpSpikes.csv',char(['RTF_',mix,'_TrialCalibResults',inc,'5spikesx',num2str(sH),'
.csv']),1,5,sH,0);
    startingDay = 128; %From line ~75 of Main.m (burn-in).
    nRuns = size(OutQMRA,2);
    nDays = size(OutQMRA{1,1},1);
```

```

    output = zeros(nRuns,3); %1st column is all new infections, 2nd is new infections
from spike days, 3rd is new infections from non-spike days.
    for i = 1:nRuns;
        output(i,1) = sum(sum(OutQMRA{1,i}(startingDay:end,1:3),2)); %Row-wise sum.
        output(i,2) = sum(sum(OutQMRA{1,i}(int32(OutQMRA{2,i}(1:5)),1:3),2));
        output(i,3) = sum(sum(OutQMRA{1,i}(setxor(startingDay:end, int32(OutQMRA{2,i}
(1:5))), 1:3),2)); %???
    end
    output(:,4) = output(:,2) + output(:,3);
    output(find(output(:,1) != output(:,4)), :) %Testing whether each row sums
properly.
    outputSum = sum(output)
    pSpikes = outputSum(2) / outputSum(1)
    output(:,5) = output(:,2) ./ output(:,1);
    Output = output(:,5);
end

Results = struct([]);
c = 1; %Initializing loop counter
inc={'Lo','Med','Hi'};
for i = 1:3;
    for j = [10,1000,100000];
        for k = ['A','B','C'];
            Output = pS(k,inc{i},j);
            eval(char(["Results.",inc{i},num2str(j),k," = Output;"]));
            disp(['Col. ',num2str(c),' ',',',inc{i},' incidence, spike height
',num2str(j),' ', mix ',k])
            c = c+1;
        end
    end
end

boxplot(struct2cell(Results)); %Need to convert struct to cell array for easy

```

```

boxplotting.
title 'Fig. S3. Proportion of infections occurring on spike days';
ylabel 'Proportion of infections';
%axis('tic[y]');
axis([0 28 0 1.1]); %Adjusting axis limits.
text(2.3,1.05,'Low incidence');
text(11.2,1.05,'Medium incidence');
text(20.5,1.05,'High incidence');
ah=get(gcf,'CurrentAxes'); %Getting handle of the axes that were just created by
boxplot().
set(ah,'XTick',1:27); %Setting tick locations manually.
set(ah,'XTickLabel','A|B|C|A|B|C|A|B|C|A|B|C|A|B|C|A|B|C|A|B|C|A|B|C|A|B|C')
%Labelling ticks manually with pathogen mixture.
hold on;
plot([9.5, 9.5], [0, 1.1], '-k'); %Now placing lines to separate boxplots by incidence
and spike height.
plot([18.5, 18.5], [0, 1.1], '-k');
plot([3.5, 3.5], [0, 1], '-k');
plot([6.5, 6.5], [0, 1], '-k');
plot([12.5, 12.5], [0, 1], '-k');
plot([15.5, 15.5], [0, 1], '-k');
plot([21.5, 21.5], [0, 1], '-k');
plot([24.5, 24.5], [0, 1], '-k');
ah2 = axes('Position',[0 0 1 1],'Visible','off'); %Creating a 2nd set of invisible axes
the size of the plot window.
axis([0 1 0 1]); %Setting the limits of the
above invisible axes.
text(0.005,0.08,'Pathogen mix:'); %Manually labeling the x axis
using the invisible axes.
text(0.005,0.04,'Spike height: 10 10^3 10^5 10 10^3 10^5 10
10^3 10^5');

print -dps -mono pInfSpike.ps; %Does not work; gives bizarrely colored output with

```

many different graphics formats. Simply grabbed a screenshot of the figure window instead.

%Resulting proportion of infections occurring on spike days.

%Lo10A = 0.101
%Lo10B = 0.0997
%Lo10C = 0.108
%Lo1000A = 0.581
%Lo1000B = 0.580
%Lo1000C = 0.693
%Lo100000A = 0.833
%Lo100000B = 0.843
%Lo100000C = 0.894

%Med10A = 0.0722
%Med10B = 0.0686
%Med10C = 0.0848
%Med1000A = 0.339
%Med1000B = 0.309
%Med1000C = 0.432
%Med100000A = 0.455
%Med100000B = 0.411
%Med100000C = 0.561

%Hi10A = 0.0490
%Hi10B = 0.0460
%Hi10C = 0.0603
%Hi1000A = 0.188
%Hi1000B = 0.167
%Hi1000C = 0.232
%Hi100000A = 0.320
%Hi100000B = 0.275

%Hi100000C = 0.318

%Just like AssignIll.m, but loops over columns of the person-pathogen matrix instead of rows, & should be faster.

%Assigns an illness duration to entries of a matrix, where rows are people and columns are pathogens

%Positive entries mean the person is infected, negative entries (or 0) mean the person has recovered

%{

COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (AssignInf.m) is part of QMRA2v5.

QMRA2v5 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRA2v5 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRA2v5. If not, see <<http://www.gnu.org/licenses/>>.

%}

%Inputs:

% PPmatrix: Matrix with 1 row per person and 1 column per pathogen (Person-Pathogen matrix)

% Responses: Vector of illness responses (probability of illness given dose), 1 entry per pathogen

```

% ResponsesNon: Vector of illness responses for people not using any device
% pUse: Probability that a person is not using any device
% durations: Vector of durations of illnesses, 1 entry per pathogen
% ImmuneTimes: Vector of durations of immunity, 1 entry per pathogen
%The output (matrixOut) is matrixIn with new illness durations assigned to some of its
entries.

function PPmatrix = AssignInf(PPmatrix,Responses,Durations,ImmuneTimes);
    sizePP = size(PPmatrix);
    Randoms = rand(sizePP); %Random #s for determining infection. One number per person
per pathogen.
    %NewlyInfected = cell(sizePP(2),1); %Initializing cell array to store index
values of people who will be newly infected.
    Immunities = ones(sizePP(1),1) * ImmuneTimes + PPmatrix;
    %Durations = ones(sizePP) * Durations;
    %Immunities = PPmatrix + ImmuneTimes; %Immunities gives days left in (infection +
immune period), or a neg. # if susceptible.
    for i = 1:sizePP(2);
        Immunities(:,i) = PPmatrix(:,i) + ImmuneTimes(i);
        NewlyInfected = intersect(find(Immunities(:,i) <= 0), find(Randoms(:,i) <
Responses(:,i))); %Gets indices of newly infected.
        PPmatrix(NewlyInfected,i) = Durations(i);
    end
end

```

%Similar to AssignIll.m, but assigns infection durations randomly.
%Assigns an illness duration to entries of a matrix, where rows are people and columns are pathogens
%Positive entries mean the person is ill, negative entries (or 0) mean the person has recovered

%{
COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (AssignInfIllRand.m) is part of QMRA2v5.

QMRA2v5 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRA2v5 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRA2v5. If not, see <<http://www.gnu.org/licenses/>>.

%}

%Inputs:

% PPmatrix: Matrix with 1 row per person and 1 column per pathogen (Person-Pathogen matrix)

% Responses: Matrix of illness responses (probability of illness given dose), 1 row per person and 1 column per pathogen

% ImmuneTimes: Vector of durations of immunity, 1 entry per pathogen

%The output (matrixOut) is matrixIn with new illness durations assigned to some of its entries.

%It requires the functions durEc(), durGi(), & durRo() in IllDurations.m.

```
function [PPmatrix,PPmatrixD] =
AssignInfIllRand(PPmatrix,PPmatrixD,Responses,ImmuneTimes,MorbidityK);
    %rand('state',28)
    sizePP = size(PPmatrix);
    Randoms = rand(sizePP);          %Random #s for determining infection. One number per
person per pathogen.
    Randoms2 = rand(sizePP); %Random #s for determining disease, as above.
    %NewlyInfected = cell(sizePP(2),1); %Initializing cell array to store index
values of people who will be newly infected.
    Immunities = ones(sizePP(1),1) * ImmuneTimes + PPmatrix; %Adjusts PPmatrix to
account for immunity.
    Durations = NaN(sizePP); %Making & populating a matrix of disease durations (same
size as PPmatrix) to assign to newly infected.
    Durations(:,1) = durEc(sizePP(1));
    Durations(:,2) = durGi(sizePP(1));
    Durations(:,3) = durRo(sizePP(1));
    for i = 1:sizePP(2); %Loop, once for each pathogen.
        %NewlyInfected = intersect(find(Immunities(:,i) <= 0), find(Randoms(:,i) <
Responses(:,i))); %Gets indices of newly infected. Note: 0 or less is susc.
        NewlyInfected = find((Immunities(:,i) <= 0 & Randoms(:,i) < Responses(:,i)) ==
1); %Should be much faster than above line.
        PPmatrix(NewlyInfected,i) = Durations(NewlyInfected,i);
        PPmatrixD(NewlyInfected,i) = 1; %Flags newly infected.
        %NewlyIll = intersect(NewlyInfected, find(Randoms2(:,i) < MorbidityK(i)));
        %PPmatrixD(NewlyIll,i) = 2;
        %---This|=====|is
        %taken from line 25 above - should run faster than prev. 2 lines.
        PPmatrixD((Immunities(:,i) <= 0 & Randoms(:,i) < Responses(:,i)) & (Randoms2(:,i)
< MorbidityK(i)),i) = 2;
```

end
end

%Similar to AssignIll.m, but assigns infection durations randomly.
%Assigns an infection duration to entries of a matrix, where rows are people and columns
are pathogens
%Positive entries mean the person is infected, negative entries (or 0) mean the person
has recovered

%{
COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (AssignInfRand.m) is part of QMRA2v5.

QMRA2v5 is free software: you can redistribute it and/or modify
it under the terms of the GNU General Public License as published by
the Free Software Foundation, either version 3 of the License, or
(at your option) any later version.

QMRA2v5 is distributed in the hope that it will be useful,
but WITHOUT ANY WARRANTY; without even the implied warranty of
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt)
along with QMRA2v5. If not, see <<http://www.gnu.org/licenses/>>.

%}

%Inputs:

% PPmatrix: Matrix with 1 row per person and 1 column per pathogen (Person-Pathogen
matrix)

% Responses: Matrix of illness responses (probability of infection given dose), 1
row per person and 1 column per pathogen

% ImmuneTimes: Vector of durations of immunity, 1 entry per pathogen

%The output (matrixOut) is matrixIn with new illness durations assigned to some of its entries.

%It requires the functions durEc(), durGi(), & durRo() in IllDurations.m.

```
function [PPmatrix] = AssignInfRand(PPmatrix,Responses,ImmuneTimes);
    %rand('state',28)
    sizePP = size(PPmatrix);
    Randoms = rand(sizePP); %Random #s for determining infection. One number per person
    per pathogen.
    Immunities = ones(sizePP(1),1) * ImmuneTimes + PPmatrix; %Adjusts PPmatrix to
    account for immunity.
    Durations = NaN(sizePP); %Making & populating a matrix of disease durations (same
    size as PPmatrix) to assign to newly infected.
    Durations(:,1) = durEc(sizePP(1));
    Durations(:,2) = durGi(sizePP(1));
    Durations(:,3) = durRo(sizePP(1));
    NewlyInfected = intersect(find(Immunities <= 0), find(Randoms < Responses));
    %Gets indices of newly infected. Note: 0 or less is susc.
    PPmatrix(NewlyInfected) = Durations(NewlyInfected);
end
```

```
%Calculates whether a given week is reported as a week with diarrhea under the reporting
scheme in the DRC Lifestraw RCT (Boisson 2010).
%It considers reduced recall of past diarrheal episodes after 2d ('remembrance') and
possible distinct diarrhea episodes in the previous 7d.
%It operates on a matrix:
%   Rows represent people, columns represent pathogens/illnesses, and entries represent
# of days remaining in the illness.
%It outputs a vector with 1 entry per person, 1 if illness is reported during the week, 0
if not.
```

```
%{
COPYRIGHT INFORMATION
```

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (CalcDiarrhWeeks.m) is part of QMRA2v5.

QMRA2v5 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRA2v5 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRA2v5. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
function vec = CalcDiarrhWeeks(InMatrix, remembrance,timeWindow,Morbidity);
    sizeInMatrix = size(InMatrix);
```



```

Randoms = rand(sizeInMatrix(1),sizeInMatrix(2));
for j = 1:sizeInMatrix(2);
    InMatrix((find(Randoms(:,j) > Morbidity(j))),j) = -9999;
end

    for i = 1:sizeInMatrix(1); %Loop over all people. Only the most recent episode
(largest entry in a row) is used to assign illness.
        %Randoms = rand(1,columns(InMatrix)); %Random numbers for determining
morbidity (these 2 lines moved upward for greater speed)
        %InMatrix(i,find(Randoms > Morbidity)) = -9999; %Apply morbidity ratio: if
asymptomatic, infection is set to -9999, and therefore not reported. Note that this
modification is not passed out of this function.
        if max(InMatrix(i,:)) >= -2;
            vec(i) = 1; %If ill during day 0, 1, or 2, assume illness is always
reported, therefore assign illness.
        elseif (max(InMatrix(i,:)) >= -timeWindow) & rand() < remembrance);
            vec(i) = 1; %Otherwise, if ill during days 3-7, randomly determine if
episode is remembered. If so, assign illness.
        else
            vec(i) = (0); %Otherwise, no illness is remembered or reported. Assign no
illness.
        end
    end
end
end

```

%Loop for obtaining estimation runs while modifying pUse, pTreat, and LRVs. Facilitates computing cluster use.

%Be sure to check that GetTrialParams.m is configured properly before running this script.

%Submit as several jobs to parallelize a calibration run (maybe not worth bothering with job array).

%{

COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (CalibrationLoopFuncCompile.m) is part of QMRA2v5.

QMRA2v5 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRA2v5 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRA2v5. If not, see <<http://www.gnu.org/licenses/>>.

%}

function[OutM OutS] =

CalibrationLoopFuncCompile(indexText,inc,calibRunsText,nSpikesText,multSpikesText,maxEcText,maxGiText,maxRoText);

%This helps with debugging, since arguments to compiled code can only be text.

index = str2num(indexText);

```

RandStream.setDefaultStream(RandStream.create('mt19937ar','seed',sum([clock index*10])));
    %Sets random stream based on clock & job index.
calibRuns = str2num(calibRunsText);
nSpikes = str2num(nSpikesText);
multSpikes = str2num(multSpikesText);
PathogensLMax = [str2num(maxEcText) str2num(maxGiText) str2num(maxRoText)];

%===Required lines for HPC
setenv MKL_DYNAMIC FALSE
%maxNumCompThreads(1); %Throws an error. Recommended, but does not seem to be
necessary.
%===
%===Parameter entry===
%switch inc; %This switch no longer needed since we're passing max pathogen
concentrations into this function.
% case 'Lo'; PathogensLMax = [6e3, 0.3, 0.025]; %From QMRA2v2/TrialCalibRuns.m.
% case 'Med'; PathogensLMax = [2e4, 0.8, 0.05];
% case 'Hi'; PathogensLMax = [1.5e5, 1.6, 0.08];
% otherwise; error('inc needs to be Lo, Med, or Hi');
%end
infile = 'nonapplicable.csv';
outfile =
['Results/TrialCalibResults',inc,nSpikesText,'spikesx',multSpikesText,'-',indexText,'.csv
'];
%===End parameter entry===

%disp(['##### Running
',num2str(size(U,2)),'*',num2str(size(T,2)),'*',num2str(size(L,2)),'+1=',num2str(combos),
' parameter combinations on ',num2str(size(tempData,1)),' parameter sets from
calibration, should have requested at least that many members in the job array. #####'])

[OutM OutS] = Main(0, 1, 0, [0 0 0], [0 0 0], PathogensLMax, calibRuns, outfile, infile,
1, nSpikes, multSpikes);

```

```
disp(['##### DONE #####'])  
end %End function.
```

```
%Beta-Poisson dose response model, using N50 (default) or beta as a parameter
%function outvar = DRbP(N50orBeta,alpha,invar,reverse='no',WhichParam='N50') %Ordinarily,
invar is dose & outvar is response.
```

```
%{
COPYRIGHT INFORMATION
```

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (DRbP.m) is part of QMRA2v5.

QMRA2v5 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRA2v5 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRA2v5. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
function outvar = DRbP(N50orBeta,alpha,invar,reverse,WhichParam) %Ordinarily, invar is
dose & outvar is response.
    if nargin == 3;
        reverse = 'no'; WhichParam = 'N50';
    end
    switch(reverse)
        case 'no'
            switch(WhichParam)
```

```

        case 'N50'
            outvar = 1-(1+(invar/N50orBeta)*(2^(1/alpha)-1)).^-alpha;
        case 'Beta'
            outvar = 1-(1+(invar/N50orBeta)).^-alpha;
        otherwise
            error(['WhichParam must be "N50" or "Beta"'])
        end
    case 'yes' %If reverse='yes', invar is response & outvar is dose.
        switch(WhichParam)
            case 'N50'
                outvar = N50orBeta * ( ((1-invar).^(-1/alpha) -1) /
(2^(1/alpha)-1) );
            case 'Beta'
                outvar = N50orBeta * ((1-invar).^(-1/alpha) -1);
            otherwise
                error(['WhichParam must be "N50" or "Beta"'])
            end
        otherwise
            error(['reverse must be "no" or "yes"'])
        end
    end
end

```

%Returns a vector of response values, determined by dose response models, for several pathogens.

%{
COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (DRchoose.m) is part of QMRA2v5.

QMRA2v5 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRA2v5 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRA2v5. If not, see <<http://www.gnu.org/licenses/>>.

%}

%Inputs are vectors with 1 entry per pathogen, all in the same order:

% nonalphas: k parameter (exponential) or N50 parameter (beta-Poisson)
% alphas: alpha parameter (beta-Poisson); NA if exponential model is desired
% Doses: Doses of pathogens received per individual (under default behavior;
see 'reverse' below)

% morbidities: Morbidity ratios: proportion of infected who are ill
% reverse: Defaults to 'no', determining proportion ill from dose. If 'yes',
determines dose from proportion ill.

%The output (outvec) is a vector containing the proportions of exposed who will fall ill.

```

%If reverse=='yes', outvec is a vector of doses calculated from invec, the proportions
ill.
%This code requires several custom functions/subroutines in the working directory:
%   DRexp.m:      Exponential dose response model
%   DRbP.m:      Beta-Poisson dose response model

%function outvec =
%DRchoose(nonalphas,alphas,invec,morbidities=1,reverse='no') %Octave
function outvec = DRchoose(nonalphas,alphas,Doses,morbidities,reverse)
    if nargin == 3;
        morbidities = 1; reverse = 'no';
    end
    if morbidities == 1;
        morbidities = ones(length(nonalphas));
    end
    switch(reverse)
        case 'no'
            for i=1:length(alphas);
                if (isnan(alphas(i))) %if alpha is NA, run exponential dose
response
                    if size(Doses)(1) == 1;
                        outvec(i) = DRexp(nonalphas(i),Doses(i)) * morbidities(i);
                    else
                        outvec(i) = DRexp(nonalphas(i),Doses(:,i)) *
morbidities(i);
                    end
                else %run beta-Poisson dose response
                    outvec(i) = DRbP(nonalphas(i),alphas(i),Doses(i)) *
morbidities(i);
                end
            end
        case 'yes'
            for i=1:length(alphas);

```



```

                if (isnan(alphas(i)))    %if alpha is NA, run exponential dose
response
                outvec(i) = DRexp(nonalphas(i),Doses(i) ./
morbidities(i),'yes','N50') ;
                else
                %run beta-Poisson dose response
                outvec(i) = DRbP(nonalphas(i),alphas(i),Doses(i) ./
morbidities(i),'yes','N50');
                end
            end
        otherwise
            error('reverse must be "no" or "yes"')
    end
end
end

```

```
%Exponential dose response model
%function outvar = DRexp(k, invar,reverse='no') %Ordinarily, invar is dose & outvar is
response.
```

```
%{
COPYRIGHT INFORMATION
```

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (DRexp.m) is part of QMRA2v5.

QMRA2v5 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRA2v5 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRA2v5. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
function outvar = DRexp(k, invar, reverse) %This works in Matlab.
    if nargin < 3;
        reverse = 'no';
    end
    switch(reverse);
        case 'no';
            outvar = 1-exp(-k * invar);
        case 'yes'; %If reverse='yes', invar is response & outvar is dose.
```

```
        outvar = log(1-invar)/-k;
    otherwise
        error(['reverse (last parameter) must be "no" or "yes"'])
    end
end
```

```
%Functions for calculating vectors of illness durations
```

```
%{  
COPYRIGHT INFORMATION
```

```
    Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").
```

```
This file (durEc.m) is part of QMRA2v5.
```

```
QMRA2v5 is free software: you can redistribute it and/or modify  
it under the terms of the GNU General Public License as published by  
the Free Software Foundation, either version 3 of the License, or  
(at your option) any later version.
```

```
QMRA2v5 is distributed in the hope that it will be useful,  
but WITHOUT ANY WARRANTY; without even the implied warranty of  
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the  
GNU General Public License for more details.
```

```
You should have received a copy of the GNU General Public License (gpl.txt)  
along with QMRA2v5. If not, see <http://www.gnu.org/licenses/>.
```

```
%}
```

```
function output = durEc(n); %Based on ...?  
    output = round(gamrnd(1.775,1.690,[n,1])); %Shape, then scale  
    output(output == 0) = 0.1; %Sets zero durations to 0.1 day instead. Will  
still function as 1 day.  
end
```

```
%Functions for calculating vectors of illness durations
```

```
%{  
COPYRIGHT INFORMATION
```

```
    Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").
```

```
This file (durGi.m) is part of QMRA2v5.
```

```
QMRA2v5 is free software: you can redistribute it and/or modify  
it under the terms of the GNU General Public License as published by  
the Free Software Foundation, either version 3 of the License, or  
(at your option) any later version.
```

```
QMRA2v5 is distributed in the hope that it will be useful,  
but WITHOUT ANY WARRANTY; without even the implied warranty of  
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the  
GNU General Public License for more details.
```

```
You should have received a copy of the GNU General Public License (gpl.txt)  
along with QMRA2v5. If not, see <http://www.gnu.org/licenses/>.
```

```
%}
```

```
function output = durGi(n); %Based on a fit of gamma dist. to limited info from Kent GP  
1988.
```

```
    output = round(gamrnd(3.206,3.431,[n,1])); %Shape, then scale  
    output(output == 0) = 0.1; %Sets zero durations to 0.1 day instead. Will
```

```
still function as 1 day.  
end
```

```
%Functions for calculating vectors of illness durations
```

```
%{  
COPYRIGHT INFORMATION
```

```
    Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").
```

```
This file (durRo.m) is part of QMRA2v5.
```

```
QMRA2v5 is free software: you can redistribute it and/or modify  
it under the terms of the GNU General Public License as published by  
the Free Software Foundation, either version 3 of the License, or  
(at your option) any later version.
```

```
QMRA2v5 is distributed in the hope that it will be useful,  
but WITHOUT ANY WARRANTY; without even the implied warranty of  
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the  
GNU General Public License for more details.
```

```
You should have received a copy of the GNU General Public License (gpl.txt)  
along with QMRA2v5. If not, see <http://www.gnu.org/licenses/>.
```

```
%}
```

```
function output = durRo(n); %Based on 4 rotavirus-infected volunteers having durations  
of 1, 2, 3, and 4 days (Kapikian 1983).  
    output = ceil(rand([n,1]) * 4);  
    output(output == 0) = 0.1; %Sets zero durations to 0.1 day instead. Will  
still function as 1 day.  
end
```

%Loop for obtaining estimation runs while modifying pUse, pTreat, and LRVs. Facilitates computing cluster use.

%Be sure to check that GetTrialParams.m is configured properly before running this script.

%Differs from EstimationLoopFuncCompile in that only a small number of parameter combinations are chosen, rather than all possible combinations.

%{
COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (EstimationLoopFuncCompileV2.m) is part of QMRA2v5.

QMRA2v5 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRA2v5 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRA2v5. If not, see <<http://www.gnu.org/licenses/>>.

%}

```
function[OutM OutS] =  
EstimationLoopFuncCompileV2(indexText,inc,mix,overallComplianceText,multConcText,nSpikesT  
ext,multSpikesText);  
Octave = size(ver('Octave'),1); %Indicator of whether the code is running under Octave  
(1) or Matlab (0).
```

```

index = str2num(indexText);
if Octave == 0;
    RandStream.setDefaultStream(RandStream.create('mt19937ar', 'seed', sum([clock
index*10])));
    %Sets random stream based on clock & job index. Doesn't work with Octave (Octave
bases the seed on the clock by default).
end
oC = str2num(overallComplianceText);
multConc = str2num(multConcText);
nSpikes = str2num(nSpikesText);
multSpikes = str2num(multSpikesText);
%baselines = str2num(baselinesText);

%===Required lines for HPC
setenv MKL_DYNAMIC FALSE
%maxNumCompThreads(1); %Throws an error. Does not seem to be necessary.
%===Parameter entry=== Note that 0 should not be included in L.
%P = [0 .1 .2]; %Vector of desired values for proportions of children never using the
device.
%N = [0 .1 .2]; %Vector of desired values for proportions of children perfectly using
the device.
L = [1 2 3 4 5]; %Vector of log reduction values desired (all marker pathogens get the
same LRV).
%Testing the code using the vectors below.
%U=[.9 1]
%T=[.9 1]
%L=[1 2]
%Constructing a matrix with all possible combos of P, N, & L
%[p n l] = ndgrid(P,N,L);
%Combos = [p(:) n(:) l(:)]; %TODO: Build functionality to check N, P,
overallCompliance, and pTreat to ensure they make sense before running.
%Building appropriate combinations of perfect compliers (P; 1st column) and noncompliers
(N; 2nd column).

```



```

Compos = [oC 0];          %Combo with max possible perfect compliers & max possible
noncompliers (same as [0 1-oC]; [oC 1-oC] is 0/0).
Compos(2,:) = [oC/2 (1-oC)/2];    %Combo with intermediate perfect/nonperfect compliers.
Compos(3,:) = [0 0];          %Combo with no perfect/nonperfect compliers (i.e., pTreat ==
oC).
ComposT = (oC - Compos(:,1)) ./ (1 - Compos(:,1) - Compos(:,2)) %TODO: doublecheck to
make sure this is right.
if sum(ComposT > 1 + eps) | sum(ComposT < 0 - eps); error('pTreat out of range (>1 or
<0); check parameter compos.');
```

```

end
Compos(:,3) = L(1); %This & subsequent 'for' loop copy the above for each possible LRV.
OutCompos = Compos;
for i = 2:size(L,2);
    NextCompos = Compos;
    NextCompos(:,3) = L(i);
    OutCompos = [OutCompos; NextCompos];
end
Baselines = zeros(1,3); %Creating baseline row. No log reduction & no perfect
compliance.
Baselines(:,2) = 1;          %Modifies above, so that 100% never use device.
Compos = [Baselines; OutCompos]; %Baseline as 1st row.
Compos' %Output results, transposed.
compos = size(Compos,1)

infile =
['RTF_',mix,'_TrialCalibResults',inc,nSpikesText,'spikesx',multSpikesText,'.csv'];
outfile =
['Results/EstResults',inc,mix,nSpikesText,'spikesx',multSpikesText,'x',multConcText,'oC',
num2str(oC*100),'-',indexText,'.csv'];
%===End parameter entry===

tempData = csvread(infile,1,1);

disp(['##### Running 3 * ',num2str(size(L,2)),'+ 1 = ',num2str(compos),' parameter

```

```

combinations on ',num2str(size(tempData,1)),' parameter sets from calibration, should
have requested at least that many members in the job array. #####')
if index == 1; %If running the baseline parameters:
    [OutM OutS] = Main(Combos(index,1), Combos(index,2), oC, [Combos(index,3)
Combos(index,3) Combos(index,3)], [0 0 0], [0 0 0], 0, outfile, infile, multConc,
nSpikes, multSpikes, 1);
else %If running the parameters from calibration:
    [OutM OutS] = Main(Combos(index,1), Combos(index,2), oC, [Combos(index,3)
Combos(index,3) Combos(index,3)], [0 0 0], [0 0 0], 0, outfile, infile, multConc,
nSpikes, multSpikes, 0);
end
disp(['##### DONE #####'])
end %End function.

```

%Loop for obtaining estimation runs while modifying pUse, pTreat, and LRVs. Facilitates computing cluster use.

%Be sure to check that GetTrialParams.m is configured properly before running this script.

%Differs from EstimationLoopFuncCompile in that only a small number of parameter combinations are chosen, rather than all possible combinations.

%{
COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (EstimationLoopFuncCompileV2PC.m) is part of QMRA2v5.

QMRA2v5 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRA2v5 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRA2v5. If not, see <<http://www.gnu.org/licenses/>>.

%}

```
function[OutM OutS] =  
EstimationLoopFuncCompileV2Untreated(indexText,inc,mix,overallComplianceText,multConcText  
,nSpikesText,multSpikesText);  
Octave = size(ver('Octave'),1); %Indicator of whether the code is running under Octave  
(1) or Matlab (0).
```

```

index = str2num(indexText);
if Octave == 0;
    RandStream.setDefaultStream(RandStream.create('mt19937ar','seed',sum([clock
index*10])));
    %Sets random stream based on clock & job index. Doesn't work with Octave (Octave
bases the seed on the clock by default).
end
oC = str2num(overallComplianceText);
multConc = str2num(multConcText);
nSpikes = str2num(nSpikesText);
multSpikes = str2num(multSpikesText);
%baselines = str2num(baselinesText);

%===Required lines for HPCC
setenv MKL_DYNAMIC FALSE
%maxNumCompThreads(1); %Throws an error. Does not seem to be necessary.
%===Parameter entry=== Note that 0 should not be included in L.
L = [1 2 3 4 5]; %Vector of log reduction values desired (all marker pathogens get the
same LRV).

infile =
['RTF_',mix,'_TrialCalibResults',inc,nSpikesText,'spikesx',multSpikesText,'.csv'];
outfile =
['Results/EstResults',inc,mix,nSpikesText,'spikesx',multSpikesText,'x',multConcText,'oC',
num2str(oC*100),'-',indexText,'.csv'];
%===End parameter entry===

tempData = csvread(infile,1,1);
disp(['##### Running ',num2str(size(L,2)),' parameter combinations on a subsample of
',num2str(size(tempData,1)),' parameter sets from calibration, should have requested at
least that many members in the job array. #####'])
[OutM OutS] = Main(1, 0, oC, [L(index) L(index) L(index)], [0 0 0], [0 0 0], 0, outfile,
infile, multConc, nSpikes, multSpikes, 0);

```

```
disp(['##### DONE #####'])  
end %End function.
```

%Loop for obtaining estimation runs while modifying pUse, pTreat, and LRVs. Facilitates computing cluster use.

%Be sure to check that GetTrialParams.m is configured properly before running this script.

%Differs from EstimationLoopFuncCompile in that only a small number of parameter combinations are chosen, rather than all possible combinations.

%{
COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (EstimationLoopFuncCompileV2Untreated.m) is part of QMRA2v5.

QMRA2v5 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRA2v5 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRA2v5. If not, see <<http://www.gnu.org/licenses/>>.

%}

```
function[OutM OutS] =  
EstimationLoopFuncCompileV2Untreated(indexText,inc,mix,overallComplianceText,multConcText  
,nSpikesText,multSpikesText);  
Octave = size(ver('Octave'),1); %Indicator of whether the code is running under Octave  
(1) or Matlab (0).
```

```

index = str2num(indexText);
if Octave == 0;
    RandStream.setDefaultStream(RandStream.create('mt19937ar','seed',sum([clock
index*10])));
    %Sets random stream based on clock & job index. Doesn't work with Octave (Octave
bases the seed on the clock by default).
end
oC = str2num(overallComplianceText);
multConc = str2num(multConcText);
nSpikes = str2num(nSpikesText);
multSpikes = str2num(multSpikesText);
%baselines = str2num(baselinesText);

%===Required lines for HPC
setenv MKL_DYNAMIC FALSE
%maxNumCompThreads(1); %Throws an error. Does not seem to be necessary.
%===Parameter entry=== Note that 0 should not be included in L.
%L = [1 2 3 4 5]; %Vector of log reduction values desired (all marker pathogens get the
same LRV).

infile =
['RTF_',mix,'_TrialCalibResults',inc,nSpikesText,'spikesx',multSpikesText,'.csv'];
outfile =
['Results/EstResults',inc,mix,nSpikesText,'spikesx',multSpikesText,'x',multConcText,'oC',
num2str(oC*100),'-1-',indexText,'.csv'];
%===End parameter entry===

tempData = csvread(infile,1,1);
nChunks = 20; %# of equal-sized chunks to break the parameter sets into. Add 1 more for
the remainder.
nSets = size(tempData,1);
chunk = floor(nSets/nChunks);
remainder = mod(nSets,nChunks);

```

```

ChunkStarts = linspace(1,nChunks*chunk+1,nChunks+1);
ChunkStarts(end+1) = nSets+1;
DataChunk = tempData(ChunkStarts(index):ChunkStarts(index+1)-1,:);
z = zeros(size(DataChunk,1),1);
DataChunk = [z DataChunk];
DataChunk = [0,0,0,0;DataChunk]; %Needs surplus 1st line because Main() will strip the
1st line off (the original R output has column names).
infile = ['TempBaselineChunk',indexText,'.csv'];
csvwrite(infile,DataChunk);

disp(['##### Running ',num2str(nSets),' parameter combinations in ',num2str(nChunks+1),'
chunks, should have requested at least that many members in the job array. #####'])
[OutM OutS] = Main(0, 1, 1, [0 0 0], [0 0 0], [0 0 0], 0, outfile, infile, multConc,
nSpikes, multSpikes, 0);

disp(['##### DONE #####'])
end %End function.

```



```
%Examining data from a single run of the QMRA lifestraw model.  
%Converting to susceptible (-9), immune (-1), or diseased (9) for each of the 3  
pathogens.
```

```
%{  
COPYRIGHT INFORMATION
```

```
    Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").
```

```
This file (Examine1Run.m) is part of QMRA2v5.
```

```
QMRA2v5 is free software: you can redistribute it and/or modify  
it under the terms of the GNU General Public License as published by  
the Free Software Foundation, either version 3 of the License, or  
(at your option) any later version.
```

```
QMRA2v5 is distributed in the hope that it will be useful,  
but WITHOUT ANY WARRANTY; without even the implied warranty of  
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the  
GNU General Public License for more details.
```

```
You should have received a copy of the GNU General Public License (gpl.txt)  
along with QMRA2v5. If not, see <http://www.gnu.org/licenses/>.
```

```
%}
```

```
%====Setting options====
```

```
startTime = 96;      %Time point at which to start looking at the data. Note that 1st  
matrix corresponds to time 0.
```

```
recode = 1;         %If 1, recode matrix entries to susc./inf./immune.
```

```
%====Finished with options, starting processing.====
```

```
PPMs = PPmatrices; %Making a copy.
```

```
switch(recode);
```

```
    case 1;
```

```

disp(['Recoding raw numbers to susc./inf./immune.'])
for i = startTime:size(PPmatrices)(2); %Starting when the actual simulation starts
(after equilibrium is reached).
    PPMs(i).KidsInt(PPMs(i).KidsInt <= -7) = -9; %7 day immune period; 0 counts as
the 1st immune day, so at -7 they are susc.
    PPMs(i).KidsInt(PPMs(i).KidsInt > 0) = 9; %Susceptible if a positive
integer.
    PPMs(i).KidsInt(abs(PPMs(i).KidsInt) != 9) = -1;%Immune if neither of the above
applies.
    PPMs(i).KidsInt(PPMs(i).KidsInt == 9) = 2; %Infected person-day marked as 2,
so as to more easily distinguish.
end
for i = startTime:size(PPmatrices)(2); %Same as above 'for' loop, but placebo.
    PPMs(i).KidsPla(PPMs(i).KidsPla <= -7) = -9;
    PPMs(i).KidsPla(PPMs(i).KidsPla > 0) = 9;
    PPMs(i).KidsPla(abs(PPMs(i).KidsPla) != 9) = -1;
    PPMs(i).KidsPla(PPMs(i).KidsPla == 9) = 2;
end
otherwise
disp(['Not recoding to susc./inf./immune, output will display raw numbers.'])
end
KidsIntStatusEc = NA(size(PPMs(1).KidsInt)(1), size(PPmatrices)(2)-(startTime-1));
KidsIntStatusGi = KidsIntStatusEc;
KidsIntStatusRo = KidsIntStatusEc;
KidsPlaStatusEc = NA(size(PPMs(1).KidsPla)(1), size(PPmatrices)(2)-(startTime-1));
KidsPlaStatusGi = KidsPlaStatusEc;
KidsPlaStatusRo = KidsPlaStatusEc;
for i = startTime:size(PPmatrices)(2); %Starting when the actual simulation starts
(after equilibrium is reached).
    for j = 1:size(PPMs(1).KidsInt)(1);
        KidsIntStatusEc(j,i-startTime+1) = PPMs(i).KidsInt(j,1);
        KidsIntStatusGi(j,i-startTime+1) = PPMs(i).KidsInt(j,2);
        KidsIntStatusRo(j,i-startTime+1) = PPMs(i).KidsInt(j,3);
    end
end

```

```

end
for j = 1:size(PPMs(1).KidsPla)(1);
    KidsPlaStatusEc(j,i-startTime+1) = PPMs(i).KidsPla(j,1);
    KidsPlaStatusGi(j,i-startTime+1) = PPMs(i).KidsPla(j,2);
    KidsPlaStatusRo(j,i-startTime+1) = PPMs(i).KidsPla(j,3);
end
end
%Now can visually inspect the 6 status matrices that have been output.
%Should be a way to collapse them also (run-length encoding?), but not yet implemented.

function outmatrix = coll(inmatrix,nMaxRuns) %Inefficient but hopefully works.
    sizeM = size(inmatrix);
    maxk = 1; %Initializing counter to determine the maximum number of runs ever seen
    during the function call.
    for i = 1:sizeM(1); %Loop over all rows
        for j = 2:sizeM(2); %Loop over each entry per row
            if j == 2; %Special procedure for first iteration, since there could
                be a transition (or not) between the 1st 2 entries.
                    k = 1; %Initiating run counter;
                    if inmatrix(i,j) != inmatrix(i,j-1);
                        if inmatrix(i,j-1) == -9; dur(k) = -1;
                        elseif inmatrix(i,j-1) == -1; dur(k) = 0.001;
                        elseif inmatrix(i,j-1) == 2; dur(k) = 1;
                        else stop('Unexpected value in matrix (not -9, -1, or 2).')
                        end
                        k = k + 1; %Increment run counter
                    if inmatrix(i,j) == -9; dur(k) = -1;
                    elseif inmatrix(i,j) == -1; dur(k) = 0.001;
                    elseif inmatrix(i,j) == 2; dur(k) = 1;
                    else stop('Unexpected value in matrix (not -9, -1, or 2).')
                    end
                else %If type of run does not change in 1st 2 entries
                    if inmatrix(i,j) == -9; dur(k) = -2;

```

```

        elseif inmatrix(i,j) == -1; dur(k) = 0.002;
        elseif inmatrix(i,j) == 2; dur(k) = 2;
        else stop('Unexpected value in matrix (not -9, -1, or 2).')
        end
    end
    elseif j == sizeM(2);    %Special procedure for last iteration - need to
drop it since it is probably incomplete.
        dur(k) = 0;
    else %If j (the column) is anything greater than 2, but not the last
column:
        if inmatrix(i,j) != inmatrix(i,j-1);    %If there is a transition,
reset the run:
            k = k + 1;
            if inmatrix(i,j) == -9; dur(k) = -1;
            elseif inmatrix(i,j) == -1; dur(k) = 0.001;
            elseif inmatrix(i,j) == 2; dur(k) = 1;
            else stop('Unexpected value in matrix (not -9, -1, or 2).')
            end
        else %If there is no transition, extend the run
            if inmatrix(i,j) == -9; dur(k) = dur(k) - 1;
            elseif inmatrix(i,j) == -1; dur(k) = dur(k) + 0.001;
            elseif inmatrix(i,j) == 2; dur(k) = dur(k) + 1;
            else stop('Unexpected value in matrix (not -9, -1, or 2).')
            end
        end
    end
end
if k > maxk;
    maxk = k    %Updates & displays maxk (largest no. runs seen so far). Use as
guide for entering nMaxRuns.
end
outmatrix(i,:) = padarray(dur,[0, nMaxRuns - length(dur)],0,'post');
end

```

```

    outmatrix(:,1) = 0; %Sets all 1st runs to 0 (they are likely to be incomplete).
    mean(outmatrix(find(outmatrix >= 1))) %Print mean duration of illness
end

function z = GraphColl(outmatrix) %Graphs output of above function.
    bins = unique(outmatrix(find(outmatrix < 0)));
    if length(bins) == 1; binsT=NaN(3,1); binsT(2)=bins; binsT(1)=binsT(2)-1;
binsT(3)=binsT(2)+1; bins=binsT; end
    subplot(2,2,1);
    hist(outmatrix(find(outmatrix < 0)),bins) %Durations of susceptibility
    bins = unique(outmatrix(find(outmatrix > 0 & outmatrix < 1)));
    if length(bins) == 1; binsT=NaN(3,1); binsT(2)=bins; binsT(1)=binsT(2)-0.001;
binsT(3)=binsT(2)+0.001; bins=binsT; end
    subplot(2,2,2);
    hist(outmatrix(find(outmatrix > 0 & outmatrix < 1)),bins) %Durations of immunity
    bins = unique(outmatrix(find(outmatrix >= 1)));
    if length(bins) == 1; binsT=NaN(3,1); binsT(2)=bins; binsT(1)=binsT(2)-1;
binsT(3)=binsT(2)+1; bins=binsT; end
    subplot(2,2,3);
    hist(outmatrix(find(outmatrix >= 1)),bins) %Durations of illness
end

%histc(PlaRo(find(PlaRo < 1 & PlaRo > 0)),[0:0.001:max(PlaRo(find(PlaRo < 1 & PlaRo >
0)))])) %Awful (but functional) way to get counts of possible values for immunity
length.

```

%This script generates a series of parameters for input into the main code, rather than stochastically generating them.

%{
COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (GetTrialParams.m) is part of QMRA2v5.

QMRA2v5 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRA2v5 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRA2v5. If not, see <<http://www.gnu.org/licenses/>>.

%}

```
reps = 1;           %# of times to run each parameter set.  
readParams = 1;    %To read parameters generated by a previous model run (1) or not  
(0).  
paramSets = 12;    %# of parameter sets to be run, if generating them  
systematically.  
%loops = paramSets * reps; %Deliberately overwrites 'loops' in the main code.  
switch(readParams);  
    case 1;  
        TrialParamsAll = dlmread(inFilename, ',', 1, 1); %Access a file returned by
```

```

ReadOutput.r. inFilename is an input to Main.m.
    paramSets = size(TrialParamsAll, 1); %Overwrites 'paramSets' above.
    disp(['Reading parameter values from ',num2str(paramSets),' trials in
',inFilename,', will winnow down to 150 if greater.'])
    if useAllParamSets == 1;
        TrialParams = TrialParamsAll;
    elseif paramSets > 150;
        selectedParamSets = randperm(paramSets);
        selectedParamSets = selectedParamSets(1:150);
        TrialParams = TrialParamsAll(selectedParamSets,:);
        paramSets = size(TrialParams,1);
    else TrialParams = TrialParamsAll;
    end
case 0;
    MinDoses = [0, 0, 0.05]; %Minimum non-zero dose. A good choice is dose that
infects 1% of population (ID1).
    ID1s = [7.5697E3, 5.0708E-1, 1.7280E-2]; %ID1 for ETEC, Giardia, & rota.
    %MinDoses = ID1s * .1; %Uncomment if ID1s are desired.
    TrialParams = zeros(paramSets,size(MinDoses, 2));
    for i = 1:length(MinDoses); %Populating all cells except the 1st row with the
minimum nonzero dose.
        TrialParams(2:paramSets,i) = MinDoses(i);
    end
    TrialParams(2,:) = MinDoses; %1st run is 0 pathogens; 2nd run is the
minimum nonzero dose.
    for i = 3:paramSets; %Uncomment the particular line desired. Comment all to
check multiple replicates of the same dose.
        %TrialParams(i,:) = MinDoses * (i-1); %Linearly increases the dose
on each model run.
        TrialParams(i,:) = 2 * TrialParams(i-1,:); %Doubles the dose on each
model run.
        %TrialParams(i,2) = 10 * TrialParams(i-1,2); %Doubles the dose for only 1
pathogen, leaving others constant.

```

```

        end
        TrialParams(:,4) = 0;    %Sets a single value for non-waterborne diarrhea
prevalence.
        %disp('Will cycle through these parameters, 1 run per set.')
        %trialPrevDiarrhBaseKids = 0 %Sets a single value for non-waterborne diarrhea
prevalence.
        %TrialParams          %Print to screen, so we see that this file was
executed & to view the trial params.
        otherwise;
            error('readParams must be 0 or 1');
    end
    %Replicating the parameter sets.
    TP = TrialParams;    %Making a copy, for use in loop below.
    if reps > 1;
        for i = 1:reps-1;
            TrialParams = [TrialParams; TP];    %Appending copies of the parameter sets.
        end
    end
    TrialParams = sortrows(TrialParams);    %Sorting so that identical parameter values are
next to each other.
    loops = size(TrialParams,1);    %Overwriting 'loops' variable in Main().
    disp(['Using ',num2str(paramSets),' parameter sets, ',num2str(reps),' times each,
totaling ',num2str(loops),' runs.'])
    if paramSets <= 25;
        disp(['Parameter sets are as follows:'])
        TP
    end
end

```


%Pulls together QMRA output files from multiple Octave threads. Runs surprisingly fast!

%{
COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (OutQMRAmerge.m) is part of QMRA2v5.

QMRA2v5 is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

QMRA2v5 is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with QMRA2v5. If not, see <<http://www.gnu.org/licenses/>>.

%}

clear all;

%First, enter the desired name of the .CSV:

filename = {'Results/MergedOutQMRA.csv'};

%Now enter as many files as necessary, each one containing the 'workspaces' from a thread.

Files = {'Results/OutQMRA20110315T130458.mat'};

rows = 0; %Initializing variable to count up total number of rows.

for i = 1:length(Files);

 eval(dispatch(['load ',char(Files(i)),' OutQMRA;'])) %Loads the 'OutQMRA' struct stored in .mat file, overwriting that object if it exists.

```

        %disp(['File ',num2str(i),' took ',num2str(), ' to run
',num2str(length(OutQMRA.CaL)),...
        '% loops ',num2str(), ' per loop.'])
        eval(dispatch(['OutQMRA',num2str(i),'=OutQMRA;'])) %Copies it and adds a numeric
suffix to the name.
        rows = rows + length(OutQMRA.EcL);
end
clear OutQMRA; %Removes the initial copy of the last file loaded.

%CSVmatrix = NA(rows,length(fieldnames(OutQMRA1))-2); %Creating the output matrix. Each
row is a QMRA iteration.
for i = 1:length(Files);
    %i %For debugging
    eval(dispatch(['OutQMRA = OutQMRA',num2str(i),';'])) %Taking 'OutQMRAx' and creating a
copy called 'OutQMRA' to work from.
    OutQMRA = rmfield(OutQMRA, 'StartTime'); OutQMRA = rmfield(OutQMRA, 'EndTime');
    CSVmatrix = OutQMRA.Fit'; %Initializing a matrix that will become a .CSV by
transposing the first structure field into it.
    for [val,key] = OutQMRA; %This special syntax allows looping over all elements of
the structure.
        %key %For debugging
        if strcmp(char(key),'Fit') == 0; %Don't do anything for the 'Fit' element
because we took care of that 2 lines before.
            CSVmatrix = [CSVmatrix, val']; %Transpose fields into columns & bind
into the matrix.
        end
    end
    eval(dispatch(['CSVmatrix',num2str(i),' = CSVmatrix;']))
    clear CSVmatrix;
end

CSVmatrix = CSVmatrix1; %Initializing output matrix.
for i = 2:length(Files);

```

```

        eval(dispatch(['CSVmatrix = [CSVmatrix; CSVmatrix',num2str(i),'];']))
end

fn=fieldnames(OutQMRA);
nFields = numel(fn);      %http://stackoverflow.com/questions/5292437/how-to-concat-cell-
array-of-strings-in-matlab
fn(1:nFields-1) = strcat(fn(1:nFields-1),{' ',''});
file = fopen(filename,'w+');
fprintf(file,'%s',disp([fn{:}]));
fclose(file);
eval(["dlmwrite('",char([filename]),"',CSVmatrix,'-append');"])
disp(['Done; .mat files have been merged and output to ',char(filename),' in
',char(pwd),' /Results/'])

```

9.5. The EITS model (chapter 5)

The transmission model (referred to as “EITSd” for short) consists of several text files containing necessary functions and subroutines; the core program is 'EITSd.m'. Simulation options are set by the choice of several values at the top of the file 'EITSd.m'. These options default to values that generate a single test run of the simulation. Other options are set when calling the function 'EITSd.m' and are described within that file; for example, the following can be submitted at the Octave (or MATLAB) prompt to do one test calibration run:

```
EITSd('C',1,'test.csv');
```

The source code is below. The filename of each of the source code files is found in the copyright information at the top of each file. Although EITSd uses some filenames that are identical to those in QMRAv13_20110414 or QMRA2v5, the content of its files differs. EITSd also requires folders named 'Graphics' and 'Results' in the working directory in order to store output.

All files in the source code that follows are part of EITSd and subject to the GNU General Public License Version 3, except for `erdrey.m`, which is part of the CONTEST toolbox at http://www.mathstat.strath.ac.uk/research/groups/numerical_analysis/contest (A. Taylor & Higham, 2009); `erdrey.m` is included at the end of the source code for completeness, by permission of the authors (Des Higham, personal communication, 24 Aug. 2012).

%Environmental transmission model of diarrheal infection transmission (main file), Kyle S. Enger, July 2012

%{
COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (EITSd.m) is part of EITSd.

EITSd is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

EITSd is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with EITSd. If not, see <<http://www.gnu.org/licenses/>>.

%}

%EITSd runs on Octave 3.2 or later.

%Also works well (and runs faster) on MATLAB; the results in chapter 5 of the accompanying dissertation were all produced with MATLAB.

%If running on MATLAB, requires the statistics toolbox, and possibly others.

%This code (EITSd.m, the core component of EITSd) is accompanied by several functions/subroutines, which are part of EITSd and need to be in the working directory:

% ApplyLRVs.m: Applies log10 reduction values to stocks of pathogens

% AssignCompliance.m: Given compliance parameters for a community, assigns specific compliance characteristics to specific households

```

% DoseResponse.m:      Applies dose response functions to individuals
% CalLoopFuncCompile.m: Code that calls EITSd() in order to facilitate parallel
processing of many differently parameterized calibration runs
% DRbP.m:              Beta-Poisson dose response model
% DRexp.m:             Exponential dose response model
% durEc.m:             Randomly pick a duration for E. coli infection
% durGi.m:             Randomly pick a duration for Giardia infection
% durRo.m:            Randomly pick a duration for rotavirus infection
% EstLoopFuncCompileBase.m: Code that calls EITSd() in order to facilitate parallel
processing of many differently parameterized baseline (no HWT) estimation runs
% EstLoopFuncCompileHWT.m: As above, but for estimation runs using HWT with imperfect
compliance
% EstLoopFuncCompileHWTPC.m: As above, but for estimation runs using HWT with perfect
compliance
% Inact.m:             Inactivates pathogens in all compartments
% PlotMicrobes.m:     Generates line charts of flows of microbes during a single
model run
% PlotPeople.m:       Generates line charts of people and their infection statuses
during a single model run
% Pooping.m:          Determines results of defecation events by infected people

%Also requires in the working directory: erdrey.m, written by Alan Taylor and Des Higham,
U. of Strathclyde,
% freely available at
http://www.mathstat.strath.ac.uk/research/groups/numerical\_analysis/contest,
% and reproduced and included with EITSd by permission of the authors, though it is not a
part of EITSd).

%Use just the 1st 3 arguments to run the model with default values for all parameters.
%CalibOrEst: Whether the model is in calibration ('C') or estimation ('E') mode.
%nRuns: Number of runs (calibration), or number of runs per parameter set (estimation).
%outFilename: Filename for output file containing results.
%inFilename: Used for estimation only. Filename root for .CSV containing parameter values

```

```

from calibration step.
%CF variables: Calibration factors; 1st value is low end of range, 2nd value is high end
of range.
%CFdecayr:          2x3 matrix: top row is low end of range of decay constant modifiers;
bottom row is high end of range. Columns are bac.-vir.-prot.
%nHHr:      Value to replace default for number of households (nHH).
%E variables: LRVs and compliance figures for various intervention scenarios in the
estimation step. Estimation necessitates input of garbage for the CF ranges above.

%dbstop(161); %Setting breakpoint for debugging. Type 'dbquit' at the debug prompt to
quit back to Octave prompt.

function [People LogHHP Log nMw nMl OutMatrix] = EITSd(CalibOrEst, nRuns, outFilename,
inFilename, CFSfr, CFHDhr, CFVr, CFDlr, CFdecayr, nHHr, ElSan, EcSan, ElHWT, EchWT,
ElHand, EchHand);
if size(ver('Octave'),1)==0; format compact; format longe; end %Fixes annoying default
display characteristics in MATLAB.
if nargin == 3; disp(['Running with default parameter values only.']); end
if CalibOrEst == 'C';
    loops = nRuns;
    OutMatrix = NaN(loops,87); %Starting output matrix, summarizing results from many
model runs, one row per run.
elseif CalibOrEst == 'E'; %Pull in parameter values from calibration runs that fit the
criteria.
    TrialParams = dlmread(inFilename,',',1,1); %Access a file returned by
ReadOutput.r. inFilename is an input to EITSd.m.
    paramSets = size(TrialParams, 1);
    if paramSets > 100 & sum([ElSan ElHWT ElHand]) > 0; %If >100 parameter sets and
an intervention is being applied, randomly sample 100 sets without replacement, to use in
estimation step. TODO: Conflicts with sampling of 150 parameter sets in calibration R
code.
        selectedParamSets = randperm(paramSets);
        selectedParamSets = selectedParamSets(1:100);
    end
end

```

```

        TrialParams = TrialParams(selectedParamSets,:);
        paramSets = size(TrialParams,1);
    end
    %Replicating the parameter sets.
    TP = TrialParams;    %Making a copy, for use in loop below.
    if nRuns > 1;
        for i = 1:nRuns-1;
            TrialParams = [TrialParams; TP];    %Appending copies of the parameter
sets.
        end
    end
    TrialParams = sortrows(TrialParams);    %Sorting on basis of ID number, so that runs
with identical CF values are next to each other.
    loops = size(TrialParams,1);
    OutMatrix = NaN(loops,88);    %Starting output matrix, summarizing results from many
model runs, one row per run. 88th column is the ID of the set of CFs input.
    disp(['Using ',num2str(paramSets),' parameter sets, ',num2str(nRuns),' times each,
totaling ',num2str(loops),' runs.'])
elseif CalibOrEst == 'D';    %Running with default parameters only.
    loops = nRuns;
end
OutMatrix(:,1) = 1:loops;    %Serially numbering rows.
Octave = size(ver('Octave'),1); %Indicator of whether the code is running under Octave
(1) or Matlab (0).
p.ProgramVersion=7; p.StartTime=datestr(now,31);
%===Parameter values (struct of parameters created simultaneously). Start with defaults
for parameters that vary in calibration step.
%loops = nRuns;    %Number of distinct runs with a given set of
parameters.
CFSf = 1E-8; p.CFSf=CFSf;    %Calibration factor: microbe transfer from surface water to
stored drinking water
CFHDh = 0.01; p.CFHDh=CFHDh; %Calibration factor: microbe transfer out of household
environment to stored water or adults (kids are higher; see wHandMouth).

```



```

CFV = 0.066; p.CFV=CFV;           %Calibration factor: microbe transfer between households
when a visit occurs (bidirectional). Rota, skin-skin, Ansari 1988.
CFDl = 1E-9; p.CFDl=CFDl;         %Calibration factor: proportion of land pathogens ingested
per adult (kids are higher; see wHandMouth).
CFdecay = [.6 .6 .6];             %Calibration factor: multiplier applied to inactivation
rates in water to convert them to inactivation rates out of water.
%Microbe-specific parameters [bacteria, viruses, protozoans]:===
Mpgf = [5E8 2E6 572000]; p.Mpgf=Mpgf; %Microbes per gram of feces (DuPont 1971, Ward
1984, Danciger 1974).
%Mpgf = [0 0 0]; p.Mpgf=Mpgf;      %For code verification (checking the development
of baseline infections).
KorN50 = [2111912, 6.171, 0.01982]; p.KorN50=KorN50; %Exponential k parameter or beta-
Poisson N50 parameter
alpha = [0.1549, 0.2531, NaN]; p.alpha=alpha; %Presence/absence of alpha
value determines beta-Poisson or exponential dose resp.
MRk = [0.214, 0.36, 0.59]; p.MRk=MRk; %Morbidity ratios for kids.
MRa = [0.214, 0.222, 0.03]; p.MRa=MRa; %Morbidity ratios for adults. Can't find a good
adult MR for E. coli, so reusing the MR for kids.
latent = [round(1.75), round(3.2), round(13.5)]; p.latent=latent; %Latent
(incubation) period.
infill = [round(3.4), round(2.5), round(18.3)]; p.infill=infill; %Duration of
infectivity (assumed the same as duration of disease). NOT USED except for determining
burn-in; see durEc(), durRo(), & durGi().
immune = [1, 1, 1]; p.immune=immune; %Immune period (assumed).
%gM = [0.6 0.6 0.6]; p.gM=gM; %Daily inactivation rate of microbes in surface
water (assume similar for soil, stored water, & hands) - see Inactivation.xls &
EITSpams.ods. CFdecay used to modify these for application to land & household
environment compartments.
%xHtoW = [0.26 0.26 0.26]; p.xHtoW=xHtoW; %Transfer from hands to water, per contact.
Half of reduction for E. coli from Pickering 2011.
%p.xHtoP=xHtoP = [0.34 0.34 0.34]; %Transfer from hands to person's mouth, per contact.
Rusin 2002 (assuming protozoa same as bac & viruses). Since so high & hand-mouth contacts
so frequent, assume all are ingested each day.

```

```

%xferVisit = 0.066; p.xferVisit=xferVisit;          %Proportion of hand pathogens that are
transferred (2-way) during a visit. Rota, skin-skin, Ansari 1988.
lSan = [3 3 3]; p.lSan=lSan;                          %LRVs attributable to sanitation. Based on
amount of feces on toilet paper (which often isn't flushed).
lHWT = [6 4 3]; p.lHWT=lHWT;                          %LRVs attributable to household water
treatment.
lHand = [0.46, 0.46, 0.46]; p.lHand=lHand; %LRVs attributable to handwashing; Luby 2001,
TTC. Slightly higher than Pickering 2011.
%Community characteristics (people & households):
%nHH = 1000; p.nHH=nHH;                               %Approximate number of households in the
simulated community
nHH = 200; p.nHH=nHH;                               %Approximate number of households in the
simulated community
mPHH = 5; p.mPHH=mPHH;                               %Mean people per household (truncated Poisson)
pKids = 0.18; p.pKids=pKids;                          %Proportion of people in the community who are <5y.
meanDeg = 5.3; p.meanDeg=meanDeg; %Mean network degree for households in 18 villages,
'passing time' network, Joe's Ecuador sites (Zelner 2012).
%Water intake and defecation output parameters:
Ws = 25; p.Ws=Ws;                                     %Size of household water storage container, in liters.
Wflow = 0; p.Wflow=Wflow; %Unused. Simple in/outflow of reservoir, daily rate,
starting with 0 (maybe 25 ft^3/sec, USGS creek measurements).
landArea = 1; p.landArea=landArea; %Unused. Community area, km^2. Presuming people only
poop within their community. Allows calc. of persons per square km.
Wdd = [1.178, 2.3]; p.Wdd=Wdd; %L H2O drunk daily. Akpata 2004, Nigerian children;
Fudge 2008, Kenyan runners (agrees w. USEPA 2011, p. 100).
fpp = [109.3, 225]; p.fpp=fpp; %Grams of feces excreted daily; Nigerian children and
adult British vegans (both with high-fiber diets).
fHands = 0.23; p.fHands=fHands; %Grams of feces on fingers after defecation
event. Based on daily feces on toilet paper. See EITSParams.ods.
pPoopH2O = 0; p.pPoopH2O=pPoopH2O; %Unused. Probability that a defecation event goes
straight into the surface water (i.e., not on land).
%Compliance parameters. 1st element is overall compliance, 2nd is compliance type:
1=alpha, 2=beta, 3=gamma.

```

```

cSan = [0, 1]; p.cSan=cSan;           %Compliance with sanitation. Defaults to no
compliance; will be varied.
cHWT = [0, 1]; p.cHWT=cHWT;          %As above, but HWT.
cHand = [0, 1]; p.cHand=cHand;       %As above, but handwashing.
%Now, basic daily rates for some key events. All rates are per day.
rH2ODrink = 6; p.rH2ODrink=rH2ODrink; %Daily drinks from stored water per person.
rPoopInf = 1; p.rPoopInf=rPoopInf;   %Defecation events, per non-ill person
rPoopIll = 3; p.rPoopIll=rPoopIll;   %Defecation events, per ill person
rVisit = 2/7; p.rVisit=rVisit;      %Visits/contact/day (each contact is a pair
of households - see adjacency list generated below, based on time spent in past week
[Zelner 2012]).
rRain = 1/14; p.rRain=rRain;         %Roughly fortnightly rainstorm
xRunoff = [0.001, 0.05]; p.xRunoff=xRunoff; %Daily transfer of pathogens from land to
surface water without rain, and with rain.
wHandMouth = 330 / 130; p.wHandMouth=wHandMouth; %Hand-mouth contacts/day for kids
divided by hand-mouth contacts/day for adults (USEPA 2011). Weights kids as being more
unhygienic.
BaseInf = [1.17 0.347 0.212]; %Baseline infections per person-year. Chosen to give 0.5
diarrheal episodes per child-year (total; half bac., 25% vir. & prot.) when distributed
randomly among the population (CalcBaselineInfectionIncidence.ods).
%===End parameter values; now simulation options===
tMax = 365; p.tMax=tMax;              %Number of simulation days desired; daysBurnIn is
then added to it.
daysBurnIn= ceil(max(immune) + max(latent) + max(infill)) * 3; p.daysBurnIn=daysBurnIn;
tMax = tMax + daysBurnIn;
pSafeStorage = 0; p.pSafeStorage=pSafeStorage; %Currently can only be 1 or 0.
Proportion of the community that has safe (household water) storage. Toggles presence of
safe storage for all HWT users in the community.
nMl = [0 0 0]; p.nMl=nMl;            %Initial numbers of microbes on the land
nMw = [0 0 0];                       %Initial number of microbes in the reservoir
if loops > 3; storeDetails = 0; else storeDetails = 1; end; %If too many loops, don't
bother storing detailed output.
SummaryResults = NaN(loops,8);        %Store summary data from all runs. TODO: fully

```

```

implement?
disp(sprintf('Running %s loops.',num2str(loops)))
for l = 1:loops;    %Allows multiple model runs. Loops once per run.
    disp(sprintf('STARTING RUN %s',num2str(l)))
    t=0;           %Initialize time counter
    LogIncInf = NaN(tMax,9); %Stores daily infection incidence (count of new infections)
for the 3 pathogens: 1:3, kids; 4:6, adults; 7:9, all.
    LogIncIll = NaN(tMax,9); %Stores daily illness incidence (count of new illnesses)
for the 3 pathogens: 1:3, kids; 4:6, adults; 7:9, all.
    Log = NaN(tMax,33);      %Logs daily status summary for all people.
    LogK = NaN(tMax,18);     %Logs daily status summary for all kids.
    LogA = NaN(tMax,18);     %Logs daily status summary for all adults.
    if storeDetails == 1;
        LogHHP = cell(tMax,2); %Stores HHs and People matrices daily.
        LogFlux = NaN(tMax,3,11); %Cube to store fluxes of microbes. z is 1:11 (1-
surface water to stored water at resupply; 2-overall visit transfer; 3-not used, but
formerly land-to-hand-to stored water at drinking; 4-rainfall; 5&6-pooing into surface
H2O & land; 7-inactivation; 8&9-kids' dose, water & hands; 10&11-adults' dose, water &
hands).
        tRain = [(1:tMax)',NaN(tMax,1)]; %Logs times of rain events.
    end
    extinct=0; equilibrium=0; %Initialize extinction & equilibrium flags.

    %===Generating the community. Yields a matrix for tracking households and a matrix
(or struct, 1 item per hh?) for tracking persons.===
    %HHs has 16 columns: counts of persons, adults, and kids (1:3); counts of microbes
on hands (4:6), stored water (7:9), food (10:12), compliance with sanitation, HWT, and
handwashing (13:15), and amount of water currently stored in the household (16).
    containerOK = 0;
    while containerOK == 0; %The 'while' loop ensures that no HH has more people than
its stored water container can supply. See the warning near the end.
        if nargin >= 9; nHH = nHHr; end %Overriding default number of households
        HHs = poissrnd(mPHH,ceil(nHH * (1 + poisspdf(0,nHH))),1); %Generates extra

```

```

households, since some will have 0 persons & will be thrown away.
    HHS = sort(HHS(find(HHS > 0)));
    nHH = size(HHS,1); %Actual number of households will probably be slightly
lower than the inputted number.
    nPeople = sum(HHS);
    nKids = round(nPeople * pKids);
    nAdults = nPeople - nKids;
    HHS(:,2) = 1; %2nd column is the count of adults. Every household has at least
1 adult. TODO: This will break if nAdults < nHH, but this is extremely unlikely.
    aA = nAdults - nHH; %Adults that still need to be assigned to a household.
    while aA > 0;
        HHS(:,3) = HHS(:,1) - HHS(:,2); %3rd column tracks the number of empty
person-slots (potential children) remaining in each household.
        candidates = find(HHS(:,3) > 0);
        picked = candidates(ceil(unifrnd(eps,size(candidates,1),1))); %This is
like Matlab's randsample(), which Octave lacks.
        HHS(picked,2) = HHS(picked,2) + 1;
        aA = aA - 1;
    end
    HHS(:,3) = HHS(:,1) - HHS(:,2); %Now, 3rd column becomes the number of
kids in each household.
    Wneeded = sum(HHS(:,2:3) .* repmat([Wdd(2),Wdd(1)],nHH,1),2); %Now error-
checking to make sure no HH exhausts their stored H2O.
    containerOK = 1;
    if sum(Wneeded > Ws) ~= 0;
        containerOK = 0;
        warning(sprintf('At least 1 household will drink > %sL daily.
Retrying...',num2str(Ws)))
    end
end
    HHS = horzcat(HHS,zeros(nHH,13)); %Adding 9 fields for the 3 marker pathogens on
hands (4:6), water (7:9), and food (10:12) in each household.
    %Also adding 3 fields to track usage (an aspect of compliance) of sanitation, HWT,

```

```

and handwashing (13:15), and 1 field to track stored water (16).
    HHs(:,13) = AssignCompliance(HHs(:,13),cSan);    %This household uses sanitation X
of the time. If X > 0, it owns a latrine.
    HHs(:,14) = AssignCompliance(HHs(:,14),cHWT);    %This household treats their water
X of the time. If X > 0, it has a HWT method.
    HHs(:,15) = AssignCompliance(HHs(:,15),cHand);    %This household washes hands X of
the time. If X > 0, it has enough soap/water.
    HHs(:,17) = pSafeStorage * (HHs(:,14) ~= 0); %All households using HWT (partially or
perfectly) have the same safe storage status.
    %HHs(:,17) = binornd(1,pSafeStorage,nHH,1); %Randomly assigning whether the
household has safe storage. TODO: binary, SS or not, village-wide?
    %Now generating matrices for tracking each adult and each child.
    Adults = zeros(nAdults,8);    %Adults, with disease status (1:3) & status counter
(4:6) for the 3 marker pathogens. 7th column designates household.
    %Disease status: -1, susceptible; 0, immune; 1, exposed; 2, infected; 3, diseased
    iA = 1;    %Counter for rows in adults.
    for i = 1:nHH; %Placing household indices in the 7th column of the matrix, so that
people can be tied to certain households.
        pop = HHs(i,2);
        Adults(iA:iA+pop-1,7) = i;
        Adults(iA:iA+pop-1,8) = HHs(i,1);    %Adds household size to each adult's record.
        iA = iA + pop;
    end
    Kids = zeros(nPeople(1) - nAdults, 8); %As above, for kids.
    iA = 1;
    kidHHindexes = find(HHs(:,3) > 0); %Gets households that have kids.
    for j = 1:size(kidHHindexes,1);    %For all households with kids...
        i = kidHHindexes(j);
        pop = HHs(i,3);
        Kids(iA:iA+pop-1,7) = i;
        Kids(iA:iA+pop-1,8) = HHs(i,1);    %Adds household size to each kid's record.
        iA = iA + pop;
    end
end

```

```

LogicalKids = logical([ones(nKids,1); zeros(nAdults,1)]); %Vector with 1 for each
kid and 0 for each adult. Used later for distinguishing adults from kids.
People = vertcat(Kids,Adults); %Combining into a single matrix, one row per
person, kids on top and adults on the bottom.
People(:,1:3) = 1; People(:,4:6) = ceil(rand(nPeople,3) * max(latent)); %Start
with everyone exposed with everything, with a random latent period; everyone will
therefore develop infection or disease.

%Now choosing particular values for parameters if calibrating.
switch(CalibOrEst);
case 'C'; %TODO: Test.
    if nargin == 3;
        1; %Do nothing, therefore use default parameter values.
    else
        checkCFHDh = max(HHs(:,2) + HHs(:,3) * wHandMouth); %The most a HH
could lose from visits is 1-(1-CFV)^max(connx/HH).
        if 10 ^ CFHDhr(2) * checkCFHDh > 1;
            warning('Upper end of CFHDh range for calibration would result
in more pathogens moving out of the household environment than exist. Adjusting.')
            CFHDhr(2) = log10(1/checkCFHDh);
            if CFHDhr(1) > CFHDhr(2); CFHDhr(1) = CFHDhr(2); end
        end
        if 10^CFVr(2) > 1 - 10^CFHDhr(2); %TODO: This doesn't seem to work.
Currently solving by using ~0.9 for CFVr(2) instead of 1.
            warning('Upper end of CFV range for calibration would result in
more pathogens moving out of the household environment than exist. Adjusting.')
            CFVr(2) = 1 - CFHDhr(2);
            if CFVr(1) > CFVr(2); CFVr(1) = CFVr(2); end
        end
        CFSf = 10^(rand(1) * (CFSfr(2) - CFSfr(1)) + CFSfr(1)); %Uniform
sampling of the surface water to stored drinking water calibration factor (log10 scale).
        CFHDh = 10^(rand(1) * (CFHDhr(2) - CFHDhr(1)) + CFHDhr(1)); %Uniform
sampling of the from-household-env. calibration factor (log10 scale).

```

```

                CFV = 10^(rand(1) * (CFVr(2) - CFVr(1)) + CFVr(1));    %Uniform
sampling of the visits calibration factor (log10 scale).
                CFDl = 10^(rand(1) * (CFDlr(2) - CFDlr(1)) + CFDlr(1));    %Uniform
sampling of the surface water to stored drinking water calibration factor (log10 scale).
                CFdecay = 10 .^ (rand(1,3) .* (CFdecayr(2,:) - CFdecayr(1,:)) +
CFdecayr(1,:));    %Uniform sampling of decay modifiers converting base decay rates to
actual decay rates (log10 scale).
                disp(sprintf('Chosen calib. factors: CFSf=%s, CFHDh=%s, CFV=%s, CFDl=
%s, CFdecayEc=%s, CFdecayRo=%s, CFdecayGi=
%s', num2str(CFSf), num2str(CFHDh), num2str(CFV), num2str(CFDl), num2str(CFdecay(1)), num2str(C
Fdecay(2)), num2str(CFdecay(3))))
            end
            case 'E'; %Using parameters from previous runs consistent with RCT. TODO:
Update code below (copied from Main.m).
                CFsetID = TrialParams(l,1);
                CFSf = TrialParams(l,2);
                CFHDh = TrialParams(l,3);
                CFV = TrialParams(l,4);
                CFDl = TrialParams(l,5);
                CFdecay = TrialParams(l,6:8);
                lSan=ElSan; cSan=EcSan; lHWT=ElHWT; cHWT=EcHWT; lHand=ElHand;
cHand=EcHand; %Reading in chosen LRV and compliance parameters for a particular
intervention scenario.
                otherwise
                error('CalibOrEst must be C (calibration) or E (estimation).');
            end %end switch block

%Now generating dosing weights for ingestion of pathogens on hands (depends on HH
composition). Kids ingest more pathogens than adults. NO LONGER NEED wtDoses since
pathogens in household environment can now persist from day to day (though they still
exponentially decay).
                %wtDoses = HHs(:,2:3) .* repmat([rHandMouth(2), rHandMouth(1)], nHH, 1); %Reversing
wtHandMouth to match cols. 2&3 in HHs.

```



```

%wtDoses(:,3) = sum(wtDoses,2);
%wtDoses(:,1:2) = wtDoses(:,1:2) ./ repmat(wtDoses(:,3),1,2) ./ HHs(:,2:3);
%Gives proportion of microbes ingested from hands by that household's set of adults
and set of children.
%wtDoses(find(isnan(wtDoses))) = 0; %Sets NaN (from dividing by 0) to 0.

chosenOnes = NaN(nPeople,5); %This matrix is used in the defecation step each day.
EffM = zeros(2,9); %This vector stores the number of 'effective'
microbes over the course of the simulation (those that contributed to a new infection,
kids [row 1] & adults [row 2]).

%Connections between households: See Zelner 2012 (in press, AJPH).
nEdges = round(nHH*meanDeg/2);
if nEdges > nHH * (nHH-1) / 2;
    disp('Community is too small to generate network of desired degree. Creating
fully connected network.')
    nEdges = nHH * (nHH-1) / 2;
end
Am = erdrey(nHH,nEdges); %As above, but random graph (Erdos-Renyi). 2nd argument is
the number of edges. Yields a sparse matrix.
[Ax Ay Av] = find(Am); %Generating adjacency list (1 row per connection).
Al = horzcat(Ax,Ay,Av); %Note that all connections in the adjacency list are 2-way,
therefore listed twice.
nA = size(Al,1); %Number of possible visits. Each connection has 2 possible
ways to visit (A visits B, or B visits A).
%Now generating all the random numbers needed during the simulation.
nRandCols = tMax+1; %Random number table (and event log) is pre-
generated based on this. One column per day.
RPoopPlace = rand(nPeople,nRandCols); %Random table for defecation location, each
person, each day. Unused if pPoopH2O==0 (the default).
%RCompHW = rand(nPeople,nRandCols); %Random table for handwashing
compliance, by household. Compliance is the same for all household members, but is
assessed individually.

```

```

    %RCompHWT = rand(nHH,nRandCols);           %Random table for household water
treatment compliance.
    %RCompSan = rand(nPeople,nRandCols);       %Random table for sanitation
compliance, by household. Compliance is the same for all household members, but is
assessed individually.
    RVisit = rand(nA,nRandCols);              %Random table for determining which visits
happen each day.
    RRain = rand(1,nRandCols);                 %Random vector for rainfall events.
    RDRtime = rand(1,nRandCols);              %Random vector for timing of dose response
events within the day.
    RDRpeople = rand(nPeople,3,nRandCols);    %Random cube for determining daily outcome
of dose response for each person & each pathogen. z coordinate is the day.
    RMR = rand(nPeople,3,nRandCols);          %Random cube for determining daily outcome
of morbidity ratios for each person & each pathogen. z coordinate is the day.
    Rbaseline = rand(nPeople,3,nRandCols);    %Random cube for determining baseline
exposures, which later turn into baseline infections. z coordinate is the day.
    %Output sanity check on community size and composition.
    disp(sprintf('Simulated community has %s people/km^2, %s households, %s people
(%s>=5y, %s<5y). Daily water demand is %s L. Smallest HH is %s, biggest is %s, mean %s
people/HH. Min connections per HH is %s, max is %s, mean is %s. Running for %s days,
preceded by %s days of burn-in, total %s
days.',num2str(nPeople/landArea),num2str(size(HHs,1)),num2str(size(People,1)),num2str(siz
e(Adults,1)),num2str(size(Kids,1)),num2str(nHH *
Ws),num2str(min(HHs(:,1))),num2str(max(HHs(:,1))),num2str(mean(HHs(:,1))),num2str(full(mi
n(sum(Am)))),num2str(full(max(sum(Am)))),num2str(full(sum(sum(Am)))/nHH),num2str(tMax-
daysBurnIn),num2str(daysBurnIn),num2str(tMax)));
    %x = disp(['Simulated community has ',num2str(nPeople/landArea),' people/km^2,
',num2str(size(HHs,1)),' households, ',num2str(size(People,1)),' people
(',num2str(size(Adults,1)),'>=5y, ',num2str(size(Kids,1)),'<5y). Smallest HH is
',num2str(min(HHs(:,1))),', biggest is ',num2str(max(HHs(:,1))),', mean
',num2str(mean(HHs(:,1))),' people/HH. Min connections per HH is
',num2str(full(min(sum(Am))),'', max is ',num2str(full(max(sum(Am))),'', mean is
',num2str(full(sum(sum(Am)))/nHH),''. Running for ',num2str(tMax-daysBurnIn),' days,

```

```

preceded by ',num2str(daysBurnIn),' days of burn-in, total ',num2str(tMax),' days.']);
%disp(x)
%==== DAILY LOOP STARTS====
tic
while t < tMax && extinct == 0 && equilibrium == 0 %Keep iterating until max
time is reached or a microbe goes extinct. The 'equilibrium' flag is not currently used.
%while t < tMax %For code verification.
    t=t+1; %Advance to the next day
    phase = 1; %Marker for debugging
    initialHHs = HHs; initialPeople = People; %Storing copies of these matrices
at the beginning of each day, for debugging.
    %if Octave == 1;
    % printf('-%s',num2str(t)); %Printing every day (for debugging)
    %else
    fprintf('%s',num2str(mod(t,10))); %Printing last digit of every day (for
debugging)
    %end
    Sus = People(:,1:3)==-1; %Updating counts of people in various states, for each
pathogen.
    nSus = sum(Sus);
    nSusK = sum(Sus(1:nKids,:));
    nSusA = sum(Sus(nKids+1:nPeople,:));
    Imm = People(:,1:3)==0;
    nImm = sum(Imm);
    nImmK = sum(Imm(1:nKids,:));
    nImmA = sum(Imm(nKids+1:nPeople,:));
    Exp = People(:,1:3)==1;
    nExp = sum(Exp);
    nExpK = sum(Exp(1:nKids,:));
    nExpA = sum(Exp(nKids+1:nPeople,:));
    Infec = People(:,1:3)==2; %Would rather use 'Inf' than 'Infec' here, but it
means +infinity to Octave/Matlab.
    nInf = sum(Infec);

```

```

nInfK = sum(Infec(1:nKids,:));
nInfA = sum(Infec(nKids+1:nPeople,:));
Ill = People(:,1:3)==3;
nIll = sum(Ill);
nIllK = sum(Ill(1:nKids,:));
nIllA = sum(Ill(nKids+1:nPeople,:));
if sum(sum([Sus;Imm;Exp;Infec;Ill]) == [nPeople nPeople nPeople]) ~= 3;
    error('Sum of all possible states ~= total number of people, for at least
1 microbe. ');
end
%Above are 3-element vectors. Now need # of people infected with anything, but
ill with nothing (nInfNotIll), and # of people ill with anything (nIllAny).
test = sum(abs(People(:,1:3)) .^ 3, 2); %This variable distinguishes 'any
infected & non-ill' from 'any ill'.
InfNotIll = test < 27 & test >= 8;
nInfNotIll = sum(InfNotIll);
nInfNotIllK = sum(InfNotIll(1:nKids));
nInfNotIllA = sum(InfNotIll(nKids+1:nPeople));
IllAny = sum(People(:,1:3)==3, 2) > 0;
nIllAny = sum(IllAny);
nIllAnyK = sum(IllAny(1:nKids));
nIllAnyA = sum(IllAny(nKids+1:nPeople));
%Check for extinction of any microbes.
phase = 2;

%The next loop is not needed since extinctions have been avoided through
assigning baseline infections.
%{
if t > daysBurnIn + 1; %This avoids an error (negative incidence) if
extinction happens during initial equilibration period.
    if sum(People(:,1) <= eps) == nPeople & sum([sum(HHs(:,[4 7
10])),nMw(1),nMl(1)]) <= 0.01;
        extinct = 1; disp(['Sim stopped at time ',num2str(t),' due to

```

```

extinction of bacteria']) %Terminates the event loop after this day is done.
    elseif sum(People(:,2) <= eps) == nPeople & sum([sum(HHs(:,[5 8
11])),nMw(2),nMl(2)]) <= 0.01;
        extinct = 2; disp(['Sim stopped at time ',num2str(t),' due to
extinction of viruses']) %Terminates the event loop after this day is done.
    elseif sum(People(:,3) <= eps) == nPeople & sum([sum(HHs(:,[6 9
12])),nMw(3),nMl(3)]) <= 0.01;
        extinct = 3; disp(['Sim stopped at time ',num2str(t),' due to
extinction of protozoa']) %Terminates the event loop after this day is done.
    end
end
%}

%if sum(sum(People(:,1:3) <= eps) == nPeople) >= 1 & (sum([sum(HHs(:,[4 7
10]),nMw(1),nMl(1))]) <= 0.01 | sum([sum(HHs(:,[5 8 11]),nMw(2),nMl(2))]) <= 0.01 |
sum([sum(HHs(:,[6 9 12]),nMw(3),nMl(3))]) <= 0.01) );
%    extinct = 1; disp(['Sim stopped early due to extinction of at least one
pathogen near time ',num2str(t)]) %Terminates the event loop after this day is done.
%end
%{
    if mod(t,50) == 0 && t >= 100; %Test whether system has equilibrated, based
on any infection (without illness). TODO: Never trips. Probably unnecessary anyway.
        typeOne = 0.1; %Test whether last 50 days & previous 50 days are drawn
from the same distribution, with this value of alpha.
        eqPass = 0; %Number of times the test has failed to reject
hypothesis of same distribution.
        if Octave==1; %Octave and Matlab have different function names for the
Wilcoxon rank-sum test.
            [pval,ztest] = u_test(Log(t-99:t-50,17),Log(t-49:t,17));
        else %if running Matlab:
            [pval,reject] = ranksum(Log(t-99:t-50,17),Log(t-
49:t,17),'alpha',typeOne);
        end
end

```

```

        if pval < type0ne;
            eqPass = 0;
        else
            eqPass = eqPass + 1;
        end
        if eqPass==2; %If the test fails to reject twice in a row, set the
equilibrium flag, terminating the event loop:
            equilibrium = 1; disp(['Sim stopped early; equilibrium apparently
reached near time ',num2str(t)])
        end
    end
    end
    %}

    %Defecation (the only source of microbes). First, partitioning shedding people
into 4 categories. TODO: Fully convert to daily fecal output? Maybe change fpp, rPoopIll,
and rPoopInf.
    phase = 3;
    chosenOnes(:,1) = IllAny & LogicalKids;           %Ill kids.
    chosenOnes(:,2) = InfNotIll & LogicalKids;       %Asymptomatic kids.
    chosenOnes(:,3) = IllAny & ~LogicalKids;         %Ill adults.
    chosenOnes(:,4) = InfNotIll & ~LogicalKids;     %Asymptomatic adults.
    chosenOnes(:,5) = ~IllAny & ~InfNotIll;         %Uninfected people.
    if unique(sum(chosenOnes,2)) ~= 1; %TODO: Check this.
        error('Multiple allocation of some people to different categories.')
    end
    nMwOld = nMw; nMlOld = nMl;
    [HHs, nMw, nMl] = Pooping(RPoopPlace(:,t), People, HHs,
Mpgf, fpp(1)*rPoopIll, fHands*rPoopIll, nMw, nMl, pPoopH2O, lHand, lSan, chosenOnes(:,1)); %Ill
kids. Note more poop on hands due to repeated defecation (rPoopIll).
    [HHs, nMw, nMl] = Pooping(RPoopPlace(:,t), People, HHs,
Mpgf, fpp(1)*rPoopInf, fHands, nMw, nMl, pPoopH2O, lHand, lSan, chosenOnes(:,2));
    %Asymptomatic kids.
    [HHs, nMw, nMl] = Pooping(RPoopPlace(:,t), People, HHs,

```

```

Mpgf, fpp(2)*rPoopIll, fHands*rPoopIll, nMw, nMl, pPoopH2O, lHand, lSan, chosenOnes(:,3)); %Ill
adults. Note more poop on hands due to repeated defecation (rPoopIll).
    [HHs, nMw, nMl] = Pooping(RPoopPlace(:,t), People, HHs,
Mpgf, fpp(2)*rPoopInf, fHands, nMw, nMl, pPoopH2O, lHand, lSan, chosenOnes(:,4));
%Asymptomatic adults.
    if storeDetails == 1;
        LogFlux(t,:,5) = nMw - nMwOld;
        LogFlux(t,:,6) = nMl - nMlOld;
    end
    postPoopHHs = HHs; %Making a copy for debugging.

%Next transfer: land to water via rain etc.
phase = 4;
if RRain(t) < rRain;
    Runoff = nMl * xRunoff(2);
    tRain(t,2) = 1; %Flagging this day as a rainy one.
else
    Runoff = nMl * xRunoff(1);
end
nMl = nMl - Runoff; %Removing pathogens from land.
nMw = nMw + Runoff; %Adding pathogens to water.
if storeDetails == 1; LogFlux(t,1:3,4) = Runoff; end

%Now quantifying all possible transfers; then apply them simultaneously (so as
to make them independent).
%Inter-household visits: quantifying microbe exchange.
phase = 5;
chosenVisits = A1(RVisit(:,t) < rVisit,:); %Choosing the visits between
households that actually happen today (conceptualized as 1 person meeting 1 other
person). Throws odd "Undefined function 'visit' for input arguments of type 'char'" error
in MATLAB compiled code, but not in interpreted code.

%The following complexity ('for' loop) is necessary to handle successive visits

```

```

on the same day by the same household.
    HU = unique([chosenVisits(:,1); chosenVisits(:,2)]); %List of households
involved in at least 1 visit
    nHHv = size(HU,1); %Number of
households involved in at least 1 visit
    MH = [HHs(HU,[1 4 5 6]), HU]; MHorig=MH; %Microbes on hands for
all households involved in at least 1 visit. Also includes # people in each household
(column 1) & the index of the household in the HHs matrix (column 5).
    for i = 1:size(chosenVisits,1); %Looping
once over each visit.
    %i
    MHrow1 = find(MH(:,5)==chosenVisits(i,1)); %Microbes on
hands for the 1st HH involved in visit i.
    MHrow2 = find(MH(:,5)==chosenVisits(i,2)); %Microbes on
hands for the 2nd HH involved in visit i.
    Mfrom1 = MH(MHrow1,2:4) ./ MH(MHrow1,1) .* CFV;
    Mfrom2 = MH(MHrow2,2:4) ./ MH(MHrow2,1) .* CFV;
    MH(MHrow1,2:4) = MH(MHrow1,2:4) - Mfrom1 + Mfrom2;
    MH(MHrow2,2:4) = MH(MHrow2,2:4) - Mfrom2 + Mfrom1;
end

    for i = 2:4;
        if abs(sum(MHorig(:,i)) - sum(MH(:,i))) > 1E-3; %Sums will not be
exactly equal due to machine imprecision (TODO: is something else causing this?).
            warning(sprintf('Imbalance in visit transfer, pathogen %s, difference
= %s',num2str(i-1),num2str( sum(MHorig(:,i))-sum(MH(:,i)) ) ))
            end
            if abs(sum(MHorig(:,i)) - sum(MH(:,i))) > 1E-1; error('Visit transfer
imbalance too severe! Halting.');
```

end

```

    NetVisitSwaps1 = MHorig(:,2:4) - MH(:,2:4); %Converting to a matrix the size
of the Hands columns of the HHs matrix.
```



```

NetVisitSwaps = sparse(repmat(HU',1,3), [ones(1,nHHv), ones(1,nHHv)*2,
ones(1,nHHv)*3], NetVisitSwaps1(:), nHH, 3); %TODO: Doublecheck.
if storeDetails == 1; LogFlux(t,1:3,2) = sum(abs(NetVisitSwaps)); end
HHs(:,4:6) = HHs(:,4:6) - NetVisitSwaps; %Applying net microbe exchange
from visits.

```

%TODO: Check here that the above transfers are positive? Currently there is a similar check within Inact().

```

%===Resupplying stored water, handling rainfall, shedding, and swapping
pathogens.===
phase = 6;
nMl = nMl + sum(HHs(:,7:9)); %Dumping out remaining water; pathogens go to the
land. This should be a very small flow.
storedWater = repmat(nMw * CFSf, nHH, 1);
nMw = nMw - sum(storedWater); %Microbes in stored water are removed from
surface water.
HHs(:,7:9) = storedWater; %Household water is resupplied, at source water
conc. of microbes.
if storeDetails == 1; LogFlux(t,1:3,1) = sum(storedWater); end;
HHs(:,7:9) = ApplyLRVs(HHs(:,7:9), LHWT, HHs(:,14)); %Determining compliance
and applying LRVs from HWT.

```

%Hand-to-stored-water transfer: applying microbe movement. Contamination of stored water by hands (from drinking, or other decanting of water). Note: not simultaneous with actual water intake (though perhaps it should be; see DR below).

```

phase = 7;
MfromHtoW = HHs(:,4:6) .* CFHDh .* repmat(~HHs(:,17),1,3); %Calculating
microbes transferred from hands to water within each HH. Do not need to consider # of
people/HH at this step (more people automatically mean more hand contamination since they
all defecate). Safe storage negates transfer (last term).
HHs(:,4:6) = HHs(:,4:6) - MfromHtoW; %Removing microbes from hands within
each household.

```

```

        HHs(:,7:9) = HHs(:,7:9) + MfromHtoW;    %Adding microbes to water within each
household.

        if storeDetails == 1; LogFlux(t,1:3,1) = LogFlux(t,1:3,1) + sum(MfromHtoW);
end %Adding drinker hand contamination to water-gatherer hand contamination.
    %{
        for i = 1:3; %Loop over pathogens
            if sum(People(:,i) ~= -1) == 0 & sum(HHs(:,i+3)) > sum(initialHHs(:,i+3));
%If all are susceptible and overall hand pathogens somehow increase:
                warning('Unexplained increase in hand compartments')
                warnHHs = HHs; warnPeople = People;
                equilibrium = 1; %Terminates simulation while allowing charts to
happen.
            end
        end
    %}

%===Inactivating pathogens in all compartments, over the first part of a day.
phase = 8;
PreSink = sum([HHs(:,4:6); HHs(:,7:9); nMw; nMl]);
[HHs, nMw, nMl] = Inact(RDRtime(t), HHs, nMw, nMl, CFdecay, Wflow, t);
if storeDetails == 1; LogFlux(t,:,7) = PreSink - sum([HHs(:,4:6); HHs(:,7:9);
nMw; nMl]); end

%Daily bookkeeping here, just before status shifts & dose response - storing
summary of simulation state and checking for problems.
phase = 9;
Log(t,1:24) = [t nSus nImm nExp nInf nIll nInfNotIll nIllAny nMw nMl];
%Storing aggregated status of all people.
Log(t,25:27) = sum(HHs(:,4:6)); %Total microbes on hands in all households.
Log(t,28:30) = sum(HHs(:,7:9)); %Total microbes in stored water in all
households.
LogK(t,1:18) = [t nSusK nImmK nExpK nInfK nIllK nInfNotIllK nIllAnyK];

```

```

%Storing aggregated status, kids only.
    LogA(t,1:18) = [t nSusA nImmA nExpA nInfA nIllA nInfNotIllA nIllAnyA];
%Storing aggregated status, adults only.
    if storeDetails == 1;
        LogHHP{t,1} = HHs; LogHHP{t,2} = People;
    end

    %For each person, increment all counters and assign durations if a status
shifts. TODO: Sum 'Shifters' and log these state transitions daily, to output incidence.
    phase = 10;
    People(:,4:6) = People(:,4:6) - 1;
    Shifters = People(:,4:6) == 0 & People(:,1:3) == 1;
        %People whose latent period has just expired...
    PeopleStates = People(:,1:3); PeopleStates(Shifters) = 2; People(:,1:3) =
PeopleStates;
        %...become infected & infectious...
    People(:,4:6) = People(:,4:6) + Shifters .* [durEc(nPeople), durRo(nPeople),
durGi(nPeople)];
        %...and are assigned durations of infection...
    LogIncInf(t,1:3) = sum(Shifters(1:nKids,:));
        %...and tallies of these are stored...
    LogIncInf(t,4:6) = sum(Shifters(nKids+1:nPeople,:));
    LogIncInf(t,7:9) = sum(Shifters);
    NewlyDiseased = Shifters & RMR(:, :, t) < [repmat(MRk, nKids, 1);
repmat(MRa, nAdults, 1)];
        %...but some of the newly infected are also randomly
diseased...
    PeopleStates = People(:,1:3); PeopleStates(NewlyDiseased) = 3; People(:,1:3) =
PeopleStates;
        %...and receive 'diseased' status...
    LogIncIll(t,1:3) = sum(NewlyDiseased(1:nKids,:));
        %...and tallies of these are stored.
    LogIncIll(t,4:6) = sum(NewlyDiseased(nKids+1:nPeople,:));
    LogIncIll(t,7:9) = sum(NewlyDiseased);
    Shifters = People(:,4:6) == 0 & (People(:,1:3) == 2 | People(:,1:3) == 3);
        %People whose period of infection/disease has just expired...
    PeopleStates = People(:,1:3); PeopleStates(Shifters) = 0; People(:,1:3) =

```

```

PeopleStates;          %...become immune...
    People(:,4:6) = People(:,4:6) + Shifters .* repmat(immune,nPeople,1);
    %...and are assigned a duration of immunity.
    Shifters = People(:,4:6) == 0 & People(:,1:3) == 0;
    %People whose immunity has just expired...
    PeopleStates = People(:,1:3); PeopleStates(Shifters) = -1; People(:,1:3) =
PeopleStates;          %...become susceptible...
    PeopleCounters = People(:,4:6); PeopleCounters(Shifters) = 0; People(:,4:6) =
PeopleCounters;       %...and their time counters are set to 0.
    %===Assessing dose response (susceptibles become exposed).===
    phase = 11;
    Doses = zeros(nPeople,9); %Dose matrix, 1 row/person. Columns 1:3, water;
columns 4:6, land; columns 7:9, household environment.
    Doses(1:nKids,1:3) = HHs(People(1:nKids,7),7:9) ./ Ws .* Wdd(1); %Dose from
water, kids.
    Doses(nKids+1:nPeople,1:3) = HHs(People(nKids+1:nPeople,7),7:9) ./ Ws .*
Wdd(2); %Dose from water, adults.
    dHHk = HHs(:,7:9) ./ Ws .* Wdd(1) .* repmat(HHs(:,3),1,3); %Determining kid
doses from stored water for each HH.
    dHHa = HHs(:,7:9) ./ Ws .* Wdd(2) .* repmat(HHs(:,2),1,3); %Determining adult
doses from stored water for each HH.
    HHs(:,7:9) = HHs(:,7:9) - dHHk - dHHa; %Actually removing the doses.

%Doses(1:nKids,4:6) = repmat(rHandMouth(1) * pLandHand * nMl, nKids,1);
Doses(1:nKids,4:6) = repmat(nMl .* CFdl .* wHandMouth, nKids,1);
Doses(nKids+1:nPeople,4:6) = repmat(nMl .* CFdl, nAdults,1);
nMl = nMl - sum(Doses(:,4:6)); %Removing land doses ingested by people from
the land.

%Doses(1:nKids,7:9) = HHs(People(1:nKids,7),4:6) .*
repmat(wtDoses(People(1:nKids,7),1),1,3); %Dose from hands, kids.
Doses(1:nKids,7:9) = HHs(People(1:nKids,7),4:6) .* CFHDh .* wHandMouth;
%Dose from household environment, kids.

```

```

%Doses(nKids+1:nPeople,7:9) = HHs(People(nKids+1:nPeople,7),4:6) .*
repmat(wtDoses(People(nKids+1:nPeople,7),2),1,3); %Dose from hands, adults.
%Doses(nKids+1:nPeople,7:9) = HHs(People(nKids+1:nPeople,7),4:6) .* CFHDh;
%Dose from household env., adults.
DosesHHenv =
[accumarray(People(:,7),Doses(:,7)),accumarray(People(:,7),Doses(:,8)),accumarray(People(
(:,7),Doses(:,9))]; %Converting dose matrix from 1 row per person to 1 row per household.
TODO: Doublecheck.
HHs(:,4:6) = HHs(:,4:6) - DosesHHenv; %Removing pathogens ingested from
household environment.

CDoseH2O = sum(Doses(1:nKids,1:3));
if storeDetails == 1; LogFlux(t,:,8) = CDoseH2O; end
CDoseL = sum(Doses(1:nKids,4:6));
if storeDetails == 1; LogFlux(t,:,9) = CDoseL; end
CDoseH = sum(Doses(1:nKids,7:9));
if storeDetails == 1; LogFlux(t,:,9) = CDoseH; end
pDoseH2OK = CDoseH2O ./ (CDoseH2O + CDoseL + CDoseH); %Proportion of total
community-wide doses from water (kids). TODO: Store for later inspection/analysis.
CDoseH2O = sum(Doses(nKids+1:nPeople,1:3));
if storeDetails == 1; LogFlux(t,:,10) = CDoseH2O; end
CDoseL = sum(Doses(nKids+1:nPeople,4:6));
if storeDetails == 1; LogFlux(t,:,11) = CDoseL; end
CDoseH = sum(Doses(nKids+1:nPeople,7:9));
if storeDetails == 1; LogFlux(t,:,11) = CDoseH; end
pDoseH2OA = CDoseH2O ./ (CDoseH2O + CDoseL + CDoseH); %Proportion of total
community-wide doses from water (adults). TODO: Store for later inspection/analysis.
%DosesT = Doses(:,1:3) + Doses(:,4:6) + Doses(:,7:9); %Summing doses from
water & hands for each person.
PeopleOld = People; %Copying People in order to assess later how many pos DR
events are about to occur.
[People EffMnew] = DoseResponse(RDRpeople(:, :, t), People, nKids, Doses, KorN50,
alpha, latent); %Determining dose response for each person. Includes latent period

```

```

assignment (TODO: randomize?).
    if t >= daysBurnIn; EffM = EffM + EffMnew; end;    %Adding today's effective
microbes to the tally.

    %Assigning baseline exposures, which will develop into infections after the
latent period expires.
    Shifters = Rbaseline(:, :, t) < repmat(BaseInf/365, nPeople, 1) & People(:, 1:3) ==
-1;    %People randomly chosen to get a baseline infection...
    PeopleStates = People(:, 1:3); PeopleStates(Shifters) = 1; People(:, 1:3) =
PeopleStates;    %...are exposed...
    PeopleCounters = People(:, 4:6); PeopleCounters(Shifters) = 0; People(:, 4:6) =
PeopleCounters;    %...their appropriate counter(s) (which are neg.) get set to 0...
    People(:, 4:6) = People(:, 4:6) + Shifters .* repmat(latent, nPeople, 1);
    %...and are assigned a latent period.
    Log(t, 31:33) = sum(Shifters);

    %===Done with dose response and baseline exposures - now inactivating over the
remainder of the day.
    phase = 12;
    PreSink = sum([HHs(:, 4:6); HHs(:, 7:9); nMw; nMl]);
    [HHs, nMw, nMl] = Inact(1-RDRtime(t), HHs, nMw, nMl, CFdecay, Wflow, t);
    if storeDetails == 1; LogFlux(t, :, 7) = LogFlux(t, :, 7) + PreSink -
sum([HHs(:, 4:6); HHs(:, 7:9); nMw; nMl]); end
    %===Day's activities are finished. Now for some bookkeeping.===
    phase = 13;
    if t == round(tMax/100);
        fprintf('\n1%% done, %s seconds elapsed, %s days so far, %s sec. per
day\n', num2str(toc), num2str(t), num2str(toc/t));
    elseif t == round(tMax/20);
        fprintf('\n5%% done, %s seconds elapsed, %s days so far, %s sec. per
day\n', num2str(toc), num2str(t), num2str(toc/t));
    elseif t == round(tMax/4);
        fprintf('\n25%% done, %s seconds elapsed, %s days so far, %s sec. per

```

```

day\n',num2str(toc),num2str(t),num2str(toc/t));
    elseif t == round(tMax/2);
        fprintf('\n50%% done, %s seconds elapsed, %s days so far, %s sec. per
day\n',num2str(toc),num2str(t),num2str(toc/t));
    elseif t == round(3*tMax/4);
        fprintf('\n75%% done, %s seconds elapsed, %s days so far, %s sec. per
day\n',num2str(toc),num2str(t),num2str(toc/t));
    end
    if Octave == 1; fflush(stdout);    end %Forces lines above to write to console.
end %Ends daily loop
%disp(x) %Redisplaying the 'sanity check' from the end of the community setup
phase.
disp(sprintf('\nRun complete, %s sec., %s sec./day, day %s, now outputting &
charting',num2str(toc),num2str(toc/tMax),num2str(t)));
tObsY = (t - daysBurnIn)/365; %Time observed, in years.
IncInfK = sum(LogIncInf(daysBurnIn:t,1:3)) ./ (nKids * tObsY);
IncInfA = sum(LogIncInf(daysBurnIn:t,4:6)) ./ (nAdults * tObsY);
IncInfP = sum(LogIncInf(daysBurnIn:t,7:9)) ./ (nPeople * tObsY);
IncIllK = sum(LogIncIll(daysBurnIn:t,1:3)) ./ (nKids * tObsY);
IncIllA = sum(LogIncIll(daysBurnIn:t,4:6)) ./ (nAdults * tObsY);
IncIllP = sum(LogIncIll(daysBurnIn:t,7:9)) ./ (nPeople * tObsY);
disp(sprintf('Child illness incidence %s, adult %s, overall %s, episodes/person-
year, bac-vir-prot.',num2str(IncIllK),num2str(IncIllA),num2str(IncIllP) ));

%Storing output from each run in OutMatrix, one run per row. 1st column is the row
number. Start with calibration variables.
OutMatrix(l,2) = CFSf;
OutMatrix(l,3) = CFHDh;
OutMatrix(l,4) = CFV;
OutMatrix(l,5) = CFDl;
OutMatrix(l,6:8) = CFdecay;
OutMatrix(l,9:11) = Mpgf;
OutMatrix(l,12:17) = [cSan cHWT cHand]; %Compliance.

```

```

    OutMatrix(l,18) = pSafeStorage;    %Whether HWT compliers are using safe storage.
    OutMatrix(l,19:27) = [lSan lHWT lHand]; %Log reduction values.
    OutMatrix(l,28:39) = [IncInfK sum(IncInfK) IncInfA sum(IncInfA) IncInfP
sum(IncInfP)]; %Incidence of infection, bac.-vir.-prot.-total, kids, adults and all
people.
    OutMatrix(l,40:51) = [IncIllK sum(IncIllK) IncIllA sum(IncIllA) IncIllP
sum(IncIllP)]; %Incidence of illness, bac.-vir.-prot.-total, kids, adults and all people.
    OutMatrix(l,52:60) = EffM(1,:);    %Microbes contributing to actual new infections
in kids, by route.
    OutMatrix(l,61:69) = EffM(2,:);    %Microbes contributing to actual new infections
in adults, by route.
    OutMatrix(l,70) = fHands;        %Number of grams of feces on hands per defecation
event.
    OutMatrix(l,71) = extinct;      %Whether 1+ microbes went extinct.
    OutMatrix(l,72) = tObsY; %Amount of time observed.
    OutMatrix(l,73:75) = [nHH nKids nAdults]; %Number of households, kids, and
adults.
    OutMatrix(l,76:78) = mean(Log(daysBurnIn:tMax,19:21)); %Mean numbers of microbes in
surface water (excluding burn-in).
    OutMatrix(l,79:81) = mean(Log(daysBurnIn:tMax,22:24)); %Mean numbers of microbes on
land (excluding burn-in).
    OutMatrix(l,82:84) = mean(Log(daysBurnIn:tMax,28:30)); %Mean numbers of microbes in
stored drinking water (excluding burn-in).
    OutMatrix(l,85:87) = mean(Log(daysBurnIn:tMax,25:27)); %Mean numbers of microbes in
household environment (excluding burn-in).
    if CalibOrEst == 'E';
        OutMatrix(l,88) = CFsetID;    %Storing the ID for the particular calibration
factor set used.
    end
    %StoreParams(p,'OutputLog'); %p is a struct holding all the parameter values.

    if storeDetails == 1 & CalibOrEst == 'C'; %Plotting a set of graphs for each run,
if only a few runs are being done.

```



```

    %Plotting people (kids):
    F2=figure('Position',[100 100 1024 768]);    %Putting graph at (100,100) on
screen with 1024x768 resolution.
    x=.05; y=.72; x1=.93; y1=.23; %Values for positioning of 1st subplot (modify y
for placing 2nd & subsequent subplots).
    subplot(3,1,1) %Plotting bacteria
    PlotPeople(LogK, tRain, x, y, x1, y1, 'bacteria', 'children <5y', daysBurnIn,
nKids);
    subplot(3,1,2) %Plotting viruses
    PlotPeople(LogK, tRain, x, y, x1, y1, 'viruses', 'children <5y', daysBurnIn,
nKids);
    subplot(3,1,3) %Plotting protozoa
    PlotPeople(LogK, tRain, x, y, x1, y1, 'protozoa', 'children <5y', daysBurnIn,
nKids);
    eval(['print Graphics/Kids',strrep(strrep(char(p.StartTime),':','-'),'
','-' ),'.png']) %Saving chart with datetime in filename w/o spaces or colons
    %Plotting people (adults):
    F3=figure('Position',[100 100 1024 768]);    %Putting graph at (100,100) on
screen with 1024x768 resolution.
    x=.05; y=.72; x1=.93; y1=.23; %Values for positioning of 1st subplot (modify y
for placing 2nd & subsequent subplots).
    subplot(3,1,1) %Plotting bacteria
    PlotPeople(LogA, tRain, x, y, x1, y1, 'bacteria', 'people >=5y', daysBurnIn,
nAdults);
    subplot(3,1,2) %Plotting viruses
    PlotPeople(LogA, tRain, x, y, x1, y1, 'viruses', 'people >=5y', daysBurnIn,
nAdults);
    subplot(3,1,3) %Plotting protozoa
    PlotPeople(LogA, tRain, x, y, x1, y1, 'protozoa', 'people >=5y', daysBurnIn,
nAdults);
    eval(['print Graphics/Adults',strrep(strrep(char(p.StartTime),':','-'),'
','-' ),'.png']) %Saving chart with datetime in filename w/o spaces or colons
    %Plotting people (everybody):

```

```

        F4=figure('Position',[100 100 1024 768]);    %Putting graph at (100,100) on
screen with 1024x768 resolution.
        x=.05; y=.72; x1=.93; y1=.23; %Values for positioning of 1st subplot (modify y
for placing 2nd & subsequent subplots).
        subplot(3,1,1) %Plotting bacteria
        PlotPeople(Log, tRain, x, y, x1, y1, 'bacteria', 'people', daysBurnIn,
nPeople);
        subplot(3,1,2) %Plotting viruses
        PlotPeople(Log, tRain, x, y, x1, y1, 'viruses', 'people', daysBurnIn, nPeople);
        subplot(3,1,3) %Plotting protozoa
        PlotPeople(Log, tRain, x, y, x1, y1, 'protozoa', 'people', daysBurnIn,
nPeople);
        eval(['print Graphics/People',strrep(strrep(char(p.StartTime),':','-'),'
','-' ),'.png']) %Saving chart with datetime in filename w/o spaces or colons
        %Plotting pathogens:
        F5=figure('Position',[100 100 1024 768]);    %Putting graph at (100,100) on
screen with 1024x768 resolution.
        x=.05; y=.55; x1=.93; y1=.42; %Values for positioning of 1st subplot (modify y
for placing 2nd & subsequent subplots).
        subplot(2,1,1) %Plotting pathogens in community.
        %set(gca,'FontSize',18); %Increases font size on axis tick labels (gca is 'get
current axes').
        set(gca,'Position',[x y-0*.5 x1 y1]);
        if Octave==1; %Octave and MATLAB handle line properties a bit differently in
graphs.
                semilogy(Log(:,1),Log(:,19)+.1,'-r','linewidth',3,Log(:,1),Log(:,20)+.1,'-
b','linewidth',3,Log(:,1),Log(:,21)+.1,'-g','linewidth',3); %All 3 of these are thick
lines
        else
                semilogy(Log(:,1),Log(:,19)+.1,'-r',Log(:,1),Log(:,20)+.1,'-
b',Log(:,1),Log(:,21)+.1,'-g','linewidth',3); %All 3 of these are thick lines
        end
        hold on;

```

```

        semilogy(Log(:,1),Log(:,22)+.1,'-r',Log(:,1),Log(:,23)+.1,'-
b',Log(:,1),Log(:,24)+.1,'-g');
        plot(tRain(:,1),tRain(:,2)+80,' x');
        title('Number of microbes in different community compartments')
        ylabel('# microbes','fontsize',20) %Axis & tick labels overlap on screen, but
output better to .PNG.
        legend('Bac., surf. H2O','Viruses, surf. H2O','Prot., surf. H2O','Bac.,
soil','Viruses, soil','Prot., soil','Time of rain events','Location','SouthEast')
        hold off;
        subplot(2,1,2) %Plotting pathogens in households.
        set(gca,'Position',[x y-1*.5 x1 y1]);
        if Octave==1;
            semilogy(Log(:,1),Log(:,28)+.1,'-r','linewidth',3,Log(:,1),Log(:,29)+.1,'-
b','linewidth',3,Log(:,1),Log(:,30)+.1,'-g','linewidth',3); %All 3 of these are thick
lines (stored water)
            else
                semilogy(Log(:,1),Log(:,28)+.1,'-r',Log(:,1),Log(:,29)+.1,'-
b',Log(:,1),Log(:,30)+.1,'-g','linewidth',3); %All 3 of these are thick lines (stored
water)
            end
            hold on;
            semilogy(Log(:,1),Log(:,25)+.1,'-r',Log(:,1),Log(:,26)+.1,'-
b',Log(:,1),Log(:,27)+.1,'-g'); %Supposed to graph total pathogens on all households'
hands.
            plot(tRain(:,1),tRain(:,2)+80,' x');
            title('Number of microbes in different household compartments')
            xlabel('Time (days)');
            ylabel('# microbes','fontsize',20) %Axis & tick labels overlap on screen, but
output better to .PNG.
            legend('Bac., stored H2O','Viruses, stored H2O','Prot., stored H2O','Bac.,
hands','Viruses, hands','Prot., hands','Time of rain events','Location','SouthEast')
            hold off;
            eval(['print Graphics/Microbes',strrep(strrep(char(p.StartTime),':','-'),'

```

```

','-'),'.png']) %Saving chart with datetime in filename w/o spaces or colons.
    %Plotting fluxes of microbes.
    F5=figure('Position',[100 100 1024 768]); %Putting graph at (100,100) on
screen with 1024x768 resolution.
    x=.05; y=.72; x1=.93; y1=.23; %Values for positioning of 1st subplot (modify y
for placing 2nd & subsequent subplots).
    subplot(3,1,1) %Plotting bacteria
    PlotMicrobes(LogFlux, tMax, tRain, x, y, x1, y1, 'bacteria', daysBurnIn);
    subplot(3,1,2) %Plotting viruses
    PlotMicrobes(LogFlux, tMax, tRain, x, y, x1, y1, 'viruses', daysBurnIn);
    subplot(3,1,3) %Plotting protozoa
    PlotMicrobes(LogFlux, tMax, tRain, x, y, x1, y1, 'protozoa', daysBurnIn);
    eval(['print Graphics/Fluxes',strrep(strrep(char(p.StartTime),':','-'),'
','-'),'.png']) %Saving chart with datetime in filename w/o spaces or colons.
    disp(['Charts done.'])
end
    disp(sprintf('Whole run took %s seconds.',num2str(toc)))
end %Ends main loop (once per run).
%outFilename = ['Results/',outFilename];
dlmwrite(outFilename, OutMatrix, '-append');
disp([' Results written to ',outFilename,'.']);
end %Ends function.

```

```
%This function applies log reduction values (LRVs) to stocks of pathogens.
%It simply reduces the input number of microbes (nMin) according to the appropriate LRVs,
and outputs the result.
%This should usually be used with row vectors, 3 values each (bacteria, viruses, and
protozoa, in that order). However, it accomodates multiple rows.
%nMininput: Vector/matrix containing number of microbes that an LRV might be applied to.
%LRVs:      Log reduction values to be applied. 1 column per microbe.
%Compliance: Proportion of all microbes to which the LRVs are being applied.
```

```
%{
COPYRIGHT INFORMATION
```

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (ApplyLRVs.m) is part of EITSd.

EITSd is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

EITSd is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with EITSd. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
function [nMoutput] = ApplyLRVs(nMininput, LRVs, Compliance);
    nRows = size(nMininput,1); %Gets number of households/people that it's acting on.
    Compliance = repmat(Compliance,1,size(LRVs,2)); %Replicating Compliance vector so
```

```
it matches the size of nMinput.  
    LRVs = repmat(LRVs,nRows,1);  
    nMoutput = nMinput .* (1-Compliance) + nMinput .* Compliance .* 10 .^ -LRVs;  
end
```

%Function to generate compliance behaviors for individual households given a particular compliance scheme.

%HHcol: the column in the household matrix that is being populated.

%cv: the compliance vector, i.e., cSan, cHWT, or cHand, 1st value being overall compliance, 2nd being compliance type.

%{

COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (AssignCompliance.m) is part of EITSd.

EITSd is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

EITSd is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with EITSd. If not, see <<http://www.gnu.org/licenses/>>.

%}

%TODO: correlate the household-level compliances assigned to the 3 main interventions?

```
function [output] = AssignCompliance(HHcol,cv);
```

```
    ct = cv(2);
```

```
    switch ct;
```

```
        case 1; %compliance type alpha: everyone perfectly complies or doesn't comply
```

```
at all
```

```

        output = rand(size(HHcol,1),1) < cv(1); %Randomly assign binary value to
compliance for each household.
        case 2; %compliance type beta: some perfectly comply, some partially comply,
some don't comply at all
            output = rand(size(HHcol,1),1);
            output(find(output < cv(1)/2)) = 1;      %Assigning perfect compliers a
value of 1
            output(find(output > 1 - ((1-cv(1)) / 2) & output < 1)) = 0; %Assigning
noncompliers a value of 0
            output(find(output > 0 & output < 1)) = cv(1);    %Assigning the remaining
partial compliers the overall compliance value
        case 3; %compliance type gamma: everyone partially complies
            output = cv(1);
        otherwise
            error('Compliance type != 1, 2, or 3, therefore invalid');
    end
end
end

```



```
%Loop for obtaining parameter values for estimation runs while modifying reservoir size,
pLandHand, etc. Facilitates computing cluster use.
%Be sure to check that GetTrialParams.m is configured properly before running this
script.
%Submit as several jobs to parallelize a calibration run (maybe not worth bothering with
job array).
```

```
%{
COPYRIGHT INFORMATION
```

```
    Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").
```

```
This file (CalLoopFuncCompile.m) is part of EITSd.
```

```
EITSd is free software: you can redistribute it and/or modify
it under the terms of the GNU General Public License as published by
the Free Software Foundation, either version 3 of the License, or
(at your option) any later version.
```

```
EITSd is distributed in the hope that it will be useful,
but WITHOUT ANY WARRANTY; without even the implied warranty of
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
GNU General Public License for more details.
```

```
You should have received a copy of the GNU General Public License (gpl.txt)
along with EITSd. If not, see <http://www.gnu.org/licenses/>.
```

```
%}
```

```
function[OutM OutS] = CalibrationLoopFuncCompile(indexText,inc,calibRunsText,jobname);
%This helps with debugging, since arguments to compiled code can only be text.
disp(class(indexText))
index = str2num(indexText);
RandStream.setDefaultStream(RandStream.create('mt19937ar','seed',sum([clock index*10])));
```

```

    %Sets random stream based on clock & job index.
    calibrRuns = str2num(calibrRunsText);
    %===Required lines for HPCC
    setenv MKL_DYNAMIC FALSE
    %maxNumCompThreads(1); %Throws an error. Recommended, but does not seem to be
    necessary.
    %===
    %===Parameter entry===
    infile = 'nonapplicable.csv';
    outfile = ['Results/',jobname,'.csv'];
    %===End parameter entry===

    %disp(['##### Running
    ',num2str(size(U,2)),'*',num2str(size(T,2)),'*',num2str(size(L,2)),'+1=',num2str(combos),
    ' parameter combinations on ',num2str(size(tempData,1)), ' parameter sets from
    calibration, should have requested at least that many members in the job array. #####'])

    if inc(1:2) == 'Hi';
    EITSd('C', calibrRuns, outfile, infile, [-11 -2.4], [-10 -1.5], [-8 -.1], [-15 -4], [-1 0
    0;2 3 3], 200); %1st try; including the max. possible levels for all ranges of transfer
    params.

    disp(['##### DONE #####'])
    end %End function.

```

```
%Determine response (infection) given a particular dose, whatever the route.  
%These arguments (R, dose, KorN50, alpha) are vectors or matrices, 1 column per pathogen.  
%Uses 3 random numbers per person to determine response.
```

```
%{  
COPYRIGHT INFORMATION
```

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (DoseResponse.m) is part of EITSd.

EITSd is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

EITSd is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with EITSd. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
```

```
function [People EffM] = DoseResponse(RDRpeople, People, nKids, Doses, KorN50, alpha,  
latent);  
    nP = size(People,1);      %Number of people (rows).  
    nAdults = nP - nKids;  
    nM = size(KorN50,2);     %Number of types of microbe (columns).  
    %marker=repmat(['B','V','P'],nP,1);    %Marker that is used to record the type(s)  
of microbe that resulted in infection.  
    DosesT = Doses(:,1:3) + Doses(:,4:6) + Doses(:,7:9);    %Summing doses from water &
```

```

hands for each person.
    Shifters = logical(zeros(nP,nM)); %Creating matrix of true/false for storing &
processing outcomes.
    DosesT = DosesT .* (People(:,1:3) == -1); %If person is not susceptible, set dose
to 0.
    for i = 1:nM %Loop over the 3 microbes
        if isnan(alpha(i))==1
            response = DRexp(KorN50(i),DosesT(:,i));
        else
            response = DRbP(KorN50(i),alpha(i),DosesT(:,i));
        end
        Shifters(:,i) = RDRpeople(:,i) < response; %Determines which people will
become infected.
    end
    %tags = char(horzcat(ones(nP,1)*88, (marker .* outcome))); %Tags for pos. dose
response events w. appropriate microbe(s) (e.g., 'FoodEatXBV').
    %event = horzcat(event,tags); %Apply the tags to the event name.

    PeopleStates = People(:,1:3); PeopleStates(Shifters) = 1; People(:,1:3) =
PeopleStates; %...People who get a positive response become exposed...
    PeopleCounters = People(:,4:6); PeopleCounters(Shifters) = 0; People(:,4:6) =
PeopleCounters; %...their appropriate counter(s) (which are negative) get set to 0...
    People(:,4:6) = People(:,4:6) + Shifters .* repmat(latent,nP,1);
    %...and are assigned a latent period.

    DosesK = Doses(1:nKids,:);
    DosesA = Doses((nKids+1):nP,:);
    ShiftersK = Shifters(1:nKids,:);
    ShiftersA = Shifters((nKids+1):nP,:);
    EffMWK = sum(DosesK(:,1:3) .* ShiftersK); %Tallying microbes that contributed to
a new infection.
    EffMLK = sum(DosesK(:,4:6) .* ShiftersK);
    EffMHK = sum(DosesK(:,7:9) .* ShiftersK);

```

```

    EffMWA = sum(DosesA(:,1:3) .* ShiftersA);    %Tallying microbes that contributed to
a new infection.
    EffMLA = sum(DosesA(:,4:6) .* ShiftersA);
    EffMHA = sum(DosesA(:,7:9) .* ShiftersA);
    EffM = [EffMWK EffMLK EffMHK; EffMWA EffMLA EffMHA];

    %People(:,1:nM) = People(:,1:nM) + 2 * outcome;    %Assigns exposure (status=1) to
those people (who are susceptible, thus have status=-1).
    %People(:,nM+1:nM+nM) = People(:,nM+1:nM+nM) .* !outcome;    %Sets counter to 0 for
those people who have a new infection incubating.
    %People(:,nM+1:nM+nM) = People(:,nM+1:nM+nM) + repmat(latent,nP,1) .* outcome;
    %Assigns the length of the latent period.
    %TODO: Assign randomly chosen latent period instead.
end

```

%Beta-Poisson dose response model, using N50 (default) or beta as a parameter

%{
COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (DRbP.m) is part of EITSd.

EITSd is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

EITSd is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with EITSd. If not, see <<http://www.gnu.org/licenses/>>.

%}

function outvar = DRbP(N50orBeta,alpha,invar,reverse,WhichParam) %Ordinarily, invar is dose & outvar is response.

if nargin == 3;

reverse = 'no'; WhichParam = 'N50';

end

switch(reverse)

case 'no'

switch(WhichParam)

case 'N50'

outvar = 1-(1+(invar/N50orBeta)*(2^(1/alpha)-1)).^-alpha;

```

        case 'Beta'
            outvar = 1-(1+(invar/N50orBeta)).^-alpha;
        otherwise
            error(['WhichParam must be "N50" or "Beta"'])
        end
    case 'yes' %If reverse='yes', invar is response & outvar is dose.
        switch(WhichParam)
            case 'N50'
                outvar = N50orBeta * ( ((1-invar).^(-1/alpha) -1) /
(2^(1/alpha)-1) );
            case 'Beta'
                outvar = N50orBeta * ((1-invar).^(-1/alpha) -1);
            otherwise
                error(['WhichParam must be "N50" or "Beta"'])
            end
        otherwise
            error(['reverse must be "no" or "yes"'])
        end
    end
end

```

```
%Exponential dose response model
%{
COPYRIGHT INFORMATION
```

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (DRexp.m) is part of EITSd.

EITSd is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

EITSd is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with EITSd. If not, see <<http://www.gnu.org/licenses/>>.

```
%}
function outvar = DRexp(k, invar, reverse) %This works in Matlab.
    if nargin < 3;
        reverse = 'no';
    end
    switch(reverse);
        case 'no';
            outvar = 1-exp(-k * invar);
        case 'yes'; %If reverse='yes', invar is response & outvar is dose.
            outvar = log(1-invar)/-k;
        otherwise
            error(['reverse (last parameter) must be "no" or "yes"'])
    end
```


end

%Functions for calculating vectors of illness durations

%{
COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (durEc.m) is part of EITSd.

EITSd is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

EITSd is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with EITSd. If not, see <<http://www.gnu.org/licenses/>>.

%}

```
function output = durEc(n);  
    output = round(gamrnd(1.775,1.690,[n,1])); %Shape, then scale. From Estrada-Garcia  
2009.  
    output(output == 0) = 1; %Sets zero durations to 1 day instead.  
end
```

```
%Functions for calculating vectors of illness durations
```

```
%{  
COPYRIGHT INFORMATION
```

```
    Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").
```

```
This file (durGi.m) is part of EITSd.
```

```
EITSd is free software: you can redistribute it and/or modify  
it under the terms of the GNU General Public License as published by  
the Free Software Foundation, either version 3 of the License, or  
(at your option) any later version.
```

```
EITSd is distributed in the hope that it will be useful,  
but WITHOUT ANY WARRANTY; without even the implied warranty of  
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the  
GNU General Public License for more details.
```

```
You should have received a copy of the GNU General Public License (gpl.txt)  
along with EITSd. If not, see <http://www.gnu.org/licenses/>.
```

```
%}
```

```
function output = durGi(n); %Based on a fit of gamma dist. to limited info from Kent GP  
1988.
```

```
    output = round(gamrnd(3.206,3.431,[n,1])); %Shape, then scale  
    output(output == 0) = 1; %Sets zero durations to 1 day instead.
```

```
end
```

```
%Functions for calculating vectors of illness durations
```

```
%{  
COPYRIGHT INFORMATION
```

```
    Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").
```

```
This file (durRo.m) is part of EITSd.
```

```
EITSd is free software: you can redistribute it and/or modify  
it under the terms of the GNU General Public License as published by  
the Free Software Foundation, either version 3 of the License, or  
(at your option) any later version.
```

```
EITSd is distributed in the hope that it will be useful,  
but WITHOUT ANY WARRANTY; without even the implied warranty of  
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the  
GNU General Public License for more details.
```

```
You should have received a copy of the GNU General Public License (gpl.txt)  
along with EITSd. If not, see <http://www.gnu.org/licenses/>.
```

```
%}
```

```
function output = durRo(n); %Based on 4 rotavirus-infected volunteers having durations  
of 1, 2, 3, and 4 days (Kapikian 1983).  
    output = ceil(rand([n,1]) * 4);  
    %output(output == 0) = 1; %Sets zero durations to 1 day instead.  
end
```

%Loop for obtaining estimation runs while modifying pUse, pTreat, and LRVs. Facilitates computing cluster use.

%Be sure to check that GetTrialParams.m is configured properly before running this script.

%Differs from EstimationLoopFuncCompile in that only a small number of parameter combinations are chosen, rather than all possible combinations.

%mix: Proportion of childhood disease that is waterborne (A, B, or C).

%{

COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (EstLoopFuncCompileBase.m) is part of EITSd.

EITSd is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

EITSd is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with EITSd. If not, see <<http://www.gnu.org/licenses/>>.

%}

```
function[OutM OutS] = EstLoopFuncCompileBase(indexText,jobname);
Octave = size(ver('Octave'),1); %Indicator of whether the code is running under Octave
(1) or Matlab (0).
index = str2num(indexText);
```

```

if Octave == 0;
    RandStream.setDefaultStream(RandStream.create('mt19937ar','seed',sum([clock
index*10])));
    %Sets random stream based on clock & job index. Doesn't work with Octave (Octave
bases the seed on the clock by default).
end
%===Required lines for HPC
setenv MKL_DYNAMIC FALSE
%maxNumCompThreads(1); %Throws an error. Does not seem to be necessary.
%===End parameter entry===
disp(['##### Running 6 parameter combinations, 2 each from runs that fit for mixes A, B,
C, and Z, no interventions, should have requested at least 6 members in the job array.
#####'])
if index == 1 | index == 2; %If running the baseline parameters:
EITSd('E',5,['Results/EstBaseA',indexText,'.csv'],'RTF_A_CalHi10b.csv',1,2,3,4,5,200,[0 0
0],[0 1],[0 0 0],[0 1],[0 0 0],[0 1]);
elseif index == 3 | index == 4;
EITSd('E',5,['Results/EstBaseB',indexText,'.csv'],'RTF_B_CalHi10b.csv',1,2,3,4,5,200,[0 0
0],[0 1],[0 0 0],[0 1],[0 0 0],[0 1]);
elseif index == 5 | index == 6;
EITSd('E',5,['Results/EstBaseC',indexText,'.csv'],'RTF_C_CalHi10b.csv',1,2,3,4,5,200,[0 0
0],[0 1],[0 0 0],[0 1],[0 0 0],[0 1]);
elseif index == 7 | index == 8;
EITSd('E',5,['Results/EstBaseZ',indexText,'.csv'],'RTF_Z_CalHi10b.csv',1,2,3,4,5,200,[0 0
0],[0 1],[0 0 0],[0 1],[0 0 0],[0 1]);
else
error('Too many jobs!')
end
disp(['##### DONE #####'])
end %End function.

```

%Loop for obtaining estimation runs while modifying pUse, pTreat, and LRVs. Facilitates computing cluster use.

%Be sure to check that GetTrialParams.m is configured properly before running this script.

%Differs from EstimationLoopFuncCompile in that only a small number of parameter combinations are chosen, rather than all possible combinations.

%mix: Proportion of childhood disease that is waterborne (A, B, or C).

%{

COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (EstLoopFuncCompileHWT.m) is part of EITSd.

EITSd is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

EITSd is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with EITSd. If not, see <<http://www.gnu.org/licenses/>>.

%}

```
function[OutM OutS] = EstLoopFuncCompileHWT(indexText,mix,overallComplianceText,jobname);
Octave = size(ver('Octave'),1); %Indicator of whether the code is running under Octave
(1) or Matlab (0).
index = str2num(indexText);
```

```

if Octave == 0;
    RandStream.setDefaultStream(RandStream.create('mt19937ar','seed',sum([clock
index*10])));
    %Sets random stream based on clock & job index. Doesn't work with Octave (Octave
bases the seed on the clock by default).
end
oC = str2num(overallComplianceText);

%===Required lines for HPCC
setenv MKL_DYNAMIC FALSE
%maxNumCompThreads(1); %Throws an error. Does not seem to be necessary.
%===Parameter entry=== Note that 0 should not be included in L.
%P = [0 .1 .2]; %Vector of desired values for proportions of children never using the
device.
%N = [0 .1 .2]; %Vector of desired values for proportions of children perfectly using
the device.
L = [1 2 3 4 5]; %Vector of log reduction values desired (all marker pathogens get the
same LRV).
%Testing the code using the vectors below.
%U=[.9 1]
%T=[.9 1]
%L=[1 2]
%Constructing a matrix with all possible combos of P, N, & L
%[p n l] = ndgrid(P,N,L);
%Combos = [p(:) n(:) l(:)]; %TODO: Build functionality to check N, P,
overallCompliance, and pTreat to ensure they make sense before running.
%Building appropriate combinations of perfect compliers (P; 1st column) and noncompliers
(N; 2nd column).
Combos = zeros(3,3); %Combo with max possible perfect compliers & max possible
noncompliers (same as [0 1-oC]; [oC 1-oC] is 0/0).
Combos(:,1) = oC;
Combos(:,2) = [1 2 3];
Combos(:,3) = L(1); %This & subsequent 'for' loop copy the above for each possible LRV.

```



```

OutCombos = Combos;
for i = 2:size(L,2);
    NextCombos = Combos;
    NextCombos(:,3) = L(i);
    OutCombos = [OutCombos; NextCombos];
end
%Baselines = zeros(1,3); %Creating baseline row. No log reduction & no perfect
compliance.
%Baselines(:,2) = 1; %Modifies above, so that 100% never use device.
Combos = OutCombos;
Combos' %Output results, transposed.
combos = size(Combos,1)

infile = ['RTF_',mix,'_CalHi10b.csv'];
outfile = ['Results/',jobname,'.csv'];
%===End parameter entry===

tempData = csvread(infile,1,1);

disp(['##### Running 3 * ',num2str(size(L,2)), ' = ',num2str(combos),' parameter
combinations on ',num2str(size(tempData,1)), ' parameter sets from calibration, should
have requested at least that many members in the job array. #####'])
%if index == 1; %If running the baseline parameters:
% [OutM OutS] = Main(Combos(index,1), Combos(index,2), oC, [Combos(index,3)
Combos(index,3) Combos(index,3)], [0 0 0], [0 0 0], 0, outfile, infile, multConc,
nSpikes, multSpikes, 1);
%else %If running the parameters from calibration:
EITSd('E',10,outfile,infile,1,2,3,4,5,200,[0 0 0],[0 1],Combos(index,3)*[1 1 1],
[Combos(index,1) Combos(index,2)], [0 0 0],[0 1]);
%end
disp(['##### DONE #####'])
end %End function.

```

%Loop for obtaining estimation runs while modifying pUse, pTreat, and LRVs. Facilitates computing cluster use.

%Be sure to check that GetTrialParams.m is configured properly before running this script.

%Differs from EstimationLoopFuncCompile in that only a small number of parameter combinations are chosen, rather than all possible combinations.

%mix: Proportion of childhood disease that is waterborne (A, B, or C).

%{

COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (EstLoopFuncCompileHWTPC.m) is part of EITSd.

EITSd is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

EITSd is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with EITSd. If not, see <<http://www.gnu.org/licenses/>>.

%}

```
function[OutM OutS] = EstLoopFuncCompileHWTPC(indexText,jobname);
Octave = size(ver('Octave'),1); %Indicator of whether the code is running under Octave
(1) or Matlab (0).
```

```
index = str2num(indexText);
```

```

if Octave == 0;
    RandStream.setDefaultStream(RandStream.create('mt19937ar','seed',sum([clock
index*10])));
    %Sets random stream based on clock & job index. Doesn't work with Octave (Octave
bases the seed on the clock by default).
end
%oC = str2num(overallComplianceText);

%===Required lines for HPCC
setenv MKL_DYNAMIC FALSE
%maxNumCompThreads(1); %Throws an error. Does not seem to be necessary.
%===Parameter entry=== Note that 0 should not be included in L.
%P = [0 .1 .2]; %Vector of desired values for proportions of children never using the
device.
%N = [0 .1 .2]; %Vector of desired values for proportions of children perfectly using
the device.
L = [1 2 3 4 5]; %Vector of log reduction values desired (all marker pathogens get the
same LRV).
%Testing the code using the vectors below.
%U=[.9 1]
%T=[.9 1]
%L=[1 2]
%Constructing a matrix with all possible combos of P, N, & L
%[p n l] = ndgrid(P,N,L);
%Combos = [p(:) n(:) l(:)]; %TODO: Build functionality to check N, P,
overallCompliance, and pTreat to ensure they make sense before running.
%Building appropriate combinations of perfect compliers (P; 1st column) and noncompliers
(N; 2nd column).
Combos = zeros(4,3); %Combo with max possible perfect compliers & max possible
noncompliers (same as [0 1-oC]; [oC 1-oC] is 0/0).
Combos(:,1) = 1;
Combos(:,2) = [1 2 3 4];
Combos(:,3) = L(1); %This & subsequent 'for' loop copy the above for each possible LRV.

```

```

OutCombos = Combos;
for i = 2:size(L,2);
    NextCombos = Combos;
    NextCombos(:,3) = L(i);
    OutCombos = [OutCombos; NextCombos];
end
%Baselines = zeros(1,3); %Creating baseline row. No log reduction & no perfect
compliance.
%Baselines(:,2) = 1; %Modifies above, so that 100% never use device.
Combos = OutCombos;
Combos' %Output results, transposed.
combos = size(Combos,1)
mixOptions='ABCZ'; %Translating the digits 1-3 into letters A-C.
mix=mixOptions(Combos(index,2));
infile = ['RTF_',mix,'_CalHil0b.csv'];
outfile = ['Results/',jobname,mix,'.csv'];
%===End parameter entry===

tempData = csvread(infile,1,1);

disp(['##### Running 4 * ',num2str(size(L,2)),' = ',num2str(combos),' parameter
combinations on ',num2str(size(tempData,1)),' parameter sets from calibration, should
have requested at least that many members in the job array. #####'])
%if index == 1; %If running the baseline parameters:
% [OutM OutS] = Main(Combos(index,1), Combos(index,2), oC, [Combos(index,3)
Combos(index,3) Combos(index,3)], [0 0 0], [0 0 0], 0, outfile, infile, multConc,
nSpikes, multSpikes, 1);
%else %If running the parameters from calibration:
EITSd('E',10,outfile,infile,1,2,3,4,5,200,[0 0 0],[0 1],Combos(index,3)*[1 1 1],[1 1],[0
0 0],[0 1]);
%end
disp(['##### DONE #####'])
end %End function.

```

%Inactivation (attenuation) of microbes in all compartments. Also includes some error checking.

%Note that CFdecay is now applied to decay in all compartments (in EITS06 and later versions).

%Formerly (in EITS05), CFdecay was only applied to land & household environments, not to surface water or stored drinking water decay.

%{
COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (Inact.m) is part of EITSd.

EITSd is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

EITSd is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with EITSd. If not, see <<http://www.gnu.org/licenses/>>.

%}

```
function [HHs, nMw, nMl] = Inact(RDRtime, HHs, nMw, nMl, CFdecay, Wflow, t);
    nHH = size(HHs,1);
    %if iscomplex(HHs) == 1;          %Works in Octave, but not MATLAB.
    if sum(sum(imag(HHs))) ~= 0;    %Works in both Octave and MATLAB.
        error('Complex number in HHs!'); %This was a problem in earlier versions of
```

```

the code.
    end
    if sum(sign([nMw nMl HHs(nHH*3+1:nHH*16)]) == -1) > 0; %Check for any problematic
negative values.
        negs = find(HHs < 0);
        [rs,cs] = find(HHs < 0);
        warning('\n%s negative values in HHs at time %s, totaling %s, rows %s, cols %s,
vals
%s.', num2str(size(negs,1)), num2str(t), num2str(sum(HHs(negs))), num2str(rs), num2str(cs), num
2str(HHs(negs)))
        %When negative values occur, they tend to be in single-person households with
repeated inter-household visits.
        if sum(HHs(negs)) > -1E-100;
            warning('Very tiny negative values (> -1E-100), setting them to 0.')
            HHs(negs) = 0; %Kludge: effectively adds some pathogens to the system to
counteract negative values; however, seldom needed.
        else
            error('Larger negative values than usual. Stopping.')
        end
        if sum(sign([nMw nMl]) == -1) > 0;
            disp(nMw); disp(nMl);
            error('Surface water or land has gone negative! Setting to 0.')
            nMw(nMw < 0) = 0; nMl(nMl < 0) = 0;
            disp(nMw); disp(nMl);
        end
        end
        nMw = nMw .* exp((-gM-Wflow) * RDRtime); %Surface water (this line
formerly used in EITS05).
        nMw = nMw .* exp(-CFdecay .* RDRtime); %Surface water.
        nMl = nMl .* exp(-CFdecay .* RDRtime); %Land. Next is household inactivation.
        for i = 1:3; %Microbe. Corresponds to the appropriate entry in gM (inactivation
rate).
            for j = [0 3]; %Compartment. 0 signifies hands (cols. 4:6 in HHs), 3 signifies

```

```

stored water (cols. 7:9 in HHs).
    if j == 0;
        HHs(:,i+3+j) = HHs(:,i+3+j) .* exp(-CFdecay(i) * RDRtime);
    elseif j == 3;
        %HHs(:,i+3+j) = HHs(:,i+3+j) .* exp(-gM(i) * RDRtime); %This line
formerly used in EITS05.
        HHs(:,i+3+j) = HHs(:,i+3+j) .* exp(-CFdecay(i) * RDRtime);
    end
end
end
end
end

```

```
%Function for generating line charts of microbe transfers.  
%Need to start the figure, set up subplots, and determine subplot positioning before  
calling this function.
```

```
%{  
COPYRIGHT INFORMATION
```

```
    Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").
```

```
This file (PlotMicrobes.m) is part of EITSd.
```

```
EITSd is free software: you can redistribute it and/or modify  
it under the terms of the GNU General Public License as published by  
the Free Software Foundation, either version 3 of the License, or  
(at your option) any later version.
```

```
EITSd is distributed in the hope that it will be useful,  
but WITHOUT ANY WARRANTY; without even the implied warranty of  
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the  
GNU General Public License for more details.
```

```
You should have received a copy of the GNU General Public License (gpl.txt)  
along with EITSd. If not, see <http://www.gnu.org/licenses/>.
```

```
%}
```

```
%LogFlux = NaN(tMax,3,8);    %Cube to store fluxes of microbes. z is 1:11 (1-surface  
water to stored water at resupply; 2-net visit transfer; 3-not used, but formerly land-  
to-hand-to stored water at drinking; 4-rainfall; 5&6-pooing into surface H2O (not done)  
& land; 7-inactivation; 8&9-kids' dose, water & hands; 10&11-adults' dose, water &  
hands).
```

```
function PlotMicrobes(Log, tMax, tRain, x, y, x1, y1, microbe, daysBurnIn);  
    if microbe(1:3) == 'bac';
```



```

        add = 0;
elseif microbe(1:3) == 'vir';
    add = 1;
elseif microbe(1:3) == 'pro';
    add = 2;
else
    error('microbe must equal 'bacteria', 'viruses', or 'protozoa'.');
end
Log = Log + 0.1;    %Avoids plotting zeros.
Y = add + 1;    %Y coordinate within LogFlux data cube (denoting microbe type).
set(gca,'Position',[x y-add*.33 x1 y1]);    %1:2; x&y of bottom L corner. 2:3; x&y
of top R corner, minus 1:2.
FluxIn = sum(Log(:,Y,5:6),3);
FluxOut = sum(Log(:,Y,7:11),3);
semilogy(1:tMax,Log(:,Y,1),'-b',1:tMax,Log(:,Y,2),'-c','linewidth',3);
hold on;
semilogy(1:tMax,Log(:,Y,4),'ob','MarkerSize',7)
semilogy(1:tMax,Log(:,Y,6),'-m','LineWidth',6)
semilogy(1:tMax,Log(:,Y,7),'-y','LineWidth',3)
semilogy(1:tMax,Log(:,Y,8),'+r','MarkerSize',3)
semilogy(1:tMax,Log(:,Y,9),'-r','LineWidth',3)
semilogy(1:tMax,Log(:,Y,10),'+y','MarkerSize',3);
semilogy(1:tMax,Log(:,Y,11),'-y',1:tMax,FluxIn,'+k',1:tMax,FluxOut,'-k');
plot(tRain(:,1),tRain(:,2) * 0.5,' x');
title(['Daily transfers of ',microbe],'fontsize',20);
ylabel(['# ',microbe],'fontsize',20)    %Axis & tick labels overlap on screen, but
output better to .PNG.
legend('Water resupply','Visits','Land-water (runoff)','Poop onto
land','Inactivation','Kid dose, water','Kid dose, hands','Adult dose, water','Adult dose,
hands','Net pos flux','Net neg flux','Time of rain events','location','southeast')
plot([daysBurnIn daysBurnIn],[0.5 max(max(max(Log)))], '-k');    %Sticking a
vertical line on the graph to denote when burn-in ends.
axis([0 tMax*1.3 0.5 max(max(FluxOut))*1.5]);

```

```
end hold off;
```

%Function for generating line charts of infection status.
%Need to start the figure, set up subplots, and determine subplot positioning before calling this function.

%{
COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (PlotPeople.m) is part of EITSd.

EITSd is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

EITSd is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with EITSd. If not, see <<http://www.gnu.org/licenses/>>.

%}

```
function PlotPeople(Log, tRain, x, y, x1, y1, microbe, host, daysBurnIn, nPeople);  
    if microbe(1:3) == 'bac';  
        add = 0;  
    elseif microbe(1:3) == 'vir';  
        add = 1;  
    elseif microbe(1:3) == 'pro';  
        add = 2;  
    else
```

```

        error('microbe must equal 'bacteria', 'viruses', or 'protozoa'.');
    end
    tMax = size(tRain,1);
    set(gca,'Position',[x y-add*.33 x1 y1]);      %1:2; x&y of bottom L corner. 2:3; x&y
of top R corner, minus 1:2.
    plot(Log(:,1),Log(:,14+add),'-r',Log(:,1),Log(:,18),'-k','LineWidth',2);
    hold on;
    plot(Log(:,1),Log(:,2+add),'g-',Log(:,1),Log(:,5+add),'-b',Log(:,1),Log(:,8+add),'-
m',Log(:,1),Log(:,11+add),'-r',Log(:,1),Log(:,17),'-k');
    %plot(Log(:,1),Log(:,2+add),'g-',Log(:,1),Log(:,5+add),'-b',Log(:,1),Log(:,8+add),'-
m',Log(:,1),Log(:,11+add),'-r',Log(:,1),Log(:,14+add),'-
r','LineWidth',2,Log(:,1),Log(:,17),'-k',Log(:,1),Log(:,18),'-k','LineWidth',2); %Works
in Octave.
    hold on;
    plot(tRain(:,1),tRain(:,2)-1,' x');
    title(['Daily infection status, ',microbe],'fontsize',20);
    ylabel(['# ',host],'fontsize',20) %Axis & tick labels overlap on screen, but output
better to .PNG.
    legend('Ill','Any illness','Susceptible','Immune','Exposed','Infected','Any
infection (no illness)','Rain events','Location','East')
    %legend('Susceptible','Immune','Exposed','Infected','Ill','Any infection (no
illness)','Any illness','Rain events','Location','East') %For commented-out plot() call
above.
    plot([daysBurnIn daysBurnIn],[0 nPeople], '-k');
    axis([0 tMax*1.3 0 nPeople+1]);
    hold off;
end

```

%Describes results of defecation events (microbes onto hands, land, and surface water) by a single person.

%{
COPYRIGHT INFORMATION

Copyright 2012 Kyle S. Enger (username "engerkyl" at the domain "msu.edu").

This file (Pooping.m) is part of EITSd.

EITSd is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

EITSd is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License (gpl.txt) along with EITSd. If not, see <<http://www.gnu.org/licenses/>>.

%}

%RPoopPlace: 1 column from the random number matrix determining whether the person poops into the water or on land.

%People: Rows from the People matrix, each row corresponding to a particular person.

%HHs: The HHs matrix, corresponding to all households.

%Mpgf: Microbes per gram of feces (3-element vector, 1 element per microbe type).

%fpp: Grams of feces excreted per defecation event (assuming 1 event per day).

Single value; run function once for each desired value.

%fHands: Grams of feces on fingers after defecation event.

%nMw: Number of microbes in the reservoir (3-element vector, 1 element per

```

microbe type).
%nMl:      As above, but number of microbes on land.
%pPoopH2O: Probability that the person defecates directly into surface water (instead
of on land).
%RCompHW: 1 column from the random number matrix determining whether the person
handwashes.
%RCompSan: 1 column from the random number matrix determining whether the person uses
sanitation.
%lHand:    LRVs from handwashing.
%lSan:     LRVs from sanitation.
%chosenOnes: Logical vector of rows in People that are being operated on.

function [HHs, nMw, nMl] = Pooping(RPoopPlace, People, HHs, Mpgf, fpp, fHands, nMw, nMl,
pPoopH2O, lHand, lSan, chosenOnes);
    nPeople = size(People,1);
    if size(chosenOnes,1) ~= nPeople; error('People and chosenOnes do not match up!');
end
    MicrobeLoad = repmat(Mpgf,nPeople,1) .* repmat(chosenOnes,1,3) .* (People(:,1:3)==2
| People(:,1:3)==3) .* repmat(fpp,nPeople,3);    %Only microbes infecting the person are
shed.
    MicrobeLoadHands = MicrobeLoad .* (fHands/fpp);
    MicrobeLoadEnv = MicrobeLoad - MicrobeLoadHands;
    PeopleCompSanHW = HHs(People(:,7),[13 15]); %Expands household-level compliance to
each person.
    MicrobeLoadHands = ApplyLRVs(MicrobeLoadHands, lHand, PeopleCompSanHW(:,2));
    %Applying handwashing LRVs to handwashing compliers' hands.
    for i = 1:max(People(:,7)); %Looping over households, to apply new hand
contamination to household 'hands' stock. Would be nice to vectorize, but not sure how.
        %PeopleInHHi = find(People(:,7)==i);
        MicrobesAdded = sum(MicrobeLoadHands(find(People(:,7)==i),:),1); %Forces sum of
each column (even if there's only 1 row).
        %MicrobesAdded = MicrobeLoadHands(find(People(:,7)==i),:);
        HHs(i,4:6) = HHs(i,4:6) + MicrobesAdded;

```

```

end
MicrobeLoadEnv = ApplyLRVs(MicrobeLoadEnv, lSan, PeopleCompSanHW(:,1)); %Applying
sanitation LRVs to sanitation compliers' feces.
PoopInH2O = RPoopPlace < pPoopH2O;
nMl = nMl + sum(MicrobeLoadEnv(~PoopInH2O,:),1); %Apply remaining microbes after
sanitation to the land.
nMw = nMw + sum(MicrobeLoadEnv(PoopInH2O,:),1); %As above, for the water.
end

```

```
function A = erdrey(n,m)
```

```
%ERDREY      Generate adjacency matrix for a G(n,m) type random graph.
%
% Input      n: dimension of matrix (number of nodes in graph).
%            m: 2*m is the number of 1's in matrix (number of edges in graph).
%            Defaults to the smallest integer larger than n*log(n)/2.
%
% Output     A: n by n symmetric matrix with the attribute sparse.
%
% Description:  An undirected graph is chosen uniformly at random from
%               the set of all symmetric graphs with n nodes and m
%               edges.
%
% Reference:   P. Erdos, A. Renyi,
%               On Random Graphs,
%               Publ. Math. Debrecen, 6 1959, pp. 290-297.
%
% Example:    A = erdrey(100,10);
```

```
%Lines 22-26 added by Kyle S. Enger to ensure proper attribution. In all other respects,
this copy of erdrey.m is identical to its original source.
```

```
%This file is part of CONTEST, a publicly available MATLAB toolbox:
```

```
http://www.mathstat.strath.ac.uk/research/groups/numerical\_analysis/contest
```

```
%See also Taylor A and Higham DJ (2009) CONTEST: A Controllable Test Matrix Toolbox for
MATLAB. ACM Transactions on Mathematical Software. 35 (4).
```

```
%This copy of erdrey.m is reproduced in Kyle S. Enger's Ph.D. dissertation by permission
of the authors (Des Higham, personal communication, 24 Aug. 2012).
```

```
%The file erdrey.m is not part of EITSd but is used by EITSd, and is included here for
completeness.
```

```
if nargin == 1
```



```

    m = ceil(n*log(n)/2);
end

nonzeros = ceil(0.5*n*(n-1)*rand(m,1));
v = zeros(n,1);
for count = 1:n
    v(count) = count*(count-1)/2;
end

I = zeros(m,1);
J = zeros(m,1);
S = ones(m,1);

for count = 1:m
    i = min(find(v >= nonzeros(count)));
    j = nonzeros(count) - (i-1)*(i-2)/2;
    I(count) = i;
    J(count) = j;
end

A = sign(sparse([I;J],[J;I],[S;S],n,n));

while nnz(A) ~= 2*m

    difference = m-nnz(A)/2;
    Inew = zeros(difference,1);
    Jnew = zeros(difference,1);
    for count = 1:difference
        index = ceil(0.5*n*(n-1)*rand);
        Inew(count) = min(find(v>=index));
        Jnew(count) = index - (Inew(count)-1)*(Inew(count)-2)/2;
    end
    I = cat(1,I,Inew);

```

```
J = cat(1,J,Jnew);  
S = ones(length(I),1);  
A = sign(sparse([I;J],[J;I],[S;S],n,n));
```

```
end
```

10. APPENDIX D: GLOSSARY

Allocation concealment: Conducting a study in such a way that assignment to a particular group is truly random, and not influenced by either the subject or the investigator. In theory, adequate allocation concealment should always be feasible.

Blinding: Conducting a study in such a way that the subject (single-blind) or the subject and the observer (double-blind) does not know which experimental group the subject belongs to.

BSF: Biosand filter, a HWT method consisting of a sand filter in which the outlet pipe starts at the bottom of the filter and exits the filter above the top layer of sand, thus allowing the sand to remain wet at all times.

CAMRA: Center for Advancing Microbial Risk Assessment. A multi-university center, it hosts the QMRWiki (<http://wiki.camra.msu.edu>).

CDC: Centers for Disease Control and Prevention (United States government agency).

CFU: Colony-forming unit. A bacterial suspension can be quantitated by spreading a sample of it on a plate of growth medium. After incubation, the number of colonies are counted. This estimates the number of colony-forming units in the original suspension. CFUs are often assumed to represent single bacterial cells, although this assumption may not be accurate if the bacteria adhere to one another.

DALY: Disability-adjusted life-year. Allows comparison of morbidity among differing diseases. Diseases are assigned 'DALY weights' based on the duration of disease, the level of disability caused by the disease, and the chance of death (the ultimate disability). The concept was developed and expanded by WHO.

Dose response equation (or curve, or model): Relationship of the mean number of pathogens ingested to the probability of a response, such as infection or disease.

Dysentery: An enteric disease characterized by intestinal inflammation and bloody stools.

EAWAG: Swiss Federal Institute for Environmental Science and Technology (German acronym).

E. coli: *Escherichia coli* bacteria. Usually a commensal inhabitant of the mammalian gut, but some strains can cause disease.

EITS model: Environmental infection transmission system model. Resembles SIR & other similar models, but directly models pathogens in the environment, in addition to states of infection by hosts.

Endemic: Describes a disease that is constantly present in a community. Hyperendemic denotes constant presence at high levels. Endemic disease levels may fluctuate over time, but rapid increases in disease are described as epidemic.

Epidemic: Temporarily increased levels of a disease in a community.

FFU: Focus-forming unit. Analogous to CFU or PFU, but applies to viruses in cell culture.

Distinct sites that are disrupted on a lawn of cells are considered to have arisen from a single FFU (and perhaps a single virion, if they do not adhere to one another).

Fomite: A physical object that can become contaminated and thus transfer pathogens between hosts.

Helminths: Worms.

HWT: Household water treatment. Synonym of POU.

Incidence, or incidence rate: Number of new cases of a disease in a population divided by time (e.g., cases of diarrhea per child per year). This is a measure of risk.

Incubation period: Time from exposure to a pathogen until the first symptoms develop. Usually longer than the prepatent period.

ID: Incidence difference. Analogous to incidence ratio (IR), but ID is obtained by subtracting the incidence in the intervention group from the incidence in the control (non-intervention) group. Also known as attributable risk.

IR: Incidence ratio, a comparison of two incidence rates. Generally the incidence in the intervention group is divided by the incidence in the control (non-intervention) group. It is a type of relative risk, representing the proportion of the incidence remaining in the population after an intervention has been applied.

LFF: LifeStraw® Family Filter, a HWT gravity-fed filtration device capable of removing viruses. It is produced and distributed by the Vestergaard Frandsen corporation.

LRV: Log₁₀ reduction value; number of factors of 10 that have been inactivated, e.g., an LRV of 2 means that 99% of microorganisms have been inactivated, LRV of 3 means that 99.9% of microorganisms have been inactivated, etc.

MATLAB: A software package (MATrix LABoratory) that is useful for programming computer simulations, among other things.

Matlab: A region of Bangladesh in which many studies of diarrhea have been conducted.

Morbidity ratio: Proportion of infected hosts that develop symptoms. Sometimes called the illness-to-infection ratio.

NGO: 'Non-governmental organization', generally a private nonprofit organization conducting human development work.

NTU: Nephelometric turbidity units, a measure of cloudiness of water. 0 is perfectly clear. The USEPA requires that municipally treated water have < 0.3 NTU in ≥ 95% of monthly samples. Water that is so cloudy that it is opaque might have an NTU of several hundred.

Odds: The number of times an outcome occurs divided by the number of times the outcome does not occur. When the odds of an outcome in two different groups are expressed as a ratio, this is termed an 'odds ratio' or OR. ORs are analogous to relative risk and approximate relative risk when the disease is uncommon (in that circumstance, number of nonoccurrences ≈ entire population), though ORs are often used in contexts where a proper

risk measure is unavailable.

ORS or ORT: Oral rehydration solution = oral rehydration salts = oral rehydration therapy.

Giving a formulation of sugar and electrolytes in clean water by mouth to an ill person in order to replace fluids, electrolytes, and energy lost to diarrhea.

Outbreak: A small or localized epidemic.

Persistent diarrhea: Diarrhea lasting longer than 14 days.

Person-time: Number of persons observed multiplied by the average amount of time during which each person was observed. Analogous to 'work-hours'.

PFU: Plaque-forming unit. Analogous to CFU, but applies to bacteriophages (viruses of bacteria) quantified on a lawn of bacteria. Round cleared areas in the lawn of bacteria are assumed to have arisen from a single bacteriophage.

POU: Point-of-use water treatment. Synonym of HWT.

Prepatent period: Time from exposure to a pathogen until the pathogen can be detected in the host. Usually shorter than the incubation period.

Prevalence, longitudinal: Amount of person-time spent ill divided by the total amount of person-time observed. This is a measure of risk.

Prevalence, point: Proportion of a population ill with a disease at a single point in time.

Preventable fraction: Proportion of disease that is prevented by a public health intervention (such as household water treatment or handwashing) in a particular population.

Rate ratio: See relative risk. Rates describe occurrence per unit time. Simple proportions are often incorrectly called rates.

Relative risk: The ratio of two risk measures, used to illustrate the magnitude of an effect. By convention, relative risks under 1 denote a protective effect, while relative risks over 1 indicate an adverse effect.

Risk: The likelihood (loosely defined) of a particular adverse outcome occurring. The number of times a particular outcome occurs divided by the total number of outcomes.

Safe storage: A common attribute of HWT methods, incorporating a storage vessel for water which has a narrow neck and (usually) a spigot, to prevent hands or other objects from (re)contaminating the water within.

SIR model: A model of infection transmission in which hosts can occupy 3 states in this order: susceptible, infectious, and removed (meaning immune or dead). Commonly modeled by a simple system of differential equations describing the rates by which hosts transfer between states, though other methods may be used. Often elaborated to include other states of infection or other orderings of states, e.g., SIS (susceptible-infectious-susceptible), SEIS (susceptible-exposed-infectious-susceptible), SEIR (susceptible-exposed-infectious-removed), etc.

SODIS: Solar disinfection, a HWT method in which contaminated water is placed in clear plastic bottles, which are then placed in the sun. Microorganisms are inactivated by a combination of UV irradiation and heating.

Sustainability: The ability of an intervention to be accepted by a community and perpetually used without any input from outside the community. Also pertains to management of resources in such a way that the resource is not depleted, e.g., using water from an aquifer at a rate no greater than its rate of recharge.

TTC: Thermotolerant coliforms. *E. coli* is an organism in this group. Commonly used as an indicator of fecal contamination of water.

UN: The United Nations.

UNICEF: The United Nations Children's Fund (originally United Nations International Children's Emergency Fund). An international development charity and agency that

particularly focuses on children and mothers.

USEPA: Environmental Protection Agency (United States government agency).

UV: Ultraviolet light, which is sometimes used to inactivate pathogens in water.

Weaning: Cessation of breastfeeding and introduction of additional foods, a process that can span months or years. Initiation of weaning (i.e., cessation of exclusive breastfeeding) is associated with sharply increased diarrhea risk.

WHO: World Health Organization.

Z-score: Number of standard deviations away from the mean. In nutrition, z-scores refer to standard distributions defining weight-for-height, height-for-age, and weight-for-age for a well-nourished reference population. Z-scores of 2 or 3 are commonly used as cutoffs for adverse nutrition outcomes.

REFERENCES

11. REFERENCES

- Abad FX, Pintó RM and Bosch A (1994) Survival of enteric viruses on environmental fomites. *Applied and Environmental Microbiology*. 60 (10), 3704–3710.
- Abba K, Sinfield R, Hart CA and Garner P (2009) Pathogens associated with persistent diarrhoea in children in low and middle income countries: systematic review. *BMC Infectious Diseases*. 9, 88.
- Adams M and Nicolaides L (1997) Review of the sensitivity of different foodborne pathogens to fermentation. *Food Control*. 8 (5-6), 227–239.
- Akinbami FO, Erinoso O and Akinwolere OA (1995) Defaecation pattern and intestinal transit in Nigerian children. *African Journal of Medicine and Medical Sciences*. 24 (4), 337–341.
- Akpata ES (2004) Fluoride ingestion from drinking water by Nigerian children aged below 10 years. *Community Dental Health*. 21 (1), 25–31.
- American Water Works Association (1999) *Waterborne Pathogens: Manual of Water Supply Practices*. Denver, CO: American Water Works Association
- Anderson RM and May RM (1991) *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford University Press
- Anon (2012) QMRAwiki. [Online] Available at: <http://camrawiki.anr.msu.edu> (accessed 26/07/12).
- Ansari SA, Sattar SA, Springthorpe VS, Wells GA and Tostowaryk W (1989) In vivo protocol for testing efficacy of hand-washing agents against viruses and bacteria: experiments with rotavirus and Escherichia coli. *Applied and Environmental Microbiology*. 55 (12), 3113–3118.
- Arnold B, Arana B, Mäusezahl D, Hubbard Alan and Colford JM (2009) Evaluation of a pre-existing, 3-year household water treatment and handwashing intervention in rural Guatemala. *International Journal of Epidemiology*. 38 (6), 1651–1661.
- Arnold BF and Colford JM (2007) Treating water with chlorine at point-of-use to improve water quality and reduce child diarrhea in developing countries: a systematic review and meta-analysis. *The American Journal of Tropical Medicine and Hygiene*. 76 (2), 354–364.
- Arnold BF, Khush RS, Ramaswamy P, London AG, Rajkumar P, Ramaprabha P, Durairaj N, Hubbard AE, Balakrishnan K and Colford JM Jr (2010) Causal inference methods to study nonrandomized, preexisting development interventions. *Proceedings of the National Academy of Sciences of the United States of America*. 107 (52), 22605–22610.
- Aronson JK (2007) Compliance, concordance, adherence. *British Journal of Clinical*

Pharmacology. 63 (4), 383–384.

- Atia AN and Buchman AL (2009) Oral rehydration solutions in non-cholera diarrhea: a review. *The American Journal of Gastroenterology*. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19550407> (accessed 04/09/09).
- Atmar RL, Opekun AR, Gilger MA, Estes Mary K, Crawford SE, Neill FH and Graham DY (2008) Norwalk virus shedding after experimental human infection. *Emerging Infectious Diseases*. 14 (10), 1553–1557.
- Awumbila M and Momsen JH (1995) Gender and the environment. Women's time use as a measure of environmental change. *Global Environmental Change: Human and Policy Dimensions*. 5 (4), 337–346.
- Ayad M, Piani AL, Barrère B, Ekouevi K and Otto J (1994) *Demographic characteristics of households*. Calverton, Maryland, USA: Macro International Inc. Available at: <http://www.measuredhs.com/publications/publication-cs14-comparative-reports.cfm>.
- Bahl R, Bhandari N, Saksena M, Strand T, Kumar GT, Bhan MK and Sommerfelt H (2002) Efficacy of zinc-fortified oral rehydration solution in 6- to 35-month-old children with acute diarrhea. *The Journal of Pediatrics*. 141 (5), 677–682.
- Banda K, Sarkar R, Gopal S, Govindarajan J, Harijan BB, Jeyakumar MB, Mitta P, Sadanala ME, Selwyn T, Suresh CR, Thomas VA, Devadason P, Kumar R, Selvapandian D, Kang Gagandeep and Balraj Vinohar (2007) Water handling, sanitation and defecation practices in rural southern India: a knowledge, attitudes and practices study. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 101 (11), 1124–1130.
- Barreto ML, Genser B, Strina Agostino, Teixeira MG, Assis AMO, Rego RF, Teles CA, Prado Matildes S, Matos SMA, Alcantara-Neves NM and Cairncross S (2010) Impact of a City-Wide Sanitation Programme in Northeast Brazil on Intestinal Parasites Infection in Young Children. *Environmental Health Perspectives*. Available at: <http://ehp03.niehs.nih.gov/article/fetchArticle.action?articleURI=info:doi/10.1289/ehp.1002058> (accessed 18/08/10).
- Barreto ML, Genser B, Strina A, Teixeira M, Assis A, Rego R, Teles C, Prado M, Matos S, Santos D, dos Santos L and Cairncross S (2007) Effect of city-wide sanitation programme on reduction in rate of childhood diarrhoea in northeast Brazil: assessment by two cohort studies. *Lancet*. 370 (9599), 1622–1628.
- Bates SJ, Trostle J, Cevallos WT, Hubbard Alan and Eisenberg JNS (2007) Relating diarrheal disease to social networks and the geographic configuration of communities in rural Ecuador. *American Journal of Epidemiology*. 166 (9), 1088–1095.
- Batterman S, Eisenberg JNS, Hardin R, Kruk ME, Lemos MC, Michalak AM, Mukherjee B, Renne E, Stein H, Watkins C and Wilson ML (2009) Sustainable control of water-related infectious diseases: a review and proposal for interdisciplinary health-based systems research. *Environmental Health Perspectives*. 117 (7), 1023–1032.

- Bern C, Martines J, de Zoysa I and Glass R I (1992) The magnitude of the global problem of diarrhoeal disease: a ten-year update. *Bulletin of the World Health Organization*. 70 (6), 705–714.
- Bhatnagar S, Bahl R, Sharma PK, Kumar GT, Saxena SK and Bhan MK (2004) Zinc with oral rehydration therapy reduces stool output and duration of diarrhea in hospitalized children: a randomized controlled trial. *Journal of Pediatric Gastroenterology and Nutrition*. 38 (1), 34–40.
- Bhutta ZA, Black R E, Brown K H, Gardner JM, Gore S, Hidayat A, Khatun F, Martorell R, Ninh NX, Penny ME, Rosado J L, Roy SK, Ruel M, Sazawal S and Shankar A (1999) Prevention of diarrhea and pneumonia by zinc supplementation in children in developing countries: pooled analysis of randomized controlled trials. Zinc Investigators' Collaborative Group. *The Journal of Pediatrics*. 135 (6), 689–697.
- Bhutta ZA, Nelson EA, Lee WS, Tarr PI, Zablah R, Phua KB, Lindley K, Bass D and Phillips A (2008) Recent advances and evidence gaps in persistent diarrhea. *Journal of Pediatric Gastroenterology and Nutrition*. 47 (2), 260–5.
- Bischoff WE, Reynolds TM, Sessler CN, Edmond MB and Wenzel RP (2000) Handwashing compliance by health care workers: The impact of introducing an accessible, alcohol-based hand antiseptic. *Archives of Internal Medicine*. 160 (7), 1017–1021.
- Bishop RF (1996) Natural history of human rotavirus infection. *Archives of Virology Supplementum*. 12, 119–128.
- Black R E (1993) Persistent diarrhea in children of developing countries. *The Pediatric Infectious Disease Journal*. 12 (9), 751–61; discussion 762–4.
- Black R E, Levine MM, Clements ML, Cisneros L and Daya V (1982) Treatment of experimentally induced enterotoxigenic *Escherichia coli* diarrhea with trimethoprim, trimethoprim-sulfamethoxazole, or placebo. *Reviews of Infectious Diseases*. 4 (2), 540–545.
- Black R E, Levine MM, Clements ML, Hughes TP and Blaser MJ (1988) Experimental *Campylobacter jejuni* infection in humans. *The Journal of Infectious Diseases*. 157 (3), 472–479.
- Black R E, Merson MH, Rowe B, Taylor PR, Abdul Alim AR, Gross RJ and Sack DA (1981) Enterotoxigenic *Escherichia coli* diarrhoea: acquired immunity and transmission in an endemic area. *Bulletin of the World Health Organization*. 59 (2), 263–268.
- Blaser MJ, Smith PD, Ravdin JI, Greenberg Harry B and Guerrant RL eds. (2002) *Infections of the Gastrointestinal Tract* (2nd edition). Philadelphia: Lippincott Williams & Wilkins
- Boehm AB (2007) Enterococci concentrations in diverse coastal environments exhibit extreme variability. *Environmental Science & Technology*. 41 (24), 8227–8232.

- Boisson S, Kiyombo M, Sthreshley L, Tumba S, Makambo J and Clasen T (2010) Field assessment of a novel household-based water filtration device: a randomised, placebo-controlled trial in the Democratic Republic of Congo. *PLoS ONE*. 5 (9), e12613.
- Boschi-Pinto C, Velebit L and Shibuya K (2008) Estimating child mortality due to diarrhoea in developing countries. *Bulletin of the World Health Organization*. 86 (9), 710–717.
- Briscoe J (1984) Intervention studies and the definition of dominant transmission routes. *American Journal of Epidemiology*. 120 (3), 449–455.
- Brownawell AM, Caers W, Gibson GR, Kendall CWC, Lewis KD, Ringel Y and Slavin JL (2012) Prebiotics and the health benefits of fiber: current regulatory status, future research, and goals. *The Journal of Nutrition*. 142 (5), 962–974.
- Brown Joe and Clasen T (2012) High adherence is necessary to realize health gains from water quality interventions. *PLoS ONE*. 7 (5), e36735.
- Brown Joe, Sobsey MD and Proum S (2007) *Use of Ceramic Water Filters in Cambodia*. UNICEF, New York and WHO, Geneva
- Brown Kenneth H, Hambidge KM and Ranum P (2010) Zinc fortification of cereal flours: current recommendations and research needs. *Food and Nutrition Bulletin*. 31 (1 Suppl), S62–74.
- Bushen OY, Kohli A, Pinkerton RC, Dupnik K, Newman RD, Sears CL, Fayer R, Lima AAM and Guerrant RL (2007) Heavy cryptosporidial infections in children in northeast Brazil: comparison of *Cryptosporidium hominis* and *Cryptosporidium parvum*. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 101 (4), 378–384.
- Buswell CM, Herlihy YM, Lawrence LM, McGuiggan JT, Marsh PD, Keevil CW and Leach SA (1998) Extended survival and persistence of *Campylobacter* spp. in water and aquatic biofilms and their detection by immunofluorescent-antibody and -rRNA staining. *Applied and Environmental Microbiology*. 64 (2), 733–741.
- Cacciò SM, Thompson RCA, McLauchlin J and Smith HV (2005) Unravelling *Cryptosporidium* and *Giardia* epidemiology. *Trends in Parasitology*. 21 (9), 430–437.
- Calloway DH, Odell AC and Margen S (1971) Sweat and miscellaneous nitrogen losses in human balance studies. *The Journal of Nutrition*. 101 (6), 775–786.
- Carter MJ (2005) Enterically infecting viruses: pathogenicity, transmission and significance for food and waterborne infection. *Journal of Applied Microbiology*. 98 (6), 1354–1380.
- CDC (2011) *CDC Health Information for International Travel 2012: The Yellow Book* (1st edition). G. W. Brunette ed. Oxford University Press, USA
- CDC (2012) *Epidemiology and Prevention of Vaccine-Preventable Diseases* (12th edition). Washington, DC: Public Health Foundation Available at:

- <http://www.cdc.gov/vaccines/pubs/pinkbook/pink-chapters.htm>.
- Centre for Affordable Water and Sanitation Technology (2008) Chlorine disinfection. Available at: <http://www.cawst.org/index.php?id=132#Chlorine>.
- Chacín-Bonilla L, Barrios F and Sanchez Y (2008) Environmental risk factors for *Cryptosporidium* infection in an island from Western Venezuela. *Memórias do Instituto Oswaldo Cruz*. 103 (1), 45–49.
- Chappell Cynthia L, Okhuysen Pablo C, Langer-Curry R, Widmer G, Akiyoshi DE, Tanriverdi S and Tzipori S (2006) *Cryptosporidium hominis*: experimental challenge of healthy adults. *The American Journal of Tropical Medicine and Hygiene*. 75 (5), 851–857.
- Checkley W, Gilman R H, Epstein LD, Suarez M, Diaz JF, Cabrera L, Black R E and Sterling CR (1997) Asymptomatic and symptomatic cryptosporidiosis: their acute effect on weight gain in Peruvian children. *American Journal of Epidemiology*. 145 (2), 156–163.
- Chen LC, Scrimshaw NS and United Nations University eds. (1983) *Diarrhea and Malnutrition: Interactions, Mechanisms, and Interventions*. New York: Plenum Press
- Cherian T, Wang S and Mantel C (2012) Rotavirus vaccines in developing countries: the potential impact, implementation challenges, and remaining questions. *Vaccine*. 30 Suppl 1, A3–6.
- Chiller TM, Mendoza CE, Lopez MB, Alvarez M, Hoekstra Robert M, Keswick BH and Luby SP (2006) Reducing diarrhoea in Guatemalan children: randomized controlled trial of flocculant-disinfectant for drinking-water. *Bulletin of the World Health Organization*. 84 (1), 28–35.
- Clasen T (2009) *Scaling up household water treatment among low-income populations*. Geneva, Switzerland: World Health Organization Available at: http://www.who.int/household_water/research/household_water_treatment/en/index.html.
- Clasen T, Bartram J, Colford JM, Luby SP, Quick R and Sobsey MD (2009) Comment on ‘Household water treatment in poor populations: is there enough evidence for scaling up now?’ *Environmental Science & Technology*. 43 (14), 5542–5544; author reply 5545–5546.
- Clasen TF, Brown Joseph and Collin SM (2006) Preventing diarrhoea with household ceramic water filters: assessment of a pilot project in Bolivia. *International Journal of Environmental Health Research*. 16 (3), 231–9.
- Clasen T, Naranjo J, Frauchiger D and Gerba CP (2009) Laboratory assessment of a gravity-fed ultrafiltration water treatment device designed for household use in low-income settings. *The American Journal of Tropical Medicine and Hygiene*. 80 (5), 819–823.
- Clasen T, Roberts IG, Rabie T, Schmidt W-P and Cairncross S (2009) Interventions to improve water quality for preventing diarrhoea. *Cochrane Database of Systematic Reviews*. (1).

- Clasen T, Schmidt W-P, Rabie Tamer, Roberts I and Cairncross S (2007) Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis. *British Medical Journal*. 334 (7597), 782.
- Cook SM, Glass R I, LeBaron CW and Ho MS (1990) Global seasonality of rotavirus infections. *Bulletin of the World Health Organization*. 68 (2), 171–177.
- Cooper BS, Pitman RJ, Edmunds WJ and Gay NJ (2006) Delaying the international spread of pandemic influenza. *PLoS Medicine*. 3 (6), e212.
- Coulliette AD, Peterson LA, Mosberg JAW and Rose Joan B (2010) Evaluation of a new disinfection approach: efficacy of chlorine and bromine halogenated contact disinfection for reduction of viruses and microcystin toxin. *The American Journal of Tropical Medicine and Hygiene*. 82 (2), 279–288.
- Cravioto A, Reyes RE, Trujillo F, Uribe F, Navarro A, De La Roca JM, Hernández JM, Pérez G and Vázquez V (1990) Risk of diarrhea during the first year of life associated with initial and subsequent colonization by specific enteropathogens. *American Journal of Epidemiology*. 131 (5), 886–904.
- Crump J, Otieno P, Slutsker L, Keswick BH, Rosen D, Hoekstra R, Vulule J and Luby SP (2005) Household based treatment of drinking water with flocculant-disinfectant for preventing diarrhoea in areas with turbid source water in rural western Kenya: cluster randomised controlled trial. *British Medical Journal*. 331 (7515), 478–481.
- Curtis VA and Cairncross S (2003) Effect of washing hands with soap on diarrhoea risk in the community: a systematic review. *The Lancet Infectious Diseases*. 3 (5), 275–81.
- Curtis VA, Cairncross S and Yonli R (2000) Domestic hygiene and diarrhoea - pinpointing the problem. *Tropical Medicine & International Health*. 5 (1), 22–32.
- Curtis VA, Danquah LO and Aunger RV (2009) Planned, motivated and habitual hygiene behaviour: an eleven country review. *Health Education Research*. 24 (4), 655–673.
- Dalton CB, Mintz ED, Wells JG, Bopp CA and Tauxe RV (1999) Outbreaks of enterotoxigenic *Escherichia coli* infection in American adults: a clinical and epidemiologic profile. *Epidemiology and Infection*. 123 (1), 9–16.
- Danciger M and Lopez M (1975) Numbers of *Giardia* in the feces of infected children. *The American Journal of Tropical Medicine and Hygiene*. 24 (2), 237–242.
- Davies GJ, Crowder M, Reid B and Dickerson JW (1986) Bowel function measurements of individuals with different eating patterns. *Gut*. 27 (2), 164–169.
- Dechesne M, Soyeux E, Loret J, Westrell T, Senstrom T, Gornik V, Koch C, Exner M, Stanger M, Agutter P, Lake R, Roser D, Ashbolt N, Dullemon Y, Hijnen W and Medema GJ (2006) *Pathogens in source water* Available at: http://www.microrisk.com/publish/cat_index_11.shtml.

- deRegnier DP, Cole L, Schupp DG and Erlandsen SL (1989) Viability of Giardia cysts suspended in lake, river, and tap water. *Applied and Environmental Microbiology*. 55 (5), 1223–1229.
- Doocy S and Burnham G (2006) Point-of-use water treatment and diarrhoea reduction in the emergency context: an effectiveness trial in Liberia. *Tropical Medicine & International Health*. 11 (10), 1542–1552.
- Duflo E, Glennerster R and Kremer M (2007) *Using randomization in development economics research: a toolkit*. London, UK: Centre for Economic Policy Research
- Dunk D (2007) A new look at bromine: a potential sleeping giant? *Water Conditioning & Purification Magazine*. 49 (10). Available at: <http://www.wcponline.com/NewsView.cfm?ID=3644>.
- DuPont H L, Chappell C L, Sterling CR, Okhuysen P C, Rose J B and Jakubowski W (1995) The infectivity of Cryptosporidium parvum in healthy volunteers. *The New England Journal of Medicine*. 332 (13), 855–859.
- DuPont H L, Formal SB, Hornick RB, Snyder MJ, Libonati JP, Sheahan DG, LaBrec EH and Kalas JP (1971) Pathogenesis of Escherichia coli diarrhea. *The New England Journal of Medicine*. 285 (1), 1–9.
- EAWAG (2004) SODIS technical notes 1 through 17. Available at: <http://www.sodis.ch/Text2002/T-Research.htm>.
- Eisenberg JNS, Hubbard Alan, Wade TJ, Sylvester MD, LeChevallier MW, Levy DA and Colford JM (2006) Inferences drawn from a risk assessment compared directly with a randomized trial of a home drinking water intervention. *Environmental Health Perspectives*. 114 (8), 1199–1204.
- Eisenberg JNS, Lei X, Hubbard AH, Brookhart M and Colford JM (2005) The role of disease transmission and conferred immunity in outbreaks: Analysis of the 1993 Cryptosporidium outbreak in Milwaukee, Wisconsin. *American Journal of Epidemiology*. 161 (1), 62–72.
- Eisenberg JNS, Moore K, Soller JA, Eisenberg D and Colford JM (2008) Microbial risk assessment framework for exposure to amended sludge projects. *Environmental Health Perspectives*. 116 (6), 727–733.
- Eisenberg JNS, Scott JC and Porco T (2007) Integrating disease control strategies: balancing water sanitation and hygiene interventions to reduce diarrheal disease burden. *American Journal of Public Health*. 97 (5), 846–852.
- Ejemot R, Ehiri J, Meremikwu M and Critchley J (2008) Hand washing for preventing diarrhoea. *Cochrane Database of Systematic Reviews*. (1). Available at: <http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD004265.pub2/abstract> (accessed 26/01/09).

- Elliott M, Stauber C, Koksai F, DiGiano F and Sobsey MD (2008) Reductions of *E. coli*, echovirus type 12 and bacteriophages in an intermittently operated household-scale slow sand filter. *Water Research*. 42 (10-11), 2662–2670.
- Enger KS, Nelson KL, Clasen T, Rose Joan B and Eisenberg JNS (2012) Linking quantitative microbial risk assessment and epidemiological data: informing safe drinking water trials in developing countries. *Environmental Science & Technology*. 46 (9), 5160–5167.
- Ensink JHJ, van der Hoek W and Amerasinghe FP (2006) *Giardia duodenalis* infection and wastewater irrigation in Pakistan. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 100 (6), 538–542.
- Erickson MC and Ortega YR (2006) Inactivation of protozoan parasites in food, water, and environmental systems. *Journal of Food Protection*. 69 (11), 2786–2808.
- Esrey SA (1996) Water, waste, and well-being: a multicountry study. *American Journal of Epidemiology*. 143 (6), 608–623.
- Esrey SA, Habicht JP and Casella G (1992) The complementary effect of latrines and increased water usage on the growth of infants in rural Lesotho. *American Journal of Epidemiology*. 135 (6), 659–666.
- Esrey SA, Potash JB, Roberts L and Shiff C (1991) Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, schistosomiasis, and trachoma. *Bulletin of the World Health Organization*. 69 (5), 609–21.
- Estrada-Garcia T, Lopez-Saucedo C, Thompson-Bonilla R, Abonce M, Lopez-Hernandez D, Santos JI, Rosado Jorge L, DuPont Herbert L and Long KZ (2009) Association of diarrheagenic *Escherichia coli* pathotypes with infection and diarrhea among Mexican children and association of atypical enteropathogenic *E. coli* with acute diarrhea. *Journal of Clinical Microbiology*. 47 (1), 93–98.
- Feachem RG and Koblinsky MA (1984) Interventions for the control of diarrhoeal diseases among young children: promotion of breast-feeding. *Bulletin of the World Health Organization*. 62 (2), 271–291.
- Fewtrell L and Colford JM (2005) Water, sanitation and hygiene in developing countries: interventions and diarrhoea - a review. *Water Science and Technology*. 52 (8), 133–142.
- Fewtrell L, Kaufmann R, Kay D, Enanoria W, Haller L and Colford JM (2005) Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. *Lancet Infectious Diseases*. 5 (1), 42–52.
- Fischer TK, Valentiner-Branth P, Steinsland H, Perch M, Santos G, Aaby P, Mølbak K and Sommerfelt H (2002) Protective immunity after natural rotavirus infection: a community cohort study of newborn children in Guinea-Bissau, west Africa. *The Journal of Infectious Diseases*. 186 (5), 593–597.

- Fischer Walker CL, Sack D and Black Robert E (2010) Etiology of diarrhea in older children, adolescents and adults: a systematic review. *PLoS Neglected Tropical Diseases*. 4 (8), e768.
- Fisman DN (2007) Seasonality of infectious diseases. *Annual Review of Public Health*. 28, 127–143.
- Flint KP (1987) The long-term survival of *Escherichia coli* in river water. *The Journal of Applied Bacteriology*. 63 (3), 261–270.
- Fudge BW, Easton C, Kingsmore D, Kiplamai FK, Onywera VO, Westerterp KR, Kayser B, Noakes TD and Pitsiladis YP (2008) Elite Kenyan endurance runners are hydrated day-to-day with ad libitum fluid intake. *Medicine and Science in Sports and Exercise*. 40 (6), 1171–1179.
- Genser B, Strina Agostino, Teles CA, Prado Matildes S and Barreto ML (2006) Risk factors for childhood diarrhea incidence: dynamic analysis of a longitudinal study. *Epidemiology*. 17 (6), 658–667.
- Gerba CP (2001) Application of quantitative risk assessment for formulating hygiene policy in the domestic setting. *The Journal of Infection*. 43 (1), 92–98.
- Ghana VAST Study Team (1993) Vitamin A supplementation in northern Ghana: effects on clinic attendances, hospital admissions, and child mortality. Ghana VAST Study Team. *Lancet*. 342 (8862), 7–12.
- Gilman R H, Marquis GS, Miranda E, Vestegui M and Martinez H (1988) Rapid reinfection by *Giardia lamblia* after treatment in a hyperendemic Third World community. *Lancet*. 1 (8581), 343–345.
- Gorter A, Sandiford P, Pauw J, Morales P, Perez R and Alberts H (1998) Hygiene behaviour in rural Nicaragua in relation to diarrhoea. *International Journal of Epidemiology*. 27 (6), 1090–1100.
- Grassly NC and Fraser C (2006) Seasonal infectious disease epidemiology. *Proceedings of the Royal Society B: Biological Sciences*. 273 (1600), 2541–2550.
- Greenberg Harry B and Estes Mary K (2009) Rotaviruses: from pathogenesis to vaccination. *Gastroenterology*. 136 (6), 1939–1951.
- Guerrant RL, Kosek M, Lima AAM, Lorntz B and Guyatt HL (2002) Updating the DALYs for diarrhoeal disease. *Trends in Parasitology*. 18 (5), 191–193.
- Guerrant RL, Oriá RB, Moore SR, Oriá MOB and Lima AAM (2008) Malnutrition as an enteric infectious disease with long-term effects on child development. *Nutrition Reviews*. 66 (9), 487–505.
- Guerrant RL, Schorling JB, McAuliffe JF and de Souza MA (1992) Diarrhea as a cause and an

- effect of malnutrition: diarrhea prevents catch-up growth and malnutrition increases diarrhea frequency and duration. *The American Journal of Tropical Medicine and Hygiene*. 47 (1 Pt 2), 28–35.
- Gundry S, Wright J and Conroy R (2004) A systematic review of the health outcomes related to household water quality in developing countries. *Journal of Water and Health*. 2 (1), 1–13.
- Haas CN, Rose J B, Gerba CP and Regli S (1993) Risk assessment of virus in drinking water. *Risk Analysis*. 13 (5), 545–552.
- Haas CN, Rose Joan B and Gerba CP (1999) *Quantitative Microbial Risk Assessment*. New York, NY: John Wiley & Sons, Inc.
- Haefner JW (2005) *Modeling Biological Systems: Principles and Applications* 2nd ed. New York: Springer
- Hall A, Hewitt G, Tuffrey V and de Silva N (2008) A review and meta-analysis of the impact of intestinal worms on child growth and nutrition. *Maternal & Child Nutrition*. 4 Suppl 1, 118–236.
- Halloran ME, Haber M, Longini I M and Struchiner CJ (1991) Direct and indirect effects in vaccine efficacy and effectiveness. *American Journal of Epidemiology*. 133 (4), 323–331.
- Halloran ME, Longini Ira M, Cowart DM and Nizam A (2002) Community interventions and the epidemic prevention potential. *Vaccine*. 20 (27-28), 3254–3262.
- Hamza H, Ben Khalifa H, Baumer P, Berard H and Lecomte JM (1999) Racecadotril versus placebo in the treatment of acute diarrhoea in adults. *Alimentary Pharmacology & Therapeutics*. 13 Suppl 6, 15–19.
- Han A and Moe K (1990) Household fecal contamination and diarrhea risk. *Journal of Tropical Medicine and Hygiene*. 93 (5), 333–336.
- Harro C, Sack D, Bourgeois AL, Walker R, DeNearing B, Feller A, Chakraborty S, Buchwaldt C and Darsley MJ (2011) A combination vaccine consisting of three live attenuated enterotoxigenic *Escherichia coli* strains expressing a range of colonization factors and heat-labile toxin subunit B is well tolerated and immunogenic in a placebo-controlled double-blind phase I trial in healthy adults. *Clinical and Vaccine Immunology*. 18 (12), 2118–2127.
- Havelaar AH, van Pelt W, Ang CW, Wagenaar JA, van Putten JPM, Gross U and Newell DG (2009) Immunity to *Campylobacter*: its role in risk assessment and epidemiology. *Critical Reviews in Microbiology*. 35 (1), 1–22.
- Hennekens CH and Buring JE (1987) *Epidemiology in Medicine* 1st ed. S. L. Mayrent ed. Boston: Little, Brown

- Henry FJ and Rahim Z (1990) Transmission of diarrhoea in two crowded areas with different sanitary facilities in Dhaka, Bangladesh. *The Journal of Tropical Medicine and Hygiene*. 93 (2), 121–126.
- Heymann DL ed. (2004) *Control of Communicable Diseases Manual* (18th edition). Washington, DC: American Public Health Association
- Hill Z, Kirkwood B and Edmond K (2004) *Family and Community Practices That Promote Child Survival, Growth and Development: A Review of the Evidence*. Geneva, Switzerland Available at: <http://www.who.int/child-adolescent-health>.
- Hrdy DB (1987) Epidemiology of rotaviral infection in adults. *Reviews of Infectious Diseases*. 9 (3), 461–469.
- Huffman SL and Combest C (1990) Role of breast-feeding in the prevention and treatment of diarrhoea. *Journal of Diarrhoeal Diseases Research*. 8 (3), 68–81.
- Hunter PR (2009) Household water treatment in developing countries: comparing different intervention types using meta-regression. *Environmental Science & Technology*. 43 (23), 8991–8997.
- Hunter PR and Thompson RCA (2005) The zoonotic transmission of Giardia and Cryptosporidium. *International Journal for Parasitology*. 35 (11-12), 1181–1190.
- Hunter PR, Zmirou-Navier D and Hartemann P (2009) Estimating the impact on health of poor reliability of drinking water interventions in developing countries. *The Science of the Total Environment*. 407 (8), 2621–2624.
- Huq A, Xu B, Chowdhury MA, Islam MS, Montilla R and Colwell RR (1996) A simple filtration method to remove plankton-associated Vibrio cholerae in raw water supplies in developing countries. *Applied and Environmental Microbiology*. 62 (7), 2508–2512.
- Hutton G and Haller Laurence (2004) *Evaluation of the Costs and Benefits of Water and Sanitation Improvements at the Global Level*. Geneva: World Health Organization Available at: http://www.who.int/water_sanitation_health/wsh0404/en/index.html (accessed 08/04/09).
- Jamison DT, Breman JG, Measham AR, Alleyne G, Claeson M, Evans DB, Jha P, Mills A and Musgrove P eds. (2006) *Disease Control Priorities in Developing Countries* (2nd edition). New York: Oxford University Press Available at: <http://www.dcp2.org/pubs/DCP>.
- Jennings V, Lloyd-Smith B and Ironmonger D (1999) Household size and the Poisson distribution. *Journal of Population Research*. 16 (1), 65–84.
- Jokipii AM and Jokipii L (1977) Prepatency of giardiasis. *Lancet*. 1 (8021), 1095–1097.
- Kaper James B, Nataro James P and Mobley HL (2004) Pathogenic Escherichia coli. *Nature*

- Reviews: Microbiology*. 2 (2), 123–140.
- Kapikian AZ, Wyatt RG, Levine MM, Black R E, Greenberg H B, Flores J, Kalica AR, Hoshino Y and Chanock RM (1983) Studies in volunteers with human rotaviruses. *Developments in Biological Standardization*. 53, 209–218.
- Keeling MJ and Eames KTD (2005) Networks and epidemic models. *Journal of the Royal Society Interface*. 2 (4), 295–307.
- Keeling MJ and Rohani P (2008) *Modeling infectious diseases in humans and animals*. Princeton: Princeton University Press Available at: <http://www.modelinginfectiousdiseases.org/>.
- Kent GP, Greenspan JR, Herndon JL, Mofenson LM, Harris JA, Eng TR and Waskin HA (1988) Epidemic giardiasis caused by a contaminated public water supply. *American Journal of Public Health*. 78 (2), 139–143.
- Kim JJ, Kuntz KM, Stout NK, Mahmud S, Villa LL, Franco EL and Goldie SJ (2007) Multiparameter calibration of a natural history model of cervical cancer. *American Journal of Epidemiology*. 166 (2), 137–150.
- Kirchhoff LV, McClelland KE, Do Carmo Pinho M, Araujo JG, De Sousa MA and Guerrant RL (1985) Feasibility and efficacy of in-home water chlorination in rural North-eastern Brazil. *The Journal of Hygiene*. 94 (2), 173–180.
- Kirkpatrick BD, Haque R, Duggal P, Mondal D, Larsson C, Peterson K, Akter J, Lockhart L, Khan S and Petri WA (2008) Association between *Cryptosporidium* infection and human leukocyte antigen class I and class II alleles. *The Journal of Infectious Diseases*. 197 (3), 474–478.
- Koopman JS (1978) Diarrhea and school toilet hygiene in Cali, Colombia. *American Journal of Epidemiology*. 107 (5), 412–20.
- Koopman JS (2004) Modeling infection transmission. *Annual Review of Public Health*. 25, 303–326.
- Kosek M, Bern Caryn and Guerrant RL (2003) The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. *Bulletin of the World Health Organization*. 81 (3), 197–204.
- Lanata CF and Mendoza W (2002) *Improving diarrhoea estimates*. [Online] Available at: http://www.who.int/child_adolescent_health/documents/diarrhoea_estimates/en/index.html (accessed 06/07/10).
- Lantagne D and Gallo W (2008) *Safe Water for the Community: A Guide for Establishing a Community-Based Safe Water System Program* (1st edition). Centers for Disease Control and Prevention Available at: <http://www.cdc.gov/safewater/resources.html>.

- Last JM ed. (1995) *A Dictionary of Epidemiology* (3rd edition). New York: Oxford University Press
- Lazzerini M and Ronfani L (2008) Oral zinc for treating diarrhoea in children. *Cochrane Database of Systematic Reviews*. (3), CD005436.
- Leclerc H, Schwartzbrod L and Dei-Cas E (2002) Microbial agents associated with waterborne diseases. *Critical Reviews in Microbiology*. 28 (4), 371–409.
- Lembcke J, Gastañaduy AS and Brown K H (1989) Prediction of total daily fecal excretion during acute childhood diarrhea. *Journal of Pediatric Gastroenterology and Nutrition*. 9 (4), 467–472.
- Levine MM, Caplan ES, Waterman D, Cash RA, Hornick RB and Snyder MJ (1977) Diarrhea caused by *Escherichia coli* that produce only heat-stable enterotoxin. *Infection and Immunity*. 17 (1), 78–82.
- Levine MM, Rennels MB, Cisneros L, Hughes TP, Nalin DR and Young CR (1980) Lack of person-to-person transmission of enterotoxigenic *Escherichia coli* despite close contact. *American Journal of Epidemiology*. 111 (3), 347–355.
- Levy K, Hubbard AE and Eisenberg JNS (2009) Seasonality of rotavirus disease in the tropics: a systematic review and meta-analysis. *International Journal of Epidemiology*. 38 (6), 1487–1496.
- Levy K, Hubbard AE, Nelson KL and Eisenberg JNS (2009) Drivers of water quality variability in northern coastal Ecuador. *Environmental Science & Technology*. 43 (6), 1788–1797.
- Lim ML and Wallace MR (2004) Infectious diarrhea in history. *Infectious Disease Clinics of North America*. 18 (2), 261–274.
- Lindesmith L, Moe C, Marionneau S, Ruvoen N, Jiang X, Lindblad L, Stewart P, LePendou J and Baric R (2003) Human susceptibility and resistance to Norwalk virus infection. *Nature Medicine*. 9 (5), 548–553.
- Li S, Eisenberg JNS, Spicknall IH and Koopman JS (2009) Dynamics and control of infections transmitted from person to person through the environment. *American Journal of Epidemiology*. 170 (2), 257–265.
- Long KZ, Rosado Jorge L and Fawzi W (2007) The comparative impact of iron, the B-complex vitamins, vitamins C and E, and selenium on diarrheal pathogen outcomes relative to the impact produced by vitamin A and zinc. *Nutrition Reviews*. 65 (5), 218–232.
- Lowbury EJ, Lilly HA and Bull JP (1964) Disinfection of hands: removal of transient organisms. *British Medical Journal*. 2 (5403), 230–233.
- Luby SP, Agboatwalla M, Raza A, Sobel J, Mint ED, Baier K, Hoekstra R M, Rahbar MH, Hassan R, Qureshi SM and Gangarosa EJ (2001) Microbiologic effectiveness of hand

- washing with soap in an urban squatter settlement, Karachi, Pakistan. *Epidemiology and Infection*. 127 (2), 237–244.
- Luby SP, Agboatwalla Mubina, Bowen A, Kenah E, Sharker Y and Hoekstra Robert M (2009) Difficulties in maintaining improved handwashing behavior, Karachi, Pakistan. *The American Journal of Tropical Medicine and Hygiene*. 81 (1), 140–145.
- Luby SP, Agboatwalla Mubina, Feikin DR, Painter J, Billhimer W, Altaf A and Hoekstra Robert M (2005) Effect of handwashing on child health: a randomised controlled trial. *Lancet*. 366 (9481), 225–233.
- Luby SP, Agboatwalla Mubina, Painter J, Altaf A, Billhimer W, Keswick Bruce and Hoekstra Robert M (2006) Combining drinking water treatment and hand washing for diarrhoea prevention, a cluster randomised controlled trial. *Tropical Medicine & International Health*. 11 (4), 479–89.
- Lunn PG (2000) The impact of infection and nutrition on gut function and growth in childhood. *The Proceedings of the Nutrition Society*. 59 (1), 147–154.
- Makutsa P, Nzaku K, Ogutu P, Barasa P, Ombeki S, Mwaki A and Quick R E (2001) Challenges in implementing a point-of-use water quality intervention in rural Kenya. *American Journal of Public Health*. 91 (10), 1571–3.
- Manz DH (2007) *BioSand water filter technology: household concrete design*. [Online] Available at: <http://manzwaterinfo.ca/index.htm>.
- Manz DH (2009) *BSF Guidance Manual 3: Basic Operation of the Concrete BioSand Water Filter*. [Online] Available at: <http://manzwaterinfo.ca/index.htm>.
- Mata LJ (1978) *The Children of Santa María Cauqué: A Prospective Field Study of Health and Growth*. Cambridge, Mass: MIT Press
- Mattison K (2011) Norovirus as a foodborne disease hazard. *Advances in Food and Nutrition Research*. 62, 1–39.
- Mäusezahl D, Christen A, Pacheco GD, Tellez FA, Iriarte M, Zapata ME, Cevallos M, Hattendorf J, Cattaneo MD, Arnold B, Smith TA and Colford JM (2009) Solar drinking water disinfection (SODIS) to reduce childhood diarrhoea in rural Bolivia: a cluster-randomized, controlled trial. *PLoS Medicine*. 6 (8), e1000125.
- McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M and Fisher P (2007) The Hawthorne effect: a randomised, controlled trial. *BMC Medical Research Methodology*. 7, 30.
- McDade T and Worthman C (1998) The weanling's dilemma reconsidered: A biocultural analysis of breastfeeding ecology. *Journal of Developmental and Behavioral Pediatrics*. 19 (4), 286–299.
- McLennan SD, Peterson LA and Rose Joan B (2009) Comparison of point-of-use technologies

- for emergency disinfection of sewage-contaminated drinking water. *Applied and Environmental Microbiology*. 75 (22), 7283–7286.
- McMahon S, Caruso BA, Obure A, Okumu F and Rheingans RD (2011) Anal cleansing practices and faecal contamination: a preliminary investigation of behaviours and conditions in schools in rural Nyanza Province, Kenya. *Tropical Medicine & International Health*. 16 (12), 1536–1540.
- Mehnert DU and Stewien KE (1993) Detection and distribution of rotavirus in raw sewage and creeks in São Paulo, Brazil. *Applied and Environmental Microbiology*. 59 (1), 140–143.
- Mensah P and Tomkins A (2003) Household-level technologies to improve the availability and preparation of adequate and safe complementary foods. *Food and Nutrition Bulletin*. 24 (1), 104–125.
- Messner MJ, Chappell C L and Okhuysen P C (2001) Risk assessment for *Cryptosporidium*: a hierarchical Bayesian analysis of human dose response data. *Water Research*. 35 (16), 3934–3940.
- Miliotis MD and Bier J eds. (2003) *International Handbook of Foodborne Pathogens*. New York: M. Dekker
- Miller SM, Fugate EJ, Craver VO, Smith JA and Zimmerman JB (2008) Toward understanding the efficacy and mechanism of *Opuntia* spp. as a natural coagulant for potential application in water treatment. *Environmental Science & Technology*. 42 (12), 4274–4279.
- Montgomery MA, Desai MM and Elimelech M (2010) Assessment of latrine use and quality and association with risk of trachoma in rural Tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 104 (4), 283–289.
- Morris SS, Cousens SN, Kirkwood BR, Arthur P and Ross DA (1996) Is prevalence of diarrhea a better predictor of subsequent mortality and weight gain than diarrhea incidence? *American Journal of Epidemiology*. 144 (6), 582–588.
- Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, Heijne J, Sadkowska-Todys M, Rosinska M and Edmunds WJ (2008) Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Medicine*. 5 (3), e74.
- Motarjemi Y, Käferstein F, Moy G and Quevedo F (1993) Contaminated weaning food: a major risk factor for diarrhoea and associated malnutrition. *Bulletin of the World Health Organization*. 71 (1), 79–92.
- Mukherjee AK, Chowdhury P, Bhattacharya MK, Ghosh M, Rajendran K and Ganguly S (2009) Hospital-based surveillance of enteric parasites in Kolkata. *BMC Research Notes*. 2, 110.
- Murray CJL and Lopez AD eds. (1996) *The Global Burden of Disease: a Comprehensive*

Assessment of Mortality and Disability from Diseases, Injuries, and Risk Factors in 1990 and Projected to 2020. Cambridge, Mass.: Published by the Harvard School of Public Health on behalf of the World Health Organization and the World Bank

- Nataro J P and Kaper J B (1998) Diarrheagenic *Escherichia coli*. *Clinical Microbiology Reviews*. 11 (1), 142–201.
- Neto RC, dos Santos LU, Sato MIZ and Franco RMB (2010) *Cryptosporidium* spp. and *Giardia* spp. in surface water supply of Campinas, southeast Brazil. *Water Science and Technology*. 62 (1), 217–222.
- Oberhelman RA, Gilman Robert H, Sheen P, Cordova J, Zimic M, Cabrera Lilia, Meza R and Perez J (2006) An intervention-control study of corralling of free-ranging chickens to control *Campylobacter* infections among children in a Peruvian periurban shantytown. *American Journal of Tropical Medicine and Hygiene*. 74 (6), 1054–1059.
- Ogunbiyi TA (1978) Whole-gut transit rates and wet stool weight in an urban Nigerian population. *World Journal of Surgery*. 2 (3), 387–393.
- Ogunjimi B, Hens N, Goeyvaerts N, Aerts M, Van Damme P and Beutels P (2009) Using empirical social contact data to model person to person infectious disease transmission: An illustration for varicella. *Mathematical Biosciences*. 218 (2), 80–87.
- Olson M, Goh J, Phillips M, Guselle N and McAllister T (1999) *Giardia* cyst and *Cryptosporidium* oocyst survival in water, soil, and cattle feces. *Journal of Environmental Quality*. 28 (6), 1991–1996.
- Onyango AO, Kenya EU, Mbithi JJN and Ng'ayo MO (2009) Pathogenic *Escherichia coli* and food handlers in luxury hotels in Nairobi, Kenya. *Travel Medicine and Infectious Disease*. 7 (6), 359–366.
- Oreskes N, Shrader-Frechette K and Belitz K (1994) Verification, validation, and confirmation of numerical models in the earth sciences. *Science*. 263 (5147), 641–646.
- Oundo JO, Kariuki SM, Boga HI, Muli FW and Iijima Y (2008) High incidence of enteroaggregative *Escherichia coli* among food handlers in three areas of Kenya: a possible transmission route of travelers' diarrhea. *Journal of Travel Medicine*. 15 (1), 31–38.
- Pancorbo OC, Evanshen BG, Campbell WF, Lambert S, Curtis SK and Woolley TW (1987) Infectivity and antigenicity reduction rates of human rotavirus strain Wa in fresh waters. *Applied and Environmental Microbiology*. 53 (8), 1803–1811.
- Parashar UD, Hummelman EG, Bresee JS, Miller MA and Glass Roger I (2003) Global illness and deaths caused by rotavirus disease in children. *Emerging Infectious Diseases*. 9 (5), 565–572.
- Parkin RT (2008) *Foundations and frameworks for human microbial risk assessment*.

- Washington, DC: United States Environmental Protection Agency (USEPA) Available at: www.epa.gov/raf/files/epa_mra_fw_comparison_report_0609.pdf.
- Patel MM, Widdowson M-A, Glass Roger I, Akazawa K, Vinjé J and Parashar UD (2008) Systematic literature review of role of noroviruses in sporadic gastroenteritis. *Emerging Infectious Diseases*. 14 (8), 1224–1231.
- Paul BD (1955) *Health, Culture, and Community; Case Studies of Public Reactions to Health Programs*. New York: Russell Sage Foundation
- Peréz Cordon G, Cordova Paz Soldan O, Vargas Vásquez F, Velasco Soto JR, Sempere Bordes L, Sánchez Moreno M and Rosales MJ (2008) Prevalence of enteroparasites and genotyping of Giardia lamblia in Peruvian children. *Parasitology Research*. 103 (2), 459–465.
- Pichichero ME, Losonsky GA, Rennels MB, Disney FA, Green JL, Francis AB and Marsocci SM (1990) Effect of dose and a comparison of measures of vaccine take for oral rhesus rotavirus vaccine. The Maryland Clinical Studies Group. *The Pediatric Infectious Disease Journal*. 9 (5), 339–344.
- Pickering AJ, Davis J, Walters SP, Horak HM, Keymer DP, Mushi D, Strickfaden R, Chynoweth JS, Liu J, Blum A, Rogers K and Boehm AB (2010) Hands, water, and health: fecal contamination in Tanzanian communities with improved, non-networked water supplies. *Environmental Science & Technology*. 44 (9), 3267–3272.
- Pickering AJ, Julian TR, Mamuya S, Boehm AB and Davis J (2011) Bacterial hand contamination among Tanzanian mothers varies temporally and following household activities. *Tropical Medicine & International Health*. 16 (2), 233–239.
- Pond K, Rueedi J and Pedley S (2004) *Pathogens in drinking water sources*. Guildford, Surrey, United Kingdom: Robens Centre for Public and Environmental Health, University of Surrey Available at: http://www.microrisk.com/publish/cat_index_11.shtml.
- Porter A (1916) An enumerative study of the cysts of Giardia (Lamblia) intestinalis in human dysenteric faeces. *Lancet*. 1, 1166–1169.
- Prado M S, Cairncross S, Strina A, Barreto ML, Oliveira-Assis AM and Rego S (2005) Asymptomatic giardiasis and growth in young children; a longitudinal study in Salvador, Brazil. *Parasitology*. 131 (Pt 1), 51–56.
- Qadri F, Svennerholm A-M, Faruque ASG and Sack RB (2005) Enterotoxigenic Escherichia coli in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. *Clinical Microbiology Reviews*. 18 (3), 465–483.
- Ramani S and Kang Gagandeep (2009) Viruses causing childhood diarrhoea in the developing world. *Current Opinion in Infectious Diseases*. 22 (5), 477–482.
- Razzolini MTP, da Silva Santos TF and Bastos VK (2010) Detection of Giardia and Cryptosporidium cysts/oocysts in watersheds and drinking water sources in Brazil urban

- areas. *Journal of Water and Health*. 8 (2), 399–404.
- Rendtorff RC (1954) The experimental transmission of human intestinal protozoan parasites. II. *Giardia lamblia* cysts given in capsules. *American Journal of Hygiene*. 59 (2), 209–220.
- Rivero-Marcotegui A, Olivera-Olmedo JE, Valverde-Visus FS, Palacios-Sarrasqueta M, Grijalba-Uche A and García-Merlo S (1998) Water, fat, nitrogen, and sugar content in feces: reference intervals in children. *Clinical Chemistry*. 44 (7), 1540–1544.
- Roberts MG and Heesterbeek JAP (2003) A new method for estimating the effort required to control an infectious disease. *Proceedings of the Royal Society B: Biological Sciences*. 270 (1522), 1359–1364.
- Rodriguez WJ, Kim HW, Brandt CD, Schwartz RH, Gardner MK, Jeffries B, Parrott RH, Kaslow RA, Smith JI and Kapikian AZ (1987) Longitudinal study of rotavirus infection and gastroenteritis in families served by a pediatric medical practice: clinical and epidemiologic observations. *The Pediatric Infectious Disease Journal*. 6 (2), 170–176.
- Rogers EM (2003) *Diffusion of Innovations* (5th edition). New York: Free Press
- Root GP (2001) Sanitation, community environments, and childhood diarrhoea in rural Zimbabwe. *Journal of Health, Population, and Nutrition*. 19 (2), 73–82.
- Rose A, Roy S, Abraham V, Holmgren G, George K, Balraj V, Abraham S, Muliyl J, Joseph A and Kang G (2006) Solar disinfection of water for diarrhoeal prevention in southern India. *Archives of Disease in Childhood*. 91 (2), 139–141.
- Rose J B, Haas CN and Regli S (1991) Risk assessment and control of waterborne giardiasis. *American Journal of Public Health*. 81 (6), 709–713.
- Rothman KJ (1986) *Modern Epidemiology* (1st edition). Boston: Little, Brown
- Roy SK, Tomkins AM, Akramuzzaman SM, Behrens RH, Haider R, Mahalanabis D and Fuchs G (1997) Randomised controlled trial of zinc supplementation in malnourished Bangladeshi children with acute diarrhoea. *Archives of Disease in Childhood*. 77 (3), 196–200.
- Schmidt W-P and Cairncross S (2009) Household water treatment in poor populations: is there enough evidence for scaling up now? *Environmental Science & Technology*. 43 (4), 986–992.
- Schmidt W-P, Genser B and Chalabi Z (2009) A simulation model for diarrhoea and other common recurrent infections: a tool for exploring epidemiological methods. *Epidemiology and Infection*. 137 (5), 644–653.
- Shediach-Rizkallah MC and Bone LR (1998) Planning for the sustainability of community-based health programs: conceptual frameworks and future directions for research, practice and policy. *Health Education Research*. 13 (1), 87–108.
- Singh G, Vajpayee P, Ram S and Shanker R (2010) Environmental reservoirs for enterotoxigenic

- Escherichia coli in south Asian Gangetic riverine system. *Environmental Science & Technology*. 44 (16), 6475–6480.
- Sjögren E, Ruiz-Palacios G and Kaijser B (1989) *Campylobacter jejuni* isolations from Mexican and Swedish patients, with repeated symptomatic and/or asymptomatic diarrhoea episodes. *Epidemiology and Infection*. 102 (1), 47–57.
- Sobsey MD (2002) *Managing water in the home: accelerated health gains from improved water supply*. [Online] Available at: http://www.who.int/water_sanitation_health/dwq/wsh0207/en/print.html (accessed 04/09/09).
- Sobsey MD and Brown Joe (2011) *Evaluating household water treatment options: health-based targets and microbiological performance specifications*. Geneva, Switzerland: World Health Organization Available at: http://www.who.int/water_sanitation_health/publications/2011/household_water/en/index.html (accessed 11/07/11).
- Sobsey MD, Stauber C, Casanova L, Brown JM and Elliott M (2008) Point of use household drinking water filtration: A practical, effective solution for providing sustained access to safe drinking water in the developing world. *Environmental Science & Technology*. 42 (12), 4261–4267.
- Solo-Gabriele HM, LeRoy Ager A, Fitzgerald Lindo J, Dubón JM, Neumeister SM, Baum MK and Palmer CJ (1998) Occurrence of *Cryptosporidium* oocysts and *Giardia* cysts in water supplies of San Pedro Sula, Honduras. *Pan American Journal of Public Health*. 4 (6), 398–400.
- Spangenberg ER, Greenwald AG and Sprott DE (2008) Will you read this article's abstract? Theories of the question–behavior effect. *Journal of Consumer Psychology*. 18 (2), 102–106.
- Sphere Project (2011) *Humanitarian Charter and Minimum Standards in Humanitarian Response* (3rd edition). Rugby, United Kingdom: Practical Action Publishing Available at: www.sphereproject.org (accessed 11/06/12).
- Stockman LJ, Fischer TK, Deming M, Ngwira B, Bowie C, Cunliffe N, Bresee J and Quick Robert E (2007) Point-of-use water treatment and use among mothers in Malawi. *Emerging Infectious Diseases*. 13 (7), 1077–80.
- Sutra S, Srisontrisuk S, Panpurk W, Sutra P, Chirawatkul A, Snongchart N and Kusowon P (1990) The pattern of diarrhea in children in Khon Kaen, northeastern Thailand: I. The incidence and seasonal variation of diarrhea. *The Southeast Asian Journal of Tropical Medicine and Public Health*. 21 (4), 586–593.
- Taylor A and Higham DJ (2008) *CONTEST: a controllable test matrix toolbox for MATLAB*. University of Strathclyde Available at: http://www.mathstat.strath.ac.uk/research/groups/numerical_analysis/contest/toolbox.

- Taylor A and Higham DJ (2009) CONTEST: A Controllable Test Matrix Toolbox for MATLAB. *ACM Transactions on Mathematical Software*. 35 (4).
- Taylor CE and Greenough WB 3rd (1989) Control of diarrheal diseases. *Annual Review of Public Health*. 10, 221–244.
- Teunis Peter F M, Moe CL, Liu P, Miller SE, Lindesmith L, Baric RS, Le Pendu J and Calderon RL (2008) Norwalk virus: how infectious is it? *Journal of Medical Virology*. 80 (8), 1468–1476.
- Teunis P F M, van der Heijden OG, van der Giessen JWB and Havelaar A (1996) *The dose-response relation in human volunteers for gastro-intestinal pathogens*. Bilthoven, The Netherlands Available at: <http://rivm.openrepository.com/rivm/handle/10029/9966> (accessed 23/07/10).
- Thompson KM, Duintjer Tebbens RJ and Pallansch MA (2006) Evaluation of response scenarios to potential polio outbreaks using mathematical models. *Risk Analysis: An Official Publication of the Society for Risk Analysis*. 26 (6), 1541–56.
- Tonglet R, Mahangaiko Lembo E, Zihindula PM, Wodon A, Dramaix M and Hennart P (1999) How useful are anthropometric, clinical and dietary measurements of nutritional status as predictors of morbidity of young children in central Africa? *Tropical Medicine & International Health*. 4 (2), 120–130.
- Toranzos GA, Gerba CP and Hanssen H (1988) Enteric viruses and coliphages in Latin America. *Toxicity Assessment: An International Journal*. 3 (5), 491–510.
- Tuite AR, Fisman DN, Kwong JC and Greer AL (2010) Optimal pandemic influenza vaccine allocation strategies for the Canadian population. *PloS One*. 5 (5), e10520.
- University of Maryland (2012) *Global Enterics Multi-Center Study (GEMS)*. [Online] Available at: <http://medschool.umaryland.edu/GEMS/> (accessed 05/04/12).
- USAID (2012) *Demographic and Health Surveys*. [Online] Available at: <http://www.measuredhs.com/> (accessed 21/05/12).
- USAID, UNICEF and World Health Organization (2005) Diarrhoea treatment guidelines including new recommendations for the use of ORS and zinc supplementation for clinic-based healthcare workers.
- USEPA (2011) *Exposure Factors Handbook: 2011 Edition*. Washington, DC: Exposure Assessment Group, Office of Health and Environmental Assessment, U.S. Environmental Protection Agency Available at: <http://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=236252>.
- USEPA (1987) Guide standard and protocol for testing microbiological water purifiers. Available at: <http://www.biovir.com/Images/pdf061.pdf>.

- USEPA (2012) *National Primary Drinking Water Regulations*. [Online] Available at: <http://water.epa.gov/drink/contaminants/index.cfm> (accessed 18/05/12).
- Valentiner-Branth P, Steinsland H, Fischer TK, Perch M, Scheutz F, Dias F, Aaby P, Mølbak K and Sommerfelt H (2003) Cohort study of Guinean children: incidence, pathogenicity, conferred protection, and attributable risk for enteropathogens during the first 2 years of life. *Journal of Clinical Microbiology*. 41 (9), 4238–4245.
- VanDerslice J and Briscoe J (1995) Environmental interventions in developing countries: interactions and their implications. *American Journal of Epidemiology*. 141 (2), 135–44.
- VanDerslice J, Popkin B and Briscoe J (1994) Drinking-water quality, sanitation, and breast-feeding: their interactive effects on infant health. *Bulletin of the World Health Organization*. 72 (4), 589–601.
- Velázquez FR, Matson DO, Calva JJ, Guerrero L, Morrow AL, Carter-Campbell S, Glass R I, Estes M K, Pickering LK and Ruiz-Palacios GM (1996) Rotavirus infections in infants as protection against subsequent infections. *The New England Journal of Medicine*. 335 (14), 1022–1028.
- Vergara M, Quiroga M, Grenon S, Pegels E, Oviedo P, Deschutter J, Rivas M, Binsztein N and Claramount R (1996) Prospective study of enteropathogens in two communities of Misiones, Argentina. *Revista do Instituto de Medicina Tropical de São Paulo*. 38 (5), 337–347.
- Vesikari T, Ruuska T, Bogaerts H, Delem A and André F (1985) Dose-response study of RIT 4237 oral rotavirus vaccine in breast-fed and formula-fed infants. *Pediatric Infectious Disease*. 4 (6), 622–625.
- Vollet JJ, Ericsson CD, Gibson G, Pickering LK, DuPont H L, Kohl S and Conklin RH (1979) Human rotavirus in an adult population with travelers' diarrhea and its relationship to the location of food consumption. *Journal of Medical Virology*. 4 (2), 81–87.
- Waddington H, Snilstveit B, White H and Fewtrell L (2009) *Water, sanitation, and hygiene interventions to combat childhood diarrhoea in developing countries*. [Online] Available at: <http://www.3ieimpact.org/page.php?pg=synthetic> (accessed 15/07/11).
- Walker RI, Steele D and Aguado T (2007) Analysis of strategies to successfully vaccinate infants in developing countries against enterotoxigenic *E. coli* (ETEC) disease. *Vaccine*. 25 (14), 2545–2566.
- Wallinga J, Teunis P and Kretzschmar M (2006) Using Data on Social Contacts to Estimate Age-specific Transmission Parameters for Respiratory-spread Infectious Agents. *American Journal of Epidemiology*. 164 (10), 936–944.
- Ward RL, Bernstein DI, Young EC, Sherwood JR, Knowlton DR and Schiff GM (1986) Human rotavirus studies in volunteers: determination of infectious dose and serological response to infection. *The Journal of Infectious Diseases*. 154 (5), 871–880.

- Ward RL, Knowlton DR and Pierce MJ (1984) Efficiency of human rotavirus propagation in cell culture. *Journal of Clinical Microbiology*. 19 (6), 748–753.
- Weaver LT (1988) Bowel habit from birth to old age. *Journal of Pediatric Gastroenterology and Nutrition*. 7 (5), 637–640.
- Wenman WM, Hinde D, Feltham S and Gurwith M (1979) Rotavirus infection in adults. Results of a prospective family study. *The New England Journal of Medicine*. 301 (6), 303–306.
- Wennerås C and Erling V (2004) Prevalence of enterotoxigenic *Escherichia coli*-associated diarrhoea and carrier state in the developing world. *Journal of Health, Population, and Nutrition*. 22 (4), 370–382.
- Wickramanayake G, Rubin A and Sproul O (1985) Effects of ozone and storage temperature on giardia cysts. *Journal American Water Works Association*. 77 (8), 74–77.
- Wilson ME (2005) Diarrhea in nontravelers: risk and etiology. *Clinical Infectious Diseases*. 41 Suppl 8, S541–546.
- Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, Gluud C, Martin RM, Wood AJG and Sterne JAC (2008) Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *British Medical Journal*. 336 (7644), 601–605.
- World Health Organization *Disease and injury regional estimates*. [Online] Available at: http://www.who.int/healthinfo/global_burden_disease/estimates_regional/en/index.html (accessed 30/03/12).
- World Health Organization (2012) *Global Burden of Disease (GBD)*. [Online] Available at: http://www.who.int/healthinfo/global_burden_disease/en/ (accessed 15/02/12).
- World Health Organization (2008) *The global burden of disease: 2004 update*. Geneva, Switzerland: WHO Press Available at: http://www.who.int/topics/global_burden_of_disease/en/ (accessed 08/10/12).
- World Health Organization (2002) *World Health Report 2002: Reducing risks, promoting healthy life*. Geneva, Switzerland: World Health Organization
- World Health Organization and UNICEF (2010) *Progress on sanitation and drinking-water: 2010 update*. Geneva, Switzerland: World Health Organization Available at: http://www.who.int/water_sanitation_health/publications/9789241563956/en/index.html (accessed 15/06/10).
- Xiao L (2010) Molecular epidemiology of cryptosporidiosis: an update. *Experimental Parasitology*. 124 (1), 80–89.
- Yeager BAC, Huttly SR, Bartolini R, Rojas M and Lanata CF (1999) Defecation practices of young children in a Peruvian shanty town. *Social Science & Medicine*. 49 (4), 531–541.

- Young CR, Ziprin RL, Hume ME and Stanker LH (1999) Dose response and organ invasion of day-of-hatch Leghorn chicks by different isolates of *Campylobacter jejuni*. *Avian Diseases*. 43 (4), 763–767.
- Zafar SN, Luby SP and Mendoza C (2010) Recall errors in a weekly survey of diarrhoea in Guatemala: determining the optimal length of recall. *Epidemiology and Infection*. 138 (2), 264–269.
- Zelner JL, Trostle J, Goldstick JE, Cevallos W, House JS and Eisenberg JNS (2012) Social connectedness and disease transmission: social organization, cohesion, village context, and infection risk in rural Ecuador. *American Journal of Public Health*.
- de Zoysa I and Feachem RG (1985) Interventions for the control of diarrhoeal diseases among young children: rotavirus and cholera immunization. *Bulletin of the World Health Organization*. 63 (3), 569–583.
- de Zoysa I, Rea M and Martines Jose (1991) Why promote breastfeeding in diarrhoeal control programmes? *Health Policy and Planning*. 6 (4), 371–379.
- Zwane AP, Zinman J, Van Dusen E, Pariente W, Null C, Miguel E, Kremer M, Karlan DS, Hornbeck R, Giné X, Duflo E, Devoto F, Crepon B and Banerjee A (2011) Being surveyed can change later behavior and related parameter estimates. *Proceedings of the National Academy of Sciences of the United States of America*. 108 (5), 1821–1826.