2010

This is to certify that the
dissertation entitled

HIERARCHICAL BAYESIAN MODELING OF HETEROGENEITY IN THE ASSOCIATION BETWEEN MILK PRODUCTION AND REPRODUCTIVE PERFORMANCE OF DAIRY COWS

presented by

NORA MARIA BELLO

has been accepted towards fulfillment
of the requirements for the

PhD degree in ANIMAL SCIENCE

Robert J. Tempelman
Major Professor's Signature

July 23, 2010
Date

HIERARCHICAL BAYESIAN MODELING OF HETEROGENEITY IN THE ASSOCIATION BETWEEN MILK PRODUCTION AND REPRODUCTIVE PERFORMANCE OF DAIRY COWS

By

Nora Maria Bello

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Animal Science

2010

ABSTRACT

HIERARCHICAL BAYESIAN MODELING OF HETEROGENEITY IN THE ASSOCIATION BETWEEN MILK PRODUCTION AND REPRODUCTIVE PERFORMANCE OF DAIRY COWS

By

Nora Maria Bello

The main objectives of this dissertation research were 1) to investigate the nature of the association between milk production and reproductive performance of dairy cows taking into consideration the *within*-herd (*cow*-level) and *between*-herd (*herd*-level) components of this association, and 2) to evaluate management factors and herd attributes as potential sources of heterogeneity in the association. A formal assessment of these objectives required the development of novel statistical methods, thereby setting an interdisciplinary foundation to this dissertation research.

First, this dissertation develops and validates hierarchical Bayesian extensions to classical bivariate linear mixed modeling of residual (cow-level) and random (herd-level) (co)variances for the joint analysis of two Gaussian outcomes. This approach involves modeling heterogeneous associations between outcomes using dispersion parameters generated from a square-root-free Cholesky reparameterization of (co)variances. These reparameterizations are unconstrained and hence can themselves be readily modeled as functions of fixed and random effects. This approach is extended further to bivariate generalized linear models, whereby modeling of heterogeneous associations between Gaussian and non-Gaussian outcomes, such as health and reproductive fitness, is facilitated using data augmentation techniques. The proposed hierarchical Bayesian

models constitute an important advancement in statistical methodology as they introduce a new dimension of heterogeneity in the study of complex biological systems, namely that of heterogeneous covariances (or correlations) between outcomes of interest.

The nature of the cow-level and herd-level associations between milk production and reproduction in dairy cows was explored by applying the aforementioned hierarchical Bayesian models to large datasets from commercial dairy farms in Michigan. Means, variances, and covariances between indicators of milk production and reproductive performance were jointly modeled as separate functions of management practices and herd attributes, with statistically important factors selected based on the Deviance Information Criterion. Evidence for heterogeneity in the association between milk production and reproduction was overwhelming. Most notably, inferred relationships were generally quite different and, in some cases, opposite in sign between the cow-level and the herd-level components. Secondly, management practices and herd attributes were identified as contributors to heterogeneity in the nature, as well as the magnitude, of the link between milk yield and reproductive performance. In particular, intensive management conditions appeared to contribute to a more favorable association in some cases (e.g., estimated herd calving interval decreased by 1.4±0.1 d per 100 kg increase in cumulative milk yield for herds using bovine somatotropin treatment) or to a partial alleviation of an overall antagonism in others (i.e. 21% greater pregnancy rates among herds implementing more frequent milking schemes). Understanding the multidimensional levels of heterogeneity in the associations between milk production and reproductive performance should have direct implications for tailoring dairy management programs that optimize overall dairy cow performance in current production systems.

To my *viejo*...

For passing on a driving curiosity for life

Salud, Pa!

ACKNOWLEDGEMENTS

First and foremost, I want to express my deepest and most sincere thank you to my major professor, Dr. Robert J. Tempelman, who bet high and against all odds when he accepted to take a reproductive physiologist with statistical aspirations as a doctoral student. I thank you for your thoughtful guidance over the years, for your unconditional support and seemingly-endless patience over the many things we have gone through together, for your respect as a colleague-to-be from day one, for sharing with me your human limitations, for showing me how to keep my feet on the ground while lifting my eyes to survey the horizon... I cannot think of a better mentor. Once again, thank you Rob!

I extend my gratitude to Dr. Ron Erskine, Dr. Roy Fogwell, Dr. Yuehua Cui and Dr. Dan Grooms for serving in my guidance committee and helping me stay focused on the target. I appreciate your patience, trust and flexibility with accommodating the innately interdisciplinary nature of my dissertation research. I am also grateful to Dr. Juan Steibel for the many thought-provoking discussions and his intellectual contributions to this work. I further thank Dr. Steve Bursian, Dr. Karen Plaut, Dr. Joe Domecq, Dr. Dennis Banks, Dr. Ron Bates and Dr. Cathy Ernst for the timely and much needed advice over the years.

I am in debt with Dr. Karen Plaut and the Department of Animal Science for the unconditional support I have received over the whole course of my graduate training at Michigan State University.

To my friends in Michigan, Rachel and Ivan Morris, Sue and Bob Carpenter, Duane Ullrey, Keith and Lynn Sterner, Ken and Liz Nobis, Angela Busta, Nuriye and Zeki Beyhan, thank you for letting me in as family and taking such good care of me. Much of my growth as a person throughout this journey, as well as the fun in the ride, I owe to you!

Finally, I want to thank my parents, Nora M. Garzon Marfort de Bello and Nicolas Bello, for being the safe reference in my life, even across cultures and thousands of miles. Also to my whole family in Argentina, especially my siblings, Coty, Ines and Pablo, thank you for your unconditional love and support throughout my long graduate training process. It is finally time to put it to work!

PREFACE

Chapter 1 in this dissertation was written in the style required for publication in the *Biometrical Journal*. Chapters 2 and 4 were written in the style required for publication in the *Journal of Dairy Science*. Chapter 3 was written in the style required for publication in *Biometrics*.

Chapter 1 corresponds to the pre-peer reviewed version of the following article: Bello N. M., J. P. Steibel and R. J. Tempelman. "Hierarchical Bayesian modeling of random and residual variance-covariance matrices in bivariate mixed effects models". Biometrical Journal 2010 June; 52(3):297-313.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

"...explain the complex visible by some simple invisible"

Jean Perrin (1870-1942), Nobel laureate, Physics

Milk yield and reproductive fertility stand as an essential dual axis for sustainability of any dairy cow operation. Indeed, milk production depends on the ability of a cow to become pregnant and give birth, thereby initiating and renewing the lactation cycle. Nevertheless, efforts to improve dairy fertility have been strongly overshadowed by focused attention on maximizing milk yield. Historical trends indicate a growing antagonistic relationship between production and fertility measures in dairy cows (Lucy, 2001). Over the past 85 years, average milk production per cow in the US increased over 4 fold, from ~1,900 kg/cow*year in 1924 to ~9,300 kg/cow*year in 2009 (http://www.nass.usda.gov/QuickStats). Meanwhile, dairy cow fertility declined to an all-time low, whereby currently approximately 1/3 of cows become pregnant to a given insemination (Norman et al., 2009). This further holds in Michigan as the state recently recorded a rolling herd average of 11,200 kg/cow and conception rates for lactating cows of approximately 35% (Dairy Herd Improvement Association, May 2010 Herd Summary DHI-202 State Reports).

Meanwhile, in recent years, the perception of a production-reproduction antagonism has been repeatedly challenged, thus triggering a passionate debate on the true nature of this association. Indeed, a growing list of studies now provides evidence for high milk yield being positively associated with fertility, whereby higher producing cows are also the ones most likely to become pregnant (Leblanc, 2010, Lopez-Gatius et al.,

2006, Peters and Pursley, 2002). Similarly, higher producing herds are commonly reported to have better reproductive performance (Laben et al., 1982, Leblanc, 2010, Nebel and McGilliard, 1993).

Concomitant with historical trends and the growing debate on the nature of the production-reproduction association, is a realization of dramatic changes in dairy herd management during the past few decades (Capper et al., 2009). Just as striking are the broad diversity and flexible dynamics of management practices currently implemented across dairy herds (Caraviello et al., 2006, Fulwider et al., 2008, Schefers et al., 2010). This implies considerable variability in the environment in which dairy cows perform, which in turn questions whether herd management may play a role as a source of heterogeneity on the association between milk production and reproductive performance of dairy cows.

Overall, the perceived controversy on the nature of the production-reproduction relationship in dairy cows and the potential role of management as a source of heterogeneity are dilemmas in urgent need for more comprehensive study. The complexity of the problem is substantial and includes multiple layers of intricacy in an inherently multivariate environment. Traditional statistical modeling strategies are not adequate to simultaneously address such multidimensional complexity. Hierarchical Bayesian models constitute a general framework and a viable alternative to approach this problem (O'Hagan, 2004). This introduction will briefly describe the key features of the Bayesian paradigm that make it uniquely suitable for our investigation and recent statistical advances that lay at the foundation of the methodology proposed to tackle the production-reproduction controversy.

1. On the relationship between milk production and dairy fertility:

1.1. The current dogma:

Greater milk production has long been associated with deteriorating reproductive performance in dairy cows. Indeed, one of the first recorded quotations on the subject dates back to a 1929 Minnesota Agricultural Experiment Station Bulletin, as cited by Hansen (2000). Later, a survey of dairy records from 1950 to 1985 in the state of New York showed that whereas milk production steadily increased from 4500 to 7500 kg/cow*year, conception rates of dairy cows decreased from 66% to 40% during the same period (Butler and Smith, 1989). More recent studies in the US raised even greater concerns as milk yields continued to increase while conception rates decreased to all-time lows (Butler, 1998, Lucy, 2001, Norman et al., 2009, Washburn et al., 2002). Similar reports from other countries (i.e. Ireland: Roche et al. (2000); United Kingdom: Royal et al. (2000); Australia: Macmillan et al. (1996); Spain: Lopez-Gatius (2003)) point towards a world-wide problem of considerable magnitude. Locally, the state of Michigan is certainly not immune to this antagonistic trend. Unless addressed, this problem seems likely to tarry indefinitely given the continued momentum of increasing milk yields and decreasing fertility levels.

Antagonistic historical trends were initially rationalized by claims that intensive selection for high milk production had led to unintended selection for low fertility. Indeed, animal breeders strongly argue for a genetic basis to the antagonism between milk production and reproduction (Castillo-Juarez et al., 2000, Hansen et al., 1983, Pryce et al., 2004, Seykora and Mcdaniel, 1983). For example, genetic correlations between

days open and 305-d milk yield ranged between 0.2 and 0.3 (Hansen et al., 1983) whereas genetic correlations between first service conception rate and mature equivalent milk yield ranged from -0.3 to -0.4 (Castillo-Juarez et al., 2000). Based on these reports, intensive selection for increasing milk productivity would result in an unintended decrease in reproductive performance. These genetic relationships certainly offer a plausible explanation for the historical trends observed. It should be noted, however, that the genetic information passed on from one generation to the other is only a fraction of the phenotype, namely heritability (Falconer, 1981). Heritability for fertility traits is consistently low across studies, with estimates commonly below 5% (Calus et al., 2005, Castillo-Juarez et al., 2000, Hansen et al., 1983, Seykora and Mcdaniel, 1983, Windig et al., 2006). In contrast, heritability for production traits varies in the range of 20 to 40% (Castillo-Juarez et al., 2000, Hansen et al., 1983, Seykora and Mcdaniel, 1983, Windig et al., 2006). Therefore, despite an antagonistic production-reproduction genetic correlation, the low heritabilities for reproductive traits raise questions about the relative importance of genetics on the chance that a cow will conceive to a given insemination. Instead, the production environment appears likely to exert a dominant role given the magnitude of the environmental component in the phenotype, as defined by one minus heritability. Adding further to the controversy, the genetic correlation between test-day milk yield and fertility traits fluctuates during lactation (Berry et al., 2003a, b), just as managerial practices do.

From a physiological perspective, energy balance has been proposed as a mechanistic link between high milk yield and poor reproductive performance (Butler, 2003). Most dairy cows in early lactation experience negative energy balance with

mobilization of adipose tissue because feed intake does not meet nutrient requirements for lactation (Bauman and Currie, 1980). The subsequent loss of body fat then signals physiological mechanisms in the reproductive endocrine cascade and can lead to disturbed reproduction (Lucy, 2003). However, studies have shown that highest producing cows are not necessarily the ones with the most extreme negative energy balance or the lowest body condition score. Rather, the ability of certain high producing cows to promptly maximize dry matter intake after parturition appears to minimize negative energy balance even with energy demands for high milk production (Lucy et al., 1992). As a result, the risk for anestrous and infertility seems associated with a finely-tuned balance between level of milk production and dry matter intake (Lucy et al., 1992), which in turn contributes to the volume of hepatic blood flow and steroid metabolism in the liver (Wiltbank et al., 2006).

1.2. Challenging the dogma:

A currently growing number of studies challenges the perception of a universal antagonistic relationship between milk production and reproduction. In fact, high milk production has been demonstrated to be positively associated with dairy cow fertility (Leblanc, 2010, Lof et al., 2007, Lopez-Gatius et al., 2006, Peters and Pursley, 2002). For instance, in cows yielding >50 kg of milk per day by 50 days in milk, the odds of pregnancy increased by a factor of 6.8 compared to cows producing below that level. As a result, an increase of 1 kg in milk yield at peak lactation was associated with an estimated decrease of 1.8 d in the interval from calving to conception (Lopez-Gatius et al., 2006). Similarly, cows with milk production above herd average had greater conception rates (45.8 vs. 33.8%) compared with their lower producing herdmates (Peters

and Pursley, 2002) and days open were reduced among cows from high compared to low producing herds (Emanuelson and Oltenacu, 1998). Favorable associations between milk production and reproduction were also apparent when the herd, rather than the cow, was evaluated as the unit of performance. For example, days open for the highest producing herds averaged one estrous cycle (i.e. ~21 days) shorter than for low producing herds (Laben et al., 1982). Similarly, high yielding herds averaged a 10-day shorter calving interval and had reduced odds of reproductive culling than low yielding herds, based on a sample of 2,700 dairy operations (Lof et al., 2007).

Further fueling this controversy, extraneous factors have been proposed to confound the nature of the relationship between milk production and reproduction. Lopez-Gatius et al. (2005a) argued that, after various management and cow factors were taken into account, no association between milk production and cow pregnancy was identified. These factors included herd, season, lactation number, insemination number, service sire and insemination technique. Two follow-up studies performed by the same research group reconfirmed these findings (Garcia-Ispierto et al., 2007, Lopez-Gatius et al., 2005b). Loss of body condition score has also been proposed to confound the association between production and reproduction; whereby after accounting for changes in cow body condition, embryonic and fetal losses were not significantly associated with milk energy output (Silke et al., 2002). Also, looking at herd as unit of performance, Windig et al. (2005) identified a subset of herds in which fertility was not associated with production level.

In summary, despite the apparently wide scope of evidence for an antagonistic association between milk yield and fertility, an increasing body of evidence is

challenging this perception in favor of a potentially favorable or neutral production-reproduction relationship. These seemingly opposite positions clearly define a polarized controversy for which there is need of further insight. My interest is to define conditions that jointly facilitate high milk yield and efficient reproduction and that will guide management actions and decisions at the farm level .

1.3. Potential role of management and herd factors:

Along with the changes in milk productivity and reproductive efficiency observed during the past century, dairy herd practices have also been modified substantially. Developments in management have affected growth, health, and lactation (Caraviello et al., 2006, Fulwider et al., 2008), thereby supporting a diverse, dynamic and vibrant dairy industry. Based on the predominant role of environment over genetics on dairy cow fertility, it may be possible that the exposure of cows to diverse management conditions affects the disparity of evidence or perceived conflict on the nature of the relationship between milk production and reproductive performance. In the following subsections, I will review management practices and herd attributes as potential candidates for sources of heterogeneity in the production-reproduction relationship in dairy cows.

1.3.1. Herd size as a historical indicator of changes in dairy management:

Herd size is perhaps one of the most conspicuous indicators of changes in herd management in the U.S. From 1965 to 2009, the number of dairy cows in the US decreased approximately 40%, from ~15 to ~9 million (http://www.nass.usda.gov/QuickStats). During the same period, the number of U.S. dairy operations decreased by roughly 94%, from ~1.1 million to ~65 thousand. Despite decreases in total cows and herds, the number of dairy herds with 500+ head has

increased by 43% since 1997 and a recent survey indicates that a 5-yr herd size goal for a representative subset of large commercial dairy farms is above 900 cows per operation (Caraviello et al., 2006). Clearly, the current trend in U.S. dairying is to consolidate into fewer and larger operations.

This shift toward larger farms is forcing a reevaluation of the traditional dairy management configuration (Lucy, 2001). For example, simply due to more cows, larger herds will require more time for completion of virtually every task in the farm including milking and movement of cows to and from their pens, mixing and delivery of feed, parturition assistance and management of transition cows, estrus detection and implementation of synchronization strategies, sorting and insemination of cows, pregnancy diagnoses and record keeping, just to name a few. The increase in time commitment and responsibilities can easily become overwhelming for former "jack-of-all-trades" small producers targeting herd expansion. As a result, responsibilities may need to be redistributed over a larger work force, whereby employees are required to become specialists on specific tasks (Bewley et al., 2001b). Under these circumstances, finding, training and retaining quality labor becomes a major issue to a successful dairy operation (Bewley et al., 2001b, Caraviello et al., 2006).

The implications of increasing herd size and subsequent changes on farm management on the association between milk production and reproductive performance of dairy cows remain unclear. Herd size has been reported as an important source of variability in milk productivity and reproductive performance of dairy cows (Cabrera et al., 2010, Fahey et al., 2002, Lof et al., 2007, Windig et al., 2006, Windig et al., 2005); however, the direction of the reported associations is not consistent.

1.3.2. Dynamic management strategies over time:

Aside from herd size, other trends in the dynamics of herd characteristics over the past few decades have been poorly recorded. A few surveys to commercial dairy farms in the Midwest (Bewley et al., 2001a, b, Fulwider et al., 2008) or across the U.S. (Caraviello et al., 2006, Jordan and Fourdraine, 1993) assess farm management practices at a point in time. If presented in a timeline, these surveys can provide some idea of progression of management practices over time, and thus serve to postulate candidate sources of heterogeneity in the association between milk production and reproduction

Adoption of synchronized breeding strategies is probably one of the hallmark tendencies in dairy management during the past 15 years. The proportion of herds incorporating this technology increased steadily from 1.9% in 1996 to approximately 20% in 2005 (Miller et al., 2007). In a survey of large commercial U.S. dairy farms, 90% indicated routine implementation of synchronization of estrus or ovulation for first and subsequent services (Caraviello et al., 2006). As a consequence, the use of natural service either as a main strategy or just for clean-up has steadily decreased. Indeed, 100% of the herds included in an early 1990's survey reported moving cows to the clean-up bull pen after 3.7 failed inseminations (Jordan and Fourdraine, 1993); conversely, only 44% of herds considered in an early 2000's survey kept a clean-up bull at all, in which case cows were moved only after 6.6 failed inseminations (Caraviello et al., 2006). Further, evidence indicates that bull service applied to 100% cows in 1944 but only to 30% of cows in 2005 (Capper et al., 2009).

Recombinant bovine somatotropin (bST) constitutes yet another manifestation of the rich dynamics of the dairy industry over the past couple decades. Introduced in the

US market in the early 1990's, bST technology enhanced productivity and productive efficiency while maintaining health and wellbeing of dairy cows (Bauman, 1999). Before consumer groups pushed bST out of the market in 2007, a survey had indicated that 71% of large U.S. commercial dairy herds supplemented cows with bST as part of their management practices (Caraviello et al., 2006).

Overall, the subsections above outline a cherry-pick of the management changes undergone by the dairy industry during the past few decades, which warrant their investigation as potential sources of heterogeneity in the association between milk production and reproduction of dairy cows. Additional components of dynamic management changes and herd attributes for further consideration include, but are not limited to, changes in herd prevalence of infectious agents that affect fertility and/or milk production, milking frequency and management of milking feeding groups.

2. Reevaluating the relationship between milk production and reproduction in dairy cows: Limitations of previous approaches

The previous section described the on-going controversy and discussed different aspects of the association between milk production and reproductive performance in dairy cows. A common, yet sometimes inadvertent, thread across such studies is an under appreciation of the dual components of the production-reproduction relationship, namely the *within*-herd (i.e. *cow*-level) versus *between*-herd (i.e. *herd*-level) components. Herds and cows constitute distinct units of performance and separate constituents of (co)variability that intertwine with each other due to the clustered nature of the data structure, to yield an overall phenotype. Whatever associations may be apparent at the

herd level (i.e., *between* herds) may not necessarily be apparent at the individual animal level (i.e., *within* herds), and vice-versa (Calus et al., 2005, Windig et al., 2005). If these distinctions are not made, any reported associations can be dangerously over-generalized or even biased! The importance of recognizing this issue was illustrated by Windig et al (2005), who showed that if herd-level information was disregarded, high milk yield was associated with poorer fertility.... but within herds, this relationship was quite diverse and fluctuated widely from strongly positive to strongly negative correlations! Such wide variation between cow and herd components may be the key to explain the conflicting evidence and current controversy on the relationship between milk production and reproductive efficiency of dairy cows. **To our knowledge, the relationship between productive and reproductive traits taking into separate consideration the dual cow and herd components has not been modeled.** Thus, it is unclear what factors, if any, and at what level, might be associated with the production-reproduction relationship in dairy cows.

Scope of inference constitutes another issue commonly overlooked in evaluating the production-reproduction controversy. Indeed, in many cases, data are collected at one or, at most, a few farms (i.e. narrow scope) but conclusions are formulated to apply across the industry (i.e. broad scope). Even among studies in which multiple herds are involved, their modeling as systematic blocking factors restricts the scope of inference (Tempelman, 2010). Narrowly scoped data, such as that obtained in one or a few herds, will likely be relevant for local decisions; however, conclusions may also be highly misleading if overly generalized to the overall dairy population across its heterogeneous managerial environments. This is specially the case if the data pertain to a research farm

that may not necessarily represent commercial herds, to which the final management conclusions are intended to apply. Under these circumstances, conflicting results between studies and contradictions between experimental data and field observations should not be surprising.

Statistical methodology that is based on univariate (i.e., single trait response) analysis represents an additional highly prevalent limitation in many production-reproduction association studies. In single-trait analyses, reproductive traits are modeled as a function of milk yield (Laben et al., 1982, Lopez-Gatius et al., 2006, Spalding et al., 1974), or, conversely, milk yield is compared between reproductively successful and unsuccessful females (Lopez-Gatius et al., 2006, Windig et al., 2005). In so doing, the prevailing assumption is that whichever trait is alternatively chosen to be the explanatory variable is measured without error and not influenced by other covariates in the model. Moreover, the univariate approach bluntly ignores the correlation between traits as a potentially important source of information, which in turn has a detrimental effect on precision of the inference (Riley et al., 2007, Sorensen et al., 2003). These limitations of univariate models are broadly recognized by animal breeders and geneticists, who in turn are more likely to implement standard multivariate analysis to appropriately estimate the genetic correlation between fertility traits and test-day milk yields (Berry et al., 2003a, b, Castillo-Juarez et al., 2000, Hansen et al., 1983, Tsuruta et al., 2009, Windig et al., 2006). Interestingly, several of these authors noted ad-hoc evidence for heterogeneity between environments in the magnitude of the genetic correlation between milk yield and fertility, when environments were described by stage of lactation (Berry et al., 2003a), herd size (Tsuruta et al., 2009) or a combination of management characteristics (Windig et al.,

2006). However, a general framework to explicitly model hierarchical heterogeneity of variance-covariance matrices in multivariate settings is lacking. Due to technical difficulties associated with positive-definiteness, the underlying premise of multivariate models is a fixed variance-covariance structure that is assumed to behave uniformly and remain constant across scenarios of risk factor combinations (Mardia et al., 1979).

In summary, there are multiple limitations in previous studies to assess the nature of the production-reproduction relationship. If these limitations are appropriately accounted for, the contradictory evidence and conflicts may be explained. The question is certainly a complex one due to the underlying hierarchy of the data structure (i.e. cow versus herd) and the need to assess management practices as potential sources of heterogeneity in the correlation between traits, as well as the technical limitations of the statistical methodology available for analysis.

3. Statistical approach:

3.1. Why Bayesian?

Analysis of hierarchical data with multiple layers of heterogeneity is generally computationally intractable if approached from classical statistical theory based on likelihood inference (Sorensen and Gianola, 2002). Alternatively, the Bayesian statistical framework is particularly suitable to these complexities due to the embedded hierarchical rationality of the Bayes paradigm and its direct inferential approach (Shoemaker et al., 1999, Sorensen and Gianola, 2002). In recent years, many areas of science and humanities have recognized the unique advantages of Bayesian statistics. Especially since the 1990's, when the development of powerful computational tools such as the simulation

intensive algorithm Markov Chain Monte Carlo (MCMC) (Gilks et al., 1996) facilitated the implementation of Bayesian methods. The modular format of MCMC provides immense flexibility to model highly hierarchical structures, such as needed to address the question on the relationship between production and reproduction in dairy cows.

3.2. Multivariate models and heterogeneous variance-covariance parameters

Joint modeling of two (or more) traits using multivariate techniques has the advantage of potentially sharper inference on both traits. The off-diagonal elements of a variance-covariance matrix represent the covariance between a pair of outcomes, which defines the nature of their relationship. Conceptually, the modeling of covariances would provide a formal methodological venue to capture heterogeneity in the association between outcomes. However, positive-definiteness constraint among variance-covariance elements imposes a very tangible technical limitation that renders a rigid structure to the multivariate variance-covariance matrix (Riley et al., 2007, Sorensen et al., 2003).

Recent developments in the medical statistical literature spawned computationally feasible and easily interpretable alternatives for flexible modeling of (co)variance elements (Pourahmadi, 1999, 2007). The key idea is a reparameterization of the variance-covariance matrix using a Cholesky-type decomposition, whereby the original matrix is decomposed into two unique matrices: a diagonal matrix and a lower triangular one. Jointly, the two new matrices retain all the information on the variation of and correlation between traits, with the clear advantage of a simpler structure that overrides positive definiteness technicalities. In fact, the reparameterized variances and covariances are unconstrained and mutually orthogonal such that each can be easily specified as a function of explanatory variables of interest (Pourahmadi, 2007). In this

dissertation, we extend this methodological approach to accommodate multiple hierarchical levels (i.e. *cow*- and *herd*-level) and layered sources of heterogeneity, as is of interest to investigate the link between milk yield and dairy cow fertility.

4. General hypothesis:

The *between*- and *within*-herd associations between milk production and reproductive performance of dairy cows are heterogeneous and depend upon management practices and herd attributes.

5. Specific Aims:

To approach the general hypothesis stated above, the core of this dissertation research is innately interdisciplinary and closely integrates concepts of animal physiology, and animal production systems with advanced elements of applied statistics.

The Specific Aims of this dissertation research are:

1) To develop and validate a hierarchical Bayesian extension to classical bivariate mixed effects methods to model heterogeneity in residual and random covariance matrices for the joint analysis of two Gaussian phenotypes;

2) To use the methodology developed in Specific Aim 1) to investigate the *within*-herd (cow-level) and *between*-herd (herd-level) associations between indicators of comprehensive (i.e. entire lactation) milk production and reproductive performance of Michigan dairy cows, including the evaluation of various herd management factors potentially affecting these associations;

3) To develop and validate a hierarchical Bayesian implementation of a bivariate generalized linear mixed-effect model for heterogeneous variance-covariance matrices in the context of a joint analysis of Gaussian and non-Gaussian traits;

4) To implement the methodology developed in Specific Aim 3) to investigate the associations between milk yield at and pregnancy outcome to first postpartum service of Michigan dairy cows, accounting for cow and herd as hierarchical units of performance and evaluating the role of management practices and herd attributes as potential sources of heterogeneity.

References

Bauman, D. E. 1999. Bovine somatotropin and lactation: from basic science to commercial application. Domest Anim Endocrinol 17(2-3):101-116.

Bauman, D. E. and W. B. Currie. 1980. Partitioning of Nutrients during Pregnancy and Lactation - a Review of Mechanisms Involving Homeostasis and Homeorhesis. Journal of Dairy Science 63(9):1514-1529.

Berry, D. P., F. Buckley, P. Dillon, R. D. Evans, M. Rath, and R. F. Veerkamp. 2003a. Genetic parameters for body condition score, body weight, milk yield, and fertility estimated using random regression models. Journal Dairy Science 86(11):3704-3717.

Berry, D. P., F. Buckley, P. Dillon, R. D. Evans, M. Rath, and R. F. Veerkamp. 2003b. Genetic relationships among body condition score, body weight, milk yield, and fertility in dairy cows. J Dairy Sci 86(6):2193-2204.

Bewley, J., R. W. Palmer, and D. B. Jackson-Smith. 2001a. Modeling milk production and labor efficiency in modernized Wisconsin dairy herds. J Dairy Sci 84(3):705-716.

Bewley, J., R. W. Palmer, and D. B. Jackson-Smith. 2001b. An overview of experiences of Wisconsin dairy farmers who modernized their operations. J Dairy Sci 84(3):717-729.

Butler, W. R. 1998. Review: effect of protein nutrition on ovarian and uterine physiology in dairy cattle. J Dairy Sci 81(9):2533-2539.

Butler, W. R. 2003. Energy balance relationships with follicular development, ovulation and fertility in postpartum dairy cows. Livestock Production Science 83(2-3):211-218.

Butler, W. R. and R. D. Smith. 1989. Interrelationships between energy balance and postpartum reproductive function in dairy cattle. Journal Dairy Science 72(3):767-783.

Cabrera, V. E., D. Solis, and J. del Corral. 2010. Determinants of technical efficiency among dairy farms in Wisconsin. Journal of Dairy Science 93(1):387-393.

Calus, M. P., J. J. Windig, and R. F. Veerkamp. 2005. Associations among descriptors of herd management and phenotypic and genetic levels of health and fertility. J Dairy Sci 88(6):2178-2189.

Capper, J. L., R. A. Cady, and D. E. Bauman. 2009. The environmental impact of dairy production: 1944 compared with 2007. Journal of Animal Science 87(6):2160-2167.

Caraviello, D. Z., K. A. Weigel, P. M. Fricke, M. C. Wiltbank, M. J. Florent, N. B. Cook, K. V. Nordlund, N. R. Zwald, and C. L. Rawson. 2006. Survey of management practices on reproductive performance of dairy cattle on large US commercial farms. J. Dairy Sci. 89(12):4723-4735.

Castillo-Juarez, H., P. A. Oltenacu, R. W. Blake, C. E. Mcculloch, and E. G. Cienfuegos-Rivas. 2000. Effect of herd environment on the genetic and phenotypic relationships among milk yield, conception rate, and somatic cell score in Holstein cattle. Journal of Dairy Science 83(4):807-814.

Emanuelson, U. and P. A. Oltenacu. 1998. Incidences and effects of diseases on the performance of Swedish dairy herds stratified by production. Journal Dairy Science 81(9):2376-2382.

Fahey, J., K. O'Sullivan, J. Crilly, and J. F. Mee. 2002. The effect of feeding and management practices on calving rate in dairy herds. Anim Reprod Sci 74(3-4):133-150.

Falconer, D. S. 1981. Introduction to Quantitative Genetics. 2nd ed. Longman, Essex, UK.

Fulwider, W. K., T. Grandin, B. E. Rollin, T. E. Engle, N. L. Dalsted, and W. D. Lamm. 2008. Survey of dairy management practices on one hundred thirteen north central and northeastern united states dairies. Journal of Dairy Science 91(4):1686-1692.

Garcia-Ispierto, I., F. Lopez-Gatius, P. Santolaria, J. L. Yaniz, C. Nogareda, and M. Lopez-Bejar. 2007. Factors affecting the fertility of high producing dairy herds in northeastern Spain. Theriogenology 67(3):632-638.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter. 1996. Introducing Markov chain Monte Carlo. Pages 1-19 in Markov Chain Monte Carlo in Practice. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, ed. Chapman and Hall, London, U.K.

Hansen, L. B. 2000. Consequences of selection for milk yield from a geneticist's viewpoint. J Dairy Sci 83(5):1145-1150.

Hansen, L. B., A. E. Freeman, and P. J. Berger. 1983. Yield and fertility relationships in dairy cattle. J Dairy Sci 66(2):293-305.

Jordan, E. R. and R. H. Fourdraine. 1993. Characterization of the management practices of the top milk producing herds in the country. J. Dairy Sci. 76(10):3247-3256.

Laben, R. L., R. Shanks, P. J. Berger, and A. E. Freeman. 1982. Factors Affecting Milk-Yield and Reproductive-Performance. Journal of Dairy Science 65(6):1004-1015.

Leblanc, S. 2010. Assessing the Association of the Level of Milk Production with Reproductive Performance in Dairy Cattle. Journal of Reproduction and Development 56:S1-S7.

Lof, E., H. Gustafsson, and U. Emanuelson. 2007. Associations between herd characteristics and reproductive efficiency in dairy herds. Journal Dairy Science 90(10):4897-4907.

Lopez-Gatius, F. 2003. Is fertility declining in dairy cattle? A retrospective study in northeastern Spain. Theriogenology 60(1):89-99.

Lopez-Gatius, F., I. Garcia-Ispierto, P. Santolaria, J. Yaniz, C. Nogareda, and M. Lopez-Bejar. 2006. Screening for high fertility in high-producing dairy cows. Theriogenology 65(8):1678-1689.

Lopez-Gatius, F., P. Santolaria, and S. Almeria. 2005a. Neospora caninum infection does not affect the fertility of dairy cows in herds with high incidence of Neospora-associated abortions. J Vet Med B Infect Dis Vet Public Health 52(1):51-53.

Lopez-Gatius, F., P. Santolaria, I. Mundet, and J. L. Yaniz. 2005b. Walking activity at estrus and subsequent fertility in dairy cows. Theriogenology 63(5):1419-1429.

Lucy, M. C. 2001. Reproductive loss in high-producing dairy cattle: where will it end? Journal of Dairy Science 84(6):1277-1293.

Lucy, M. C. 2003. Mechanisms linking nutrition and reproduction in postpartum cows. Reprod Suppl 61:415-427.

Lucy, M. C., C. R. Staples, W. W. Thatcher, P. S. Erickson, R. M. Cleale, J. L. Firkins, J. H. Clark, M. R. Murphy, and B. O. Brodie. 1992. Influence of Diet Composition, Dry-Matter Intake, Milk-Production and Energy-Balance on Time of Postpartum Ovulation and Fertility in Dairy-Cows. Animal Production 54:323-331.

Macmillan, K. L., I. J. Lean, and C. T. Westwood. 1996. The effects of lactation on the fertility of dairy cows. Aust Vet J 73(4):141-147.

Mardia, K. V., J. T. Kent, and J. M. Bibby. 1979. Multivariate analysis. Probability and Mathematical Statistics. Academic Press, Inc., San Diego, CA.

Miller, R. H., H. D. Norman, M. T. Kuhn, J. S. Clay, and J. L. Hutchison. 2007. Voluntary waiting period and adoption of synchronized breeding in dairy herd improvement herds. Journal Dairy Science 90(3):1594-1606.

Nebel, R. L. and M. L. McGilliard. 1993. Interactions of high milk yield and reproductive performance in dairy cows. J. Dairy Sci. 76(10):3257-3268.

Norman, H. D., J. R. Wright, S. M. Hubbard, R. H. Miller, and J. L. Hutchison. 2009. Reproductive status of Holstein and Jersey cows in the United States. Journal of Dairy Science 92(7):3517-3528.

O'Hagan, A. 2004. Bayesian statistics: principles and benefits. in Bayesian Statistics and Quality Modelling in the Agro-Food Production Chain. Vol. 3. M. A. J. S. van Boekel, A. Stein, and A. H. C. van Bruggen, ed. Wageningen University and Research Center, Germany.

Peters, M. W. and J. R. Pursley. 2002. Fertility of lactating dairy cows treated with Ovsynch after presynchronization injections of PGF2 alpha and GnRH. Journal of Dairy Science 85(9):2403-2406.

Pourahmadi, M. 1999. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. Biometrika 86(3):677-690.

Pourahmadi, M. 2007. Cholesky decompositions and estimation of a covariance matrix: orthogonality of variance-correlation parameters. Biometrika 94(4):1006-1013.

Pryce, J. E., M. D. Royal, P. C. Garnsworthy, and I. L. Mao. 2004. Fertility in the high-producing dairy cow. Livestock Production Science 86(1-3):125-135.

Riley, R. D., K. R. Abrams, P. C. Lambert, A. J. Sutton, and J. R. Thompson. 2007. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. Statistics in Medicine 26(1):78-97.

Roche, J. F., D. Mackey, and M. D. Diskin. 2000. Reproductive management of postpartum cows. Anim Reprod Sci 60-61:703-712.

Royal, M. D., A. O. Darwash, A. P. E. Flint, R. Webb, J. A. Woolliams, and G. E. Lamming. 2000. Declining fertility in dairy cattle: changes in traditional and endocrine parameters of fertility. Animal Science 70:487-501.

Schefers, J. M., K. A. Weigel, C. L. Rawson, N. R. Zwald, and N. B. Cook. 2010. Management practices associated with conception rate and service rate of lactating Holstein cows in large, commercial dairy herds. Journal of Dairy Science 93(4):1459-1467.

Seykora, A. J. and B. T. Mcdaniel. 1983. Heritabilities and Correlations of Lactation Yields and Fertility for Holsteins. Journal of Dairy Science 66(7):1486-1493.

Shoemaker, J. S., I. S. Painter, and B. S. Weir. 1999. Bayesian statistics in genetics: a guide for the uninitiated. Trends Genet 15(9):354-358.

Silke, V., M. G. Diskin, D. A. Kenny, M. P. Boland, P. Dillon, J. F. Mee, and J. M. Sreenan. 2002. Extent, pattern and factors associated with late embryonic loss in dairy cows. Animal Reproduction Science 71(1-2):1-12.

Sorensen, D. and D. Gianola. 2002. Likelihood, Bayesian and MCMC Methods in Quantitative Genetics. Statistics for Biology and Health. Springer Verlag, New York.

Sorensen, P., M. S. Lund, B. Guldbrandtsen, J. Jensen, and D. Sorensen. 2003. A comparison of bivariate and univariate QTL mapping in livestock populations. Genetics Selection Evolution 35(6):605-622.

Spalding, R. W., R. W. Everett, and R. H. Foote. 1974. Fertility in New York artificially inseminated Holstein herds in dairy herd improvement. J Dairy Sci 58:718-723.

Tempelman, R. J. 2010. Addressing scope of inference for global genetic evaluation of livestock. in Proceedings of the 47th Meeting of the Brazilian Animal Science Society. Salvador, Bahia, Brazil.

Tsuruta, S., I. Misztal, C. Huang, and T. J. Lawlor. 2009. Bivariate analysis of conception rates and test-day milk yields in Holsteins using a threshold-linear model with random regressions. Journal Dairy Science 92(6):2922-2930.

Washburn, S. P., W. J. Silvia, C. H. Brown, B. T. McDaniel, and A. J. McAllister. 2002. Trends in reproductive performance in Southeastern Holstein and Jersey DHI herds. Journal of Dairy Science 85(1):244-251.

Wiltbank, M., H. Lopez, R. Sartori, S. Sangsritavong, and A. Gumen. 2006. Changes in reproductive physiology of lactating dairy cows due to elevated steroid metabolism. Theriogenology 65(1):17-29.

Windig, J. J., M. P. Calus, B. Beerda, and R. F. Veerkamp. 2006. Genetic correlations between milk production and health and fertility depending on herd environment. Journal Dairy Science 89(5):1765-1775.

Windig, J. J., M. P. Calus, and R. F. Veerkamp. 2005. Influence of herd environment on health and fertility and their relationship with milk production. Journal Dairy Science 88(1):335-347.

CHAPTER 1

Hierarchical Bayesian Modeling of Random and Residual Variance-Covariance Matrices in Bivariate Mixed Effects Models

This is the pre-peer reviewed version of the following article: Bello N. M., J. P. Steibel and R. J. Tempelman. "Hierarchical Bayesian modeling of random and residual variance-covariance matrices in bivariate mixed effects models". Biometrical Journal 2010 June; 52(3):297-313.

ABSTRACT

Bivariate mixed effects models are often used to jointly infer upon covariance matrices for both random effects (**u**) and residuals (**e**) between two different phenotypes in order to investigate the architecture of their relationship. However, these (co)variances themselves may additionally depend upon covariates as well as additional sets of exchangeable random effects that facilitate borrowing of strength across a large number of clusters. We propose a hierarchical Bayesian extension of the classical bivariate mixed effects model by embedding additional levels of mixed effects modeling of reparameterizations of **u**-level and **e**-level (co)variances between two traits. These parameters are based upon a recently popularized square-root free Cholesky decomposition and are readily interpretable, each conveniently facilitating a generalized linear model characterization. Using MCMC methods, we validate our model based on a simulation study and apply it to a joint analysis of milk yield and calving interval phenotypes in Michigan dairy cows. This analysis indicates that the

e-level relationship between the two traits is highly heterogeneous across herds and depends upon systematic herd management factors.

1. Introduction

Multivariate mixed effects models have been routinely used to investigate the architecture of relationships between two or more traits at several different levels, specifically (co)variance matrices for different sets of random (**u**) effects and residual (**e**) effects. In animal breeding, for example, co(variance) matrices for random genetic effects are specified in addition to that for residual effects to investigate how the phenotypic relationships between corresponding traits can be partitioned into random family or cluster effects (**u**) and residual effects (**e**) (Thompson and Meyer, 1986).

We are specifically interested in the joint analysis of milk production and reproductive efficiency of dairy cows. These two classes of phenotypes help define the necessary foundation for a successful dairy farm. Although antagonistic correlations (e.g., higher milk production leading to poorer fertility) have been generally reported, there are enough discrepancies across studies to suggest the need for modeling (co)variances as functions of covariates that characterize dairy management effects or herd environments (Laben et al., 1982; Lopez-Gatius et al., 2006; Lucy, 2001; Washburn et al., 2002). If these associations are significant, it may be possible to identify and recommend management strategies that mitigate the antagonistic relationship between the two traits. We consider the relationship between two representative traits using **u**-level (co)variances between clusters, e.g.,

herds, and **e**-level (co)variances between measurement units, e.g., cows within herds, hypothesizing that **u**-level and **e**-level (co)variance matrices are heterogeneous and may depend upon different systematic factors.

Modeling of both fixed and random effects influencing heterogeneity on residual variances has been previously developed in univariate models (Kizilkaya and Tempelman, 2005; Mulder, Bijma and Hill, 2007; Ros et al., 2004). However, work on explicit structural modeling of covariance matrices as functions of covariates has been limited because of necessary positive semi-definite constraints. To facilitate this issue at the **e**-level, Pourahmadi (1999) proposed a square root free Cholesky transformation of the (co)variance matrix for time ordered responses (i.e., longitudinal data) such that (co)variances are reparameterized as generalized autoregressive parameters (GARP) and innovation variances. We observe that the same transformation could also be applied to **u**-level (co)variances and that the resulting parameters can be readily specified as linked functions of linear models. Hence, multifactorial sources of heterogeneity on co(variances) can be modeled at both the **u**-level and **e**-level, recognizing that (co)variance matrices between observed phenotypes (i.e., at the **y**-level) on two or more traits could be separately affected by each of the two components. We also propose that the **e**-level GARP and innovation variances be modeled not only as functions of systematic (i.e., fixed) effects, but also of exchangeable cluster-specific random effects that can be characterized by a distribution, similar to those specified for fixed and random effects in classical mixed effects models. In a Bayesian model, all unknown parameters are considered to be random effects. Nevertheless, from a Bayesian

perspective, a fixed effects factor might be defined as one where each of its effects is specified with independent non-informative or vaguely informative prior distributions (Robinson, 1991; Sorensen and Gianola, 2002). A mixed effects model representation at deeper levels of a hierarchical Bayesian model is even more critical as it should facilitate efficient shrinkage estimation for cluster effects, each cluster, e.g. herd, characterized by many levels, each with a relatively limited number of measurement units or subjects, e.g., cows.

We believe our proposed model could be considered for a number of other applications in which the joint evaluation of multiple outcomes of interest is currently restricted by the assumption of constant correlations between traits across environments. For example, multivariate biomedical meta-analyses as well as multi-center medical studies typically involve the joint analysis of two or more types of outcomes, with medical centers then defining the clusters and patients defining the subjects (Riley et al., 2007); heterogeneous co(variances) across centers could then lead to suboptimal inferences on treatment effects under the assumption of constant co(variance) matrices. In the context of plant breeding and variety testing, flexible modeling of heterogeneous variance-covariance structures of yield responses in multi-environment trials of genetically related strains could be used to elucidate factors involved in genotype-by-environment interactions (Crossa et al., 2006; Piepho et al., 2008). In fact, a more general multivariate approach to genetic mapping of quantitative trait loci is likely to advance our understanding of complex phenotypic traits in plants, animals and humans (Mauricio, 2001; Sorensen et al., 2003). Studies where randomization occurs at the level of the cluster (Eldridge et al.,

2004), e.g., geographical areas, communities, schools, or worksites, rather than at the level of the individual within cluster, may also benefit from multifactorial modeling of (co)variances. For pedagogical purposes and for the intent of our dairy application of interest, our presentation is based on the modeling of two different responses or traits; however, our work could certainly be generalized to multivariate time-ordered data of any dimension as originally considered in Pourahmadi and Daniels (2002).

The objectives of our study are 1) to develop a hierarchical Bayesian extension to classical bivariate mixed effects modeling of residual (**e**) and random (**u**) covariance matrices for the joint analysis of two phenotypes, 2) to further validate the properties of our method implemented using Markov Chain Monte Carlo (MCMC) based on a simulation study, and 3) to apply our method to a joint analysis of milk production and reproduction of first-lactation dairy cows in Michigan. Whenever possible, we strive to choose prior density specifications that are conditionally conjugate (Gelman, 2006) in order to expedite Gibbs sampling steps in our MCMC algorithm (Gelfand and Smith, 1990).

2. Methods

2.1. Hierarchical Bayesian model construction

We start with the conventional bivariate linear mixed model

$$y_{ij} = \mathbf{x}_{ij}^{(1)}{}'\boldsymbol{\beta}_i + \mathbf{z}_j{}'\mathbf{u}_i + e_{ij} \tag{1}$$

where y_{ij} is the observation for trait i (i=1, 2) on subject j (j=1,...,n), $\boldsymbol{\beta}_i$ is a $p_i^{(1)}$ x 1 vector of unknown fixed location parameters for factors (e.g., parity, year, calving

season, etc.) unique to trait i; $\mathbf{u}_i$ is a q x 1 vector of unknown classical random effects (e.g., herd or contemporary group, etc.) unique to trait i and e_{ij} is the corresponding residual. Also, $\mathbf{x}_{ij}^{(1)}$ and $\mathbf{z}_j'$ are known incidence row vectors for subject j. For pedagogical reasons, we assume the same single random effects factor of clusters, e.g. herds, is common to both traits and for all subsequent random effects modeling presented thereafter (i.e. $\mathbf{z}_j'$ is the same for both traits and all levels of the hierarchical data structure modeled). Independent bivariate normal densities are assumed for each subject-specific pair of residuals $\mathbf{e}_{.j} = \left[e_{1,j} \;\; e_{2,j}\right]'$ on the two traits with $\mathrm{E}\left(\mathbf{e}_{.j}\right) = \mathbf{0}$ and $\mathrm{var}\left(\mathbf{e}_{.j}\right) = \mathbf{R}_j$ where

$$\mathbf{R}_j = \begin{bmatrix} \sigma_{e_1,j}^2 & \sigma_{e_{12},j} \\ \sigma_{e_{12},j} & \sigma_{e_2,j}^2 \end{bmatrix}. \tag{2}$$

From a Bayesian perspective, the elements of $\boldsymbol{\beta}_i$ are typically considered to be classical fixed effects (Sorensen and Gianola, 2002); that is, parameters whose elements would not be considered to be exchangeable random variables. Typically, we might specify subjective prior densities on fixed effects such as, for example, $\boldsymbol{\beta}_i \,|\, \boldsymbol{\beta}_i^{\mathbf{0}}, \mathbf{V}_i^{(\boldsymbol{\beta})} \sim N\left(\boldsymbol{\beta}_i^{\mathbf{0}}, \mathbf{V}_i^{(\boldsymbol{\beta})}\right)$, with hyperparameters $\boldsymbol{\beta}_i^{\mathbf{0}}$ and $\mathbf{V}_i^{(\boldsymbol{\beta})}$ being specified as known for i=1,2. Bounded uniform priors are also commonly considered (Sorensen and Gianola, 2002) as, typically, enough data is available to infer upon

elements of $\boldsymbol{\beta}_i$ with any reasonable noninformative prior distribution in large field studies (Gelman, 2006).

Denote $\mathbf{u}_{.k} = \left[u_{1,k} \quad u_{2,k}\right]'$ where $u_{i,k}$ denotes element k of $\mathbf{u}_i$ and is the random effect of cluster k ($1 \leq k \leq q$) for trait i. We specify independent structural bivariate normal prior densities on each $\mathbf{u}_{.k}$ with $E(\mathbf{u}_{.k}) = 0$ and $\text{var}(\mathbf{u}_{.k}) = \mathbf{G}_k$ such that:

$$\mathbf{G}_k = \begin{bmatrix} \sigma^2_{u_1,k} & \sigma_{u_{12},k} \\ \sigma_{u_{12},k} & \sigma^2_{u_2,k} \end{bmatrix}. \tag{3}$$

We reparameterize the variance-covariance matrices by implementing a square-root-free Cholesky decomposition to each $\mathbf{R}_j$ and $\mathbf{G}_k$ (co)variance matrix. Hence, we rewrite $\mathbf{R}_j$ in Equation (2) as:

$$\mathbf{R}_j = \begin{bmatrix} \sigma^2_{e_1,j} & \varphi^{(e)}_j \sigma^2_{e_1,j} \\ \varphi^{(e)}_j \sigma^2_{e_1,j} & \sigma^2_{e_{2|1},j} + \left(\varphi^{(e)}_j\right)^2 \sigma^2_{e_1,j} \end{bmatrix}. \tag{4}$$

Here $\varphi^{(e)}_j$ represents the subject-specific **e**-level regression coefficient of $e_{2,j}$ on $e_{1,j}$; that is, $e_{2,j} = \varphi^{(e)}_j e_{1,j} + e_{2|1,j}$ where $e_{2|1,j} \sim N\left(0, \sigma^2_{e_{2|1,j}}\right)$ is conditionally independent of $e_{1,j}$. Similarly, we rewrite $\mathbf{G}_k$ in Equation (3) as:

$$\mathbf{G}_k = \begin{bmatrix} \sigma^2_{u_1,k} & \varphi_k^{(u)}\sigma^2_{u_1,k} \\ \varphi_k^{(u)}\sigma^2_{u_1,k} & \sigma^2_{u_{2|1},k} + \left(\varphi_k^{(u)}\right)^2 \sigma^2_{u_1,k} \end{bmatrix},$$

(5)

where $\varphi_k^{(u)}$ represents the cluster-specific **u**-level regression coefficient of $u_{2,k}$ on $u_{1,k}$; that is, $u_{2,k} = \varphi_k^{(u)} u_{1,k} + u_{2|1,k}$ where $u_{2|1,k} \sim N\left(0, \sigma^2_{u_{2|1,k}}\right)$ is conditionally independent of $u_{1,k}$.

Using the conventions established by Pourahmadi (1999) and Daniels and Pourahmadi (2002), $\sigma^2_{u_{2|1,k}}$ and $\sigma^2_{e_{2|1,j}}$ might be referred to as the random effect and residual *innovation* variances on trait i = 2 specific to cluster k and subject j, respectively. Alternatively, we prefer the term *conditional* variances due to the between-trait conditional independence of residuals and random effects that is implied by the Cholesky decomposition in Equations (4) and (5). With these reparameterizations, Equation (1) does not change for trait i = 1 since it is specified as the first trait, and hence its random or residual effects are not conditioned upon those of any other trait. However, for trait i = 2, Equation (1) would be rewritten as:

$$y_{2j} = \mathbf{x}_{2,j}^{(1)}{}'\boldsymbol{\beta}_2 + \mathbf{z}_j{}'\left(\boldsymbol{\Psi}^{(u)}\mathbf{u}_1 + \mathbf{u}_{2|1}\right) + \varphi_j^{(e)} e_{1,j} + e_{2|1,j}.$$

where $\mathbf{u}_{2|1} = \left\{u_{2|1,k}\right\}_{k=1}^{q}$ is a q x 1 vector of random effects on trait 2 conditional on trait 1 and $\boldsymbol{\Psi}^{(u)}$ is a diagonal matrix with diagonal elements

$\boldsymbol{\varphi}^{(u)} = \left[\varphi_1^{(u)} \;\; \varphi_2^{(u)} \;\; \ldots \;\; \varphi_q^{(u)}\right]'$. It should then be apparent that $\varphi_k^{(u)} = \sigma_{u_{12},k} / \sigma^2_{u_1,k}$ and $\varphi_j^{(e)} = \sigma_{e_{12},j} / \sigma^2_{e_1,j}$. That is, $\varphi_k^{(u)}$ can be interpreted as the conditional change in $u_{2,k(j)}$, and hence in $y_{2,j}$, for every unit change in $u_{1,k(j)}$ where $k(j)$ defines the cluster k associated with subject j. Similarly, $\varphi_j^{(e)}$ can be interpreted as the conditional change in $e_{2,j}$, and hence in $y_{2,j}$, for every unit change in $e_{1,j}$. Hence, we refer to parameters $\varphi_k^{(u)}$ and $\varphi_j^{(e)}$ as the **u**-level and **e**-level *regression coefficients*, respectively, for our two trait application, rather than as GARP by Pourahmadi (1999). Note that $\mathbf{R}_j$ and $\mathbf{G}_k$ are guaranteed to be positive definite for any respective values of $\varphi_j^{(e)}$ and $\varphi_k^{(u)}$ (Pourahmadi, 1999), thereby facilitating their specification as a linear function of covariates and/or random effects.

We subsequently describe the generalized linear modeling of heterogeneous variances and covariances. We first specify a linear mixed effects model on each subject-specific $\varphi_j^{(e)}$:

$$\varphi_j^{(e)} = \mathbf{x}_j^{(2)\prime} \boldsymbol{\gamma}_e + \mathbf{z}_j' \mathbf{m}. \tag{6}$$

Here, $\boldsymbol{\gamma}_e$ represents a $p^{(2)}$ x 1 vector of unknown fixed effects whereas **m** represents a q x 1 vector of unknown cluster-specific random effects as before but

such that $\mathbf{m} \sim N\left(\mathbf{0}, \mathbf{I}\sigma_m^2\right)$. Furthermore, $\mathbf{x}_j^{(2)\prime}$ is a known row incidence vector. Note that the effects considered in $\boldsymbol{\gamma}_e$ do not necessarily need to mirror those considered for location parameters $\boldsymbol{\beta}_i$; that is, it is not necessary that $\mathbf{x}_j^{(2)\prime} = \mathbf{x}_{ij}^{(1)\prime}$ for either $i = 1$ or $i = 2$.

We similarly specify a linear model on each cluster-specific $\varphi_k^{(u)}$:

$$\varphi_k^{(u)} = \mathbf{x}_k^{(3)\prime}\boldsymbol{\gamma}_u \tag{7}$$

where $\boldsymbol{\gamma}_u$ represents a $p^{(3)}$ x 1 vector of unknown fixed effects with $\mathbf{x}_k^{(3)\prime}$ being the associated known row incidence vector.

We also model the conditional residual variances $\sigma_{e_{1,j}}^2$ and $\sigma_{e_{2|1,j}}^2$ as multiplicative functions of fixed and random effects (Cardoso, Rosa and Tempelman, 2005; Kizilkaya and Tempelman, 2005), expressing the log-linked relationships as follows:

$$\log\left(\boldsymbol{\sigma}_{e_i}^2\right) = \mathbf{x}_{ij}^{(4)\prime}\log\left(\boldsymbol{\tau}_{e_i}\right) + \mathbf{z}_j'\log\left(\mathbf{v}_{e_i}\right);\ i = 1, 2|1. \tag{8}$$

Here $\boldsymbol{\sigma}_{e_i}^2 = \left\{\sigma_{e_{i,j}}^2\right\}_{j=1}^{n}$ represents the n x 1 vector of subject-specific conditional residual variances with $i = 2|1$ referring to the corresponding parameter for trait 2 conditioned upon that for trait 1. Also, $\boldsymbol{\tau}_{e_i}$ represents a $p_i^{(4)}$ x 1 vector of

unknown fixed effects whereas $\mathbf{v}_{e_i} = \left\{ v_{e_i,k} \right\}_{k=1}^{q}$ represents a q x 1 vector of unknown cluster-specific random effects. Furthermore, $\mathbf{x}_{ij}^{(4)\prime}$ is a known incidence row vector. To obtain conditional conjugacy (Gelman, 2006), we adopt independent inverted gamma (*IG*) prior densities for each trait-specific set of random effects on the conditional residual variances:

$$v_{e_i,k} \mid \eta_{e_i} \sim IG\left(\eta_{e_i}, \eta_{e_i} - 1\right); \; i = 1, 2|1,$$

such that $\mathrm{E}\left(v_{e_i,k} \mid \eta_{e_i}\right) = 1$ and $\mathrm{var}\left(v_{e_i,k} \mid \eta_{e_i}\right) = \left(\eta_{e_i} - 2\right)^{-1}$ for k = 1,...,q (Cardoso et al., 2005; Kizilkaya and Tempelman, 2005). Note that the structural prior on each of $\mathbf{v}_{e_1}$ and $\mathbf{v}_{e_2}$, as well as that previously specified on $\mathbf{m}$, allows for borrowing of information across levels or clusters for each random effects factor in a manner similar to what the Gaussian prior density does for the vectors of classical random effects $\mathbf{u}_1$ and $\mathbf{u}_2$ (Robinson, 1991).

We also introduce heterogeneity in the $\mathbf{u}$-level variances by modeling the conditional variances of the random effects as multiplicative functions of fixed effects, such that the logarithmic expression of this relationship is also linear:

$$\log\left(\boldsymbol{\sigma}_{u_i}^2\right) = \mathbf{x}_{ik}^{(5)\prime} \log\left(\boldsymbol{\tau}_{u_i}\right); \; i = 1, 2|1. \tag{9}$$

Here, $\boldsymbol{\sigma}^2_{\mathbf{u}_i} = \left\{ \sigma^2_{u_{i,k}} \right\}_{k=1}^{q}$ represents the vector of cluster-specific conditional random effects variances whereas $\boldsymbol{\tau}_{u_i}$ represents a $p_i^{(5)}$ x 1 vector of unknown fixed effects. Furthermore, $\mathbf{x}_{ik}^{(5)\prime}$ is a known incidence row vector.

For all remaining prior density specifications, we treat all hyperparameters as known, again striving to choose priors that are conditionally conjugate to facilitate Gibbs sampling steps. First we adopt subjectively-specified Gaussian prior densities on the fixed effects influencing heterogeneity of the **e**-level and **u**-level regression coefficients, i.e., $\boldsymbol{\gamma}_e \sim N\left(\boldsymbol{\mu}_{\boldsymbol{\gamma}}^{(e)}, \mathbf{V}_{\boldsymbol{\gamma}}^{(e)}\right)$, $\boldsymbol{\gamma}_u \sim N\left(\boldsymbol{\mu}_{\boldsymbol{\gamma}}^{(u)}, \mathbf{V}_{\boldsymbol{\gamma}}^{(u)}\right)$, although again bounded uniform priors might be specified as well. We further specify an $IG(\alpha_m, \beta_m)$ prior on σ_m^2. Independent inverted-gamma priors are also placed on elements of $\boldsymbol{\tau}_{e_i} = \left\{ \tau_{e_{il}} \right\}_{l=1}^{p_i^{(4)}}$ and $\boldsymbol{\tau}_{u_i} = \left\{ \tau_{u_{il}} \right\}_{l=1}^{p_i^{(5)}}$, specifically, $\tau_{e_{il}} \sim IG\left(\alpha_i^{(e)}, \beta_i^{(e)}\right)$, l=1,2,..., $p_i^{(4)}$, and $\tau_{u_{il}} \sim IG\left(\alpha_i^{(u)}, \beta_i^{(u)}\right)$, l=1,2,..., $p_i^{(5)}$, for i = 1, 2|1, as these priors are conditionally conjugate when the elements of the corresponding row incidence vectors pertain to the intercept or are dummy variables for classification factors (Kizilkaya and Tempelman, 2005). Again, we characterize elements of $\boldsymbol{\gamma}_e$, $\boldsymbol{\gamma}_u$, $\boldsymbol{\tau}_{e_1}$, $\boldsymbol{\tau}_{e_{2|1}}$, $\boldsymbol{\tau}_{u_1}$ and $\boldsymbol{\tau}_{u_{2|1}}$ as being

fixed effects since we don't consider elements within those vectors as being exchangeable (Sorensen and Gianola, 2002).

We assign a vaguely informative, though proper, prior density (Cardoso et al., 2005; Kizilkaya and Tempelman, 2005) to the hyperparameters characterizing the distribution of the random effects for the **e**-level heteroskedasticity, as follows:

$$\eta_{e_i} \sim p\left(\eta_{e_i}\right) \propto \left(1+\eta_{e_i}\right)^{-2}; \text{ for } \eta_{e_i} > 0 \text{ and } i = 1, 2|1. \tag{10}$$

As shown previously by Albert (1988), this prior defines a uniform prior density U(0,1) on the transformed variable $\varsigma = g\left(\eta_{e_i}\right) = \left(1+\eta_{e_i}\right)^{-1}$. Then, by change of variables, $f_{\eta_{e_i}} = f_{\varsigma}\left(g^{-1}\left(\eta_{e_i}\right)\right)\left|\frac{\partial}{\partial \eta_{e_i}} g^{-1}\left(\eta_{e_i}\right)\right| = \left(1+\eta_{e_i}\right)^{-2}$ where f denotes the probability density function.

2.2. Inference

Our inference for the proposed hierarchical Bayesian model is based on MCMC. The joint posterior distribution of all unknowns as well as the full conditional densities (FCD) for these unknowns, as necessary for implementing MCMC, are presented in the Appendix. Regular identifiability constraints are required on all fixed effects parameters, namely $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, γ_e, γ_u, $\boldsymbol{\tau}_{e_1}$, $\boldsymbol{\tau}_{e_{2|1}}$, $\boldsymbol{\tau}_{u_1}$ and $\boldsymbol{\tau}_{u_{2|1}}$ in order to remove hypersensitivity to their respective prior specifications (Gelfand and Sahu, 1999) for example, only t-1 indicator or dummy variables is required for a classification factor with t levels (Kutner et al., 2005). We thereby adapt the corner

parameterization (Clayton, 1996; Kizilkaya and Tempelman, 2005), also known as the set-to-zero restriction (Milliken and Johnson, 2009), whereby an overall intercept is always specified and the effect corresponding to one arbitrarily chosen level of each fixed effects factor is "zeroed out" or removed. This parameterization is also popularized in SAS linear models software (Littell, Freund and Spector, 1991). Certainly, an alternative full rank parameterization, such as the sum-to-zero restriction (e.g., Kaufman and Sain, 2010), could have been considered. Although Bayesian inference for these two alternative parameterizations would not be strictly invariant to the same non-informative prior distributions, the information provided by the data to the posterior densities of same estimable linear combinations of the fixed effects under either parameterization (and many possible others) would be identical (Gelfand and Sahu, 1999).

3. Simulation Study

We validate our proposed model using a simulation study for which our focus was on inference on γ_e, γ_u and σ_m^2. Two correlated response variables were simulated to mimic milk yield and calving interval for approximately 50,000 subjects (e.g., cows) distributed across 200 clusters (e.g., herds)`

within each replicated dataset. The number of subjects (or cows) per cluster (or herd) was drawn from a discretized gamma distribution based on the mean and variance of cluster sizes obtained from an actual dataset to be described later. We considered three different broad scenarios or correlation architectures between traits that might be plausible for a number of disparate applications. These 3 scenarios

differed in terms of general sign of the **e**-level and **u**-level regression coefficients, namely: A) *same sign*: positive **u**-level and **e**-level coefficients; B) *opposite sign*: negative **u**-level and positive **e**-level coefficients; C) *zero correlation*: zero **u**-level and **e**-level coefficients. We also considered 4 different values for the variance component σ_m^2: I) $\sigma_m^2 = 0$; II) $\sigma_m^2 = 0.1$; III) $\sigma_m^2 = 1$; and IV) $\sigma_m^2 = 10$. Ten replicate datasets were simulated for each of the 12 possible populations as defined by the factorial of 3 different correlation architectures with 4 different values of σ_m^2. The same two levels of a single fixed effects factor were considered, where applicable, for all location parameters, conditional residual and random effects variance components, and **e**-level and **u**-level regression coefficients. In other words, the corresponding incidence row vectors for all fixed effects terms were identical such that all covariates were cluster-specific (e.g., herd-specific); i.e. $\mathbf{x}_{1j}^{(1)\prime} = \mathbf{x}_{2j}^{(1)\prime} = \mathbf{x}_{j}^{(2)\prime} = \mathbf{x}_{k(j)}^{(3)\prime} = \mathbf{x}_{1j}^{(4)\prime} = \mathbf{x}_{2j}^{(4)\prime} = \mathbf{x}_{1,k(j)}^{(5)\prime} = \mathbf{x}_{2,k(j)}^{(5)\prime}$, with the first element set equal to 1 to specify an intercept and the second element being a Bernoulli (0,1) random draw with probability of 0.25 to partially mimic an unbalanced design structure as based on the corner parameterization. We used arbitrary 2 x 1 specifications for $\boldsymbol{\gamma}_e = \begin{bmatrix} \gamma_{e_1} & \gamma_{e_2} \end{bmatrix}'$ and $\boldsymbol{\gamma}_u = \begin{bmatrix} \gamma_{u_1} & \gamma_{u_2} \end{bmatrix}'$ from Equations (6) and (7) to create the intended correlation architectures such that $\boldsymbol{\gamma}_e = \boldsymbol{\gamma}_u = \mathbf{0}$ in scenario C; these specifications are provided in Table 1. We also set $\boldsymbol{\tau}_{e_1} = \begin{bmatrix} \tau_{e_1,1} & \tau_{e_1,2} \end{bmatrix}' = \begin{bmatrix} 176 & 220 \end{bmatrix}'$ and

$\boldsymbol{\tau}_{e_{2|1}} = \left[\tau_{e_{2|1},1} \;\; \tau_{e_{2|1},2}\right]' = \left[9{,}100 \;\; 13{,}000\right]'$ per Equation (8)

and $\boldsymbol{\tau}_{u_1} = \left[\tau_{u_1,1} \;\; \tau_{u_1,2}\right]' = \left[150 \;\; 100\right]'$ and

$\boldsymbol{\tau}_{u_{2|1}} = \left[\tau_{u_{2|1},1} \;\; \tau_{u_{2|1},2}\right]' = \left[900 \;\; 600\right]'$ per Equation (9) for all simulated datasets. Similarly, the same hyperparameter values, $\eta_{e_1} = 8$ and $\eta_{e_{2|1}} = 4$, were used for all datasets to specify the degree of heterogeneity in conditional residual variances across clusters for traits 1 and 2, respectively. In all cases, flat unbounded priors were specified on $\boldsymbol{\gamma}_e$, $\boldsymbol{\gamma}_u$ and σ_m^2, as well as for $\boldsymbol{\beta}_i$, i = 1, 2 and for $\boldsymbol{\tau}_{u_i}$ and $\boldsymbol{\tau}_{e_i}$, i = 1, 2|1. However, additional caution should be used for smaller datasets such that informative priors might be needed to ensure propriety of the joint posterior density.

For the analysis of each of the 120 simulated datasets, the length of the MCMC chain was 100,000 cycles after a burn-in period of 1,000 cycles. Convergence diagnostics for all relevant parameters (i.e. parameters with non-exchangeable priors) was monitored graphically, and also following Raftery and Lewis (1992). For all elements of $\boldsymbol{\gamma}_e$ and $\boldsymbol{\gamma}_u$, and for σ_m^2, we assessed frequentist properties based on the 95% highest posterior density interval (HPD).

The Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) is commonly used to infer upon evidence for fixed and random sources of heterogeneity on residual variances by comparing quality of fit between competing hierarchical models (Ibanez-Escriche et al., 2008; Ros et al., 2004). In this study, we

validated the DIC as a means to test for the importance of σ_m^2 on the e-level regression coefficient in the bivariate context. Two competing models were evaluated: a full model (M_1) that included cluster-specific e-level regressions (i.e., $\sigma_m^2 > 0$) and a null model (M_0) that did not (i.e., $\sigma_m^2 = 0$). The difference between the two corresponding DIC values, respectively DIC_1 and DIC_0, were used to draw conclusions on the importance of σ_m^2. Smaller values of DIC are indicative of improved model fit, such that positive values of (DIC_0 - DIC_1) would suggest M_1 to be the better fitting model and thus indicate evidence of non-zero σ_m^2. Generally, DIC differences exceeding 7 are believed to indicate a decisive difference in model fit (Spiegelhalter et al., 2002).

For all 90 replicated datasets in which $\sigma_m^2 > 0$, values of (DIC_0 - DIC_1) were all greater than +7, thereby always correctly selecting the full model. Moreover, as expected, the value of (DIC_0 - DIC_1) increased with greater values of σ_m^2 but showed no pattern between the different correlation architectures. Ranges of (DIC_0 - DIC_1) values were [11, 98] for $\sigma_m^2 = 0.1$; [522, 1378] for $\sigma_m^2 = 1.0$; and [4658, 14175] for $\sigma_m^2 = 10$. For 29 of the 30 replicated datasets where $\sigma_m^2 = 0$, the absolute values of (DIC_0 - DIC_1) were less than 7, with the range being [-3.9, 5.1]. The remaining dataset had a DIC difference of 9, thereby incorrectly choosing the full model, at least based on the rule of thumb provided by Spiegelhalter et al.

(2002). We then believe these results validate DIC and Spiegelhalter's rule as a reliable model choice criterion for a decision rule on σ_m^2.

3.1.Posterior inference on random regression parameters:

Table 1 presents the minimum and maximum values for each of the upper and lower boundaries of the 95% HPD of the posterior distribution for γ_{e_1}, γ_{e_2}, γ_{u_1}, γ_{u_2} and σ_m^2 across the 10 replicates for each of the 12 simulation populations considered. Coverage probabilities for the **e**- and **u**-regression parameters across the entire simulation study was near frequentist expectation as the replicate-specific 95% HPD included the true parameter value in 537 out of 570 cases (based on 120 replicated datasets times 4 fixed effects parameters, namely γ_{e_1}, γ_{e_2}, γ_{u_1} and γ_{u_2}; plus 90 cases on σ_m^2 for datasets involving non-zero σ_m^2). This result also partly validates that for reasonably sized datasets, such as those simulated in our study, unbounded flat priors on these same parameters may be relatively innocuous; however, proper priors should generally be considered to ensure propriety of the posterior density.

For each simulated population, posterior means (not shown) of γ_{e_1}, γ_{e_2}, σ_m^2, γ_{u_1} and γ_{u_2}, were evaluated for bias with respect to their true values using a one-sample non-parametric Wilcoxon Rank Sum Test and a one-sample *t*-test assuming normality. Based on a Type I error rate of 5% for each parameter, these tests did not support biased estimation of posterior means for any regression

parameters for any of the simulated populations (not shown). As expected, posterior means of γ_{u_1} and γ_{u_2}, were more variable and their 95% HPD were wider than for γ_{e_1} and γ_{e_2}, as there is typically greater uncertainty for inferences on dispersion parameters characterizing random effects as opposed to those for residuals. Furthermore, Table 1 illustrates that increasing values of σ_m^2 had a detrimental effect on the precision of inference on γ_{e_1} and γ_{e_2}. Nevertheless, the correlation architecture, as manifested by the three different combinations of values for γ_{e_1}, γ_{e_2}, γ_{u_1} and γ_{u_2} did not seem to influence the width of the 95% HPD for any of those parameters. Overall, we noticed no difference in inferential performance between scenarios A, B and C.

4. Application to Dairy Data

4.1. Data description

The two traits of interest were milk yield (kg. x 100) adjusted to 305 day lactation lengths and calving interval (days) defined as the interval from the first calving to second calving in primiparous dairy cows. Data on 49,789 first-lactation cow records from 578 Michigan dairy herds from 2005 to 2007 were provided by the National Dairy Herd Improvement Association (DHIA, Raleigh, NC). Random clusters were characterized by 1,408 herd-years or contemporary groups, being defined as the cluster of animals managed within the same herd and year. All subsequent random effects modeling for this example is based on this cluster definition. Complete data was not available for all herd-year clusters and, when

available, it was scrutinized to ensure a minimum cluster size of 25 lactation records per herd-year.

Some classical fixed effects (i.e., $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$) factors were considered for both traits, including the effects of 4 calving seasons (Winter: December to February; Spring: March to May; Summer: June to August; and Fall: September to November) and 3 years (2005, 2006, 2007). Additionally for $\boldsymbol{\beta}_1$ (i.e., milk production), we considered the fixed effects of 3 levels of bovine somatotropin (bST) supplementation: non-users (0% of the herd enrolled), intermediate users (>0-50% of the herd enrolled), and committed users (≥50% of the herd enrolled), as well as the fixed effects of 2 different levels of milking frequency (2 times per day or 2X, versus 3 or more times per day or 3^{+}X). Both of these factors are only recorded at the herd level and reflect potentially different herd management strategies.

We used an ad-hoc approach (Bello, Erskine and Tempelman, 2009) to select candidate sources of systematic heterogeneity to model on the **e**-level and **u**-level relationships (i.e., $\boldsymbol{\gamma}_e$, $\boldsymbol{\gamma}_u$) between milk production and reproductive performance although we emphasize that the chosen factors are not intended to represent a comprehensive list. We modeled $\varphi_j^{(e)}$ as a function of the fixed effects ($\boldsymbol{\gamma}_e$) of milking frequency in the herd whereas $\varphi_k^{(u)}$ was modeled as a function of the fixed effects ($\boldsymbol{\gamma}_u$) of bST supplementation. To be consistent with these specifications, the fixed effects specifications for the corresponding **e**-level and **u**-level conditional variances were mirrored accordingly. That is, $\boldsymbol{\tau}_{e_1}$ and $\boldsymbol{\tau}_{e_{2|1}}$ were specified by the

herd milking frequency factor such that $\mathbf{x}_j^{(2)\prime} = \mathbf{x}_{1j}^{(4)\prime} = \mathbf{x}_{2j}^{(4)\prime} = [1 \quad 1]$ or $[1 \quad 0]$ depending on whether cow j was milked twice or more than twice daily, respectively. Similarly, $\boldsymbol{\tau}_{u_1}$ and $\boldsymbol{\tau}_{u_{2|1}}$ were specified by the level of bST supplementation such that $\mathbf{x}_k^{(3)\prime} = \mathbf{x}_{1k}^{(5)\prime} = \mathbf{x}_{2k}^{(5)\prime} = [1 \ 1 \ 0]$, $[1 \ 0 \ 1]$, or $[1 \ 0 \ 0]$ for non-users, intermediate users and committed users, respectively. Furthermore, random cluster effects were also modeled for **e**-level conditional variances, $\mathbf{v}_{e_1}$ and $\mathbf{v}_{e_{2|1}}$ as per Equation (8) with independent inverted-gamma priors having hyperparameters η_{e_1} and $\eta_{e_{2|1}}$, respectively, and their own prior specifications as in Equation (10). Prior densities for all remaining parameters were specified as indicated previously for the simulation study. Also, as with the simulation study, two competing models were fitted to the data: a full model fitting herd-year as a random cluster-specific source of **e**-level heterogeneity (**m**) on $\varphi_j^{(e)}$ with $\mathbf{m} \sim N(\mathbf{0}, \mathbf{I}\,\sigma_m^2)$ and a reduced model ignoring this source of heterogeneity (i.e., $\sigma_m^2 = 0$). For each of the two competing models, we ran one long MCMC chain (100,000 saved cycles after 1,000 cycles of burn-in), using the same convergence diagnostics on all parameters with non-exchangeable priors, as described in the simulation study. For each parameter of interest, we summarize the posterior density using posterior means, posterior standard deviations, and 95% HPD. In addition, we report the effective sample size (ESS) as a measure of the number of effectively independent samples amongst the 100,000 dependent MCMC samples (Sorensen et al., 1995).

4.2. Results

As previously mentioned, model choice and thus evaluation of hierarchical heterogeneity of the **e**-level relationship was based on DIC. The DIC for the full model was 36.2 units less than that for the reduced model, implying that σ_m^2 or variation in cluster or herd-year effects on the **e**-level relationship between 305-d milk yield and calving interval among first parity cows is significant. Hence, we base all of our subsequent inference on a full model that includes a mixed model specification for each subject-specific (or cow-specific) $\varphi_j^{(e)}$.

Based on this full model, posterior means, posterior standard deviations, 95% HPD and ESS for MCMC inference on **e**-level ($\boldsymbol{\gamma}_e$ and σ_m^2) and **u**-level ($\boldsymbol{\gamma}_u$) hyperparameters are summarized in Table 2. The ESS indicated sufficient number of MCMC iterations, although mixing for σ_m^2 appeared to be substantially hampered relative to the other parameters. Such slower mixing was also observed in the simulation study and was not surprising given that σ_m^2 specifies the deepest, and thus least informative, level in the hierarchical model. It appears that, in general, the **e**-level relationship between 305-d milk yield and projected calving interval differed substantially in magnitude from the **u**-level relationship. The overall **e**-level relationship, based on the posterior mean of $\frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j^{(2)\prime}\boldsymbol{\gamma}_e$, was of 0.55 d longer projected calving interval per 100 kg increase in 305-d milk yield and appeared to be

significantly different from zero (95% HPD = [0.49, 0.62]). In contrast, the posterior density of $\frac{1}{q}\sum_{k=1}^{q}\mathbf{x}_k^{(3)\prime}\boldsymbol{\gamma}_u$ indicated the overall **u**-level relationship did not depart significantly from zero (95% HPD = [-0.11, 0.46]). Hence cows with higher milk yields tended to have poorer reproductive efficiency than cows with lower milk yields, but there was no strong evidence that higher producing herds had better or worse reproductive performance than lower producing herds.

At the **e**-level, the respective posterior means ± posterior standard deviation for $\gamma_{e,3^+X} = [1 \quad 0]\boldsymbol{\gamma}_e$ between the two traits for cows in 3^+X milking herds was 0.45±0.05 d/100kg compared to 0.66±0.04 d/100 kg for $\gamma_{e,2X} = [1 \quad 1]\boldsymbol{\gamma}_e$ pertaining to cows in 2X milking herds. A 95% HPD on their difference $(\gamma_{e,2X} - \gamma_{e,3^+X})$ was [0.08, 0.34], thereby indicating a more favorable relationship between 305-d milk yield and calving interval for cows with more frequent milking. However, at the **u**-level, the data did not support any evidence of bST usage influencing the relationship between the two traits, as the 95% HPD of all pairwise differences between the three levels overlapped with zero (results not shown). As also seen in the simulation study, uncertainty in inference was greater for parameters determining the between-trait correlation for random (**u**) effects than that for residual (**e**) effects, as illustrated by the differences in widths of the corresponding 95% HPD (Table 2).

The posterior inference summary on σ_m^2 is also reported in Table 2. Based on DIC results presented above, we concluded strong evidence for $\sigma_m^2 > 0$. Hence, the **e**-level relationship between milk yield and reproduction is heterogeneous across clusters or herd-years. Assuming that **m** is multivariate normal and that σ_m^2 is equal to its posterior mean of 0.09, one might anticipate a range of $\pm 2\sqrt{0.09} = 1.2$ d per 100 kg between the most extreme herd-year effects, using the Empirical Rule (Ott and Longnecker, 2001). Therefore, centered on an overall posterior mean of 0.55 d/100 kg as described earlier, we expect different clusters to range from -0.05 to 1.15 d of calving interval for every 100 kg increase of 305-d milk yield. Hence, it is possible for some herds to have no overall **e**-level relationship between the two traits, whereas other herds may have highly unfavorable relationships.

The evidence for heterogeneity of conditional variances was considerable. Posterior means, posterior standard deviations, 95% HPD and ESS for **e**-level and **u**-level conditional variances for the two traits are presented in Table 3. The **e**-level heteroskedasticity was prominent for both traits. For example, there was strong evidence for greater **e**-level or between-cow variability on both milk yield and reproductive performance on 3^+X milking herds compared to 2X milking herds. Defining $\sigma^2_{e_i,2X} = \exp\left(\begin{bmatrix}1 & 1\end{bmatrix} \log\left(\boldsymbol{\tau}_{e_i}\right)\right)$ and $\sigma^2_{e_i,3^+X} = \exp\left(\begin{bmatrix}1 & 0\end{bmatrix}\log\left(\boldsymbol{\tau}_{e_i}\right)\right)$ for $i = 1, 2|1$, the 95% HPD for the variance ratio $\left(\sigma^2_{e_i,2X} / \sigma^2_{e_i,3^+X}\right)$ between the

two levels did not include 1, being [0.62, 0.69] for milk yield (i=1) and [0.79, 0.94] for calving interval conditional on milk yield (i=2|1).

The magnitude of heterogeneity of the **e**-level or between-cow variance across clusters for each trait was also considerable, as indicated by the concentration of the 95% HPD on relatively large values of $\sigma_{v,e_i} = \frac{1}{\sqrt{\eta_{e_i} - 2}}$ (Table 3), which incidentally also defines the **e**-level coefficient of variation (CV) for conditional variances between clusters (Kizilkaya and Tempelman, 2005). That is, the posterior means for σ_{v,e_1} and $\sigma_{v,e_{2|1}}$ (Table 3) indicate that the CV of cluster-specific residual variances is roughly 32% and 77%, respectively. Indeed, the largest and smallest herd-year specific posterior means for elements of $\mathbf{v}_{e_1}$ were 2.72 and 0.44, respectively, for 305-d milk yield, meaning that there is an estimated 6-fold change between the most extreme herds for between-cow variability. Residual heteroskedasticity across clusters was even more noticeable for calving interval (again conditional on milk yield) as the largest and smallest posterior means of elements of $\mathbf{v}_{e_{2|1}}$ were estimated to be 6.05 and 0.21, respectively, leading to an estimated fold change of 28.

At the **u**-level, our analysis (Table 3) indicated that milk yield was significantly more variable between herds with an indecisive strategy on bST supplementation (>0-50% of the herd enrolled) compared with herds that were either committed to bST supplementation (≥50% of the herd enrolled) or that were not bST users at all (0% of the herd enrolled). In contrast, between-herd variation on calving interval conditional on milk yield was significantly greater among herds that used

bST supplementation at either level, compared to those that did not supplement at all with bST (0% of the herd enrolled).

5. Discussion:

In this study, we present a hierarchical Bayesian extension to classical bivariate mixed effects modeling that provides a general framework for investigating sources of heterogeneity for residual or subject level (**e**) and random or cluster level (**u**) (co)variances between two traits of interest. Using simulation, we validated the proposed hierarchical Bayesian model which is based on a recently developed (co)variance matrix reparameterization (Pourahmadi et al., 1999). We also validated the use of the DIC to choose between models that differ by the specification of cluster-specific random effects on the residual relationships between two traits. We then applied the model to investigate a currently critical dairy cattle management issue as it pertains to the covariance matrix architecture between milk production and reproductive fitness, specifically how herd management and environmental covariates may influence the random (i.e., herd) and residual (i.e., cow) level (co)variances.

The Cholesky-based reparameterization proposed by Pourahmadi (1999) alleviates the concern for checking positive definiteness constraints and, based on desirable orthogonality properties of the transformation (Pourahmadi, 2007), facilitates independent hierarchical modeling for each of the resulting parameters. From a multivariate applications standpoint, factors influencing $\varphi_j^{(e)}$ and $\varphi_k^{(u)}$ may be of greatest interest because they determine the subject and cluster specific

relationships, respectively, between traits in an unconstrained and easily interpretable manner. As previously noted by Pourahmadi (1999), these two sets of parameters imply a temporal order among response variables, such that inference on the constituent fixed effects ($\boldsymbol{\gamma}_e$ and $\boldsymbol{\gamma}_u$) and random effects (**m**) is also inherently order-dependent. In some subject-matter contexts, this order dependency embedded in the model may be a limitation for application to some inferential problems. In our case, however, the temporal argument is naturally based on the sequence of physiological events in a dairy cow. In a dairy production system, cows are already milking at the time reproductive management is implemented (Ensminger, 1993), thus implying milk production to be a factor that potentially influences reproductive performance. Conceptually, our model can be extended to $t > 2$ traits for more standard longitudinal data analysis applications as in Pourahmadi (1999); however, the number of different linear model components will increase to $3t + t(t\text{-}1)$ from the 8 different generalized linear models (i.e., on y_{1j}, y_{2j}, $\sigma^2_{e_{1,j}}$, $\sigma^2_{e_{2|1,j}}$, $\sigma^2_{u_{1,k}}$, $\sigma^2_{u_{2|1,k}}$, $\varphi_j^{(e)}$ and $\varphi_k^{(u)}$ considered within this paper.

The results from our dairy cattle application were very intuitive. However, up until this point, we knew of no other formal method to infer upon factors that systematically affect the relationships between two traits, and, more specifically, how this relationship is differentially driven by cluster-specific random versus residual effects and their component covariate effects. Our application suggests that the antagonistic relationship (high milk production associated with poorer reproductive performance) is primarily driven at the residual or cow level, but that

the degree of this relationship depended upon a common management practice, namely daily milking frequency. The additional mixed model extension on modeling variability (σ_m^2) in this relationship implied further that the residual relationship between the two traits is significantly heterogeneous across herds such that some herds may not even have an antagonistic relationship between the two traits. These results warrant further investigation of other management practices and herd-related factors to help unveil other potential sources of heterogeneity in the production-reproduction relationship across herds. That is, herds with inferences unusually distal to zero for their respective elements in **m** might be investigated retrospectively to explore any potentially new important management and environmental factors that affect $\varphi_j^{(e)}$. As our analysis did not consider a comprehensive set of factors, our estimates of σ_m^2 are likely to be somewhat inflated because of other potentially important covariates that were not modeled. A more comprehensive analysis based on a larger dataset and simultaneous fitting of several fixed effects is forthcoming in future animal science publications.

We believe that some future work on hierarchical modeling of heterogeneous covariances is merited. Firstly, an alternative (co)variance parameterization to Pourahmadi (1999) was proposed by Chen and Dunson (2003) and may be worth exploring further as an alternative framework for mixed effects modeling of (co)variances. One potentially attractive idea from Chen and Dunson (2003) and Kinney and Dunson (2007) is the use of Bayesian model averaging across a large number of candidate models as opposed to the use of DIC to choose between models.

Second, we were surprised to note a large estimated rank correlation of 0.68 between the posterior means of elements of $\mathbf{u}_2$ and the corresponding elements of $\mathbf{v}_{2|1}$, indicating that herd-years with longer calving intervals were also herd-years with more variable calving intervals among cows. A lognormal specification on $\mathbf{v}_{2|1}$ would facilitate a formal multivariate Gaussian prior between $\mathbf{u}_2$ and $\log(\mathbf{v}_{2|1})$ and might better capture this relationship (Ibanez-Escriche et al., 2008; Ros et al., 2004; Sorensen and Waagepetersen, 2003).

Third, we find that the conditional framework of the Bayesian paradigm is particularly appealing to the proposed bivariate linear mixed model due to its naturally embedded hierarchical rationality. We recognize that an analytical likelihood-based implementation might circumvent the computational expense of MCMC and any potential concerns about specification of prior distributions on parameters of interest. Along these lines, the *h*-likelihood approach proposed within the framework of double hierarchical generalized linear models (DHGLM) appears to be an attractive starting point (Lee and Nelder, 2001, 2006). Development of methodological extensions within the DHGLM framework may allow for joint modeling and likelihood-based inference on means, variances and covariances in a multivariate context, as well as introduction of random effects on their linear predictors.

Fourth, the reader might note that random effects specifications were not specified for the linear model on $\sigma^2_{u_{1,k}}$, $\sigma^2_{u_{2|1,k}}$ (Equation 9) or $\varphi_k^{(u)}$ (Equation 7). Since herd would then represent the experimental unit rather than cluster in both

cases, specifying additional random herd effects on the right side of both Equations (7) and (9) would model overdispersion due to unknown random effects. Conceptually, an additional "residual" term might be also considered for $\sigma^2_{e_{1,j}}$, $\sigma^2_{e_{2,j}}$ (Equation 8) or $\varphi_j^{(e)}$ (Equation 6), in order to account for overdispersion at the subject (i.e. cow) level (Cardoso et al., 2005; Foulley et al., 2004). In essence, these specifications would determine the marginal prior densities of **u** and **e** to be heterogeneous Student *t* rather than Gaussian distributions, thereby conferring some outlier-robustness properties (Cardoso, Rosa and Tempelman, 2007). Nevertheless, these extensions would create considerable increases in computational time using MCMC.

Finally, a note regarding the particularly large dataset size used in this study is in order: The size of the dataset was intended to allow for powerful inference across the deepest levels of the proposed hierarchical model while mimicking the data structure of the dairy cow application. However, the authors acknowledge that the performance of this complex of a hierarchical model in more modest-sized datasets should be investigated further.

To facilitate computational efficiency, we used the open-source free software `R` (R Development Core Team, 2008) incorporating the sparse linear algebra package `SparseM` (Koenker and Ng, 2009). Investigation of some of the methodological extensions suggested above may require coding the MCMC algorithm in a lower

level programming language (such as FORTRAN or C++) in order to make computing tractable and more efficient.

Computer code in R is readily available as supporting information from the Journal's website: http://www.biometrical-journal.com

6. Summary

Linear mixed effects modeling of (co)variances, and thus of relationships between traits of interest are possible for both random and residual effects based on a recently popularized covariance matrix decomposition. Hence, researchers should be able to further fine-tune inference on the architecture of correlations between traits by modeling (co)variances as functions of additional fixed and random effects. Using MCMC techniques, we validate the proposed methodology with a simulation study and demonstrate its applicability by addressing the question of heterogeneous relationships between milk production and reproductive performance of dairy cows.

Acknowledgements: We thank Dr. John Clay and his staff at the Dairy Records Management Systems, Raleigh, NC for providing the DHIA dataset used in this study. This study was partially funded by the Elwood Kirkpatrick Dairy Research Endowment, the Michigan Milk Producers Association, the College of Agriculture and Natural Resources and the Department of Animal Science at Michigan State University.

Conflict of Interests Statement: The authors have declared no conflict of interest.

Appendix: Full conditional densities (FCD)

Write the data for the two traits on subject j as $\mathbf{y} = \begin{bmatrix} \mathbf{y}_{1j} & \mathbf{y}_{2j} \end{bmatrix}'$ such that the entire data vector is $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1' & \mathbf{y}_2' & \mathbf{y}_3' & \dots & \mathbf{y}_n' \end{bmatrix}'$. Furthermore, write fixed and random design matrices for the two traits specific to animal j, respectively as $\mathbf{X}_j^{(1)} = \begin{bmatrix} \mathbf{x}_{1j}^{(1)'} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{2j}^{(1)'} \end{bmatrix}$ and $\mathbf{Z}_j = \begin{bmatrix} \mathbf{z}_j' & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_j' \end{bmatrix}$; $j = 1,2,\dots n$. Hence, the corresponding overall design matrices can be written as $\mathbf{X}^{(1)} = \begin{bmatrix} \mathbf{X}_1^{(1)'} & \mathbf{X}_2^{(1)'} & \dots & \mathbf{X}_n^{(1)'} \end{bmatrix}'$ and $\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1' & \mathbf{Z}_2' & \dots & \mathbf{Z}_n' \end{bmatrix}'$ linking $\mathbf{y}$ to $\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1' & \boldsymbol{\beta}_2' \end{bmatrix}'$ and $\mathbf{u} = \begin{bmatrix} \mathbf{u}_1' & \mathbf{u}_2' \end{bmatrix}'$, respectively. We also specify $\mathrm{var}(\mathbf{e}) = \boldsymbol{\Sigma}_e = \bigoplus_{j=1}^{n} \mathbf{R}_j$ where $\mathbf{e} = \begin{bmatrix} \mathbf{e}_1' & \mathbf{e}_2' & \mathbf{e}_3' & \dots & \mathbf{e}_n' \end{bmatrix}'$ and $\oplus$ denotes the direct sum operator (Searle, 1982) such that it should be readily noted that $\boldsymbol{\Sigma}_e^{-1} = \bigoplus_{j=1}^{n} \mathbf{R}_j^{-1}$. We similarly define $\boldsymbol{\Sigma}_g = \mathrm{var}(\mathbf{u})$ noting that $\boldsymbol{\Sigma}_g^{-1}$ can be readily determined by rearranging elements of $\bigoplus_{k=1}^{q} \mathbf{G}_k^{-1}$ by animals within traits rather than by traits within animals. It can then be noted using mixed model theory (Sorensen and Gianola, 2002) that the joint FCD of $\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\beta}' & \mathbf{u}' \end{bmatrix}'$ is multivariate Gaussian:

$$\boldsymbol{\theta} \sim N\left(\left(\mathbf{W}'\boldsymbol{\Sigma}_e^{-1}\mathbf{W}+\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\right)^{-1}\left(\mathbf{W}'\boldsymbol{\Sigma}_e^{-1}\mathbf{y}+\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\theta}_0\right),\left(\mathbf{W}'\boldsymbol{\Sigma}_e^{-1}\mathbf{W}+\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\right)^{-1}\right) \quad \text{[A1]}$$

for $\mathbf{W}=\left[\mathbf{X}^{(1)} \;\; \mathbf{Z}\right]$, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}=\mathrm{diag}\left(\mathbf{V}_1^{(\beta)} \;\; \mathbf{V}_2^{(\beta)} \;\; \boldsymbol{\Sigma}_g\right)$, and $\boldsymbol{\theta}_0=\mathrm{diag}\left(\boldsymbol{\beta}_1^{0\,\prime} \;\; \boldsymbol{\beta}_2^{0\,\prime} \;\; \mathbf{0}_{2q\,\mathrm{x}\,1}{}'\right)'$. There are a number of different alternative strategies for sampling from elements of $\boldsymbol{\theta}$, including single site or univariate Gibbs updates (Wang, Rutledge and Gianola, 1994) and block sampling strategies (GarciaCortes and Sorensen, 1996) that exploit the sparsity (i.e., high frequency of zero elements) in $\left(\mathbf{W}'\boldsymbol{\Sigma}_e^{-1}\mathbf{W}+\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\right)^{-1}$. Note then that draws of $\mathbf{u}_{2|1}$ can then simply be determined as $\mathbf{u}_2-\boldsymbol{\Psi}^{(u)}\mathbf{u}_1$ whereas draws of $\mathbf{e}_{2|1}$ can be determined as a vector with elements $\left\{e_{2,j}-\varphi_j^{(e)}e_{1,j}\right\}$.

Similar developments can be used to demonstrate that the FCD of $\left[\boldsymbol{\gamma}_e{}' \;\; \mathbf{m}'\right]'$, is multivariate Gaussian except that one makes the following substitutions in [A1]: $\left[\left[\mathbf{x}_1^{(2)} \;\; \mathbf{x}_2^{(2)} \;\; \ldots \;\; \mathbf{x}_n^{(2)}\right]' \;\; \left[\mathbf{z}_1 \;\; \mathbf{z}_2 \;\; \ldots \;\; \mathbf{z}_n\right]'\right]$ for $\mathbf{W}$, $\mathrm{diag}\left(\sigma^2_{e_{2|1},j}/e^2_{1,j}\right)$ for $\boldsymbol{\Sigma}_e$, $\mathrm{diag}\left(\mathbf{V}_{\gamma}^{(e)},\mathbf{I}\sigma_m^2\right)$ for $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, $\left(\boldsymbol{\mu}_{\gamma}^{(e)\,\prime} \;\; \mathbf{0}_{q\,\mathrm{x}\,1}{}'\right)'$ for $\boldsymbol{\theta}_0$, and a n x 1 vector with elements $\left\{y_{2j}-\mathbf{x}_{2,j}^{(1)}{}'\boldsymbol{\beta}_2-\mathbf{z}_j{}'\left(\boldsymbol{\Psi}^{(u)}\mathbf{u}_1+\mathbf{u}_{2|1}\right)\right\}$ for $\mathbf{y}$. Similarly, the FCD for $\boldsymbol{\gamma}_u$ is also multivariate Gaussian making the following

substitution for terms in [A1]: $\left[\mathbf{x}_1^{(3)} \;\; \mathbf{x}_2^{(3)} \;\; \ldots \;\; \mathbf{x}_n^{(3)}\right]'$ for $\mathbf{W}$, $\mathrm{diag}\left(\sigma^2_{e_{2|1},j} \Big/ \left(\mathbf{z}'_j \mathbf{u}_1\right)^2\right)$ for $\boldsymbol{\Sigma}_e$, $\mathbf{V}_{\boldsymbol{\gamma}}^{(u)}$ for $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, and $\left\{y_{2j} - \mathbf{x}_{2,j}^{(1)}{}'\boldsymbol{\beta}_2 - \mathbf{z}_j{}'\mathbf{u}_{2|1} - \varphi_j^{(e)} e_{1j}\right\}$ for $\mathbf{y}$.

The FCD for σ_m^2 can be readily demonstrated to be inverse gamma with parameters $(q/2)+\alpha_m$ and $\left((1/2)\mathbf{m}'\mathbf{m}\right)+\beta_m$. Similarly, the FCD of elements $\tau_{e_{i,l}}$ of $\boldsymbol{\tau}_{e_i} = \left\{\tau_{e_i,l}\right\}_{l=1}^{p_i^{(4)}}$ (i=1, 2|1) that correspond to the intercept and effects of levels of different classification factors that are not zeroed out can be demonstrated to be inverse gamma (Kizilkaya and Tempelman, 2005) with parameters $\left(n_{e_i,l}^{(\tau)} \Big/ 2\right)+\alpha_i^{(e)}$ and

$$(1/2)\sum_{j=1}^{n} \mathrm{I}\left(x_{ijl}^{(4)}=1\right)\left(e_{i,j}^2 \Bigg/ \left(\prod_{l'=1,l'\neq l}^{p_i^{(4)}} \left(\tau_{e_i,l'}\right)^{x_{ijl'}^{(4)}} \prod_{k=1}^{q}\left(v_{e_i,l'}\right)^{z_{jk}}\right)\right)+\beta_i^{(e)}.$$

Here $x_{ijl}^{(4)}$ denotes element l of $\mathbf{x}_{ij}^{(4)}$ and z_{jk} denotes element k of $\mathbf{z}_j$. Furthermore, $n_{e_i,l}^{(\tau)} = \sum_{j=1}^{n} \mathrm{I}\left(x_{ijl}^{(4)}=1\right)$ with $\mathrm{I}\left(x_{ijl}^{(4)}=1\right)$ being an indicator variable taking value 1 if $x_{ijl}^{(4)}=1$ and 0 otherwise. For elements that represent continuous covariates, rather

than dummy variables for classification factors or the intercept, a Metropolis-Hastings update is required (Cardoso et al., 2005). The FCD for elements of $\mathbf{v}_{e_i} = \left\{ v_{e_i,k} \right\}_{k=1}^{q}$ can similarly be seen to be inverse gamma with parameters

$$\left(n_{e_i,k}^{(v)} \Big/ 2 \right) + \eta_{e_i} \quad \text{and} \quad (1/2) \sum_{j=1}^{n} \mathrm{I}\left(z_{jk} = 1 \right) \left(e_{i,j}^{2} \Big/ \prod_{l=1}^{p_i^{(4)}} \left(\tau_{e_i,l} \right)^{x_{ijl}^{(4)}} \right) + \eta_{e_i} - 1$$

where $n_{e_i,k}^{(v)} = \sum_{j=1}^{n} \mathrm{I}\left(z_{jk} = 1 \right)$ for $\mathrm{I}\left(z_{jk} = 1 \right)$ being an indicator variable that takes value 1 if $z_{jk} = 1$ and 0 otherwise.

The FCD for the hyperparameter η_{e_i} for i (i=1, 2|1) does not have a recognizable form:

$$p\left(\eta_{e_i} \mid ELSE, \mathbf{y} \right) \propto$$
$$\left(\eta_{e_i} - 1 \right)^{\eta_{e_i} q} \left(\Gamma\left(\eta_{e_i} \right)^{q} \right)^{-1} \exp\left(-\left(\eta_{e_i} - 1 \right) \sum_{k=1}^{q} \mathbf{v}_{k,e_i}^{-1} \right) \left(\prod_{k=1}^{q} \left(\mathbf{v}_{k,e_i} \right)^{-\left(\eta_{e_i} + 1 \right)} \right) p\left(\eta_{e_i} \right)$$

Hence, sampling for η_{e_i} requires a Metropolis-Hastings step. In this case, we sampled from the FCD of $\zeta_{e_i} = \log\left(\eta_{e_i} \right)$ using a Metropolis algorithm (Chib and Greenberg, 1995) with a normal approximation to the FCD as the proposal density. That is, the proposal density is Gaussian with mean equal to the mode of the FCD function and a variance equal to the negative of the inverse Hessian of the FCD function evaluated at the previously sampled value for ζ_{e_i} . For this purpose, we

use a Newton-type algorithm based on a line search (Schnabel, Koontz and Weiss, 1985).

Finally, the FCD for elements $\tau_{u_{i,l}}$ of $\boldsymbol{\tau}_{u_i} = \left\{\tau_{u_i,l}\right\}_{l=1}^{p_i^{(5)}}$ (i=1, 2|1) can be shown to follow an inverse gamma density with parameters $\left(n_{u_i,l}^{(\tau)}/2\right) + \alpha_i^{(u)}$ and $\frac{1}{2}\sum_{k=1}^{q} \mathrm{I}\left(x_{ikl}^{(5)} = 1\right)\left(u_{i,k}^2 \Big/ \prod_{l'=1, l' \neq l}^{p_i^{(5)}} \left(\tau_{u_i,l'}\right)^{x_{ikl'}^{(5)}}\right) + \beta_i^{(u)}$. Here $x_{ikl}^{(5)}$ denotes element l of $\mathbf{x}_{ik}^{(5)}$ and $n_{u_i,l}^{(\tau)} = \sum_{k=1}^{q} \mathrm{I}\left(x_{ikl}^{(5)} = 1\right)$ with $\mathrm{I}\left(x_{ikl}^{(5)} = 1\right)$ being an indicator variable with value of 1 if $x_{ikl}^{(5)} = 1$ and 0 otherwise.

Table 1.1.a Minimum and maximum of the upper and lower boundaries of the 95% highest posterior density (HPD) interval, as well as true values used for simulation, for parameters defining heterogeneity in the residual (e) level and random (u) level regression coefficients, namely γ_e, σ_m^2 and γ_u, across 10 replicates for each of 12 simulation populations defined by all factorial combinations of 3 correlation architectures (A, B, C) with 2 (I, II) of 4 different values of σ_m^2 as indicated in text.

Scenarios		**A. Same sign:** positive u-level and e-level coefficients			**B. Opposite sign:** negative u-level and positive e-level coefficients			**C. Zero correlation:** zero u-level and e-level coefficients		
	Regression Parameters †	**True**	**Lower HPD [Min, Max]**	**Upper HPD [Min, Max]**	**True**	**Lower HPD [Min, Max]**	**Upper HPD [Min, Max]**	**True**	**Lower HPD [Min, Max]**	**Upper HPD [Min, Max**
I. $\sigma_m^2 = 0$	γ_{u_1}	0.3	[-1.81, 0.14]	[0.52, 2.22]	-0.3	[-2.17, -0.80]	[-0.21, 0.69]	0	[-1.44, -0.08]	[0.17, 1.75]
	γ_{u_2}	0.1	[-0.64, 0.07]	[0.34, 1.01]	-0.5	[-1.18, -0.32]	[-0.16, 0.59]	0	[-0.82, -0.24]	[0.05, 0.55]
	γ_{e_1}	0.6	[0.44, 0.58]	[, 0.60, 0.74]	0.6	[0.44, 0.57]	[0.59, 0.73]	0	[-0.10, 0.01]	[0.05, 0.17]
	γ_{e_2}	0.5	[0.37, 0.48]	[0.53, 0.65]	0.5	[0.37, 0.47]	[0.53, 0.63]	0	[-0.16, -0.04]	[0.00, 0.13]
	σ_m^2	0	-	-	0	-	-	0	-	-
II. $\sigma_m^2 = 0.1$	γ_{u_1}	0.3	[-1.03, 0.20]	[0.74, 1.81]	-0.3	[-1.49, -0.62]	[0.05, 0.99]	0	[-1.47, 0.44]	[0.22, 2.43]
	γ_{u_2}	0.1	[-0.45, -0.03]	[0.48, 0.86]	-0.5	[-1.37, -0.38]	[-0.37, 0.53]	0	[-0.89, -0.07]	[0.10, 0.81]
	γ_{e_1}	0.6	[0.35, 0.57]	[0.55, 0.80]	0.6	[0.40, 0.54]	[0.61, 0.76]	0	[-0.19, -0.03]	[0.00, 0.19]
	γ_{e_2}	0.5	[0.30, 0.46]	[0.53,0.70]	0.5	[0.32, 0.53]	[0.52, 0.77]	0	[-0.20, -0.02]	[0.00, 0.20]
	σ_m^2	0.1	[0.02, 0.08]	[0.11, 0.23]	0.1	[0.02, 0.07]	[0.12, 0.21]	0.1	[0.01, 0.06]	[0.10, 0.21]

† $\gamma_{u_1}, \gamma_{u_2}$ correspond to u-level regression coefficients for levels 1 and 2, respectively, of a fixed effect factor.

$\gamma_{e_1}, \gamma_{e_2}$ correspond to e-level regression coefficients for levels 1 and 2, respectively, of a fixed effect factor.

σ_m^2 corresponds to the parameter defining random between-herd heterogeneity among the e-level regression parameters.

Table 1.1.b Minimum and maximum of the upper and lower boundaries of the 95% highest posterior density (HPD) interval, as well as true values used for simulation, for parameters defining heterogeneity in the residual (e) level and random (u) level regression coefficients, namely γ_e, σ_m^2 and γ_u, across 10 replicates for each of 12 simulation populations defined by all factorial combinations of 3 correlation architectures (A, B, C) with 2 (III, IV) of 4 different values of σ_m^2 as indicated in text.

Scenarios		A. **Same sign:** positive u-level and e-level coefficients			B. **Opposite sign:** negative u-level and positive e-level coefficients			C. **Zero correlation:** zero u-level and e-level coefficients		
	Regression Parameters †	**True**	**Lower HPD [Min, Max]**	**Upper HPD [Min, Max]**	**True**	**Lower HPD [Min, Max]**	**Upper HPD [Min, Max]**	**True**	**Lower HPD [Min, Max]**	**Upper HPD [Min, Max**
III. $\sigma_m^2 = 1$	γ_{u_1}	0.3	[-0.87, 0.04]	[0.38, 1.45]	-0.3	[-1.57, -0.59]	[-0.04, 1.00]	0	[-1.53, -0.20]	[0.13, 1.77]
	γ_{u_2}	0.1	[-0.95, -0.14]	[0.04, 0.72]	-0.5	[-1.44, -0.41]	[-0.42, 0.62]	0	[-0.90, -0.27]	[0.09, 0.65]
	γ_{e_1}	0.6	[0.09, 0.52]	[0.48, 0.88]	0.6	[0.15, 0.65]	[0.51, 1.04]	0	[-0.41, 0.08]	[-0.01, 0.47]
	γ_{e_2}	0.5	[-0.01, 0.45]	[0.38, 0.82]	0.5	[0.03, 0.47]	[0.40, 0.86]	0	[-0.47, 0.07]	[-0.06, 0.47]
	σ_m^2	1	[0.57, 0.86]	[1.08, 1.52]	1	[0.51, 0.78]	[0.97, 1.36]	1	[0.42, 0.82]	[0.80, 1.47]
IV. $\sigma_m^2 = 10$	γ_{u_1}	0.3	[-1.52, 0.21]	[0.66, 2.14]	-0.3	[-1.71, -0.70]	[0.12, 0.86]	0	[-1.26, -0.30]	[0.52, 1.48]
	γ_{u_2}	0.1	[-0.90, -0.16]	[-0.06, 0.79]	-0.5	[-1.35, -0.94]	[-0.42, 0.02]	0	[-0.85, 0.23]	[0.10, 1.05]
	γ_{e_1}	0.6	[-0.16, 0.58]	[0.84, 1.50]	0.6	[-0.24, 0.68]	[0.79, 1.67]	0	[-0.96, -0.04]	[0.03, 0.88]
	γ_{e_2}	0.5	[-0.30, 0.44]	[0.70, 1.35]	0.5	[-0.31, 0.52]	[0.71, 1.51]	0	[-0.94, -0.07]	[0.05, 0.85]
	σ_m^2	10	[6.2, 9.6]	[10.1, 14.9]	10	[6.7, 10.5]	[10.8, 16.5]	10	[6.5, 8.4]	[10.4, 13.1]

† $\gamma_{u_1}, \gamma_{u_2}$ correspond to u-level regression coefficients for levels 1 and 2, respectively, of a fixed effect factor.

$\gamma_{e_1}, \gamma_{e_2}$ correspond to e-level regression coefficients for levels 1 and 2, respectively, of a fixed effect factor.

σ_m^2 corresponds to the parameter defining random between-herd heterogeneity among the e-level regression parameters.

Table 1.2. Posterior mean (PMEAN), posterior standard deviation (PSD), 95% highest posterior density (HPD) intervals and effective sample size (ESS) on residual (**e**) level (namely, γ_e and σ_m^2) and random (**u**) level (namely, γ_u) regression parameters between milk yield at 305 days-in-lactation and calving interval in Michigan first lactation dairy cows.

Regression parameters †	PMEAN	PSD	95% HPD	ESS
$\gamma_{u,0\%bST}$, d/100 kg	0.16 [x]	0.17	[-0.17, 0.49]	28 549
$\gamma_{u,>0-50\%bST}$, d/100 kg	0.17 [x]	0.20	[-0.22, 0.56]	28 959
$\gamma_{u,>50\%bST}$, d/100 kg	0.15 [x]	0.19	[-0.22, 0.51]	28 409
$\gamma_{e,2X}$, d/100 kg	0.66 [a]	0.04	[0.57, 0.74]	79 573
$\gamma_{e,3^+X}$, d/100 kg	0.45 [b]	0.05	[0.36, 0.54]	61 687
σ_m^2,(d/100 kg)^2	0.09	0.03	[0.03, 0.16]	612

([x]) and ([a,b]) Letters indicate significant differences (two-tailed Bayesian P-value < 0.05) between management practices within the **u**-level and **e**-level regression parameters, respectively,

† $\gamma_{u,0\%bST} = [1 \quad 1 \quad 0]\gamma_u$, $\gamma_{u,>0-50\%bST} = [1 \quad 0 \quad 1]\gamma_u$ and $\gamma_{u,>50\%bST} = [1 \quad 0 \quad 0]\gamma_u$ are the random (**u**) level regression parameters between milk yield at 305 days-in-lactation and calving interval for herds that had 0%, >0 to 50% and >50% of their cows enrolled for supplementation with bovine somatotropin (bST), respectively.

$\gamma_{e,2X} = [1 \quad 1]\gamma_e$ and $\gamma_{e,3^+X} = [1 \quad 0]\gamma_e$ are the residual (**e**) level regression parameters between milk yield at 305 days-in-lactation and calving interval for cows in herds with twice a day and three times a day (or greater) milking frequency, respectively.

σ_m^2 is the parameter defining random between-herd heterogeneity among the **e**-level regression parameters.

Table 1.3. Posterior mean (PMEAN), posterior standard deviation (PSD), 95% highest posterior density (HPD) intervals, and effective sample size (ESS) for residual (**e**) level and random (**u**) level conditional variances for milk yield at 305 days-in-milk and calving interval in Michigan first lactation dairy cows.

Variance Components †	PMEAN	PSD	95% HPD	ESS
Milk yield at 305 days in milk				
$\sigma^2_{u_1,0\%bST}$, (100 kg)^2	109 [x]	15	[81, 140]	3 863
$\sigma^2_{u_1,<50\%bST}$, (100 kg)^2	164 [y]	8	[148, 181]	52 986
$\sigma^2_{u_1,\geq 50\%bST}$, (100 kg)^2	119 [x]	14	[92, 148]	63 551
$\sigma^2_{e_1,2X}$, (100 kg)^2	181 [a]	3	[175, 186]	19 318
$\sigma^2_{e_1,3+X}$, (100 kg)^2	276 [b]	67	[263, 289]	3 591
σ_{v,e_1}	0.32	0.02	[0.29, 0.35]	5 910
Calving interval				
$\sigma^2_{u_{2\|1},0\%bST}$, days^2	297 [x]	63	[181, 422]	1 689
$\sigma^2_{u_{2\|1},>0-50\%bST}$, days^2	759 [y]	59	[646, 877]	18 159
$\sigma^2_{u_{2\|1},\geq 50\%bST}$, days^2	695 [y]	113	[490, 928]	26 160
$\sigma^2_{e_{2\|1},2X}$, days^2	9 126 [a]	235	[8 669, 9 584]	5 878
$\sigma^2_{e_{2\|1},3+X}$, days^2	10 593 [b]	429	[9 769, 11 454]	1 250
$\sigma_{v,e_{2\|1}}$	0.77	0.05	[0.68, 0.87]	5 294

([x,y]) and ([a,b]) Letters indicate significant differences (two-tailed Bayesian P-value < 0.05) between management practices within the **u**-level and **e**-level factors, respectively, for each trait.

† $\sigma^2_{u_i,0\%bST} = \exp\left(\left[1 \quad 1 \quad 0\right] \log\left(\boldsymbol{\tau}_{u_i}\right)\right)$, $\sigma^2_{u_i,>0-50\%bST} = \exp\left(\left[1 \quad 0 \quad 1\right] \log\left(\boldsymbol{\tau}_{u_i}\right)\right)$ and $\sigma^2_{u_i,>50\%bST} = \exp\left(\left[1 \quad 0 \quad 0\right] \log\left(\boldsymbol{\tau}_{u_i}\right)\right)$ are the random (**u**) level conditional variances for milk yield at 305 days-in-lactation (i = 1) and calving interval (i = 2|1) for herds that had 0%, >0 to 50% and >50% of the herd enrolled for supplementation with bovine somatotropin (bST), respectively.

$\sigma^2_{e_i,2X} = \exp\left(\left[1 \quad 1\right] \log\left(\boldsymbol{\tau}_{e_i}\right)\right)$ and $\sigma^2_{e_i,3^+X} = \exp\left(\left[1 \quad 0\right] \log\left(\boldsymbol{\tau}_{e_i}\right)\right)$ are the residual (**e**) level conditional variances for milk yield at 305 days-in-milk (i = 1) and calving interval (i = 2|1) for herds with twice a day and three times a day (or greater) milking frequency, respectively.

σ_{v,e_i} is the **e**-level coefficient of variation for conditional variances between clusters.

References

Albert, J. H. (1988). Computational methods using a Bayesian hierarchical generalized linear model. *Journal of the American Statistical Association* **83**, 1037-1044.

Bello, N. M., Erskine, R. J., and Tempelman, R. J. (2009). Heterogeneous relationship between milk production and reproduction in dairy cows: Preliminary evidence. In *Proceedings of the Annual Joint Meeting of the American Dairy Science Association and the American Society of Animal Science*. Montreal, Quebec, Canada.

Cardoso, F. F., Rosa, G. J., and Tempelman, R. J. (2005). Multiple-breed genetic inference using heavy-tailed structural models for heterogeneous residual variances. *Journal of Animal Science* **83**, 1766-1779.

Cardoso, F. F., Rosa, G. J., and Tempelman, R. J. (2007). Accounting for outliers and heteroskedasticity in multibreed genetic evaluations of postweaning gain of Nelore-Hereford cattle. *Journal of Animal Science* **85**, 909-918.

Chen, Z., and Dunson, D. B. (2003). Random effects selection in linear mixed models. *Biometrics* **59**, 762-769.

Chib, S., and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *American Statistician* **49**, 327-335.

Clayton, D. G. (1996). Generalized Linear Mixed Models. In *Markov chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (eds), 275-302. London: Chapman and Hall.

Crossa, J., Burgueno, J., Cornelius, P. L., McLaren, G., Trethowan, R., and Krishnamachari, A. (2006). Modeling genotype x environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. *Crop Science* **46**, 1722-1733.

Daniels, M. J., and Pourahmadi, M. (2002). Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika* **89**, 553-566.

Eldridge, S. M., Ashby, D., Feder, G. S., Rudnicka, A. R., and Ukoumunne, O. C. (2004). Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clinical Trials* **1**, 80-90.

Ensminger, M. E. (1993). *Dairy Cattle Science*, Third edition: Interstate Publishers Inc.

Foulley, J. L., Sorensen, D., Robert-Granie, C., and Bonaiti, B. (2004). Heteroskedasticity and Structural Models for Variances. *Journal of Indian Society of Agricultural Statistics* **57**, 64-70.

GarciaCortes, L. A., and Sorensen, D. (1996). On a multivariate implementation of the Gibbs sampler. *Genetics Selection Evolution* **28**, 121-126.

Gelfand, A. E., and Sahu, K. (1999). Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association* **94**, 247-253.

Gelfand, A. E., and Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association* **85**, 398-409.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**, 515-533.

Ibanez-Escriche, N., Sorensen, D., Waagepetersen, R., and Blasco, A. (2008). Selection for environmental variation: a statistical analysis and power calculations to detect response. *Genetics* **180**, 2209-2226.

Kaufman, C. G., and Sain, S. R. (2010). Bayesian Functional ANOVA Modeling Using Gaussian Process Prior Distributions. *Bayesian Analysis* **5**, 123-149.

Kinney, S. K., and Dunson, D. B. (2007). Fixed and random effects selection in linear and logistic models. *Biometrics* **63**, 690-698.

Kizilkaya, K., and Tempelman, R. J. (2005). A general approach to mixed effects modeling of residual variances in generalized linear mixed models. *Genetic Selection and Evolution* **37**, 31-56.

Koenker, R., and Ng, P. (2009). SPARSEM: Sparse Linear Algebra. R package Version 0.80.

Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2005). *Applied Linear Statistical Models*, Fifth edition. New York, NY: McGraw-Hill/Irwin Publishers.

Laben, R. L., Shanks, R., Berger, P. J., and Freeman, A. E. (1982). Factors Affecting Milk-Yield and Reproductive-Performance. *Journal of Dairy Science* **65**, 1004-1015.

Lee, Y., and Nelder, J. A. (2001). Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika* **88**, 987-1006.

Lee, Y., and Nelder, J. A. (2006). Double hierarchical generalized linear models. *Journal of the Royal Statistical Society Series C-Applied Statistics* **55**, 139-167.

Littell, R. C., Freund, R. J., and Spector, P. C. (1991). *SAS System for Linear Models*, Third edition: SAS Institute.

Lopez-Gatius, F., Garcia-Ispierto, I., Santolaria, P., Yaniz, J., Nogareda, C., and Lopez-Bejar, M. (2006). Screening for high fertility in high-producing dairy cows. *Theriogenology* **65**, 1678-1689.

Lucy, M. C. (2001). Reproductive loss in high-producing dairy cattle: where will it end? *Journal of Dairy Science* **84**, 1277-1293.

Mauricio, R. (2001). Mapping quantitative trait loci in plants: Uses and caveats for evolutionary biology. *Nature Reviews Genetics* **2**, 370-381.

Milliken, G. A., and Johnson, D. E. (2009). *Analysis of Messy Data - Volume 1: Designed Experiments*, Second edition: Chapman and Hall/CRC Press.

Mulder, H. A., Bijma, P., and Hill, W. G. (2007). Prediction of breeding values and selection responses with genetic heterogeneity of environmental variance. *Genetics* **175**, 1895-1910.

Ott, R. L., and Longnecker, M. (2001). *An Introduction to Statistical Methods and Data Analysis*, Fifth edition. United States of America: Duxbury.

Piepho, H. P., Mohring, J., Melchinger, A. E., and Buchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* **161**, 209-228.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* **86**, 677-690.

Pourahmadi, M. (2007). Cholesky decompositions and estimation of a covariance matrix: orthogonality of variance-correlation parameters. *Biometrika* **94**, 1006-1013.

Pourahmadi, M., and Daniels, M. J. (2002). Dynamic conditionally linear mixed models for longitudinal data. *Biometrics* **58**, 225-231.

R Development Core Team (2008). R: A language and environment for statistical computing. Version 2.8.1. Vienna, Austria: R Foundation for Statistical Computing.

Raftery, A. E., and Lewis, S. (eds) (1992). *How Many Iterations in the Gibbs Sampler?* Oxford, UK: Oxford University Press.

Riley, R. D., Abrams, K. R., Lambert, P. C., Sutton, A. J., and Thompson, J. R. (2007). An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Statistics in Medicine* **26**, 78-97.

Robinson, G. K. (1991). That BLUP is a good thing - the estimation of random effects. *Statistical Science* **6**, 15-51.

Ros, M., Sorensen, D., Waagepetersen, R., *et al.* (2004). Evidence for genetic control of adult weight plasticity in the snail Helix aspersa. *Genetics* **168**, 2089-2097.

Schnabel, R. B., Koontz, J. E., and Weiss, B. E. (1985). A Modular System of Algorithms for Unconstrained Minimization. *Acm Transactions on Mathematical Software* **11**, 419-440.

Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*, First edition: Wiley-Interscience.

Sorensen, D., and Gianola, D. (2002). *Likelihood, Bayesian and MCMC Methods in Quantitative Genetics*. New York: Springer Verlag.

Sorensen, D., and Waagepetersen, R. (2003). Normal linear models with genetically structured residual variance heterogeneity: a case study. *Genetical Research* **82**, 207-222.

Sorensen, D. A., Andersen, S., Gianola, D., and Korsgaard, I. (1995). Bayesian-Inference in Threshold Models Using Gibbs Sampling. *Genetics Selection Evolution* **27**, 229-249.

Sorensen, P., Lund, M. S., Guldbrandtsen, B., Jensen, J., and Sorensen, D. (2003). A comparison of bivariate and univariate QTL mapping in livestock populations. *Genetics Selection Evolution* **35**, 605-622.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society - Series B* **64**, 1-34.

Thompson, R., and Meyer, K. (1986). A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livestock Production Science* **15**, 299-313.

Wang, C. S., Rutledge, J. J., and Gianola, D. (1994). Bayesian-Analysis of Mixed Linear-Models Via Gibbs Sampling with an Application to Litter Size in Iberian Pigs. *Genetics Selection Evolution* **26**, 91-115.

Washburn, S. P., Silvia, W. J., Brown, C. H., McDaniel, B. T., and McAllister, A. J. (2002). Trends in reproductive performance in Southeastern Holstein and Jersey DHI herds. *Journal of Dairy Science* **85**, 244-251.

CHAPTER 2

Management-Based Heterogeneity in the Association between and the Variability of Milk Production and Calving Interval of Dairy Cows

ABSTRACT

Inferences on the nature of the association between milk production and reproductive performance of dairy cows have been conflicting. This problem may relate to an underappreciation of the differences between the within-herd (i.e. *cow level*) versus across-herd (i.e. *herd level*) components of this relationship. Furthermore, these associations may depend upon various management factors. We recently developed a bivariate hierarchical Bayesian approach to model heterogeneity in variances and covariances for both cow- and herd level components between two traits as a function of various explanatory factors. The objectives of this study were to apply this model 1) to investigate the nature of the relationship between 305d milk yield (MY) and calving interval (CI) of Michigan dairy cows, and 2) to evaluate various herd management factors as potential sources of heterogeneity in this relationship. Data consisted of 124,079 lactation records from 541 Michigan dairy farms. Means, variances and covariances between MY and CI were jointly modeled as separate functions of various management practices and herd attributes, with the final model chosen using the Deviance Information Criterion. Herds heavily involved with bST (>50% of their cows) had a favorable association between the two traits with an estimated change in herd CI of -1.37±0.13 d per 100kg increase in herd MY. Within herds, higher producing cows had overall poorer reproductive performance than lower producing cows, with CI increasing by 0.51±0.01 d

per 100kg increase in MY. This antagonism was particularly more pronounced if cows were milked twice-a-day (0.57±0.01d/100kg) as compared to thrice-a-day (or greater) (0.45±0.01d/100kg); furthermore, significant differences in the MY-CI association were also evident between years and seasons. The cow level association between MY and CI was significantly variable across herds (between-herd standard deviation = 0.17±0.01d/100kg), thus supporting future retrospective investigation of other management sources of heterogeneity on within-herd association between MY and CI. Understanding the factors that influence the between-herd and within-herd associations between MY and dairy fertility is critical to tailoring dairy management programs that optimize overall dairy performance.

Keywords: dairy cow, herds, milk production, reproduction, management.

INTRODUCTION

Milk yield and reproduction are among the most important broad categories of phenotypes for successful dairy production. Historical trends indicate declining reproductive performance with increasing milk yield (Butler and Smith, 1989; Chapter 1; Hare et al., 2006; Lucy, 2001; Norman et al., 2009). The implications of these trends have raised major concern regarding the long-term sustainability of the dairy industry. Recent studies have challenged this general assertion of antagonism between milk yield and reproductive performance. Specifically some data support favorable associations between milk production and reproduction, whereby higher producing cows were more likely to become pregnant (Emanuelson and Oltenacu, 1998; Lopez-Gatius et al., 2006) and higher producing herds had the lower average number of days open and the shortest

average intervals between calvings (Laben et al., 1982; Lof et al., 2007). Overall, these conflictive results indicate the need for a more comprehensive strategy to investigate management practices and herd attributes that might influence the association between milk yield and reproduction.

A common problem with previous studies is an under appreciation that associations within herd (*cow level*) may be inherently different from associations between herds (i.e. *herd level*) (Calus et al., 2005; Windig et al., 2005). Herds and cows constitute different units of performance; the relationship between two or more outcomes at both levels intertwines with each other to yield an overall phenotype. If these distinctions between cows and herds are not made, any reported associations may be overly generalized or even biased (Windig et al., 2005)!

Some association studies were based on univariate (i.e., single trait response) analyses, whereby reproductive traits are modeled as a function of milk yield (Laben et al., 1982; Lopez-Gatius et al., 2006; Spalding et al., 1974), or, conversely, milk yield is compared between reproductively successful and unsuccessful females (Lopez-Gatius et al., 2006; Windig et al., 2005). In such single-trait models, however, the prevailing assumptions are that the trait chosen to be the explanatory variable is measured without error and it is not influenced by other independent variables in the model. Clearly, these assumptions are tenuous and likely to lead to potentially biased or misleading inferences.

Hierarchical multivariate models provide a general framework to explicitly study associations between outcomes in the form of covariances and partition the components of these associations (Sorensen and Gianola, 2002); e.g., herd and cow; an obvious third component not addressed in this paper is genetic. We recently developed a hierarchical

bivariate Bayesian model that further specifies the different components of association between two traits as functions of fixed and random effects (Bello et al., 2010; Chapter 1). In this study, we apply our recently developed methodology to investigate the cow level and herd level associations between 305-day cumulative milk yield (MY) and calving interval (CI) of Michigan dairy cows, including the evaluation of various management factors and herd attributes that may be involved in these associations.

MATERIALS AND METHODS

Data Description

Data files for test-day records for Michigan dairy farms enrolled in the Dairy Herd Improvement program were obtained from Dairy Records Management Systems (DHIA; Raleigh, NC). Lactation records from first, second and third parity Holstein cows that calved between January 2005 and December 2006 were specifically extracted. Herds were required to have at least 25 cows per year. Lactation records were required to be based on at least 5 test-dates per lactation, and herds were required to have yearly average test-day intervals not greater than 45 days. All records were required to be complete for cow and herd identification as well as for the response variables of interest and potentially important explanatory variables, as described later. After editing, the total number of lactation records available for analysis was 124,079, corresponding to 98,950 cows from 541 dairy herds, commensurate with 987 herd-year clusters or contemporary groups.

The dependent variables considered in this study were y_1 = MY, expressed in kg; and y_2 = CI, expressed in days. These variables were specifically selected to be long-

term cumulative summaries of performance for milk production and reproduction throughout lactation. Here, CI was defined as the number of days between 2 consecutive calvings. For lactation records for which a subsequent calving date was not available (i.e. cows that did not become pregnant; namely, 24% of lactation records), CI was calculated based on the last recorded breeding date plus 280 days of average gestation and considered to be right-censored. Four calving seasons were defined based on the month of calving when lactation was initiated: Fall, from September through November; Winter, from December through February; Spring, from March through May; and Summer, from June through August. Information on selected management practices and herd-year descriptors was also gathered from the DHIA dataset as potential explanatory variables. These included herd milking frequency (i.e., 2 times per day , or 2X, versus 3 or more times per day, or $3^{+}X$), herd usage of bovine somatotropin (bST) i.e. non-users, with 0% of the herd enrolled; intermediate users, with >0-50% of the herd enrolled; and committed users, with ≥50% of the herd enrolled), individual cow supplementation with bST during a lactation (i.e. yes or no), herd size (expressed on the log base 10 scale and as a deviation from its mean) and herd expansion (expressed as the percentage change in herd size from the preceding year). In addition, the use of synchronized breeding (yes or no) was considered, as defined on a herd-year basis and using the adjusted Chi-square categorization method proposed by Miller et al. (2007), whereby herds were classified as either having synchronized breeding or not. Deciding which of these factors were to be incorporated as explanatory variables, and at what level of the hierarchical model, was based on a sequential model selection approach described below.

Animal Care and Use Committee approval was not obtained for this study because the data were obtained from an existing performance records database.

Model specification and posterior inference

We implement our recently developed Bayesian approach to model heterogeneity of cow level and herd level variance-covariance matrices under a bivariate linear mixed model using Markov chain Monte Carlo (MCMC) methods; more details on our procedure can be found in Bello et al. (2010; Chapter 1). We assume that pairs of records on MY and CI are available on each of n cows such that the data vector $\mathbf{y}_1 = \left\{y_{1,i}\right\}_{i=1}^{n}$ for MY and the data vector $\mathbf{y}_2$ for CI are both n x 1. We formally accommodate the right-censored nature of CI using data augmentation, as presented by Sorensen et al. (1998). Briefly, the n x 1 data vector for CI, $\mathbf{y}_2 = \left[\mathbf{y}_{21}' \quad \mathbf{y}_{22}'\right]'$ is composed of $\mathbf{y}_{21}$ as a n_1 x 1 vector of uncensored observations (i.e., a subsequent calving was indeed recorded) and $\mathbf{y}_{22} = \left\{y_{22,i}\right\}_{i=1}^{n_2}$ as a n_2 x 1 vector of right-censored values as described earlier. In other words, the elements of $\mathbf{y}_{22}$ are known to be less than or equal to the corresponding actual but unknown CI, say $\mathbf{y}_{22}^{*}$. Following Sorensen et al. (1998), we "augment" the data with MCMC samples from $\mathbf{y}_{22}^{*}$ with $\mathbf{y}_{22}^{*} = \left\{y_{22,i}^{*}\right\}_{i=1}^{n_2}$, $y_{22,i}^{*} \geq y_{22,i}$ to account for the uncertainty in $\mathbf{y}_{22}^{*}$. That is, at each MCMC cycle, we replace $\mathbf{y}_{22}$ with samples of $\mathbf{y}_{22}^{*}$ and write the new augmented

data vector as $\mathbf{y}_2^{*\prime} = \left[\mathbf{y}_{21}' \quad \mathbf{y}_{22}^{*\prime}\right] = \left\{y_{2,i}^{*}\right\}_{i=1}^{n}$, which we then use in our bivariate linear mixed model (Bello et al., 2010; Chapter 1):

$$\begin{bmatrix} y_{1,i} \\ y_{2,i}^{*} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{1i}'\boldsymbol{\beta}_1 + \mathbf{z}_i'\mathbf{u}_1 + e_{1,i} \\ \mathbf{x}_{2i}'\boldsymbol{\beta}_2 + \mathbf{z}_i'\mathbf{u}_2 + e_{2,i} \end{bmatrix}. \qquad [1]$$

Here, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are, respectively, p_1 x 1 and p_2 x 1 vectors of selected classical fixed effects whereas $\mathbf{u}_1$ and $\mathbf{u}_2$ are each q x 1 vectors of classical random effects of herd-year with subscripts denoting trait (1 for MY, 2 for CI). Similarly, $e_{1,i}$ and $e_{2,i}$ are residual effects on the corresponding response variables as specific to the ith record. The jth herd-year specific random (co)variance matrix and the ith cow-specific residual (co)variance matrix are defined, respectively, as

$$\mathbf{G}_j = \text{var}\begin{bmatrix} u_{1,j} \\ u_{2,j} \end{bmatrix} = \begin{bmatrix} \sigma^2_{u_{1,j}} & \sigma_{u_{12},j} \\ \sigma_{u_{12},j} & \sigma^2_{u_{2,j}} \end{bmatrix} \text{ and } \mathbf{R}_i = \text{var}\begin{bmatrix} e_{1,i} \\ e_{2,i} \end{bmatrix} = \begin{bmatrix} \sigma^2_{e_{1,i}} & \sigma_{e_{12},i} \\ \sigma_{e_{12},i} & \sigma^2_{e_{2,i}} \end{bmatrix}. \quad [2]$$

Here $\sigma^2_{u_{1,j}}$ and $\sigma^2_{u_{2,j}}$ are the random effects (i.e., herd-year level) variances in MY and CI, respectively, and $\sigma_{u_{12},j}$ is the corresponding random effects covariance between the two traits, specific to the jth herd. Similarly, $\sigma^2_{e_{1,i}}$ and $\sigma^2_{e_{2,i}}$ represent the residual (i.e., cow level) variances for MY and CI, respectively, with $\sigma_{e_{12,i}}$ being the

corresponding residual covariance between the traits for the ith cow. This defines then a multivariate heteroskedastic model.

Following Bello et al. (2010; Chapter 1), we specify the pair of residuals on the ith cow record as follows

$$\mathbf{e}_i = \begin{bmatrix} e_{1,i} \\ e_{2,i} \end{bmatrix} = \begin{bmatrix} e_{1,i} \\ e_{1,i}\varphi_i^{(e)} + e_{2|1,i} \end{bmatrix} = \begin{bmatrix} 0 \\ e_{1,i} \end{bmatrix} \varphi_i^{(e)} + \begin{bmatrix} e_{1,i} \\ e_{2|1,i} \end{bmatrix}. \qquad [3]$$

Here $\varphi_i^{(e)}$ represents the cow specific (residual) association of $e_{2,i}$ on $e_{1,i}$, such that $e_{2|1,i}$ is the conditional residual for CI given MY, being independent of $e_{1,i}$ with $e_{2|1,i} \sim N\left(0, \sigma^2_{e_{2|1,i}}\right)$. Hence, we rewrite $\sigma_{e_{12},i} = \varphi_i^{(e)} \sigma^2_{e_1,i}$ and $\sigma^2_{e_2,i} = \sigma^2_{e_{2|1},i} + \left(\varphi_i^{(e)}\right)^2 \sigma^2_{e_1,i}$.

Similarly, we specify the following relationship for the pair of random effects on the jth herd:

$$\mathbf{u}_{.j} = \begin{bmatrix} u_{1,j} \\ u_{2,j} \end{bmatrix} = \begin{bmatrix} u_{1,j} \\ u_{1,j}\varphi_j^{(u)} + u_{2|1,j} \end{bmatrix} = \begin{bmatrix} 0 \\ u_{1,j} \end{bmatrix} \varphi_j^{(u)} + \begin{bmatrix} u_{1,j} \\ u_{2|1,j} \end{bmatrix}. \qquad [4]$$

Here $\varphi_j^{(u)}$ represents the herd-year specific (random effects) association of $u_{2,j}$ on $u_{1,j}$, such that $u_{2|1,j}$ is the conditional random effect on CI given MY corresponding to

herd-year j, being independent of $u_{1,j}$ with $u_{2|1,j} \sim N\left(0, \sigma^2_{u_{2|1,j}}\right)$. Hence, we rewrite

$$\sigma_{u_{12},j} = \varphi_j^{(u)} \sigma^2_{u_1,j} \text{ and } \sigma^2_{u_2,j} = \sigma^2_{u_{2|1},j} + \left(\varphi_j^{(u)}\right)^2 \sigma^2_{u_1,j}.$$

Note that $\varphi_j^{(u)} = \dfrac{\sigma_{u_{12},j}}{\sigma^2_{u_1,j}}$ can be interpreted as the conditional change in $u_{2,j(i)}$, and hence in $y^*_{2,i}$, for every unit change in $u_{1,j(i)}$, where $j(i)$ defines the jth herd-year associated with cow i. Similarly, $\varphi_i^{(e)} = \dfrac{\sigma_{e_{12},i}}{\sigma^2_{e_1,i}}$ can be interpreted as the conditional change in $e_{2,i}$, and hence in $y^*_{2,i}$, for every unit change in $e_{1,i}$. These association coefficients hence describe the relationship between milk production and reproductive performance at two different levels.

In modeling sources of heterogeneity on cow level associations, we specify the following linear mixed effects model:

$$\varphi_i^{(e)} = \mathbf{x}'_{3i}\boldsymbol{\gamma}_e + \mathbf{z}'_i\mathbf{m}_e. \qquad [5]$$

Here, $\boldsymbol{\gamma}_e$ represents a p_3 x 1 vector of unknown fixed effects, with $\mathbf{x}'_{3i}$ being the known row incidence vector, and $\mathbf{m}_e$ represents a q x 1 vector of unknown random herd-specific effects on the residual association such that $\mathbf{m}_e \sim N\left(\mathbf{0}, \mathbf{I}\sigma^2_{m_e}\right)$. As with classical specifications in typical mixed effects models (e.g., Equation [1]), the term "fixed

effects" here pertains to the effects of systematic management factors that can be subsequently inferred upon in other studies whereas "random effects" pertains to the effects of potentially exchangeable factors that can be characterized by a distribution (Robinson, 1991). We similarly specify a linear mixed model on each herd-specific association:

$$\varphi_j^{(u)} = \mathbf{x}_{4j}^{'}\boldsymbol{\gamma}_u + \mathbf{w}_j^{'}\mathbf{m}_u \qquad [6]$$

where $\boldsymbol{\gamma}_u$ represents a p_4 x 1 vector of unknown fixed effects with $\mathbf{x}_{4j}^{'}$ being a known row incidence vector. Furthermore, $\mathbf{m}_u$ represents a r x 1 vector of unknown random county-specific effects such that $\mathbf{m}_u \sim N\left(\mathbf{0}, \mathbf{I}\sigma_{m_u}^2\right)$ and with $\mathbf{w}_j^{'}$ being the corresponding known row incidence vector for herd j. Typically, $\mathbf{m}_u$ would represent random effects on broader classifications than that for $\mathbf{m}_e$ since the experimental unit (i.e., herd-year) in Equation [6] is larger than that (i.e., cow) in Equation [5]; subsequently $r < q$.

We also model the conditional cow level or residual variances $\sigma_{e_1,i}^2$ and $\sigma_{e_{2|1},i}^2$ as multiplicative functions of fixed and random effects (Foulley et al., 1990), expressing the logarithm of this relationship as follows:

$$\begin{aligned} \log\left(\sigma_{e_1,i}^2\right) &= \mathbf{x}_{5i}^{'}\log\left(\boldsymbol{\tau}_{e_1}\right) + \mathbf{z}_i^{'}\log\left(\mathbf{v}_{e_1}\right) \\ \log\left(\sigma_{e_{2|1},i}^2\right) &= \mathbf{x}_{6i}^{'}\log\left(\boldsymbol{\tau}_{e_{2|1}}\right) + \mathbf{z}_i^{'}\log\left(\mathbf{v}_{e_{2|1}}\right) \end{aligned} \qquad [7]$$

Here $\boldsymbol{\tau}_{e_1}$ and $\boldsymbol{\tau}_{e_{2|1}}$ represent $p_5 \times 1$ and $p_6 \times 1$ vectors of fixed effects, respectively, with $\mathbf{x}'_{5i}$ and $\mathbf{x}'_{6i}$ being known incidence vectors. Furthermore, $\mathbf{v}_{e_1} = \left\{ v_{e_{1,j}} \right\}_{j=1}^{q}$ and $\mathbf{v}_{e_{2|1}} = \left\{ v_{e_{2|1,j}} \right\}_{j=1}^{q}$ each represent $q \times 1$ vectors of unknown random herd-specific effects, specified with independent gamma priors $v_{e_1,i} \mid \eta_{e_1} \sim IG\left(\eta_{e_1}, \eta_{e_1} - 1\right)$ and $v_{e_{2|1},i} \mid \eta_{e_{2|1}} \sim IG\left(\eta_{e_{2|1}}, \eta_{e_{2|1}} - 1\right)$, as in Bello et al. (2010; Chapter 1). That is, $\mathrm{E}\left(v_{e_1,i} \mid \eta_{e_1}\right) = \mathrm{E}\left(v_{e_{2|1},i} \mid \eta_{e_{2|1}}\right) = 1$ with $\mathrm{var}\left(v_{e_1,i} \mid \eta_{e_1}\right) = \dfrac{1}{\eta_{e_1} - 2}$ and $\mathrm{var}\left(v_{e_{2|1},i} \mid \eta_{e_{2|1}}\right) = \dfrac{1}{\eta_{e_{2|1}} - 2}$ such that η_{e_1} and $\eta_{e_{2|1}}$ more or less function as unknown "variance components" to be estimated from the data.

Similarly, we model the conditional herd-specific random effects variances $\sigma^2_{u_1,j}$ and $\sigma^2_{u_{2|1},j}$ as multiplicative functions of fixed and random effects, expressing the logarithm of this relationship as:

$$\begin{aligned} \log\left(\sigma^2_{u_1,j}\right) &= \mathbf{x}'_{7j} \log\left(\boldsymbol{\tau}_{u_1}\right) + \mathbf{w}'_j \log\left(\mathbf{v}_{u_1}\right) \\ \log\left(\sigma^2_{u_{2|1},j}\right) &= \mathbf{x}'_{8j} \log\left(\boldsymbol{\tau}_{u_{2|1}}\right) + \mathbf{w}'_j \log\left(\mathbf{v}_{u_{2|1}}\right) \end{aligned} \qquad \textbf{[8]}$$

Here $\boldsymbol{\tau}_{u_1}$ and $\boldsymbol{\tau}_{u_{2|1}}$ represent $p_7 \times 1$ and $p_8 \times 1$ vectors of fixed effects, respectively, with $\mathbf{x}'_{7j}$ and $\mathbf{x}'_{8j}$ being known incidence vectors. Furthermore, $\mathbf{v}_{u_1} = \left\{ v_{u_{1,k}} \right\}_{k=1}^{r}$ and $\mathbf{v}_{u_{2|1}} = \left\{ v_{u_{2|1,k}} \right\}_{k=1}^{r}$ represent r x 1 vectors of unknown random county-specific effects, each which are specified with different independent inverted-gamma priors: $v_{u_1,k} \mid \eta_{u_1} \sim IG\left(\eta_{u_1}, \eta_{u_1} - 1\right)$ and $v_{u_{2|1},k} \mid \eta_{u_{2|1}} \sim IG\left(\eta_{u_{2|1}}, \eta_{u_{2|1}} - 1\right)$, as in Bello et al. (2010; Chapter 1). As with η_{e_1} and $\eta_{e_{2|1}}$, η_{u_1} and $\eta_{u_{2|1}}$ are unknown and need to be estimated from the data.

Note then that there are a total of 8 different "mixed model" specifications or submodels embedded in our hierarchical model; these submodels include the two classical specifications (one per trait) for location parameters in Equation [1], another in each of Equations [5] and [6], and two in each of Equation [7] and Equation [8]. We consider Equations [5] and [7] as representing the cow level modeling of residual associations and variances, whereas Equations [6] and [8] represent the herd-year level modeling of random effects associations and variances in our hierarchical model.

Prior specifications for all parameters were identical to those specified in the simulation study of Bello et al (2010; Chapter 1); i.e., flat unbounded priors were specified on $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\gamma}_e$, $\boldsymbol{\gamma}_u$, $\sigma^2_{m_e}$, $\sigma^2_{m_u}$, $\boldsymbol{\tau}_{e_1}$, $\boldsymbol{\tau}_{e_{2|1}}$, $\boldsymbol{\tau}_{u_1}$, and $\boldsymbol{\tau}_{u_{2|1}}$, whereas $\eta \sim p(\eta) \propto (1+\eta)^{-2}$ priors were specified on each of η_{e_1}, $\eta_{e_{2|1}}$, η_{u_1} and $\eta_{u_{2|1}}$ Furthermore, standard linear model restrictions are imposed on elements of the "fixed"

effects (β_1, β_2, γ_e, γ_u, τ_{e_1}, $\tau_{e_{2|1}}$, τ_{u_1}, and $\tau_{u_{2|1}}$) to ensure identifiability of the parameters following Bello et al. (2010; Chapter 1). The length of the MCMC chain for each model under investigation was 200,000 cycles after a burn-in period of 5,000 cycles. The only difference from the implementation presented in Bello et al. (2010; Chapter 1) is that the unobserved values for $\mathbf{y}_{22}^{*}$ are generated for each MCMC cycle using data augmentation (Tanner, 1993) because of right-censoring, such that MCMC samples for all other parameters are based on the augmented data $\mathbf{y}_2^{*} = \left[\mathbf{y}_{21}' \;\; \mathbf{y}_{22}^{*}{}'\right]'$. Convergence of the MCMC chain and sampling diagnostics were monitored graphically and following Raftery and Lewis (1992), as also in Bello et al. (2010; Chapter 1). We summarize posterior densities for each parameter of interest using posterior means, posterior standard deviations, and the 95% highest posterior density intervals (HPD). In addition, we report the effective sample size (ESS) as a measure of the number of effectively independent samples or Monte Carlo error amongst the 200,000 dependent MCMC samples (Sorensen et al., 1995).

Model selection

As in Bello et al. (2010; Chapter 1), we used the Deviance Information Criteria (DIC) (Spiegelhalter et al., 2002) as a measure of model fit to compare competing models. Smaller values of DIC indicate better fit, and, generally, DIC differences exceeding 7 are believed to indicate a decisive difference in model fit (Spiegelhalter et al., 2002). Table 1 lists the systematic fixed effects factors and covariates (i.e., management practices and herd attributes) that were considered for inclusion into the cow level and herd-year level of the hierarchical model. As indicated earlier, we also included

the random effects of herd-years for cow level modeling, as per Equations [5] and [7], and for the classical random effects $\mathbf{u}_1$ and $\mathbf{u}_2$ in Equation [1]. Similarly, the random effects of counties were specified for herd-year level modeling, as per Equations [6] and [8].

The classical fixed factors for modeling location parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ always included the effects of parity, calving season, year and individual bST treatment. As we were not specifically interested in $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ *per se*, all the aforementioned factors fitted on $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ were specified in all models to ensure inferences on other parameters were robust to model misspecification. That is, our primary objective was to identify sources of heterogeneity on the relationship between MY and CI (per Equations [5] and [6]) as well as on the variability (per Equations [7] and [8]) of the two traits in dairy cows.

Model selection was conducted in a forward stepwise manner, such that each factor and covariate was evaluated one at a time for model inclusion based on their contribution to model fit using DIC; this stepwise DIC strategy is similar to that implemented by Daniels and Zhao (2003). We started by selecting the best-fitting univariate models, one for each of MY and CI, as per Kizilkaya and Tempelman (2005), before investigating factors that influenced herd-year level or cow level associations between traits. For each trait, selection for factors and covariates influencing the cow level variances in Equation [7] and the herd-year level variances in Equation [8] consisted of four steps. The first step involved including the factor or covariate for cow level variance in the model that led to the largest decrease in DIC and restarting the process with respect to remaining factors or covariates until none led to a DIC decrease

of 7 or greater. The second step involved whether to include or not include random herd-year effects for cow level variance depending on a DIC decrease of less than or greater than 7 respectively. Steps 3 and 4 mirrored Steps 1 and 2 except they pertained to selecting the best fitting model for the herd-year level variance with the random effects being defined by county in that case. The procedural details for these steps and the resulting chosen factors are outlined in Tables 2 and 3 for MY and CI, respectively. It should also be noted that model selection for MY and CI were both based on the use of the first equation in each of Equations [7] and [8]; that is, there is obviously no conditioning on another trait in a series of univariate analyses.

The selected univariate models, as based on the aforementioned procedure, were then connected as a null bivariate model (i.e., based on only overall herd-year level and cow-level association specifications) to further investigate factors influencing cow level and herd-year level associations between MY and CI per Equations [5] and [6]. Table 4 outlines forward selection details and final outcomes using DIC (>7) on the selection of fixed and random effects on the cow level and herd-year level associations between MY and CI. We believe our model selection strategy on these associations is robust as inferences upon them are already conditioned upon important sources of heterogeneity on cow level and herd level variances, as per Tables 2 and 3.

In the final selected model, inferences were directed upon the marginal means for levels of each fixed effects factor expressed at the average covariate value for any significant covariate and averaging across levels of other fixed effects factors chosen for a particular submodel; i.e., in Equations [5],[6], [7], and [8]. In addition, these marginal means were exponentiated to represent variances on the observed scale per Equations [7]

and [8] as also in Bello et al. (2010; Chapter 1). These marginal means are analogous to the least squares means popularized, for example, in SAS linear models software (Milliken and Johnson, 2009 p. 226). We subsequently refer to the corresponding posterior means and posterior standard deviations as estimated means and their standard errors, respectively. Throughout the paper, statistical significance for fixed effects parameters was established based on whether or not the HPD of a difference between two parameters included 0 (in the case of parameters defining Equations [5] and [6]) or the ratio between two parameters included 1 (in the case of parameters defining Equations [7] and [8]). For any comparison of interest between two parameters, say generically θ_1 and θ_2, we also report the Bayesian *P*-value defined as: *P*-value $=2\times\min\left(\Pr\left(\theta_1-\theta_2\geq 0\,|\,\mathbf{y}\right),\Pr\left(\theta_1-\theta_2<0\,|\,\mathbf{y}\right)\right)$.

RESULTS AND DISCUSSION

Cow level and herd level associations between 305-d cumulative milk yield and calving interval: Inference on sources of heterogeneity.

Model selection based on stepwise DIC indicated strong evidence for milking frequency, year, calving season, and herd expansion being linked to the cow level association between 305-d cumulative MY and CI. Additionally, variability between herd-year clusters in the cow level associations was also evident (Table 4). Only bST supplementation was identified as a source of heterogeneous association at the herd-year level (Table 4). A summary of the marginal posterior inference for statistically

significant key parameters describing the association between 305-d cumulative MY and CI is shown in Table 5 and further described subsequently.

The nature of the herd-year level association between MY and CI differed with the bST supplementation strategy implemented in the herd (P<0.0001). In herds where most cows (>50% of the herd) were subjected to bST supplementation, CI decreased by an estimated 1.37±0.13 d for every 100 kg increase in MY, thereby yielding a strongly favorable association (Table 5). In contrast, the association between MY and CI was essentially null in herds that used bST in less than 50% of their cows or that did not use bST at all, as the corresponding HPD were highly concentrated around zero (Table 5). This observed difference between levels of bST usage may not necessarily reflect a direct role of the technology; rather, bST usage may be considered a proxy for general level of herd management. Indeed, successful adoption of bST technology necessitates a higher level of herd management (Bauman, 1992). Inclusion of bST into management may also entail, for example, targeted nutritional programs, proactive transition-cow management, standardized milking practices and frequent herd health evaluations. Furthermore, cows need to be injected with bST on a regular schedule, thus creating opportunities for additional surveillance of individual cows, early diagnosis of potential problems, and prompt attention when needed.

At the cow level, an overall antagonism was observed between MY and CI. Specifically, every 100 kg increase in MY translated into 0.51±0.01 d longer CI (95% HPD = [0.49, 0.53]), as indicated by posterior inference on $\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^{(2)\prime}\gamma_e$. This antagonistic association between MY and CI was alleviated by approximately 20%

among cows in $3^{+}X$ milking herds compared to cows in more traditional 2X schemes ($P<0.0001$; Table 5). The estimated association for cows in $3^{+}X$ milking herds was 0.45±0.02 d/100 kg compared to an estimate of 0.57±0.01 d/100 kg for cows in traditional 2X milking schemes. Adoption of management practices to enhance milk yield , such as thrice-a-day milking, is an ongoing trend in the US dairy industry (Ruegg, 2001). Therefore, as with bST supplementation, the improved MY-CI relationship observed among cows in $3^{+}X$ milking herds may not necessarily be attributed to a direct physiological effect of milking frequency. Instead, milking frequency may be yet another general indicator of a more specialized and intensive level of overall cow management. In support of this interpretation, our data support congruity in the adoption of specialized management practices, whereby herds involved with bST supplementation were up to 3 times more likely to implement $3^{+}X$ milking compared to non-bST herds (Chi-square test; $P<0.0001$).

During the year 2006, the estimated antagonism of the cow level association between MY and CI worsened by ~24% compared to the year 2005 ($P<0.0001$; Table 5). Investigation of the reasons for such marked differences between 2005 and 2006 should consider, among other factors, weather (e.g. ambient temperatures, rainfall), forage quality and conditions of the milk markeft, especially regarding cull cow prices and milk prices.

For cows calving during the summer, the estimated antagonism between MY and CI was alleviated by approximately 8 to 16% compared to cows calved during any other season ($P<0.05$; Table 5). These results may seem counterintuitive due to the well-

documented negative effects of heat stress on milk yield and fertility of dairy cows (Rensis and Scaramuzzi, 2003; West, 2003). We emphasize that the association parameter that quantifies the relationship between MY and CI does so in relative terms (i.e. change in days of CI per unit change in MY). Therefore, our finding of alleviated antagonism among Summer-calved cows may be indicative of disproportionately more pronounced and longer-lasting effects of heat stress on MY as compared to CI. Indeed, in Summer-calved cows, the classical least squares mean estimate of MY was lower by approximately 600 to 800 kg compared to cows calved during any other season ($P<0.0001$). In contrast, CI for Summer-calved cows was actually shortened (i.e., improved) by 12 to 19 days compared to other seasons ($P<0.0001$). Heat stress disturbs milk yield during the early rise and peak of lactation. Thus, the overall scale of the lactation curve is likely to be compromised, resulting in a smaller cumulative lactation yield. In contrast, most Summer-calved cows would most likely not be eligible for breeding until either late Summer or Fall, when ambient temperature and humidity are more moderate and would likely exert less of an effect on reproductive performance.

During the process of bivariate model selection (Table 4), herd expansion was selected by stepwise DIC as an important explanatory covariate on the cow level association between MY and CI. However, in the final model (after random herd-year effects were included), the 95% HPD of the corresponding parameter included zero, thereby indicating no evidence for a link between herd expansion and the MY-CI association ($P=0.22$).

Beyond the aforementioned statistically significant fixed effects, there was also evidence for the cow level association between MY and CI differing from herd-year to

herd-year groups. This inference was facilitated by our treating herd-year groups as random blocking effects rather than as fixed effects, as is typically the case in genetic studies. In doing so, we allowed for borrowing of information across herd-years (Tempelman, 2004), resulting in better precision estimates for herd-year effects (Tempelman, 2010). The magnitude of the variability between herd-years is quantified by $\sigma^2_{m_e}$ and was estimated as 0.030±0.005 (d per 100 kg)^2 (Table 5). This indicates that unidentified effects of management or environment still influence additional differences between herds in the cow level association between MY and CI. In order to provide a meaningful interpretation of $\sigma^2_{m_e}$, we consider the empirical rule (Ott and Longnecker, (2001). For normally distributed herd-specific effects ($\mathbf{m}_e$), one might anticipate the cow level association to have a range of ± 2 $\sqrt{\hat{\sigma}^2_{m_e}} = \pm 0.35$ or a span of 0.70 d/100 kg between the most extreme herds. Assuming then an average cow level association of ~0.5 d/100 kg across herds, as consistent with our study, herds might be expected to have within-herd associations ranging from 0.15 to 0.85 d of CI per 100 kg increase in MY. That is, in some herds, cows would be expected to display only a mildly unfavorable MY-CI association (0.15 d/100kg), whereas the situation could be considerably more adverse (0.85 d/100kg) in other herds; of course these ranges would shift further depending upon baseline fixed effects (e.g. 2X versus 3^+X milking frequency). Previous evidence supports this interpretation, whereby the association between milk production and reproductive performance was shown to vary from herd to herd (Windig et al., 2006; Windig et al., 2005) and the magnitude of this association depended upon the specific

herd environment (Castillo-Juarez et al., 2000). Thus, further investigation of additional management practices is warranted, as these may help further explain the significant variability ($\sigma^2_{m_e}$) for cow level production-reproduction association across herds. Other potential sources of differences between herds on the magnitude of the cow level association may include herd-specific disease prevalence (Emanuelson and Oltenacu, 1998), herd-specific differences in cow response to treatments (LeBlanc, 2008), criteria for allocating cows into management groups throughout lactation (Berry et al., 2003a; Tsuruta et al., 2009) and herd-specific success of synchronization programs (Stevenson et al., 2008), among others.

Overall, these results indicate that under intensive management conditions, the association between milk production and reproductive performance is favorable at the herd level and also, partially alleviated of its overall antagonism at the cow level. This apparent dichotomy between the cow- and herd-year levels of the production-reproduction association may be indicative of different mechanisms underlying the association within-herds and between-herds. That is, specialized intensive management practices that encourage higher herd productivity may also facilitate better than average herd reproductive performance simply by channeling opportunities for attentive and responsive observation of cows. Within those herds, however, cows with the highest milk yields still appear to be at greater physiological risk for reproductive failure. In particular, elevated steroid metabolism in the liver of high-producing cows has been proposed as a critical physiological mechanism underlying the antagonism between lactational and reproductive physiology of individual cows (Wiltbank et al., 2006). So, even though intensively-managed high producing herds may better manage reproduction and partially

alleviate the physiological conflict, an antagonistic association between milk yield and reproduction is still the net outcome among cows within a herd.

Variability in Milk Production and Reproductive Performance.

Tables 6 and 7 summarize the posterior inference for cow level and herd level variance components on MY and CI (conditional on MY), respectively. For both outcomes, the evidence for heterogeneity of variance components, namely heteroskedasticity, was substantial as both fixed effects and random effects affected the cow level variance components (as per Equation [7]) and the herd-year level variance components (as per Equation [8]), as noted previously in Tables 2 and 3.

Overall, MY was about 65% more variable among multiparous compared to primiparous cows ($P<0.0001$). Also, cow-to-cow variability in milk production differed between calving seasons ($P<0.01$); cows that calved in the fall had the most variable MY, whereas cows that calved in the spring had the most uniform MY. In addition, larger herd sizes were associated with increased cow-to-cow variability for milk yield, whereby the variance on MY increased by 69% as the number of cows per herd grew by a factor of 10 ($P<0.0001$; Table 6 and Figure 1). This is consistent with results from Tsuruta (2009), who reported that large herds were characterized by increased variability as well as greater means in level of milk productivity, thus reflecting a common scaling phenomenon. From a technical efficiency perspective, number of cows in the herd constitutes the number one determinant of mean productivity level (Cabrera et al., 2010), such that greater variability in large herds may be innately unavoidable. Evaluation of this scaling phenomenon was not an objective of this study and thus, was not examined

with our statistical model (i.e. herd size was not fitted within the classical fixed effects in $\boldsymbol{\beta}_1$, thus we cannot formally assess changes in mean MY as a function of herd size).

Evidence for heteroskedasticity in CI, conditional on MY, was substantial both at the cow and at the herd levels. Identifiable factors associated with CI heteroskedasticity included bST usage, year, calving season and geographic locations. Inference on heterogeneity of CI variance unconditional on MY, as derived from Equations (3) and (4), was of the same nature to the conditional case (not shown). Thus, for the purpose of consistency between parameters modeled and reported, conclusions regarding heteroskedasticity of CI are presented conditional on MY.

We determined that usage of bST was a two-pronged source of heteroskedasticity in the reproductive performance of dairy cows. First, bST treatment of individual cows was associated with twice as large variability in CI as compared to no bST treatment ($P<0.0001$). Second, herd level of bST supplementation was also identified as a source of conditional heteroskedasticity on CI, whereby cow-to-cow variability was decreased by half in herds with most cows (>50% of the herd) subjected to bST supplementation as compared to herds that either did not use bST or herds that did so only partially (i.e.<50% of the herd) ($P<0.0001$). We reconcile this inference (regarding level of herd bST usage) with the aforementioned result (on bST treatment of individual cows) as follows. Use of bST allows farmers to prolong the lactation of milking cows in a herd despite a non-pregnant reproductive status (Bauman, 1992). Individual cows under bST treatment then have additional opportunities to become pregnant over a longer period of time, thereby leading to greater variability in CI. In contrast, if not treated with bST, cows undergo a less persistent physiological decrease in milk production that naturally constrains

breeding to a considerably narrower window of time (i.e. less variable CI). Alternatively, on a herd level basis, consistent performance of cows subjected to high management standards, as implied by bST adoption, is a reasonable expectation.

The conditional variance in CI between herds (herd level) and within herds (cow level) was ~100% (P=0.03) and ~30% (P<0.0001) greater, respectively, in 2006 compared to 2005. Reproductive efficiency in a dairy farm is a long-term outcome that can be easily influenced by daily management decisions. The greater variation observed in reproductive performance for cows calving in 2006 may be partially explained by differential commercial priorities and breadth of strategies implemented by dairy farmers to deal with the challenging market conditions of the time (Thomas, 2006).

Variability in within-herd CI, conditional on MY, was significantly dependent upon calving season (Table 7). Cows that calved in Spring or Summer were more consistent in their subsequent CI compared to cows that calved during Fall and Winter (P<0.0001). Most Spring-calved cows will become eligible for breeding during Summer, at a time when fertility is impaired. Furthermore, Summer service rates are decreased due to typically low estrous detection rates (Rensis and Scaramuzzi, 2003) or perhaps, a conscious decision of the farm manager to withhold services during periods of intense heat stress. As a result, Spring-calved cows will naturally cluster for breeding with Summer-calved cows during early Fall, when heat stress is receding and longer intervals of postpartum clean-up (DIM) are more conducive to establishing a pregnancy. It is important to note, nevertheless, that these inferences on differences on variability are conditional on MY; given that season was also an important source of residual

heteroskedasticity for MY, this implies even greater marginal or unconditional differences in variability for CI between seasons when not adjusting for MY.

For both MY and CI (conditional on MY), differences between herds in the magnitude of the variation among their cows were considerable. This heterogeneity was modeled for each MY and CI by specifying the corresponding random herd-year specific effects with coefficient of variation (CV) $\sigma_{v,e_i} = \frac{1}{\sqrt{\eta_{e_i} - 2}}$ (Kizilkaya and Tempelman, 2005). The 95% HPD and posterior means for σ_{v,e_1} and $\sigma_{v,e_{2|1}}$ indicated that the CV of *within*-herd variances was significant and roughly 31% and 102% for MY and CI, respectively (Tables 6 and 7). Indeed, the cow level variances in MY for the most extremely variable herd was estimated to be 6.8 times greater than for the least variable herd, as per the ratio of the corresponding posterior mean variances, namely 3.39 and 0.50 respectively (expressed relative to a typically variable herd; = 1). For CI, again conditional on MY, the largest and smallest herd-specific relative variances, as per their posterior means, were 8.6 and 0.18, respectively, i.e., a ratio of 47. This is consistent with previous evidence on substantial variation between herds in their reproductive performance (Morton, 2010). This suggests that if the determining factors were identified and modified, it would be possible to attain important improvements in the consistency of dairy reproduction.

Herd level CI showed geographic patterns of conditional heteroskedasticity across Michigan counties. Figure 2 maps county-specific relative variances for herd CI in the 65 counties that were represented in the DHIA dataset out of a possible total of 83 Michigan counties. Generally, counties across the state had fairly consistent CI across herds (i.e.

average or below-average relative variation). However, there appeared to be pockets of substantially large between-herd reproductive performance along the Eastern and Southern Upper Peninsula and a few scattered counties in the Lower Peninsula of Michigan, where the relative variance between herds for those counties was up to 6 times greater than among herds in a typically variable county (reference = 1). Consistent herd reproductive management relies upon frequent and periodic pregnancy checks that allow for timely intervention in non-pregnant cows. These checks are usually conducted by food animal veterinarians. Shortage of food supply veterinary medicine professionals is a national problem recognized by the American Veterinary Science Association (Prince et al., 2006) and applies to private veterinary practice as well as the public, industrial and academic sectors. Insufficient availability of food veterinary professionals in locations distant from large population centers (e.g. Michigan Upper Peninsula) or alternatively, a temporary shortage in pocket areas throughout Michigan, may play a role in the observed geographic pattern of inconsistencies in herd reproductive performance. Further evaluation will be needed.

In summary, the evidence for heterogeneity of variances in dairy cow data is overwhelmingly strong and spans multiple management factors as well as dual hierarchical levels (i.e. cow and herd). We believe that heteroskedasticity of such magnitude calls for serious consideration of explicit modeling of variances as a standard procedure in dairy research. Disregarding such heteroskedasticity is likely to oversimplify inference and result in misleading implications in delineating guidelines for dairy herd management. A similar recommendation could be made in the context of meta-analysis studies, which appear to be increasingly popular in the current dairy

science literature. This is particularly so if studies used in a meta-analysis are inherently rather different from each other.

CONCLUSIONS

In this study, we consider the conflicting association between milk production and reproduction in Michigan dairy cows using recently developed hierarchical Bayesian technology (Bello et al., 2010; Chapter 1). We revealed that the nature of the production-reproduction association needs to be partitioned into components, whereby a favorable link among intensively-managed herds coexists with an overall antagonism among cows within herds. Moreover, management practices and unidentified herd-specific factors appear to be potential sources of heterogeneity in this association.

This study provides novel formal evidence that the concept of "one-size-fits-all" does not apply to the relationship between milk production and reproductive performance of dairy cows. Instead, it is apparent that milk yield and reproduction relate to each other in a complex multidimensional (i.e. cow- and herd-levels) and multifactorial manner that intertwines physiological mechanisms at the cow level with managerial decisions at the herd level. Given the statistical significance of management practices and random herd-specific effects, it would appear that the link between production and reproduction is, at least partially, manageable and can thus be altered or optimized. Indeed, the association between milk production and reproductive performance was enhanced, or at least alleviated of an overall antagonism under conditions of intensive management, as characterized by implementation of bST technology and increased milking frequency. More research will be needed to ascertain management scenarios under which milk

production and reproductive performance of dairy cows can be jointly optimized. For instance, it would be desirable to investigate interactions between fixed effects (i.e. management practices) as potential sources of heterogeneity in the association between MY and CI. As an example, one might be interested in assessing whether the effect of milking frequency differs between primiparous and multiparous cows in order to tailor management strategies to each group accordingly. An additive genetic component may also be of interest to evaluate a potentially inheritable constituent of the production-reproduction association (Berry et al., 2003b; Tsuruta et al., 2009). Additional extensions in the statistical methodology would be required to accommodate these questions.

Further investigation is needed to better understand the interplay of management and environmental factors that lead to consistency of dairy cow performance. An appreciation of the sources of heteroskedasticity for milk yield can provide insight in deriving strategies for consistent cash flow to the farmer and uniform input volume into the dairy processing industry. In turn, consistent reproductive performance is of main interest for long-term planning and investment of dairy enterprises, as it pertains for example to herd expansion.

IMPLICATIONS

This work supports adoption of specialized management practices (such as bST technology and milking frequency) towards more intensive production systems as a potential venue to jointly optimize milk production and reproductive performance of dairy cows. These same management strategies are also the foundation for sustainability in modern dairy farming. Indeed, technological management tools have been shown to

enhance efficiency of dairy production (Cabrera et al., 2010) in an environmentally friendly manner (Capper et al., 2008). Moreover, technology-driven production efficiency combined with controlled environmental impact will be critical to mitigating the food economic challenges of this century (Simmons, 2009) and thus should be given careful consideration.

Finally, our results suggest considerable complexity of dairy production systems, probably due to delicate interactions between the unique physiology of dairy cows and the variety of production management systems within which such physiology is managed. As we shift the paradigm in agriculture from single issues to a comprehensive systems approach, it becomes imperative to understand the individual and combined contributions of each management piece to the multifactorial nature of integrated performance of dairy cattle. The developing field of livestock production epidemiology provides a unique opportunity to so channel dynamic interactions between veterinary medicine, animal science and applied statistics towards this end. The potential outcomes of synergizing a thorough appreciation for physiological mechanisms in individual animals with an in-depth understanding of the structure and dynamics of livestock production systems from a strong quantitative foundation are compelling for a comprehensive understanding of complex biological systems such as the integral performance of dairy cows.

ACKNOWLEDGMENTS

We thank Dr. John Clay and his staff at the Dairy Records Management Systems, Raleigh, NC for providing the DHIA dataset used in this study. This study was partially funded by the Elwood Kirkpatrick Dairy Research Endowment, the Michigan Milk

Producers Association, the College of Agriculture and Natural Resources and the Department of Animal Science at Michigan State University.

Table 2.1. List of fixed effects (classification factors and linear regression on covariates) tested as explanatory variables for heterogeneity of cow and herd level (co)variances on milk yield and calving interval.

Cow level (co)variability
• Parity (Primi- vs. Multiparous)
• Calving season (Winter, Spring, Summer, Fall)
• Year (2005, 2006)
• Milking frequency (2 vs 3+ times per day)
• Individual cow treatment with bovine somatotropin during lactation (Yes/No)
• Level of herd supplementation with bovine somatotropin (0%, >0 to 50% and >50% of the herd)
• Reproductive management practices: Use of synchronization strategies (Yes/No)
• Herd size (number of head as covariate)
• Herd expansion (% change in herd size from preceding year as covariate)

Herd-year level (co)variability
• Calving season (Winter, Spring, Summer, Fall)
• Year (2005, 2006)
• Milking frequency (2 vs 3+ times per day)
• Level of herd supplementation with bovine somatotropin (0%, >0 to 50% and >50% of the herd)
• Reproductive management practices: Use of synchronization strategies (Yes/No)
• Herd size (number of head as covariate)
• Herd expansion (% change in herd size from preceding year as covariate)

Table 2.2. Sequential details of the forward model selection procedure implemented on variance components for a univariate model on cumulative milk production at 305-days-in-milk of Michigan dairy cows. Selection of fixed and random effects into the model was based on model fit as determined by Deviance Information Criteria (DIC).

	DIC difference	
Factors and covariates entering the model:	**Relative to Null Model**	**Relative to Model in Preceding Step**
Null Model, consisting of: • Fixed effects on the mean (parity, calving season, year and individual cow treatment with bovine somatotropin during lactation, as per Table 1); and • Random clustering effect of *herd-year* on the mean.	**0**	.
Step 1: Evaluation of fixed effects on the *cow-to-cow* (cow-level) variance		
1.1) Parity (Primi- vs. Multiparous)	-3474	-3474
1.2) Calving season (Winter, Spring, Summer, Fall)	-3601	-127
1.3) Herd size (number of heads)	-3685	-84
No additional effects entered the model	.	.
Step 2: Evaluation of random effects on the *cow-to-cow* (cow-level) variance		
2.1) Clustering effect of herd-year	-8581	-4896
Step 3: Evaluation of fixed effects on the variance between *herd-year* clusters (Herd level)		
No effect entered the model	.	.
Step 4: Evaluation of random effects on the variance between *herd-year* clusters (Herd level)		
No effect entered the model	.	.

Table 2.3. Sequential details of the forward model selection procedure implemented on variance components for a univariate model on calving interval of Michigan dairy cows. Selection of fixed and random effects into the model was based on model fit as determined by Deviance Information Criteria (DIC).

	DIC difference	
Factors and covariates entering the model:	**Relative to Null Model**	**Relative to Model in Preceding Step**
Null Model, consisting of: • Fixed effects on the mean (parity, calving season, year and individual cow treatment with bovine somatotropin (bST) during lactation, as per Table 1); and • Random clustering effect of *herd-year* on the mean.	**0**	.
Step 1: Evaluation of fixed effects on the *cow-to-cow* (cow-level) variance		
1.1) Individual cow treatment with bST (Yes/No)	-683	-683
1.2) Calving season (Winter, Spring, Summer, Fall)	-1112	-429
1.3) Level of herd supplementation with bovine somatotropin	-1142	-30
1.4) Year (2005, 2006)	-1151	-9
No additional effects entered the model	.	.
Step 2: Evaluation of random effects on the *cow-to-cow* (cow-level) variance		
2.1) Clustering effect of herd-year	-11434	-10283
Step 3: Evaluation of fixed effects on the variance between *herd-year* clusters (Herd-level)		
3.a) Year (2005, 2006)	-11448	-14
No additional effects entered the model	.	.
Step 4: Evaluation of random effects on the variance between *herd-year* clusters (Herd-level)		
4.a) County within Michigan	-11500	-52
No additional effects entered the model	.	.

Table 2.4. Sequential details of the forward model selection procedure implemented on the Cholesky-reparameterized covariances (expressed as regression coefficients) between cumulative 305-d milk yield and calving interval of Michigan dairy cows. Selection of fixed and random effects into the model was based on model fit as determined by Deviance Information Criteria (DIC).

	DIC difference	
Factors and covariates entering the model:	**Relative to Null Model**	**Relative to Model in Preceding Step**
Null Model, consisting of: • Univariate model on cumulative milk yield at 305-days-in-milk, as selected in Table 2. • Univariate model on calving interval, as selected in Table 3. • Covariances between traits are modeled as homogeneous and estimated accordingly.	0	.
Step 1: Evaluation of fixed effects on the *cow*-level regression coefficient		
1.1) Milking frequency (2 vs 3+ times per day)	-67	-67
1.2) Year (2005, 2006)	-106	-39
1.3) Calving season (Winter, Spring, Summer, Fall)	-135	-29
1.4) Herd expansion (% change in herd size from preceding year)	-154	-19
No additional effect entered the model	.	.
Step 2: Evaluation of random effects on the *cow*-level regression coefficient		
2.1) Clustering effect of herd-year	-397	- 243
No additional effects entered the model	.	.
Step 3: Evaluation of fixed effects on the *herd*-level regression coefficient		
3.1) Level of herd supplementation with bovine somatotropin	-409	-12
No effects entered the model	.	.
Step 4: Evaluation of random effects on the *herd*-level regression coefficient		
No effects entered the model	.	.

Table 2.5. Marginal mean estimates (PMEAN), standard errors (PSD), and 95% highest posterior density intervals (HPD) and effective sample sizes (ESS) for statistically significant cow-level and herd-level associations between cumulative milk yield at 305 days in milk (MY) and calving interval (CI) in Michigan dairy cows.

Association between MY and CI	PMEAN	PSD	95%HPD	ESS
Herd Level Associations				
Level of herd bST usage				
bST_{Herd} = 0%, d/100 kg	0.01 [x]	0.06	[-0.11, 0.12]	3,569
bST_{Herd} = >0-50%, d/100 kg	0.07 [x]	0.12	[-0.17, 0.31]	1,484
bST_{Herd} = >50%, d/100 kg	-1.37 [y]	0.13	[-1.63, -1.11]	1,892
Cow Level Associations				
Milking frequency				
2X, d/100 kg	0.57 [a]	0.01	[0.55, 0.60]	132,743
3^{+}X, d/100 kg	0.45 [b]	0.02	[0.41, 0.49]	173,621
Year				
2005, d/100 kg	0.46 [a]	0.02	[0.42, 0.49]	174,312
2006, d/100 kg	0.57 [b]	0.02	[0.53, 0.60]	134,345
Season				
Winter, d/100 kg	0.54 [c,e]	0.02	[0.50, 0.58]	152,327
Spring, d/100 kg	0.50 [c]	0.02	[0.47, 0.54]	159,199
Summer, d/100 kg	0.46 [d]	0.02	[0.42, 0.49]	158,110
Fall, d/100 kg	0.55 [e]	0.02	[0.51, 0.58]	151,897
Herd Expansion				
10% change in herd size, d/100 kg	-0.0082	0.0067	[-0.0213, 0.0049]	174,056
Variability between Herds				
$\sigma^2_{m_e}$,(d/100 kg)^2	0.030	0.005	[0.021, 0.039]	8,739

([x,y]) Letters indicate significant differences (P < 0.0001) between levels of the management factor on the *herd*-level regression parameter.

([a,b]) and ([c,d,e]) Letters indicate significant differences (P < 0.0001 and P < 0.05, respectively) between levels of each management factor on the *cow*-level regression parameter.

$\sigma^2_{m_e}$ defines random *herd*-specific heterogeneity on the *cow*-level regression coefficients.

Table 2.6. Posterior means (PMEAN), posterior standard deviations (PSD), 95% highest posterior density intervals (HPD) and effective sample size (ESS) for *cow-to-cow* (i.e. residual level) and *between-herd* (i.e. random level) variances for cumulative milk yield at 305 days-in-milk in Michigan dairy cows.

Variance Components $(100\ kg)^2$ †	PMEAN	PSD	95%HPD	ESS
Herd-Level Variances				
Between-herd variance, $(100\ kg)^2$	1,208	57	[1102, 1323]	172,743
Cow-Level (Cow-to-Cow) Variances				
Parity				
Primiparous, $(100\ kg)^2$	1,333 [a]	23	[1288, 1377]	3,486
Multiparous, $(100\ kg)^2$	2,199 [b]	37	[2125, 2271]	3,359
Season				
Winter, $(100\ kg)^2$	1,678 [c]	30	[1619, 1737]	3,998
Spring, $(100\ kg)^2$	1,598 [d]	28	[1543, 1656]	4,071
Summer, $(100\ kg)^2$	1,763 [e]	32	[1701, 1825]	3,780
Fall, $(100\ kg)^2$	1,817 [f]	36	[1754, 1881]	3,341
Herd Size				
10X change in herd size, $(100\ kg)^2$	1.69	0.05	[1.59,1.80]	2,478
Between Herds				
Coefficient of Variation σ_{v,e_1}	0.31	0.01	[0.29, 0.34]	21,662

([a, b]) and ([c, d, e, f]) Letters indicate significant differences ($P<0.0001$ and $P<0.01$) in *cow-to-cow* variation between levels of each management factor.

σ_{v,e_1} is the *cow*-level coefficient of variation for the conditional variance between herd-year clusters.

Table 2.7. Posterior means (PMEAN), posterior standard deviations (PSD), 95% highest posterior density intervals (HPD) and effective sample size (ESS) for *cow*-level (i.e. residual) and *herd*-level (i.e. random) conditional variances for calving interval in Michigan dairy cows.

Variance Components †	PMEAN	PSD	95%HPD	ESS
Herd-Level (Between-Herd) Variances				
Year				
2005, days^2	3,864 ^{x}	807	[2618, 5349]	6,213
2006, days^2	8,189 ^{y}	4,661	[3911, 14216]	16,208
Between Counties				
Coefficient of variation $\sigma_{v,u_{2\|1}}$	3.67	4.59	[0.78, 9.23]	7,229
Cow-Level (Cow-to-Cow) Variances				
Cow Treatment with bST				
No, days^2	11,560 ^{a}	445	[10704, 12454]	1,233
Yes, days^2	28,390 ^{b}	1257	[26019, 30955]	1,428
Season				
Winter, days^2	20,060 ^{a}	774	[18568, 21591]	1,366
Spring, days^2	16,450 ^{b}	634	[15241, 17717]	1,333
Summer, days^2	16,350 ^{b}	630	[15165, 17617	1,329
Fall, days^2	19,940 ^{a}	774	[18476, 21488]	1,219
Level of Herd bST usage				
bST_{Herd} = 0%, days^2	22,720 ^{a}	868	[21059, 24448]	2,002
bST_{Herd} = >0-50%, days^2	24,210 ^{a}	1,510	[21353, 27254]	1,730
bST_{Herd} => 50%, days^2	10,830 ^{b}	697	[9468, 12181]	447
Year				
2005, days^2	16,000 ^{a}	687	[14672, 17351]	1,976
2006, days^2	20,510 ^{b}	875	[18805, 22222]	957
Between Herds				
Coefficient of variation $\sigma_{v,e_{2\|1}}$	1.03	0.09	[0.87, 1.20]	2,935

($^{x,\ y}$) Letters indicate significant differences (P=0.03) in the *between-herd* variance between levels of the management factor.

($^{a,\ b,\ c,\ d}$) Letters indicate significant differences (P<0.0001) in the *cow-to-cow* variance between levels of each management factor.

$\sigma_{v,u_{2|1}}$ is the *herd*-level coefficient of variation for the conditional variances between counties.

$\sigma_{v,e_{2|1}}$ is the *cow*-level coefficient of variation for the conditional variances between herd-year clusters.

Figure 2.1: Cow-to-cow variance estimates (black line) for cumulative milk yield at 305-days-in-milk in Michigan dairy cows expressed as a function of herd size (in the log base 10 scale along the x-axis) and prediction for herd-specific cow-to-cow variances (grey dots).

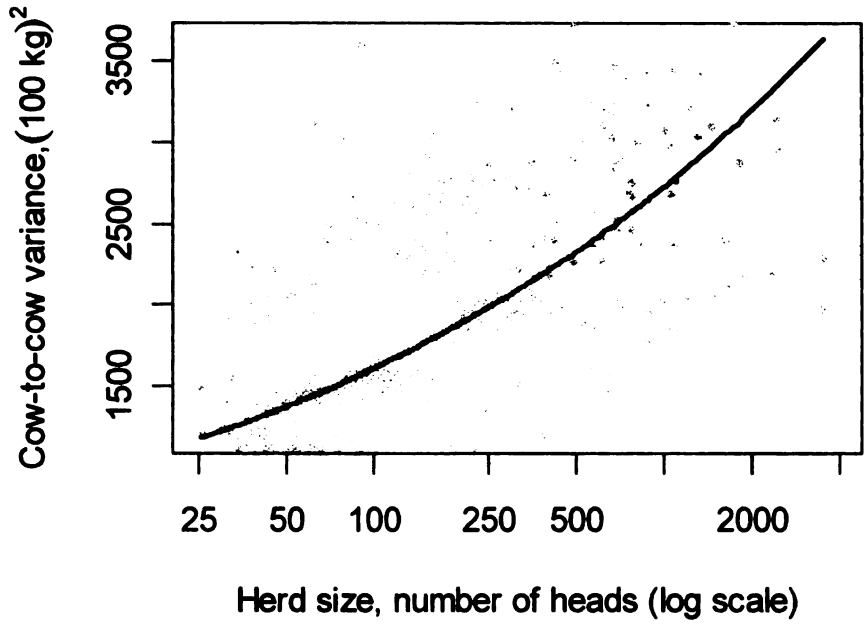
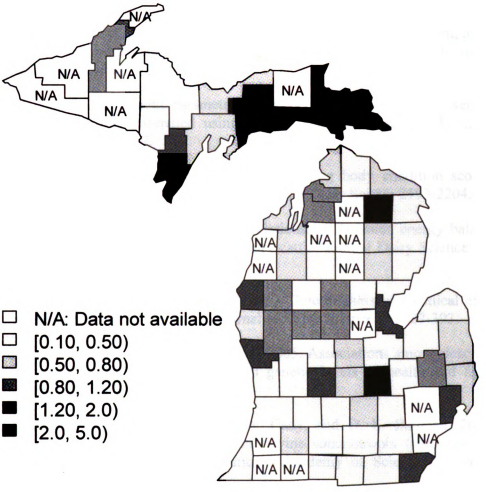


Figure 2.2: County map of Michigan representing county-specific *between*-herd variances in calving interval, relative to a typical county variance (reference = 1).



References

Bauman, D. E. 1992. Bovine somatotropin: Review of an emerging animal technology. Journal Dairy Science 75: 3432-3451.

Bello, N. M., J. P. Steibel, and R. J. Tempelman. 2010. Hierarchical modeling of random and residual variance-covariance matrices in bivariate mixed effects models. Biometrical Journal. 52: 297–313.

Berry, D. P. et al. 2003a. Genetic parameters for body condition score, body weight, milk yield, and fertility estimated using random regression models. Journal Dairy Science 86: 3704-3717.

Berry, D. P. et al. 2003b. Genetic relationships among body condition score, body weight, milk yield, and fertility in dairy cows. J Dairy Sci 86: 2193-2204.

Butler, W. R., and R. D. Smith. 1989. Interrelationships between energy balance and postpartum reproductive function in dairy cattle. Journal Dairy Science 72: 767-783.

Cabrera, V. E., D. Solis, and J. del Corral. 2010. Determinants of technical efficiency among dairy farms in wisconsin. Journal of Dairy Science 93: 387-393.

Calus, M. P., J. J. Windig, and R. F. Veerkamp. 2005. Associations among descriptors of herd management and phenotypic and genetic levels of health and fertility. J Dairy Sci 88: 2178-2189.

Capper, J. L., E. Castaneda-Gutierrez, R. A. Cady, and D. E. Bauman. 2008. The environmental impact of recombinant bovine somatotropin (rbst) use in dairy production. Proceedings of the National Academy of Sciences of the United States of America 105: 9668-9673.

Castillo-Juarez, H., P. A. Oltenacu, R. W. Blake, C. E. Mcculloch, and E. G. Cienfuegos-Rivas. 2000. Effect of herd environment on the genetic and phenotypic relationships among milk yield, conception rate, and somatic cell score in holstein cattle. Journal of Dairy Science 83: 807-814.

Daniels, M. J., and Y. D. Zhao. 2003. Modelling the random effects covariance matrix in longitudinal data. Statistics in Medicine 22: 1631-1647.

Emanuelson, U., and P. A. Oltenacu. 1998. Incidences and effects of diseases on the performance of swedish dairy herds stratified by production. Journal Dairy Science 81: 2376-2382.

Foulley, J. L., D. Gianola, M. S. Cristobal, and S. Im. 1990. A method for assessing extent and sources of heterogeneity of residual variances in mixed linear-models. Journal of Dairy Science 73: 1612-1624.

Hare, E., H. D. Norman, and J. R. Wright. 2006. Trends in calving ages and calving intervals for dairy cattle breeds in the united states. Journal Dairy Science 89: 365-370.

Kizilkaya, K., and R. J. Tempelman. 2005. A general approach to mixed effects modeling of residual variances in generalized linear mixed models. Genetic Selection and Evolution 37: 31-56.

Laben, R. L., R. Shanks, P. J. Berger, and A. E. Freeman. 1982. Factors affecting milk-yield and reproductive-performance. Journal of Dairy Science 65: 1004-1015.

LeBlanc, S. J. 2008. Postpartum uterine disease and dairy herd reproductive performance: A review. Vet J 176: 102-114.

Lof, E., H. Gustafsson, and U. Emanuelson. 2007. Associations between herd characteristics and reproductive efficiency in dairy herds. Journal Dairy Science 90: 4897-4907.

Lopez-Gatius, F. et al. 2006. Screening for high fertility in high-producing dairy cows. Theriogenology 65: 1678-1689.

Lucy, M. C. 2001. Reproductive loss in high-producing dairy cattle: Where will it end? Journal of Dairy Science 84: 1277-1293.

Miller, R. H., H. D. Norman, M. T. Kuhn, J. S. Clay, and J. L. Hutchison. 2007. Voluntary waiting period and adoption of synchronized breeding in dairy herd improvement herds. Journal Dairy Science 90: 1594-1606.

Milliken, G. A., and D. E. Johnson. 2009. Analysis of messy data - volume 1: Designed experiments. Second ed. Chapman and Hall/CRC Press.

Morton, J. M. 2010. Interrelationships between herd-level reproductive performance measures based on intervals from initiation of the breeding program in year-round and seasonal calving dairy herds. Journal of Dairy Science 93: 901-910.

Norman, H. D., J. R. Wright, S. M. Hubbard, R. H. Miller, and J. L. Hutchison. 2009. Reproductive status of holstein and jersey cows in the united states. Journal of Dairy Science 92: 3517-3528.

Ott, R. L., and M. Longnecker. 2001. An introduction to statistical methods and data analysis. Fifth ed. Duxbury, United States of America.

Prince, J. B., D. M. Andrus, and K. P. Gwinner. 2006. Future demand, probable shortages, and strategies for creating a better future in food supply veterinary medicine. Journal of the American Veterinary Medicine Association 229: 57-69.

Raftery, A. E., and S. Lewis (Editors). 1992. How many iterations in the gibbs sampler? Bayesian statistics. Oxford University Press, Oxford, UK, 763-773 pp.

Rensis, F. D., and R. J. Scaramuzzi. 2003. Heat stress and seasonal effects on reproduction in the dairy cow--a review. Theriogenology 60: 1139-1151.

Robinson, G. K. 1991. That blup is a good thing - the estimation of random effects. Statistical Science 6: 15-51.

Ruegg, P. L. 2001. Health and production management in dairy herds. In: O. M. Radostis (ed.) Herd health: Food animal production medicine. W.B. Saunders Company.

Simmons, J. 2009. Food economics and consumer choice.

Sorensen, D., and D. Gianola. 2002. Likelihood, bayesian and mcmc methods in quantitative genetics. Springer Verlag, New York.

Sorensen, D. A., S. Andersen, D. Gianola, and I. Korsgaard. 1995. Bayesian-inference in threshold models using gibbs sampling. Genetics Selection Evolution 27: 229-249.

Sorensen, D. A., D. Gianola, and I. R. Korsgaard. 1998. Bayesian mixed-effects model analysis of a censored normal distribution with animal breeding applications. Acta Agriculturae Scandinavica Section a-Animal Science 48: 222-229.

Spalding, R. W., R. W. Everett, and R. H. Foote. 1974. Fertility in new york artificially inseminated holstein herds in dairy herd improvement. J Dairy Sci 58: 718-723.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society - Series B 64: 1-34.

Stevenson, J. S. et al. 2008. Detection of anovulation by heatmount detectors and transrectal ultrasonography before treatment with progesterone in a timed insemination protocol. J Dairy Sci 91: 2901-2915.

Tanner, M. A. 1993. Tools for statistical inference. Springer-Verlag.

Tempelman, R. J. 2004. Experimental design and statistical methods for classical and bioequivalence hypothesis testing with an application to dairy nutrition studies. Journal of Animal Science 82 E-Suppl: E162-172.

Tempelman, R. J. 2010. Addressing scope of inference for global genetic evaluation of livestock Proceedings of the 47th Meeting of the Brazilian Animal Science Society, Salvador, Bahia, Brazil.

Thomas, C. 2006. Dairy market outlook report - monthly updates since 1999. Technical bulletin published by Michigan State University Extension: http://www.dairyteam.msu.edu/.

Tsu

Wes

Wilh

Win

Win

Tsuruta, S., I. Misztal, C. Huang, and T. J. Lawlor. 2009. Bivariate analysis of conception rates and test-day milk yields in holsteins using a threshold-linear model with random regressions. Journal Dairy Science 92: 2922-2930.

West, J. W. 2003. Effects of heat-stress on production in dairy cattle. Journal of Dairy Science 86: 2131-2144.

Wiltbank, M., H. Lopez, R. Sartori, S. Sangsritavong, and A. Gumen. 2006. Changes in reproductive physiology of lactating dairy cows due to elevated steroid metabolism. Theriogenology 65: 17-29.

Windig, J. J., M. P. Calus, B. Beerda, and R. F. Veerkamp. 2006. Genetic correlations between milk production and health and fertility depending on herd environment. Journal Dairy Science 89: 1765-1775.

Windig, J. J., M. P. Calus, and R. F. Veerkamp. 2005. Influence of herd environment on health and fertility and their relationship with milk production. Journal Dairy Science 88: 335-347.

CHAPTER 3

Hierarchical Bayesian Modeling of Heterogeneous Cluster and Subject Level Associations between Continuous and Binary Outcomes

SUMMARY. The augmentation of categorical outcomes with underlying Gaussian variables in bivariate generalized mixed effects models has facilitated the joint modeling of continuous and binary response variables. These models typically assume that random effects and residual effects (co)variances are homogeneous across all clusters and subjects, respectively. However, it seems likely in certain situations that these dispersion parameters may themselves be affected by systematic effects. We propose a hierarchical Bayesian extension of bivariate generalized linear models whereby (co)variances are specified as linear combinations of fixed and random effects following a square-root free Cholesky reparameterization that relaxes traditional positive semi-definite constraints on the reparameterized (co)variances. Using MCMC-based inference, we test the proposed model by simulation and apply it to a dairy cattle dataset in which the random and residual effects (co)variances between milk production and fertility of dairy cows are modeled as functions of fixed effects, as defined by management factors, as well as random cluster effects.

KEY WORDS: Bayesian; Bivariate Model; Cholesky decomposition; Generalized linear mixed model; Heterogeneous covariance.

1. Introduction

Joint generalized linear modeling of mixed outcomes has been of sustained interest in agricultural and biomedical research, with the bivariate model for a continuous and a binary response being of particular interest (Liu, Daniels and Marcus, 2009; O'Malley, Normand and Kuntz, 2003; Tsuruta et al., 2009; Wu, Gianola and Weigel, 2009). Due to non-zero covariances between outcomes, joint modeling of this nature permits information to be shared between Gaussian and non-Gaussian responses, thereby providing greater inferential efficiency for location parameters (i.e., treatment effects) than separate analyses for each outcome (Riley et al., 2007; Teixeira-Pinto and Normand, 2009). This is particularly true for treatment or risk factors specified for binary outcomes (Gueorguieva and Agresti, 2001; McCulloch, 2008). However, a prevailing, and potentially limiting, assumption of these models is that the variance-covariance structure is homogeneous across treatments or risk factors.

In some disciplines, parameters that specify the covariance or association between continuous and binary outcomes may be of equal or even greater interest than conventional treatment effects or risk factor effects on each outcome. This is particularly true in quantitative genetics whereby bivariate generalized linear mixed models (GLMM) are used to investigate associations at two or more levels; that is, between random genetic or cluster effects and between residual effects for the two outcomes or traits (Janss and Foulley, 1993; Tsuruta et al., 2009). An overriding motivation for these investigations into between-trait associations is based on agricultural sustainability; e.g., what implications do increasing meat and milk production, generally continuous outcomes, have for fitness and reproductive performance, generally binary outcomes? Along this

line of thought, our work is motivated by concerns that increasing levels of milk production per cow may have unfavorable implications for fertility of dairy cattle over time (Hare, Norman and Wright, 2006; Lucy, 2001). However, the associations between milk production and reproduction in dairy cattle have not always been antagonistic; sometimes neutral (i.e., zero covariances) or even favorable associations have been reported (Emanuelson and Oltenacu, 1998; Lof, Gustafsson and Emanuelson, 2007; Lopez-Gatius et al., 2006). Given such conflictive results between studies, we investigate whether the between-trait associations may depend upon various systematic effects by hierarchically extending the model for additional generalized linear model specifications. Based on the underlying data hierarchy, we specify the between-trait associations on random cluster effects (e.g., herds) and the between-trait associations on residual subject effects (e.g., cows) separately, recognizing that these relationships may be quite different and even opposite in sign (Bello, Steibel and Tempelman, 2010).

We recently developed a Bayesian procedure for modeling random cluster and residual subject effect associations between two continuous traits based on a multivariate Gaussian likelihood specification (Bello et al., 2010), whereby a linear model was implemented on parameters derived from a square-root free Cholesky decomposition of (co)variances for random cluster-specific effects and for residual subject-specific effects. In this paper, we further extend this methodology for the joint analysis of a continuous and a binary outcome, recognizing that GLMM extensions for the Bayesian analysis of binary responses may require additional care and study, particularly as it pertains to identifiability of parameters (Teixeira-Pinto, A., and Normand, 2009) and the potential impact of vaguely specified prior densities (Natarajan and Kass, 2000).

The objectives of this article are to present, validate, and demonstrate a hierarchical Bayesian extension of a bivariate GLMM in which random cluster and residual subject variance-covariance matrices are, in turn, modeled as functions of *fixed* and *random* effects. Although Bayesian inference implies a probability distribution on all unknown parameters, such that they are all genuinely random effects, a Bayesian interpretation of a fixed effects factor might be one where each of its effects is specified with independent non-informative or vaguely informative prior distributions with known hyperparameters (Bello et al., 2010; Sorensen and Gianola, 2002). Conversely, we characterize those factors whose levels could be considered to be exchangeable as random effects, which are specified by a structural prior whose hyperparameters are estimated from the data.

The article is organized as follows. We review the bivariate linear/probit mixed effects model in Section 2, reparameterize the random (cluster) and residual (subject) effects variance-covariance matrices into readily interpretable conditional variance and unconstrained association parameters in Section 3, and describe the MCMC-based hierarchical Bayesian implementation of the proposed bivariate GLMM in Section 4. Section 5 presents alternative interpretations of the association parameters in the observed scale. We validate our proposed method using an extensive simulation study in Section 6 and illustrate its application on a dairy cattle dataset in Section 7. Section 8 presents further discussion, followed by brief conclusions in Section 9.

2. The Bivariate Generalized Linear Mixed Model

Let y_{1j} be the observed continuous response and y_{2j} be the observed binary categorical outcome on subject j; $j=1,\ldots,n$. As elucidated in Albert and Chib (1993) and

G

ur

re

G

Ha

u

res

ve

th

cl

in

su

ou

wi

wit

Gueorguieva and Agresti (2001) amongst others, y_{2j} is assumed to be determined by an underlying normally distributed variable y_{2j}^{*} such that $y_{2j} = I\left(y_{2j}^{*} > 0\right)$ for $I(.)$ representing the indicator function. This is equivalent to specifying a probit link in a GLMM for y_{2j}. The underlying bivariate GLMM is then written as follows:

$$\begin{bmatrix} y_{1j} \\ y_{2j}^{*} \end{bmatrix} = \begin{bmatrix} \mu_{1j} + e_{1j} \\ \mu_{2j} + e_{2j} \end{bmatrix} \quad \text{where} \quad \begin{bmatrix} \mu_{1j} \\ \mu_{2j} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{1j}'\boldsymbol{\beta}_1 + \mathbf{z}_j'\mathbf{u}_1 \\ \mathbf{x}_{2j}'\boldsymbol{\beta}_2 + \mathbf{z}_j'\mathbf{u}_2 \end{bmatrix}. \tag{1}$$

Here, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are vectors of classical fixed effects whereas $\mathbf{u}_1 = \{u_{1k}\}_{k=1}^{q}$ and $\mathbf{u}_2 = \{u_{2k}\}_{k=1}^{q}$ are vectors of classical random effects for the two outcomes, respectively, for each of q clusters. Also, $\mathbf{x}_{1j}'$, $\mathbf{x}_{2j}'$ and $\mathbf{z}_j'$ are known incidence vectors specific to subject j whereas e_{1j} and e_{2j} are random residual effects unique to the jth subject. To simplify presentation, we assume that the same single random clustering factor is common to both traits and is the basis for all random effects modeling in this paper; i.e., $\mathbf{z}_j'$ is the same for both responses in Equation (1) and for all subsequent random effects specifications. Furthermore, we assume that complete pairs of outcomes are available on all n subjects. However, neither assumption is a restriction with the use of our method.

Independent prior bivariate Gaussian densities, $f\left(\mathbf{u}_{.k} \mid \mathbf{G}_k\right)$ and $f\left(\mathbf{e}_{.j} \mid \mathbf{R}_j\right)$, both with null means, are respectively specified on each cluster-specific pair of random effects

$\mathbf{u}_{.k} = \left[u_{1k} \;\; u_{2k}\right]'$ and on each subject-specific pair of residuals $\mathbf{e}_{.j} = \left[e_{1j} \;\; e_{2j}\right]'$ for the two outcomes, with (co)variances defined by

$$\mathbf{G}_k = \begin{bmatrix} \sigma^2_{u_1k} & \sigma_{u_{12}k} \\ \sigma_{u_{12}k} & \sigma^2_{e_2k} \end{bmatrix} \text{ and } \mathbf{R}_j = \begin{bmatrix} \sigma^2_{e_1j} & \sigma_{e_{12}j} \\ \sigma_{e_{12}j} & \sigma^2_{e_2j} \end{bmatrix}. \tag{2}$$

Note that random effects variances and covariances are unique to each cluster k as are residual variances and covariances to each subject j.

For a bivariate GLMM, the joint density $f\left(\mathbf{y}_1, \mathbf{y}_2^* \mid \boldsymbol{\beta}_1, \mathbf{u}_1, \boldsymbol{\beta}_2, \mathbf{u}_2, \left\{\mathbf{R}_j\right\}_{j=1}^n\right)$ for the *complete* data, including $\mathbf{y}_1 = \left\{y_{1j}\right\}_{j=1}^n$ and the augmented data $\mathbf{y}_2^* = \left\{y_{2j}^*\right\}_{j=1}^n$, can be written as the *factorization* or product of the marginal density for $\mathbf{y}_1$:

$$f\left(\mathbf{y}_1 \mid \boldsymbol{\beta}_1, \mathbf{u}_1, \left\{\sigma^2_{e_1j}\right\}_{j=1}^n\right) = \prod_{j=1}^n N\left(\mu_{1j}, \sigma^2_{e_1j}\right) \tag{3}$$

with the density function of $\mathbf{y}_2^*$ conditioned on $\mathbf{y}_1$:

$$f\left(\mathbf{y}_2^* \mid \mathbf{y}_1, \boldsymbol{\beta}_1, \mathbf{u}_1, \boldsymbol{\beta}_2, \mathbf{u}_2, \left\{\varphi_j^{(e)}\right\}_{j=1}^n\right) = \prod_{j=1}^n N\left(\mu_{2|1j}, \sigma^2_{e_{2|1}j} = 1\right) \tag{4}$$

(Catalano and Ryan, 1992; Janss and Foulley, 1993). Here, $\mu_{2|1j} = \mu_{2j} + \varphi_j^{(e)} e_{1j}$ with $\sigma^2_{e_{2|1}j} = \sigma^2_{e_2j} - \left(\varphi_j^{(e)}\right)^2 \sigma^2_{e_1j}$ being constrained to $\sigma^2_{e_{2|1}j} = 1 \; \forall j$ to ensure identifiability of the remaining parameters, as necessary for a probit model specification

on y_{2j}. Furthermore, $\varphi_j^{(e)} = \sigma_{e_{12}j} / \sigma_{e_1 j}^2$ represents the residual subject level association coefficient for e_{2j} on e_{1j}. In other words, the probability for a success outcome on the binary trait conditional on the continuous trait can be written as $\Pr\left(y_{2j} = 1 \mid y_{2j}^*, y_{1j}\right) = \Pr\left(y_{2j}^* > 0 \mid y_{1j}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{u}_1, \mathbf{u}_2, \mathbf{R}_j\right) = \Phi\left(\mu_{2|1j}\right)$ with $\Phi(.)$ defining the standard normal cdf. Then, the distribution of the observed outcome $\mathbf{y}_2$ conditioned on $\mathbf{y}_1$ is specified as:

$$f\left(\mathbf{y}_2 \mid \mathbf{y}_1, \boldsymbol{\beta}_1, \mathbf{u}_1, \boldsymbol{\beta}_2, \mathbf{u}_2, \left\{\varphi_j^{(e)}\right\}_{j=1}^n\right) = \prod_{j=1}^n \left(\Phi\left(\mu_{2|1j}\right)\right)^{y_{2j}} \left(1 - \Phi\left(\mu_{2|1j}\right)\right)^{\left(1-y_{2j}\right)} \tag{5}$$

3. Reparameterization of Variances and Covariances

The factorization presented in Equations (3) and (4) is equivalent to a square-root-free Cholesky decomposition for $\mathbf{R}_j$ and which we also use for $\mathbf{G}_k$ (Bello et al., 2010); i.e.,

$$\mathbf{R}_j = \begin{bmatrix} \sigma_{e_1 j}^2 & \varphi_j^{(e)} \sigma_{e_1 j}^2 \\ \varphi_j^{(e)} \sigma_{e_1 j}^2 & 1 + \left(\varphi_j^{(e)}\right)^2 \sigma_{e_1 j}^2 \end{bmatrix}$$

$$\text{and} \quad \mathbf{G}_k = \begin{bmatrix} \sigma_{u_1 k}^2 & \varphi_k^{(u)} \sigma_{u_1 k}^2 \\ \varphi_k^{(u)} \sigma_{u_1 k}^2 & \sigma_{u_{2|1} k}^2 + \left(\varphi_k^{(u)}\right)^2 \sigma_{u_1 k}^2 \end{bmatrix}. \tag{6}$$

Note that $\varphi_k^{(u)} = \sigma_{u_{12}k} / \sigma_{u_1 k}^2$ represents the random cluster level association of u_{2k} on u_{1k}, such that $u_{2k} = \varphi_k^{(u)} u_{1k} + u_{2|1k}$ where $u_{2|1k} \sim N\left(0, \sigma_{u_{2|1}k}^2\right)$ is conditionally independent of u_{1k}. Similarly, $e_{2j} = \varphi_j^{(e)} e_{1j} + e_{2|1j}$ where $e_{2|1j} \sim N(0,1)$ is conditionally independent of e_{1j}. We thereby rewrite the linear model for y_{2j}^* in Equation (1) as:

$$y_{2j}^* = \mathbf{x}_{2j}'\boldsymbol{\beta}_2 + \mathbf{z}_j'\left(\boldsymbol{\Psi}^{(u)}\mathbf{u}_1 + \mathbf{u}_{2|1}\right) + \varphi_j^{(e)} e_{1j} + e_{2|1j} \tag{7}$$

where $\mathbf{u}_{2|1} = \left\{u_{2|1k}\right\}_{k=1}^{q}$ and $\boldsymbol{\Psi}^{(u)}$ is a q x q diagonal matrix with elements $\left\{\varphi_k^{(u)}\right\}_{k=1}^{q}$.

Note that $\varphi_j^{(e)}$ and $\varphi_k^{(u)}$, as well as the logarithms of $\sigma_{e_1 j}^2$, $\sigma_{u_1 k}^2$, and $\sigma_{u_{2|1}k}^2$ can be completely unconstrained, yet $\mathbf{R}_j$ and $\mathbf{G}_k$ will still guarantee to be positive semi-definite. Hence, this decomposition facilitates modeling each of these 5 components as linear functions of covariates (Pourahmadi, 1999). Consider, in particular, three special cases: 1) $\varphi_j^{(e)} \to 0$, 2) $\varphi_k^{(u)} \to 0$, and 3) $\sigma_{u_{2|1}k}^2 \to 0$ and $\varphi_k^{(u)} = \varphi^{(u)}\ \forall\ k$. For Case 1), this defines $\sigma_{e_{12}j}$ = 0 such that residual effects for the two outcomes within subject j are independent. Similarly for Case 2), this defines $\sigma_{u_{12}k}$ = 0 such that random effects

for the two outcomes on cluster k are independent. For Case 3), this equates to specifying that both outcomes share a common random effect or latent variable for each cluster with only a difference in scale: $u_{2k} = \varphi^{(u)} u_{1k}$; this is a popular specification in bivariate mixed outcome models (Catalano and Ryan, 1992; Teixeira-Pinto and Normand, 2009). By further specifying heterogeneity of these parameters, each with their own mixed effects submodels (see next section), our proposed methodology has the potential to enhance modeling flexibility beyond many existing parametric bivariate continuous-binary approaches.

4. Heterogeneous (Co)Variance Modeling

We specify linear (mixed) models on $\varphi_j^{(e)}$ and $\varphi_k^{(u)}$ as follows:

$$\varphi_j^{(e)} = \mathbf{x}_{3j}'\gamma_e + \mathbf{z}_j'\mathbf{m}, \tag{8}$$

$$\varphi_k^{(u)} = \mathbf{x}_{4k}'\gamma_u. \tag{9}$$

Here γ_e and γ_u represent vectors of unknown fixed effects with subjectively specified priors $f(\gamma_e)$ and $f(\gamma_u)$. Conversely, $\mathbf{m} = \{m_k\}_{k=1}^{q}$ represents a vector of unknown cluster-specific random effects with a structural Gaussian prior $f\left(\mathbf{m} \mid \sigma_m^2\right) = N\left(\mathbf{0}, \mathbf{I}\sigma_m^2\right)$ such that, in turn, an inverted gamma (*IG*) prior distribution $f\left(\sigma_m^2\right) = IG\left(\alpha_m, \beta_m\right)$, with known α_m and β_m, is specified on σ_m^2 (Bello et al.,

2010); i.e. $\mathrm{E}\left(\sigma_m^2\right) = \beta_m / \left(\alpha_m - 1\right)$. Finally, $\mathbf{x}_{3j}^{'}$ and $\mathbf{x}_{4k}^{'}$ are known row incidence vectors specific to the jth subject and the kth cluster, respectively.

We also model the logarithm of variance components $\sigma_{e_1 j}^2$, $\sigma_{u_1 k}^2$ and $\sigma_{u_{2|1} k}^2$ as separate linear combinations of fixed and/or random effects (Foulley et al., 1990) as follows:

$$\log\left(\sigma_{e_1 j}^2\right) = \mathbf{x}_{5j}^{'} \log\left(\boldsymbol{\tau}_{e_1}\right) + \mathbf{z}_j^{'} \log\left(\mathbf{v}_{e_1}\right), \tag{10}$$

$$\log\left(\sigma_{u_1 k}^2\right) = \mathbf{x}_{6k}^{'} \log\left(\boldsymbol{\tau}_{u_1}\right), \text{ and} \tag{11}$$

$$\log\left(\sigma_{u_{2|1} k}^2\right) = \mathbf{x}_{7k}^{'} \log\left(\boldsymbol{\tau}_{u_{2|1}}\right). \tag{12}$$

Here, $\boldsymbol{\tau}_{e_1}$, $\boldsymbol{\tau}_{u_1}$ and $\boldsymbol{\tau}_{u_{2|1}}$ represent vectors of fixed effects with subjective priors $f\left(\boldsymbol{\tau}_{e_1}\right)$, $f\left(\boldsymbol{\tau}_{u_1}\right)$, and $f\left(\boldsymbol{\tau}_{u_{2|1}}\right)$, respectively. Conversely, $\mathbf{v}_{e_1} = \left\{v_{e_1 k}\right\}_{k=1}^{q}$ represents a vector of cluster-specific random effects on the subject-specific residual variance, such that $v_{e_1 k} \sim IID\ f\left(v_{e_1 k} \mid \eta_{e_1}\right) = IG\left(\eta_{e_1}, \eta_{e_1} - 1\right)$. In turn, $\eta_{e_1} \sim f\left(\eta_{e_1}\right) \propto \left(1 + \eta_{e_1}\right)^{-2}$, for $\eta_{e_1} > 0$ (Bello et al., 2010). Also, $\mathbf{x}_{5j}^{'}$, $\mathbf{x}_{6k}^{'}$, and $\mathbf{x}_{7k}^{'}$ are corresponding known incidence row vectors.

For subsequent brevity, we refer to Equation (3) as $f\left(\mathbf{y}_1\right)$ and Equation (4) as $f\left(\mathbf{y}_2^* \middle| \mathbf{y}_1\right)$. The joint posterior density of all unknown parameters, namely $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\mathbf{u}_1$,

$\mathbf{u}_2$, $\boldsymbol{\gamma}_e$, $\mathbf{m}$, σ_m^2, $\boldsymbol{\gamma}_u$, $\boldsymbol{\tau}_{e_1}$, $\boldsymbol{\tau}_{u_1}$, $\boldsymbol{\tau}_{u_{2|1}}$, $\mathbf{v}_{e_1}$ and η_{e_1}, given $\mathbf{y}_1$ and $\mathbf{y}_2$, can then be written as the product of all density specifications presented thus far; i.e., $f(\mathbf{y}_1)$, $f\left(\mathbf{y}_2^* \middle| \mathbf{y}_1\right)$, $f(\boldsymbol{\beta}_1)$, $f(\boldsymbol{\beta}_2)$, $\prod_{k=1}^{q} f(\mathbf{u}_{.k} \mid \mathbf{G}_k)$, $f(\boldsymbol{\gamma}_e)$, $f\left(\mathbf{m} \mid \sigma_m^2\right)$, $f\left(\sigma_m^2\right)$, $f(\boldsymbol{\gamma}_u)$, $f\left(\boldsymbol{\tau}_{u_1}\right)$, $f\left(\boldsymbol{\tau}_{u_{2|1}}\right)$, $f\left(\boldsymbol{\tau}_{e_1}\right)$, $\prod_{k=1}^{q} f\left(v_{e_1 k} \mid \eta_{e_1}\right)$, and $f\left(\eta_{e_1}\right)$. The Markov chain Monte Carlo (MCMC) strategy for generating samples from this joint posterior density is identical to what is presented in Bello et al. (2010) with just two exceptions. First, as previously noted, we constraint $\sigma_{e_{2|1}j}^2 = 1 \ \forall j$ to ensure parameter identifiability (alternatively, we could have set $\sigma_{e_2 j}^2 = 1 \ \forall j$ and conducted inference on $\sigma_{e_{2|1}j}^2$). Secondly, the augmented data y_{2j}^* needs to be sampled from truncated normal distributions at each MCMC iteration depending upon the value of y_{2j}; i.e.,

$y_{2j}^* \mid y_{2j} = 1, \boldsymbol{\beta}, \mathbf{u} \sim N\left(\mu_{2|1j}, 1\right) I\left(y_{2j}^* > 0\right)$ or

$y_{2j}^* \mid y_{2j} = 0, \boldsymbol{\beta}, \mathbf{u} \sim N\left(\mu_{2|1j}, 1\right) I\left(y_{2j}^* < 0\right)$ (Albert and Chib, 1993). It would be also important to note that standard identifiability constraints on "fixed" effects parameters such as $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\gamma}_e$, $\boldsymbol{\gamma}_u$, $\boldsymbol{\tau}_{e_1}$, $\boldsymbol{\tau}_{u_1}$ and $\boldsymbol{\tau}_{u_{2|1}}$ are also necessary (Gelfand and Sahu, 1999), as previously described in more detail (Bello et al., 2010).

5. Interpretation of Associations on the Observed Scale

As previously demonstrated by Bello et al. (2010), the interpretation of $\varphi_j^{(e)}$ and $\varphi_k^{(u)}$ on the underlying continuous variable scale is rather straightforward. Recall that $\mu_{2|1j} = \mu_{2j} + \varphi_j^{(e)} e_{1j}$. That is, $\varphi_j^{(e)}$ can be simply interpreted as the conditional change in y_{2j}^*, for each unit increase in e_{1j}. Similarly, $\varphi_k^{(u)}$ represents the conditional change in y_{2j}^* for each unit increase in $u_{1k(j)}$ where *k(j)* denotes the cluster associated with subject *j*. However, for the bivariate continuous-binary model, the probit link forces a more complex interpretation of cluster and subject level associations on the observed outcome scale as they depend upon baseline values of other parameters (McCulloch, 2008). For example, suppose that μ_2 is specified such that $\Phi(\mu_2)$ represents a baseline incidence rate for the binary outcome. For a particular residual increase of e_1 units on the continuous outcome, the conditional expected incidence rate for a particular value of $\varphi_j^{(e)}$ becomes $\Phi\left(\mu_2 + \varphi_j^{(e)} e_1\right)$; i.e., the expected differential in incidence rates is $\Delta_e = \Phi\left(\mu_2 + \varphi_j^{(e)} e_1\right) - \Phi(\mu_2)$.

The elegance of MCMC lies in its ability to provide the posterior density of any function of the model parameters. Given Equation (8) and recognizing that $\mathrm{E}\left(m_k \mid \sigma_m^2\right) = 0$, we substitute estimable linear functions $\mathrm{E}\left(\varphi_j^{(e)}\right) = \mathbf{k}_e' \boldsymbol{\gamma}_e$ of interest for $\varphi_j^{(e)}$ in the specified differential above. For example, we could set $\mathbf{k}_e' = \mathbf{x}_{3j}'$ for

any j or, more generally, any linear combination thereof $\mathbf{k}_e^{'}=\sum_{j=1}^{n} c_j \mathbf{x}_{3j}^{'}$ for known scalar c_j. Conditional on a baseline probit mean of μ_2 on the binary outcome and a residual increment of e_1 on the continuous outcome, we might investigate the posterior density of Δ_e as a measure of the residual subject-level association on the observed scale as it depends upon $\mathbf{k}_e^{'}\boldsymbol{\gamma}_e$ as being:

$$\Delta_e = \Phi\left(\mu_2 + \left(\mathbf{k}_e^{'}\boldsymbol{\gamma}_e\right)e_1\right) - \Phi\left(\mu_2\right) \tag{13}$$

Similarly, given Equation (9), we could specify a random cluster-level association between the two outcomes as an estimable linear function, $\mathbf{k}_u^{'}\boldsymbol{\gamma}_u$, of $\boldsymbol{\gamma}_u$ relative to a baseline probit mean of μ_2 on the binary outcome and a random effects increment of u_1 on the continuous outcome:

$$\Delta_u = \Phi\left(\mu_2 + \left(\mathbf{k}_u^{'}\boldsymbol{\gamma}_u\right)u_1\right) - \Phi\left(\mu_2\right) \tag{14}$$

An appropriate value for μ_2 in both equations (13) and (14) above might be $\mu_2 = 1/n\sum_{j=1}^{n} \mathbf{x}_{2j}^{'}\boldsymbol{\beta}_2$. Similarly, "typical" residual subject-level and random cluster-level standard deviations could be substituted for e_1 and u_1, respectively. That is, using Equations (10) and (11), one might use $\tilde{\sigma}_e = \sqrt{\exp\left(1/n\sum_{j=1}^{n} \mathbf{x}_{5j}^{'} \log\left(\boldsymbol{\tau}_{e_1}\right)\right)}$ for e_1

in Equation (13), recognizing that $\mathrm{E}\left(\mathbf{v}_{e_1}\right)=\mathbf{1}$, and

$\tilde{\sigma}_u=\sqrt{\exp\left(1/q\sum_{k=1}^{q}\mathbf{x}_{6k}^{'}\log\left(\boldsymbol{\tau}_{u_1}\right)\right)}$ for u_1 in Equation (14).

6. Simulation Study

6.1 Design

Using simulation, we investigated the inferential properties on key parameters defining between-trait associations, namely, γ_e, γ_u and σ_m^2, in our proposed bivariate GLMM. We considered 3 different types of correlation architectures with respect to average cluster-level $\bar{\varphi}^{(u)}=1/q\sum_{k=1}^{q}\varphi_k^{(u)}$ and subject-level $\bar{\varphi}^{(e)}=1/n\sum_{j=1}^{n}\varphi_j^{(e)}$ associations: A) *same sign*: $\mathrm{sgn}\left(\bar{\varphi}^{(u)}\right)=-1$, $\mathrm{sgn}\left(\bar{\varphi}^{(e)}\right)=-1$; B) *opposite sign*: $\mathrm{sgn}\left(\bar{\varphi}^{(u)}\right)=-1$, $\mathrm{sgn}\left(\bar{\varphi}^{(e)}\right)=+1$; and C) *no association*: $\bar{\varphi}^{(u)}=0$, $\bar{\varphi}^{(e)}=0$. We also considered 5 different values of σ_m^2: I) $\sigma_m^2=0$; II) $\sigma_m^2=0.01$; III) $\sigma_m^2=0.1$; IV) $\sigma_m^2=1$; and V) $\sigma_m^2=10$. Each of the 15 scenarios (3 correlation architectures x 5 values of σ_m^2) was represented by 10 replicated datasets. Within each dataset, correlated Gaussian and binary outcomes were simulated for each of n ~40,000 subjects unevenly distributed among q = 200 clusters, similar to that described in Bello et al. (2010). The same two-level fixed effect factor was specified at the subject level for each of $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$,

γ_e and τ_{e_1}; i.e. $\mathbf{x}'_{1j} = \mathbf{x}'_{2j} = \mathbf{x}'_{3j} = \mathbf{x}'_{5j}$. Similarly, the same two-level fixed effects factor was specified for each of the cluster-level terms γ_u, τ_{u_1} and $\tau_{u_{2|1}}$; i.e., $\mathbf{x}'_{4k} = \mathbf{x}'_{6k} = \mathbf{x}'_{7k}$. Assignment of subjects and clusters to the two levels of the fixed effect factor was based on a Bernoulli distribution with probability 0.25 to partly mimic an unbalanced design structure. We specified $\gamma_e = \begin{bmatrix} \gamma_{e_1} & \gamma_{e_2} \end{bmatrix}'$ and $\gamma_u = \begin{bmatrix} \gamma_{u_1} & \gamma_{u_2} \end{bmatrix}'$ using the values indicated for each of scenarios A, B and C in Table 1. Furthermore, we defined $\tau_{e_1} = \begin{bmatrix} \tau_{e_1,1} & \tau_{e_1,2} \end{bmatrix}' = \begin{bmatrix} 0.75 & 1 \end{bmatrix}'$, $\tau_{u_1} = \begin{bmatrix} \tau_{u_1,1} & \tau_{u_1,2} \end{bmatrix}' = \begin{bmatrix} 0.2 & 0.4 \end{bmatrix}'$, $\tau_{u_{2|1}} = \begin{bmatrix} \tau_{u_{2|1},1} & \tau_{u_{2|1},2} \end{bmatrix}' = \begin{bmatrix} 0.5 & 0.4 \end{bmatrix}'$ and $\eta_{e_1} = 5$ for all simulation scenarios. We also specified flat unbounded priors on $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, γ_e, σ_m^2, γ_u, τ_{e_1}, τ_{u_1} and $\tau_{u_{2|1}}$. We certainly recognize it as best practice to specify proper priors on all parameters in order to ensure a proper joint posterior density. However, we were particularly interested in how resilient inference might be using the proposed model with noninformative priors, realizing that the interpretation of the results from our simulation study could otherwise be tainted by subjective specifications.

6.2 Convergence Diagnostics

A single MCMC chain was run for each model (see later) fitted to each of the 150 simulated datasets (15 scenarios x 10 replicates each). In all cases, we saved 50,000 iterations after a burn-in period of 8,000 iterations. Convergence diagnostics for all parameters with subjective or non-exchangeable priors were based on trace plots and on

Raftery and Lewis (1992). Of particular concern and interest were the mixing properties of the parameters specified with noninformative priors and at the deepest level of the model hierarchy, namely γ_{e_1}, γ_{e_2}, γ_{u_1}, γ_{u_2} and σ_m^2. We also monitored data augmentation samples for three randomly selected elements of $\mathbf{y}_2^*$ in all MCMC runs. Web Figures 1-4 show representative trace plots for elements of $\boldsymbol{\gamma}_e$, $\boldsymbol{\gamma}_u$, $\boldsymbol{\tau}_{e_1}$, $\boldsymbol{\tau}_{u_1}$ and $\mathbf{y}_2^*$ from one simulated dataset, as well as representative trace plots for σ_m^2 considering each true value of σ_m^2. Chain convergence did not appear to be an issue in any situation.

6.3 Model Choice

For each of the 150 simulated datasets, we fitted a null model (M_0) that constrained $\sigma_m^2 = 0$ and a competing full model (M_1) that specified cluster-specific subject-level associations; i.e. $\sigma_m^2 > 0$. As criteria for model choice, we applied the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) and the pseudo-Bayes Factor (pBF) (Gelfand and Dey, 1994). Note that both DIC and pBF require specification of the joint bivariate data likelihood $f(\mathbf{y}_1,\mathbf{y}_2) = f(\mathbf{y}_1) f(\mathbf{y}_2|\mathbf{y}_1)$, similar to Janss and Foulley (1993), with $f(\mathbf{y}_1)$ and $f(\mathbf{y}_2|\mathbf{y}_1)$ provided in Equations (3) and (5), respectively.

To draw conclusions upon the statistical significance of σ_m^2, we use 1) the difference between the corresponding model DIC values, namely $\mathrm{DIC}_\Delta = \mathrm{DIC}_{\mathrm{M}_0} - \mathrm{DIC}_{\mathrm{M}_1}$, and 2) the ratio of the corresponding pBF of M_1 relative to M_0 expressed on the $\log_{10}$

scale, which we refer to as $\log_{10}$pBF. Values of DIC_Δ and $\log_{10}$ pBF greater than zero would support the choice of M_1 over M_0, and thus, indicate evidence for non-zero σ_m^2; conversely, negative values would support M_0. We use a $|\mathrm{DIC}_\Delta|$ of 7 or greater (Spiegelhalter et al., 2002) and a $\log_{10}$ pBF of 2 or greater (Kass and Raftery, 1995) to conclude upon a decisive difference in fit between the two models.

For all 30 datasets where $\sigma_m^2 = 0$, DIC_Δ fell within a range of [-4.2, 1.6] and $\log_{10}$ pBF fell within a range of [-1.1, 0.3] such that neither M_1 or M_0 were decisively chosen. Both selection criteria impressively favored the correct model (M_1) for all 120 datasets where $\sigma_m^2 > 0$. As expected, the magnitude of DIC_Δ and $\log_{10}$ pBF increased with greater values of σ_m^2, indicating greater statistical power to detect non-zero σ_m^2. More specifically, the range of DIC_Δ values was [32, 247] for $\sigma_m^2 = 0.01$; [663, 2220] for $\sigma_m^2 = 0.1$; [6291, 12437] for $\sigma_m^2 = 1$; and [19984, 33833] for $\sigma_m^2 = 10$. Similarly, the range of $\log_{10}$ pBF was [6, 53] for $\sigma_m^2 = 0.01$; [144, 481] for $\sigma_m^2 = 0.1$; [1364, 2698] for $\sigma_m^2 = 1$; and [4336, 7338] for $\sigma_m^2 = 10$. The three different correlation architectures did not appear to influence decisions on model choice. We believe that these results validate the use of DIC and pBF as model selection criteria for the proposed bivariate GLMM.

We also considered the relative contributions of the marginal $f(\mathbf{y}_1)$ and conditional $f(\mathbf{y}_2|\mathbf{y}_1)$ components of the joint likelihood $f(\mathbf{y}_1,\mathbf{y}_2)$ on the model choice criteria.

When $\sigma_m^2 > 0$, we noticed that large DIC_Δ and $\log_{10}$ pBF for detecting σ_m^2 were overwhelmingly driven by the $f(\mathbf{y}_2|\mathbf{y}_1)$ component (results not shown). This was to be expected as σ_m^2 specifies the degree of heterogeneity on $\varphi_j^{(e)}$ which, in turn, links the model fit of the binary outcome to that of the more informative continuous outcome. As σ_m^2 increased, the $f(\mathbf{y}_1)$ component of $f(\mathbf{y}_1,\mathbf{y}_2)$ also did contribute to overall model fit such that when $\sigma_m^2 = 10$, $f(\mathbf{y}_1)$ contributed to DIC_Δ by between 10 to 22 points, and to $\log_{10}$ pBF by approximately 2 to 5 orders of magnitude in favor of M_1 over M_0. Nevertheless, the impact of $f(\mathbf{y}_1)$on the two model choice criteria was dwarfed (<0.1%) relative to that of $f(\mathbf{y}_2|\mathbf{y}_1)$. Thus, a bivariate Gaussian-binary analysis that considers heterogeneous associations does sharpen inference for the binary response, though there is a small improvement in fit for the Gaussian outcome as well.

6.4 Inference on Heterogeneous Associations

Table 1 provides details on the minimum and maximum of the upper and lower boundaries of the 95% highest posterior density (HPD) interval of the posterior distributions of γ_{e_1}, γ_{e_2}, γ_{u_1}, γ_{u_2} and σ_m^2 across the 10 replicates for each of the 15 scenarios considered. For these five parameters, the coverage probability of the 95% HPD ranged from 95.3 to 96.7% across 150 replicates (or 120 for σ_m^2 where $\sigma_m^2 > 0$). These coverage probabilities were consistent with probabilistic expectation as they were not determined to be significantly different from the nominal value, neither did they

differ with correlation architecture or magnitude of σ_m^2. For completeness, we also report coverage probabilities for parameters defining heterogeneity of (conditional) cluster-level and subject-level variances, namely elements of $\boldsymbol{\tau}_{u_1}$, $\boldsymbol{\tau}_{u_{2|1}}$, $\boldsymbol{\tau}_{e_1}$ and η_{e_1}. For each of these seven parameters, HPD coverage satisfactorily matched nominal coverage as the 95% HPD coverage probability ranged from 92.0 to 99.3% across 150 replicates.

We evaluated potential bias of the posterior mean for each of γ_{e_1}, γ_{e_2}, γ_{u_1}, γ_{u_2} and σ_m^2 under each of the 15 scenarios with a one-sample *t*-test and a one-sample non-parametric Wilcoxon Rank Sum Test, using the true parameter values as the null values for those tests. Based on a Bonferroni-corrected (as per the 15 different scenarios) Type I error rate of 5%, we found no evidence of bias of posterior means for any of these parameters (results not shown). Similar results were encountered when the posterior mean of each element of $\boldsymbol{\tau}_{u_1}$, $\boldsymbol{\tau}_{u_{2|1}}$, $\boldsymbol{\tau}_{e_1}$ and η_{e_1} was evaluated for bias (results not shown).

Precision of inference was greater for γ_{e_1} and γ_{e_2} compared to γ_{u_1} and γ_{u_2}, as illustrated by the narrower 95% HPD of the former in Table 1. Indeed, less inferential precision is to be expected at the random cluster-level compared to the residual subject-level since γ_{u_1} and γ_{u_2} are further removed from the data ($\mathbf{y}_1$ and $\mathbf{y}_2$) compared to γ_{e_1} and γ_{e_2} in the model hierarchy. Note that the precision of estimation on γ_{e_1} and γ_{e_2} (but not on γ_{u_1} and γ_{u_2}) worsened as σ_m^2 increased, as indicated by their wider 95% HPD.

n

as

na

th

wa

ma

7.

Wa

rep

sta

her

Ass

rep

To

prin

τ_{e}

day

(i.e.

man

(i.e.

Overall, we believe this simulation study validates the proposed bivariate GLMM as a tool to infer upon heterogeneous residual subject-level and random cluster-level associations between Gaussian and binary outcomes. The use of model selection criteria, namely DIC and pBF, worked as intended. These results did not appear to depend upon the nature of the subject-level and cluster-level correlation architectures (i.e. A, B, C) that were considered. Furthermore, 95% HPD were close to nominal coverage and posterior means were seemingly unbiased.

7. Application

We revisit the motivating example regarding the association between milk production and reproductive performance of dairy cows. We received data on milk yield and pregnancy status following first postpartum insemination for n = 39,917 cows from q = 319 dairy herds in Michigan recorded during 2006 from the National Dairy Herd Improvement Association (Raleigh, NC). Here, cow defines the residual subject level whereas herd represents the clustering factor that identifies the random level of the hierarchical model. To illustrate the use of our proposed method, we selected cow parity, categorized as either primiparous or multiparous, as the basis for subject-level fixed effects on $\boldsymbol{\gamma}_e$ and $\boldsymbol{\tau}_{e_1}$ (i.e. $\mathbf{x}'_{3j} = \mathbf{x}'_{5j}$) and herd milking frequency, recorded as 2 vs. 3 or more times per day (i.e. 2X vs. 3^+X), as the basis for cluster-level fixed effects on $\boldsymbol{\gamma}_u$, $\boldsymbol{\tau}_{u_1}$ and $\boldsymbol{\tau}_{u_{2|1}}$ (i.e. $\mathbf{x}'_{4k} = \mathbf{x}'_{6k} = \mathbf{x}'_{7k}$). Both of these factors are common determinants of management and performance in conventional dairy farms. For all sets of random effects (i.e. $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{m}$ and $\mathbf{v}_{e_1}$), we specified herd as the common clustering factor. In

addition, location parameters (i.e. $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$) for both traits included the fixed effects of parity and calving season (Winter: December to February; Spring: March to May; Summer: June to August; and Fall: September to November), as well as polynomial functions of the covariate "days in milk" (i.e. 5th order Legendre polynomials for $\boldsymbol{\beta}_1$, as reviewed by Schaeffer (2004), and linear and quadratic terms for $\boldsymbol{\beta}_2$). Specifications of prior densities, convergence diagnostics and model choice criteria were as indicated for the simulation study.

In comparing the two models with $\sigma_m^2 = 0$ (M_0) versus $\sigma_m^2 > 0$ (M_1), neither DIC (DIC_Δ = 1.7) nor pBF ($\log_{10}$ pBF = 0.35) provided sufficient evidence to reject M_0. Thus, we base our subsequent inference on the null model (M_0), adopting $\varphi_j^{(e)} = \mathbf{x}_{3j}'\boldsymbol{\gamma}_e$ and assuming no significant variation between herds in the residual subject-level association between traits.

Our inferential interest is focused on $\boldsymbol{\gamma}_e = \left[\gamma_{e,primiparous} \quad \gamma_{e,multiparous}\right]$ and $\boldsymbol{\gamma}_u = \left[\gamma_{u,2\mathrm{X}} \quad \gamma_{u,3^+\mathrm{X}}\right]$, where the element subscripts denote the levels of the corresponding fixed effects factors. Table 2 summarizes posterior inference on $\boldsymbol{\gamma}_e$ and $\boldsymbol{\gamma}_u$ using posterior means, posterior standard deviations, 95% HPD and effective sample size, the latter denoting the effective number of independent samples after accounting for autocorrelation in the MCMC samples (Sorensen et al., 1995).

At the subject-level, we determined no evidence for a difference in the between-trait association among cows of first (primiparous) versus subsequent (multiparous) lactations

as the 95% HPD on $\left(\gamma_{e,primiparous} - \gamma_{e,multiparous}\right)$ included 0. Thus, we infer upon the overall subject-level association between milk yield and pregnancy outcome to first postpartum insemination using $\bar{\varphi}^{(e)} = 1/n\sum_{j=1}^{n}\mathbf{x}_{3j}'\gamma_e$. The posterior mean $\bar{\varphi}^{(e)}$ was found to be 0.0021 with a 95% HPD of [0.0006, 0.0036] such that there is some evidence of an overall positive residual association between the two traits. At the cluster-level, the between-trait association differed with milking frequency practices, such that the posterior mean for $\gamma_{u,3^+\mathrm{X}}$ was significantly more negative than that for $\gamma_{u,2\mathrm{X}}$ (Table 2) as the 95% HPD on their mean difference $\left(\gamma_{u,2\mathrm{X}} - \gamma_{u,3^+\mathrm{X}}\right)$ did not include 0 (i.e., 95% HPD=[0.0038, 0.0302]). Therefore, it appeared that the cluster-level association between the two outcomes was antagonistic for 3^+X milking herds whereas there was no significant between-trait association for 2X milking herds.

Figure 1 illustrates the posterior densities of Δ_e (Equation 13) or the residual subject-level associations on the observed scale using $\mu_2 = 1/n\sum_{j=1}^{n}\mathbf{x}_{2j}'\boldsymbol{\beta}_2$ and $e_1 = \tilde{\sigma}_e$ for each of the two parities (i.e., $\mathbf{k}_e'$ = [1 0] for $\Delta_{e,primiparous}$ and $\mathbf{k}_e'$ = [0 1] for $\Delta_{e,multiparous}$), and the posterior densities for Δ_u (Equation 14) or the cluster-level associations on the observed scale using $\mu_2 = 1/n\sum_{j=1}^{n}\mathbf{x}_{2j}'\boldsymbol{\beta}_2$ and $u_1 = \tilde{\sigma}_u$ for each of the two milking frequencies (i.e., $\mathbf{k}_u'$ = [1 0] for $\Delta_{u,2\mathrm{X}}$ and $\mathbf{k}_u'$ = [0 1] for $\Delta_{u,3^+\mathrm{X}}$). Recall then that Δ_e indicates the change in cow pregnancy rate per $\tilde{\sigma}_e$ kg

(posterior mean=7.6) of increased cow milk production whereas Δ_u indicates the change in herd pregnancy rate per $\tilde{\sigma}_u$ kg (posterior mean=4.7) of increased herd milk production. The 95% HPD on $\left(\Delta_{e,primiparous} - \Delta_{e,multiparous}\right)$ included 0, consistent with previously reported inference on $\left(\gamma_{e,primiparous} - \gamma_{e,multiparous}\right)$ whereas the 95% HPD on $\left(\Delta_{u,2\ X} - \Delta_{u,3^+X}\right)$ was [0.0086, 0.0526], again consistent with the 95% HPD on $\left(\gamma_{u,2X} - \gamma_{u,3^+X}\right)$ not including 0. Note, for example, from Figure 1 that the density for $\Delta_{u,3^+X}$ is strongly concentrated on -0.04. This implies that as a herd's milk production increases by one average cluster-level standard deviation, namely $\tilde{\sigma}_u = 4.7$ kg, relative to an overall mean production (posterior mean of $\mu_1 = 40.8$ kg), herd pregnancy rates drop by 4 percentual points relative to a baseline herd fertility, namely $\Phi(\mu_2) = 34\%$ pregnancy rate to first postpartum insemination.

Overall, it appeared that within herds (residual subject-level), higher producing cows were also more likely to become pregnant at first insemination regardless of parity. In contrast, herds (random cluster-level) with greater milk yields had generally lower pregnancy rates, but only if under a 3^+X milking scheme. This antagonism was not apparent amongst 2X milking herds, whereby the cluster-level association was estimated as null based on the zero-overlapping 95% HPD for $\gamma_{u,2X}$ (Table 2). Thus, a favorable subject (cow) level association was counteracted by a factor-level-specific

random between-cluster antagonism, thus suggesting differing underlying mechanisms in the between-trait association among herds and among cows.

For completeness, Web Table 1 summarizes posterior inference on (conditional) variances.

8. Discussion

Motivated by a practice problem in animal agriculture, we propose a hierarchical Bayesian extension that allows for mixed effects modeling of heterogeneous random (cluster-level) and residual (subject-level) (co)variance matrices in the joint analysis of a continuous and a binary distributed trait within a bivariate GLMM. Inferences based on this model help understand the underlying mechanisms that alter associations between mixed outcomes on different levels (e.g. cluster, subject), especially where evidence regarding the nature of these associations is contradictory. The proposed model is implemented to investigate the nature of the production-reproduction association in dairy cattle. Results indicate that the nature of the between-herd (cluster-level) association differs across management practices (i.e. milking frequency) and across strata level (i.e. cow versus herd). Disregarding this multifactorial-multidimensional heterogeneity could lead to overly-generalized, even biased, conclusions on the association between outcomes, which in turn, would be likely to have negative implications for optimizing overall performance of the dairy business.

Our data application was not intended to be comprehensive; there are many other potentially important covariates that may affect the production-reproduction association in dairy cows, the investigation of which is forthcoming in future publications. For this

purpose, a Bayesian model averaging approach to selecting important factors and covariates (e.g. Chen and Dunson, 2003) would be a useful extension to our model. It would also be of interest to incorporate random effects other than herd clusters into the proposed model. A relevant candidate might be genetic effects since it appears that the genetic correlation or association between traits may inherently depend upon other factors (Tsuruta et al., 2009).

It has been realized that multivariate GLMM permit the sharing of information between outcomes, thus sharpening inference on outcome-specific location parameters when compared to univariate GLMM analysis (Riley et al., 2007). However, the possibility that, for example, residual associations may be heterogeneous across clusters (i.e. $\sigma_m^2 > 0$) or depend upon treatment effects may require a modeling approach similar to that proposed in this paper to ensure that the sharing of information between outcomes is not distorted by the assumption of homogeneous associations. This consideration is particularly relevant for meta-analyses, whereby the underlying assumption of homogeneous (co)variances within studies (subject-level) and between studies (cluster-level) is likely to be too restrictive.

A multivariate extension of the proposed GLMM would be very appealing to simultaneously accommodate more than 2 mixed outcomes, with the necessary identifiability restrictions for additional binary outcomes. Of particular interest might be the application to joint analysis of continuous and binary longitudinal data. Additional parsimonious constraints on the variance-covariance matrix may be required, particularly with increasing dimensionality. Covariance structures such as those proposed in

antedependence models may constitute a viable modeling option (Jaffrezic, Thompson and Hill, 2003).

Our proposed model is based on a convenient and widely used data augmentation technique for probit models (Albert and Chib, 1993) whereby the binary outcomes $\mathbf{y}_2$ are believed to be determined by corresponding underlying normally distributed variables $\mathbf{y}_2^*$. Since $\mathbf{y}_1$ and $\mathbf{y}_2^*$ are then multivariate normal, this facilitates a Cholesky-type reparameterization of the random cluster-level and residual subject-level (co)variance matrices and the modeling approach that we previously developed for a bivariate continuous model (Bello et al., 2010). Similarly, data augmentation equips our methodology with the necessary flexibility to be easily extended to other mixed outcome models, whereby observed non-Gaussian outcomes are also determined by (functions of) underlying normally distributed random variables (McCulloch, 2008). In Web Appendix A, we describe adjustments to the proposed methodology that enables ordered categorical, count and censored responses to be modeled for heterogeneous covariances, conditional upon a Gaussian outcome. Furthermore, our implementation of data augmentation is readily extendable to recover information lost with missing observations for $\mathbf{y}_2$, if needed (Najita, Li and Catalano, 2009; Tanner, 1993). Conversely, if the prevalence of missing data were extensive, parameter identifiability may become a greater issue with a model that has an extensive hierarchical specification such as that proposed in this paper.

As discussed by Bello et al. (2010), our bivariate specification is not invariant to order. For example, we could have alternatively developed a model where the continuous outcome (now $\mathbf{y}_2$) was specified to be conditioned upon the binary outcome (now $\mathbf{y}_1$)

similar to O'Malley et al. (2003). The approach would be then to augment the likelihood with an underlying normally distributed variable $\mathbf{y}_1^*$ that predetermines $\mathbf{y}_1$; furthermore, we would restrict $\sigma^2_{e_1 j} = 1 \; \forall j$ rather than $\sigma^2_{e_{2|1} j}$ or $\sigma^2_{e_2 j}$. For our particular application, the bivariate continuous-binary model was the logical choice as the continuous production outcome precedes, and hence is believed to be somewhat causal to, the binary pregnancy event.

Finally, the parameters γ_e, γ_u, and σ^2_m arguably constitute the deepest levels of the model hierarchy, such that the data is least informative (i.e. little inferential power) on these parameters, particularly if the analysis comprises binary data. Undoubtedly, relatively large size datasets will be required for inference on heterogeneous covariances between mixed outcomes based on the proposed bivariate GLMM. While we did not intend to make recommendations on sample size and statistical power, the large size (i.e., thousands of observations) of the simulated datasets was intended to be representative of datasets commonly encountered in large field studies using extensively parameterized hierarchical models (Bello et al., 2010; O'Malley et al., 2003; Tsuruta et al., 2009). We did intend, however, to use non-informative priors on all "fixed effects" parameters in this study to investigate robustness to prior specifications for large datasets and found that coverage probabilities of 95% HPD closely followed nominal coverage across plausible simulation scenarios. Nevertheless, we would routinely recommend the use of informative or reference priors (Natarajan and Kass, 2000) in most applications.

9. Conclusions

We present a Bayesian approach to modeling heterogeneous residual subject-level and random cluster-level (co)variances between a continuous and a binary outcome in the context of a bivariate GLMM. This methodology can be readily applied to the study of complex biological phenomena in many subject-matter applications for which there may be two or more mixed outcomes and one might suspect heterogeneity in the association between the outcomes as a function of covariates and/or random cluster effects. The proposed model constitutes an enhancement in current statistical methodology in that it introduces a new dimension of heterogeneity, namely that of covariances among multivariate mixed outcomes.

ACKNOWLEDGEMENTS: This study was partially funded by the Elwood Kirkpatrick Dairy Research Endowment, the Michigan Milk Producers Association, the College of Agriculture and Natural Resources and the Department of Animal Science at Michigan State University

Table 3.1.a *Minimum and maximum of the upper and lower boundaries of the 95% highest posterior density (HPD) interval as well as [illegible]*

Table 3.1.a Minimum and maximum of the upper and lower boundaries of the 95% highest posterior density (HPD) interval, as well as true values used for simulation, for parameters defining heterogeneity in the residual subject (**e**) level and random cluster (**u**) level associations, namely γ_e, σ_m^2 and γ_u, across 10 replicates for each of 15 simulation populations defined by all factorial combinations of 3 correlation architectures (A, B, C) with 3 (I, II, III) of 5 different values of σ_m^2 as indicated in text.

Scenarios		**A. Same sign:** negative **u**-level and **e**-level coefficients			**B. Opposite sign:** negative **u**-level and positive **e**-level coefficients			**C. Zero correlation:** zero **u**-level and **e**-level coefficients		
	Parameters †	**True**	**Lower HPD [Min, Max]**	**Upper HPD [Min, Max]**	**True**	**Lower HPD [Min, Max]**	**Upper HPD [Min, Max]**	**True**	**Lower HPD [Min, Max]**	**Upper HPD [Min, Max]**
I. $\sigma_m^2 = 0$	γ_{u_1}	-0.35	[-1.01, -0.45]	[-0.33, 0.19]	-0.35	[-0.85, -0.44]	[-0.16, 0.18]	0	[-0.60, -0.05]	[0.06, 0.58]
	γ_{u_2}	-0.25	[-0.86, -0.26]	[-0.04, 0.39]	-0.25	[-0.70, -0.20]	[0.02, 0.48]	0	[-0.80, 0.06]	[0.01, 0.73]
	γ_{e_1}	-0.2	[-0.24, -0.20]	[-0.18, -0.14]	0.2	[0.15, 0.19]	[0.21, 0.25]	0	[-0.06, 0.00]	[0.00, 0.05]
	γ_{e_2}	-0.1	[-0.13, -0.10]	[-0.10, -0.06]	0.1	[0.06, 0.10]	[0.10, 0.14]	0	[-0.02, 0.00]	[0.01, 0.03]
	σ_m^2	0	–	–	0	–	–	0	–	–
II. $\sigma_m^2 = 0.01$	γ_{u_1}	-0.35	[-1.09, -0.42]	[-0.45, 0.30]	-0.35	[-1.00, -0.54]	[-0.35, 0.04]	0	[-0.60, -0.19]	[-0.04, 0.47]
	γ_{u_2}	-0.25	[-0.88, -0.31]	[-0.17, 0.42]	-0.25	[-0.99, -0.27]	[-0.13, 0.41]	0	[-0.53, -0.11]	[0.00, 0.61]
	γ_{e_1}	-0.2	[-0.27, -0.21]	[-0.20, -0.14]	0.2	[0.14, 0.20]	[0.21, 0.27]	0	[-0.06, -0.01]	[0.01, 0.06]
	γ_{e_2}	-0.1	[-0.16, -0.11]	[-0.10, -0.06]	0.1	[0.04, 0.08]	[0.10, 0.14]	0	[-0.04, 0.00]	[0.01, 0.06]
	σ_m^2	0.01	[0.002, 0.011]	[0.010, 0.027]	0.01	[0.004, 0.008]	[0.012, 0.021]	0.01	[0.003, 0.009]	[0.014, 0.023]
III. $\sigma_m^2 = 0.1$	γ_{u_1}	-0.35	[-0.70, -0.34]	[-0.19, 0.31]	-0.35	[-0.91, -0.53]	[-0.24, 0.02]	0	[-0.56, -0.12]	[0.05, 0.47]
	γ_{u_2}	-0.25	[-0.86, -0.25]	[-0.12, 0.36]	-0.25	[-0.84, -0.33]	[-0.16, 0.31]	0	[-0.68, -0.15]	[0.02, 0.34]
	γ_{e_1}	-0.2	[-0.31, -0.22]	[-0.17, -0.11]	0.2	[0.10, 0.17]	[0.22, 0.29]	0	[-0.09, -0.01]	[0.01, 0.11]
	γ_{e_2}	-0.1	[-0.20, -0.11]	[-0.08, -0.01]	0.1	[0.00, 0.09]	[0.12, 0.20]	0	[-0.08, 0.00]	[0.03, 0.11]
	σ_m^2	0.1	[0.04, 0.10]	[0.09, 0.18]	0.1	[0.06, 0.10]	[0.12, 0.17]	0.1	[0.06, 0.09]	[0.10, 0.16]

Table 3.1.b Minimum and maximum of the upper and lower boundaries of the 95% highest posterior density (HPD) interval, as well as true values used for simulation, for parameters defining heterogeneity in the residual subject (e) level and random cluster (u) level associations, namely γ_e, σ_m^2 and γ_u, across 10 replicates for each of 15 simulation populations defined by all factorial combinations of 3 correlation architectures (A, B, C) with 2 (IV, V) of 5 different values of σ_m^2 as indicated in text.

Scenarios		A. **Same sign:** negative **u**-level and **e**-level coefficients			B. **Opposite sign:** negative **u**-level and positive **e**-level coefficients			C. **Zero correlation**: zero **u**-level and **e**-level coefficients		
	Parameters †	**True**	**Lower HPD [Min, Max]**	**Upper HPD [Min, Max]**	**True**	**Lower HPD [Min, Max]**	**Upper HPD [Min, Max]**	**True**	**Lower HPD [Min, Max]**	**Upper HPD [Min, Max]**
IV. $\sigma_m^2 = 1$	γ_{u_1}	-0.35	[-0.82, -0.36]	[-0.20, 0.25]	-0.35	[-1.01, -0.33]	[-0.39, 0.28]	0	[-0.58, -0.01]	[0.14, 0.73]
	γ_{u_2}	-0.25	[-1.07, -0.44]	[-0.23, 0.37]	-0.25	[-0.88, -0.32]	[-0.10, 0.29]	0	[-0.63, 0.00]	[0.19, 0.74]
	γ_{e_1}	-0.2	[-0.45, -0.22]	[-0.15, 0.09]	0.2	[-0.06, 0.20]	[0.26, 0.50]	0	[-0.30, -0.06]	[0.03, 0.25]
	γ_{e_2}	-0.1	[-0.36, -0.12]	[-0.07, 0.19]	0.1	[-0.18, 0.11]	[0.14, 0.41]	0	[-0.30, -0.09]	[0.03, 0.22]
	σ_m^2	1	[0.63, 1.00]	[1.04, 1.64]	1	[0.61, 1.00]	[1.01, 1.62]	1	[0.61, 0.84]	[1.03, 1.39]
V. $\sigma_m^2 = 10$	γ_{u_1}	-0.35	[-0.79, -0.41]	[-0.19, 0.24]	-0.35	[-0.95, -0.48]	[-0.36, 0.35]	0	[-0.70, -0.02]	[-0.14, 0.67]
	γ_{u_2}	-0.25	[-0.74, -0.23]	[0.01, 0.61]	-0.25	[-1.04, -0.22]	[-0.32, 0.94]	0	[-1.01, -0.10]	[0.04, 0.72]
	γ_{e_1}	-0.2	[-1.20, -0.26]	[-0.31, 0.54]	0.2	[-0.57, 0.23]	[0.35, 1.21]	0	[-0.81, -0.12]	[0.25, 1.01]
	γ_{e_2}	-0.1	[-1.05, -0.16]	[-0.17, 0.64]	0.1	[-0.71, 0.09]	[0.22, 1.06]	0	[-0.81, -0.09]	[0.21, 1.03]
	σ_m^2	10	[5.4, 9.1]	[9.0, 15.8]	10	[6.9, 9.7]	[11.9, 16.4]	10	[6.4, 10.3]	[10.8, 17.6]

† $\gamma_{u_1}, \gamma_{u_2}$ correspond to random cluster-level association parameters for levels 1 and 2, respectively, of a fixed effect factor.

$\gamma_{e_1}, \gamma_{e_2}$ correspond to residual subject-level association parameters for levels 1 and 2, respectively, of a fixed effect factor.

σ_m^2 corresponds to the parameter defining random between-cluster heterogeneity among the residual subject-level regression parameters.

Table 3.2. Posterior mean (PMEAN), posterior standard deviation (PSD), 95% highest posterior density (HPD) intervals and effective sample size (ESS) of MCMC samples on residual subject level and random cluster level association parameters (namely, γ_e and γ_u expressed in the underlying liability scale, respectively) between milk yield and pregnancy outcome at 1st postpartum insemination in Michigan dairy cows.

Regression parameters †	PMEAN	PSD	95% HPD	ESS
$\gamma_{u,2X}$, liability scale/kg	-0.0068	0.0043	[-0.0152, 0.0017]	4000
$\gamma_{u,3^+X}$, liability scale/kg	-0.0239	0.0048	[-0.0332, -0.0141]	1939
$\gamma_{e,primiparous}$, liability scale/kg	0.0028	0.0014	[0.0002, 0.0055]	20172
$\gamma_{e,multiparous}$, liability scale/kg	0.0015	0.0009	[-0.0003, 0.0033]	18151

† $\gamma_{u,2X}$ and $\gamma_{u,3^+X}$ are the random cluster level association parameters between milk yield and pregnancy outcome at first postpartum insemination for herds with twice a day (2X) and three times a day (or greater; 3^+X) milking frequency, respectively.

† $\gamma_{e,primiparous}$ and $\gamma_{e,multiparous}$ are the residual subject level association parameters between milk yield and pregnancy outcome at first postpartum insemination for cows in their first (primiparous) or subsequent (multiparous) lactation, respectively.

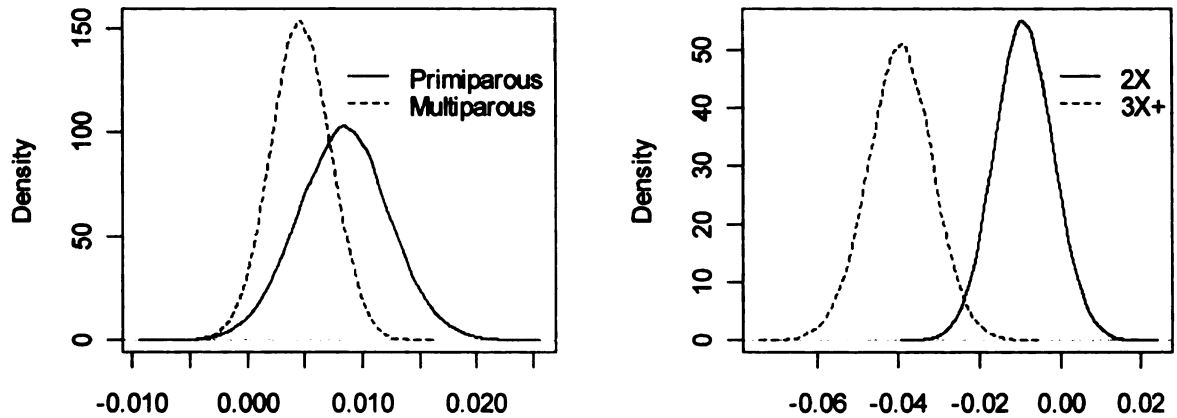
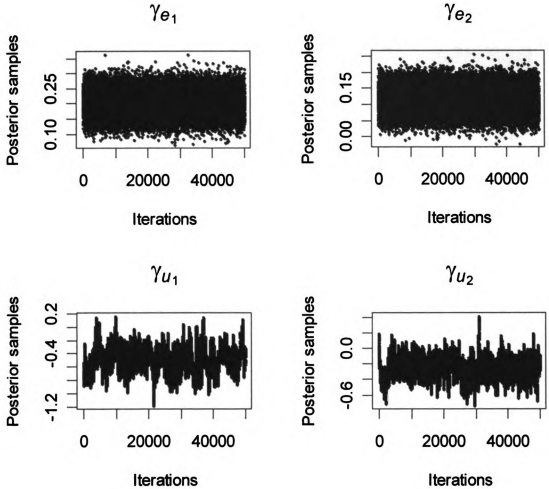
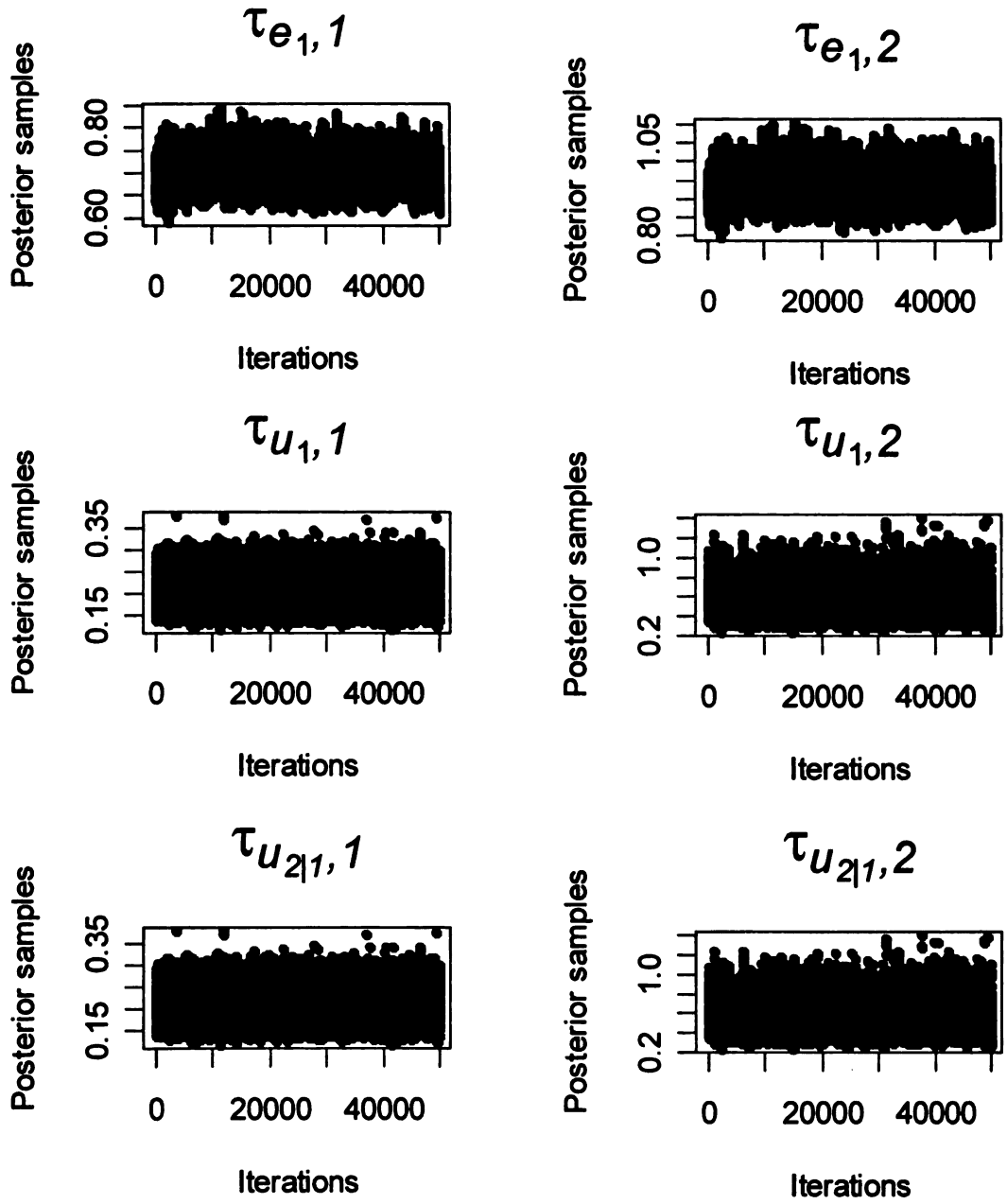


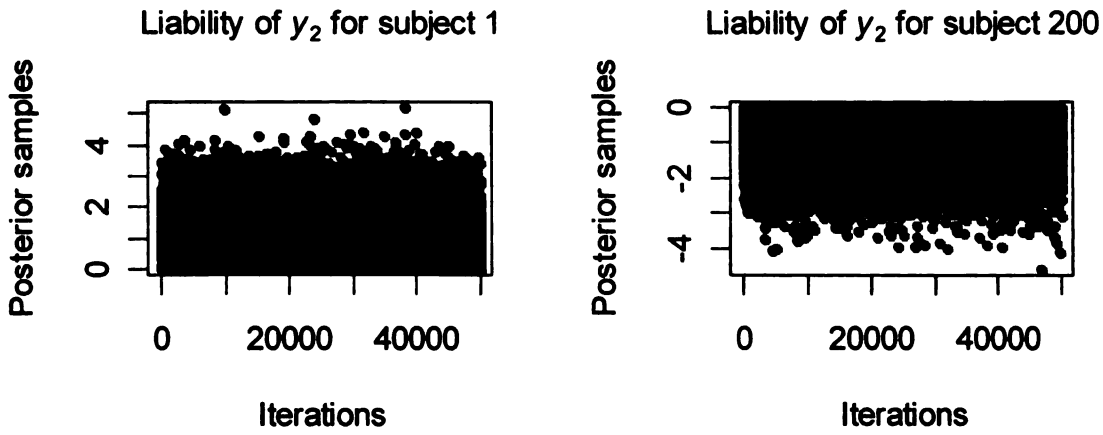
Figure 3.1: Posterior density of the differential on the conditional probability of pregnancy success to first postpartum insemination. The left panel illustrates the posterior density for the residual subject level differential, namely $\Delta_e = \Phi\left(\mu_2 + \left(\mathbf{k}_e'\gamma_e\right)e_1\right) - \Phi\left(\mu_2\right)$, for cows in their first (primiparous) or subsequent (multiparous) lactation. The right panel depicts the posterior density for the random cluster level differential, namely $\Delta_u = \Phi\left(\mu_2 + \left(\mathbf{k}_u'\gamma_u\right)u_1\right) - \Phi\left(\mu_2\right)$, for herds with twice a day (2X) and three times a day (or greater; 3^+X) milking frequency. Baseline values of μ_2, e_1 and u_1 used in the figures were obtained as described in Section 5 of the text.



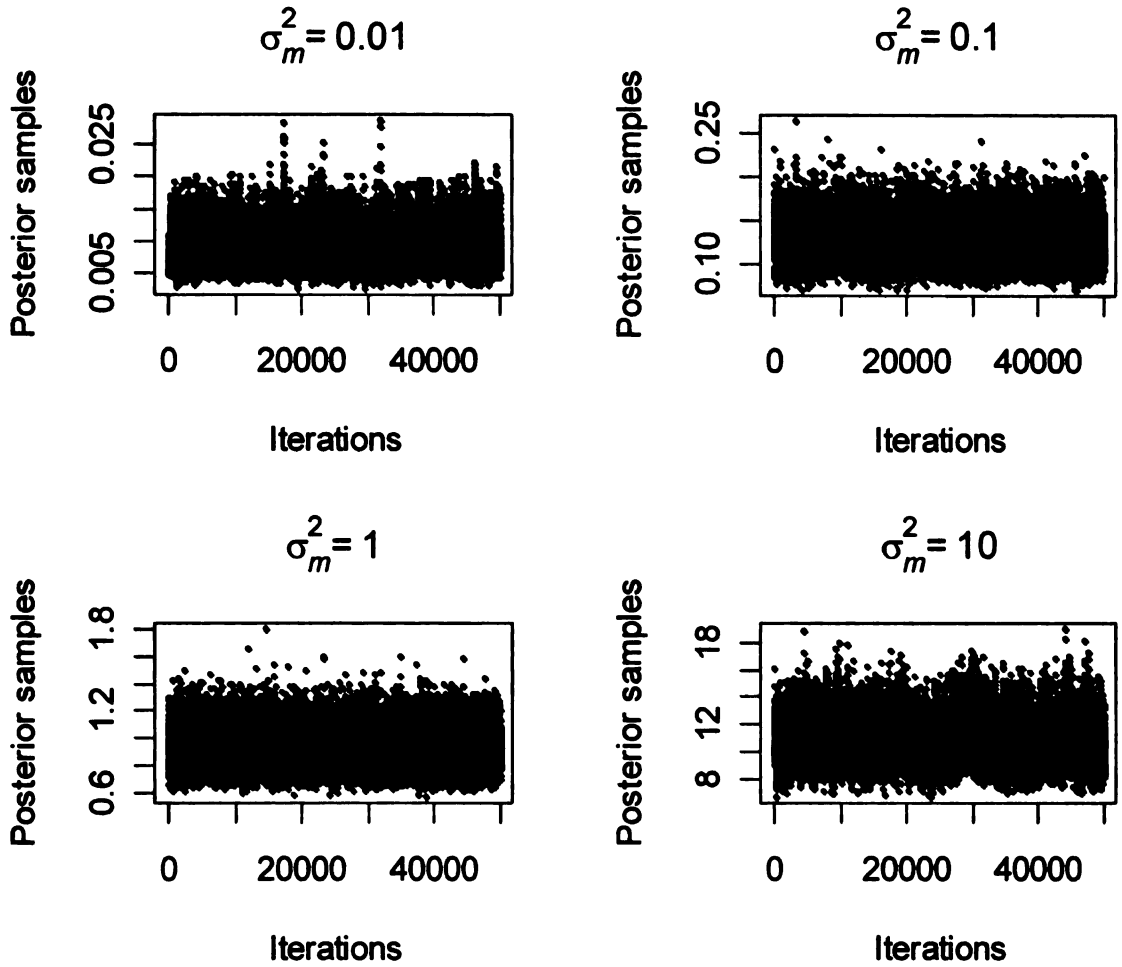
Web Figure 3.1. Trace plots for residual subject-level and random cluster-level association parameters, namely $\gamma_e = \begin{bmatrix} \gamma_{e_1} & \gamma_{e_2} \end{bmatrix}$ and $\gamma_u = \begin{bmatrix} \gamma_{u_1} & \gamma_{u_2} \end{bmatrix}$ respectively, evaluated at each of the two levels of the simulated fixed effect factor. These trace plots correspond to one selected simulated dataset and are provided to illustrate mixing of the Markov Chain Monte Carlo while sampling from the posterior density of the corresponding parameter. Chain convergence did not appear to be an issue for any of these parameters.



Web Figure 3.2. Trace plots for residual subject-level and random cluster-level (conditional) variances evaluated at each of the two levels of the simulated fixed effect factor, namely $\boldsymbol{\tau}_{e_1} = \left[\tau_{e_1,1} \quad \tau_{e_1,2}\right]$, $\boldsymbol{\tau}_{u_1} = \left[\tau_{u_1,1} \quad \tau_{u_1,2}\right]$ and $\boldsymbol{\tau}_{u_{2|1}} = \left[\tau_{u_{2|1},1} \quad \tau_{u_{2|1},2}\right]$. These trace plots correspond to one selected simulated dataset and are provided to illustrate mixing of the Markov Chain Monte Carlo while sampling from the posterior density of the corresponding parameter. Chain convergence did not appear to be an issue for any of these parameters.



Web Figure 3.3. Trace plots for the underlying normally distributed variables y^*_{2j} corresponding to the binary response y_{2j} for subjects j=1 and j=200. These trace plots correspond to one selected simulated dataset and are provided to illustrate mixing of the Markov Chain Monte Carlo while sampling from the posterior density of the corresponding underlying normally distributed variables. Chain convergence did not appear to be an issue for either scenario.



Web Figure 3.4. Trace plots for the parameter defining cluster variability for the residual subject-level association between traits, namely σ_m^2, corresponding to one simulated dataset in each of the scenarios considered for $\sigma_m^2 > 0$, namely $\sigma_m^2 = 0.01$, $\sigma_m^2 = 0.1$, $\sigma_m^2 = 1$ and $\sigma_m^2 = 10$. Plots are provided to illustrate Markov Chain Monte Carlo mixing and sampling from the posterior density of σ_m^2. The plots do not provide indication of convergence problems for any of the scenarios considered.

Web Table 3.1. Posterior mean (PMEAN), posterior standard deviation (PSD), 95% highest posterior density (HPD) intervals and effective sample size (ESS) of MCMC samples for residual subject-level and random cluster-level conditional variances for milk yield and pregnancy outcome at 1st postpartum insemination in Michigan dairy cows. As noted in the table, the residual (cow) level variability in milk yield was greater for multiparous compared to primiparous cows, whereas the random (herd) level variability for milk yield and for pregnancy outcome did not differ between herds with a 3^{+}X versus a 2X milking frequency.

Variance Components †	PMEAN	PSD	95% HPD	ESS
Milk yield (kg^2)				
$\sigma^2_{u_1,2X}$	19.8 [x]	2.1	[16.1, 24.1]	22780
$\sigma^2_{u_1,3^+X}$	27.2 [x]	5.0	[18.2, 37.1]	5229
$\sigma^2_{e_1,primiparous}$	40.1 [a]	1.0	[38.2, 42.1]	2956
$\sigma^2_{e_1,multiparous}$	76.1 [b]	1.8	[72.5, 79.6]	2281
σ_{v,e_1}	0.39	0.03	[0.34, 0.44]	5429
Pregnancy outcome (underlying normal scale)				
$\sigma^2_{u_{2\|1},2X}$	0.048 [x]	0.008	[0.034, 0.063]	8 305
$\sigma^2_{u_{2\|1},3^+X}$	0.029 [x]	0.007	[0.016, 0.042]	2 926

([x,y]) and ([a,b]) Letters indicate significant differences (two-tailed Bayesian P-value<0.0001) between management practices within the random (herd) level and residual (cow) level factors, respectively, for each trait.

† $\sigma^2_{u_i,2X} = \exp\left([1 \quad 0]\log\left(\tau_{u_i}\right)\right)$ and $\sigma^2_{u_i,3^+X} = \exp\left([0 \quad 1]\log\left(\tau_{u_i}\right)\right)$ are the random (herd) level conditional variances for milk yield ($i = 1$) and pregnancy outcome ($i = 2|1$) for herds with twice a day (2X) and three times a day (or greater; 3^{+}X) milking frequency, respectively.

$\sigma^2_{e_1,primiparous} = \exp\left([1 \quad 0]\log\left(\tau_{e_1}\right)\right)$ and $\sigma^2_{e_1,multiparous} = \exp\left([0 \quad 1]\log\left(\tau_{e_1}\right)\right)$ are the residual (cow) level variances for milk yield for cows in their first (primiparous) or subsequent (multiparous) lactation, respectively.

$\sigma_{v,e_1} = \frac{1}{\sqrt{\eta_{e_1} - 2}}$ is the e-level coefficient of variation for conditional variances between clusters.

Web Appendix A:

In this appendix, we present extensions of the proposed bivariate GLMM to model heterogeneous covariances between a Gaussian outcome and a non-Gaussian response other than binary, namely ordered categorical, right censored and count Poisson data.

An ordered categorical response y_{2j} can be easily implemented in the proposed bivariate GLMM through data augmentation in a similar way to that described for a binary response, whereby the outcome is determined by discretizing an underlying Gaussian variable y_{2j}^{*}, as proposed by Harville and Mee (1984). That is, for an ordinal trait with l=1,..., L ≥ 3 categories, $y_{2j} = l$ if $\tau_{l-1} < y_{2j}^{*} \leq \tau_l$ where $-\infty = \tau_0 \leq \tau_1 \leq \ldots \leq \tau_{L-1} \leq \tau_L = \infty$ are threshold parameters that define boundaries between categories. A priori, threshold parameters can be considered to be jointly distributed as order statistics from a uniform distribution in the interval $[\tau_0, \tau_L]$. The posterior FCD of the threshold parameters can be shown to have independent uniform distributions of the form $\left(\min\left(\mathbf{y}_2^{*} \mid \mathbf{y}_2 = l+1\right) - \max\left(\mathbf{y}_2^{*} \mid \mathbf{y}_2 = l\right)\right)^{-1}$. As with the binary case, issues of parameter identifiability require constraints on one threshold parameter and on $\sigma^2_{e_{2|1}}$, usually $\tau_1 = 0$ and $\sigma^2_{e_{2|1}} = 1$. Alternatively, $\sigma^2_{e_{2|1}}$ can be explicitly modeled (rather than constrained) provided that an additional threshold, say $\tau_2 > \tau_1$, is fixed (Sorensen et al., 1995). In such case, it would also be possible to infer heterogeneous $\sigma^2_{e_{2|1}}$ with two or more categories (Kizilkaya and Tempelman, 2005). A posteriori, the liabilities follow independent normal distributions truncated at the bin

thresholds, as previously described for the binary case in Section 6. The remaining parameters of the model have FCD identical to those described in the bivariate Gaussian linear model (Bello et al., 2010).

Based on extensions from previous work by Sorensen et al. (1998), data augmentation can also readily accommodate censored responses in the proposed bivariate GLMM. Let the data on the jth subject be $\left(y_{2j},\delta_{2j}\right)$, where δ_{2j} is a censoring indicator and y_{2j} represents the observed value of $Y_{2j}=\min\left(Y_{2j}^{*},C_{2j}\right)$ such that Y_{2j}^{*} is normally distributed and C_{2j} is the point of censoring. For observations for which $Y_{2j}^{*}>C_{2j}$, the response is considered censored $\left(\delta_{2j}=1\right)$ and an augmented variables y_{2j}^{*} is generated from a normal distribution with mean $\left(\mathbf{x}_{(1)j}'\boldsymbol{\beta}_2+\mathbf{z}_j'\mathbf{u}_2+\varphi_j^{(e)}e_{1j}\right)$ and variance $\sigma_{e_{2|1}}^2=\sigma_{e_2}^2-\left(\varphi_j^{(e)}\right)^2\sigma_{e_1}^2$ truncated on the left at the censoring point C_{2j}. The augmented variables y_{2j}^{*} are then fed into the bivariate GLMM, as presented in Equation (1). The remaining model parameters have identical FCD to the bivariate Gaussian linear model (Bello et al., 2010).

For an outcome y_{2j} assumed to follow a Poisson distribution with parameter $\lambda_{2j}>0$, we implement the lognormal link function as per Foulley et al. (1987), whereby we define the auxiliary variable $\eta_{2j}=\ln\left(\lambda_{2j}\right)$. For the proposed bivariate

GLMM, we consider that a priori $\eta_{2j} \sim N\left(\mathbf{x}_{(2)j}'\boldsymbol{\beta}_2 + \mathbf{z}_j'\mathbf{u}_2 + \varphi_j^{(e)} e_{1j}, \sigma_{2|1}^2\right)$. The FCD for the auxiliary variable η_{2j} on the Poisson outcome y_{2j}, conditional on the data and all other model parameters (labeled *ELSE* in the equation below), has the form:

$$f\left(\eta_{2j} \mid ELSE\right) \propto$$

$$\exp\left(-\exp\left(\eta_{2j}\right)\right)\left(\exp\left(\eta_{2j}\right)\right)^{y_{2j}} \exp\left(-\frac{\left(\eta_{2j} - \mathbf{x}_{(2)j}'\boldsymbol{\beta}_2 - \mathbf{z}_j'\mathbf{u}_2 - \varphi_j^{(e)} e_{1j}\right)^2}{\sigma_{2|1}^2}\right) \tag{A1}$$

This density is not immediately recognizable; thus, generation of η_{2j} would require a Metropolis-Hastings (MH) step. An independence chains MH implementation would appear appropriate since the left part of Equation (A1) resembles the kernel of a normal distribution. As an alternative, a random walk MH would also be feasible. Thus generated, the auxiliary variables η_{2j} can then be used in place of y_{2j}^* in the bivariate GLMM presented in Equation (1). The remaining model parameters can be shown to have identical FCD to those in the bivariate Gaussian linear model (Bello et al., 2010), so that the remaining MCMC implementation is straightforward.

References

Albert, J. H., and Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association* **88**, 669-679.

Bello, N. M., Steibel, J. P., and Tempelman, R. J. (2010). Hierarchical Modeling of Random and Residual Variance-Covariance Matrices in Bivariate Mixed Effects Models. *Biometrical Journal.* **52**, 297–313.

Catalano, P. J., and Ryan, L. M. (1992). Bivariate Latent Variable Models for Clustered Discrete and Continuous Outcomes. *Journal of the American Statistical Association* **87**, 651-658.

Chen, Z., and Dunson, D. B. (2003). Random effects selection in linear mixed models. *Biometrics* **59**, 762-769.

Emanuelson, U., and Oltenacu, P. A. (1998). Incidences and effects of diseases on the performance of Swedish dairy herds stratified by production. *Journal Dairy Science* **81**, 2376-2382.

Foulley, J. L., Gianola, D., Cristobal, M. S., and Im, S. (1990). A Method for Assessing Extent and Sources of Heterogeneity of Residual Variances in Mixed Linear-Models. *Journal of Dairy Science* **73**, 1612-1624.

Foulley, J. L., Gianola, D., and Im, S. (1987). Genetic Evaluation of Traits Distributed as Poisson-Binomial with Reference to Reproductive Characters. *Theoretical and Applied Genetics* **73**, 870-877.

Gelfand, A. E., and Dey, D. K. (1994). Bayesian Model Choice - Asymptotics and Exact Calculations. *Journal of the Royal Statistical Society Series B-Methodological* **56**, 501-514.

Gelfand, A. E., and Sahu, K. (1999). Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association* **94**, 247-253.

Gueorguieva, R. V., and Agresti, A. (2001). A correlated probit model for joint modeling of clustered binary and continuous responses. *Journal of the American Statistical Association* **96**, 1102-1112.

Hare, E., Norman, H. D., and Wright, J. R. (2006). Trends in calving ages and calving intervals for dairy cattle breeds in the United States. *Journal Dairy Science* **89**, 365-370.

Harville, D. A., and Mee, R. W. (1984). A Mixed-Model Procedure for Analyzing Ordered Categorical-Data. *Biometrics* **40**, 393-408.

Jaffrezic, F., Thompson, R., and Hill, W. G. (2003). Structured antedependence models for genetic analysis of repeated measures on multiple quantitative traits. *Genetical Research* **82**, 55-65.

Janss, L. L. G., and Foulley, J. L. (1993). Bivariate Analysis for One Continuous and One Threshold Dichotomous Trait with Unequal Design Matrices and an Application to Birth-Weight and Calving Difficulty. *Livestock Production Science* **33**, 183-198.

Kass, R. E., and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association* **90**, 773-795.

Kizilkaya, K., and Tempelman, R. J. (2005). A general approach to mixed effects modeling of residual variances in generalized linear mixed models. *Genetic Selection and Evolution* **37**, 31-56.

Liu, X. F., Daniels, M. J., and Marcus, B. (2009). Joint Models for the Association of Longitudinal Binary and Continuous Processes With Application to a Smoking Cessation Trial. *Journal of the American Statistical Association* **104**, 429-438.

Lof, E., Gustafsson, H., and Emanuelson, U. (2007). Associations between herd characteristics and reproductive efficiency in dairy herds. *Journal Dairy Science* **90**, 4897-4907.

Lopez-Gatius, F., Garcia-Ispierto, I., Santolaria, P., Yaniz, J., Nogareda, C., and Lopez-Bejar, M. (2006). Screening for high fertility in high-producing dairy cows. *Theriogenology* **65**, 1678-1689.

Lucy, M. C. (2001). Reproductive loss in high-producing dairy cattle: where will it end? *Journal of Dairy Science* **84**, 1277-1293.

McCulloch, C. (2008). Joint modelling of mixed outcome types using latent variables. *Statistical Methods in Medical Research* **17**, 53-73.

Najita, J. S., Li, Y., and Catalano, P. J. (2009). A novel application of a bivariate regression model for binary and continuous outcomes to studies of fetal toxicity. *Journal of the Royal Statistical Society Series C-Applied Statistics* **58**, 555-573.

Natarajan, R., and Kass, R. E. (2000). Reference Bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association* **95**, 227-237.

O'Malley, A. J., Normand, S. L. T., and Kuntz, R. E. (2003). Application of models for multivariate mixed outcomes to medical device trials: coronary artery stenting. *Statistics in Medicine* **22**, 313-336.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* **86**, 677-690.

Raftery, A. E., and Lewis, S. (eds) (1992). *How Many Iterations in the Gibbs Sampler?* Oxford, UK: Oxford University Press.

Riley, R. D., Abrams, K. R., Lambert, P. C., Sutton, A. J., and Thompson, J. R. (2007). An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Statistics in Medicine* **26**, 78-97.

Schaeffer, L. R. (2004). Application of random regression models in animal breeding. *Livestock Production Science* **86**, 35-45.

Sorensen, D., and Gianola, D. (2002). *Likelihood, Bayesian and MCMC Methods in Quantitative Genetics*. New York: Springer Verlag.

Sorensen, D. A., Andersen, S., Gianola, D., and Korsgaard, I. (1995). Bayesian-Inference in Threshold Models Using Gibbs Sampling. *Genetics Selection Evolution* **27**, 229-249.

Sorensen, D. A., Gianola, D., and Korsgaard, I. R. (1998). Bayesian mixed-effects model analysis of a censored normal distribution with animal breeding applications. *Acta Agriculturae Scandinavica Section a-Animal Science* **48**, 222-229.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society - Series B* **64**, 1-34.

Tanner, M. A. (1993). *Tools for Statistical Inference*: Springer-Verlag.

Teixeira-Pinto, A., and Normand, S. L. T. (2009). Correlated bivariate continuous and binary outcomes: Issues and applications. *Statistics in Medicine* **28**, 1753-1773.

Tsuruta, S., Misztal, I., Huang, C., and Lawlor, T. J. (2009). Bivariate analysis of conception rates and test-day milk yields in Holsteins using a threshold-linear model with random regressions. *Journal Dairy Science* **92**, 2922-2930.

Wu, X. L., Gianola, D., and Weigel, K. (2009). Bayesian joint mapping of quantitative trait loci for Gaussian and categorical characters in line crosses. *Genetica* **135**, 367-377.

CHAPTER 4

Cows and Herds Constitute Distinct Hierarchical Levels of the Association between Milk Yield and Pregnancy Outcome in Dairy Cows

ABSTRACT

In this study, we investigate heterogeneity in the association between milk yield (MY) and pregnancy outcome (PO) in dairy cows, formally accounting for the *within*-herd (i.e. cow-level) and *between*-herd (i.e. herd-level) hierarchical components of the association. Our ultimate purpose is to provide a general framework for insight on the ongoing controversy on the production-reproduction relationship. We implement our recently developed bivariate hierarchical generalized linear mixed model (Chapter 3) to infer upon heterogeneity in the covariances between MY and PO focusing on these performance outcomes at first postpartum insemination. We also evaluate management practices and herd attributes that may contribute to explain such association heterogeneity. Data consisted of 89,105 DHIA cow records from 379 dairy herds in Michigan. Our hierarchical model naturally accounts for cows and herds as separate levels of the association between MY and PO. The model also considers means, variances and covariances between MY and PO as separate functions of various management practices and herd attributes. Final inferences were based on a best-fitting model selected using Deviance Information Criterion. Within herds, MY and PO were apparently not linked to each other, as the association parameter did not significantly depart from zero across cows in a herd. Among herds, the relationship between MY and PO was antagonistic and dependent on management practices that determine a baseline

level of fertility for the herd. Hence, herds with greater milk yields at the time of first insemination had impaired pregnancy rates, but within such herds, cows with higher daily yields were not any more or less likely to become pregnant than lower yielding herdmates. Michigan counties differed in the magnitude of the herd-level association, thus indicating that regional environmental conditions or management practices may partially alleviate the herd-level antagonism between MY and PO. In summary, the nature of the association between MY and PO in dairy cows is highly heterogeneous due to the hierarchical duality of cow and herd components and to management factors potentially involved in such heterogeneity. Hence, we conclude that the production-reproduction link is not a one-size-fits-all concept. Further research will be required to delineate scenarios conducive to jointly optimize MY and PO performance of dairy cows in commercial farms during first postpartum insemination.

Keyword: dairy cow, herd, milk yield, pregnancy outcome, management.

INTRODUCTION

The nature of the association between milk yield and fertility in dairy cows remains one of the most controversial issues in dairy production systems. The practical implications of this controversy are critical to the efficiency of dairy cow performance and ultimately to sustainability of the dairy business. Indeed, the initiation and renewal of a lactation cycle is determined by the ability of a cow to become pregnant and calve repeatedly during her lifetime, thereby defining a closely intertwined and dynamic relationship between milk production and reproduction. Many reports indicate a progressing antagonism between milk yield and reproductive traits, attributed in part to a

prolonged unilateral focus on maximizing milk yield while failing to give proper attention to fertility (Butler and Smith, 1989, Hare et al., 2006, Lucy, 2001). However, evidence is currently mounting that disputes the perception of a general antagonistic link and rather supports either no relationship or favorable associations between milk production and dairy reproduction (Emanuelson and Oltenacu, 1998, Leblanc, 2010, Lopez-Gatius et al., 2006). A comprehensive assessment is urgently required to clarify the production-reproduction controversy and help delineate management strategies that optimize both aspects of dairy cow performance.

Understanding the shortcoming of the methodologies used this far to assess the production-reproduction association may help explain the current controversy. First, a general under-appreciation of cows and herds as separate units of performance is a common deficiency in many studies that hinders appropriate recognition of sources of variation (i.e. cow-level and herd-level, respectively). Furthermore, herds have been historically incorporated into statistical models as fixed effects as opposed to random blocking factors. Failure to model herds as random blocks narrows the scope of inference by limiting the breadth of inquiry (Tempelman, 2010). Indeed, ignoring the presence of hierarchical sources of random variation has been recognized as "one of the most common and serious mistakes in statistical analysis of data" (Littell et al., 2002) and can lead to severely biased parameter estimates that fail to detect significant sources of variation (St-Pierre, 2001). Herds, along with the cows managed within those herds, represent hierarchical components of the production-reproduction association, each contributing to overall variation in distinctive manners not necessarily alike (Chapter 2).

An additional source of confusion in the production-reproduction controversy is the choice of statistical analysis. Single-trait analyses, despite their high prevalence, are inappropriate to investigate associations between two or more performance outcomes that are of joint interest, as is the case with milk yield and reproduction. The reason for this is the underlying assumption that whichever outcome is used as an explanatory variable is not influenced by other covariates in the model. This is certainly a highly questionable assumption given the *outcome* nature of the responses of interest. Instead, the infrequently-implemented multiple-trait analysis is appropriate in jointly evaluating multiple outcomes, whereby the magnitude of relationship between these outcomes can be explicitly captured through covariance parameters.

In this manuscript, we formally address the aforementioned limitations by means of a recently developed bivariate generalized hierarchical modeling framework that enables inference on heterogeneous covariances (Chapter 3). Development of this statistical methodology makes the study herein particularly timely to evaluate heterogeneous associations in the production-reproduction controversy. In this study, we investigate the cow-level (i.e. *within*-herd) and herd-level (i.e. *between*-herd) associations between daily milk yield (MY) and pregnancy outcome (PO) in Michigan dairy cows. We also assessed management practices and herd attributes as potential sources of heterogeneity on these associations. Our focus was on performance at first postpartum insemination, as this point in lactation is a unique opportunity to evaluate the production-reproduction association when daily yield is at or near maximum and fertility is first tested.

MATERIALS AND METHODS

Data Description

Data files for test-day records for Michigan dairy farms enrolled in the Dairy Herd Improvement program were obtained from Dairy Records Management Systems (DHIA; Raleigh, NC). Lactation records from first, second and third parity Holstein cows that calved between January 2005 and December 2006 were extracted. Herds were required to have at least 25 total cows per year and at least one breeding reported for a minimum of 50% of the herd each year. Only first postpartum inseminations occurring between 30 and 200 days after calving were considered for analysis. To ensure quality of milk yield data, yield records were required for the test-dates immediately prior and immediately after the recorded date of first insemination. The interval between these test-dates was required to be no greater than 35 days. All lactation records used in this study were required to have complete records on cow and herd identification as well as for the response variables of interest and potentially important explanatory variables, as described later. After editing, the total number of lactation records available for analysis was 89,105 corresponding to 74,745 cows from 379 dairy herds.

Dependent variables considered in this study were y_1 = daily MY at first postpartum insemination, expressed in kg; and y_2 = PO to first postpartum insemination, expressed in a binary scale (i.e. pregnant/not pregnant). Here, MY at first insemination was computed as a linear interpolation between yield records on the test-date immediately before and the test-date immediately after first postpartum insemination for a given lactation. A cow was considered pregnant to first postpartum insemination if no subsequent inseminations were recorded on that lactation and first insemination was

followed by calving after a gestation length within the range of 280 d ± 3 SD. The estimated standard deviation for gestation length, SD = $\sqrt{39.2} = 6.3$ d, was computed as the square root of the corresponding estimated phenotypic variance (Jamrozik et al., 2005). Records for second or later postpartum inseminations or a gestation length greater than specified above, were considered an indication of failure to establish a pregnancy to first postpartum insemination. If gestation length was shorter than the lower threshold specified above, the record was considered technically flawed (i.e., date recording error) and excluded from analysis. For cows recorded to have only one postpartum insemination, and that were removed from the herd for non-reproductive reasons before their expected calving date, information on transrectal pregnancy diagnosis (whenever available) was used to assess pregnancy outcome to first insemination. First postpartum insemination is the earliest opportunity to establish a pregnancy after the voluntary waiting period. Therefore, success to first insemination could be considered a suitable indicator of fertility. Alternatively, with increased number of postpartum inseminations, the subset of cows that remain eligible for breeding may have a higher prevalence of reproductive, health or nutritional problems, which may in turn confound the evaluation of fertility and its association with milk yield. Furthermore, in standard management conditions, first postpartum insemination is expected to occur in close proximity to the highest daily MY records of a lactation. These circumstances render first postpartum insemination as a particularly relevant point in time to examine the nature of the association between milk production and fertility in dairy cows. Similarly, we perceive considerable potential for impact of management practices and herd issues around this time postpartum.

Four calving seasons were defined based on the month of calving when lactation was initiated: Fall, from September through November; Winter, from December through February; Spring, from March through May; and Summer, from June through August. Information on selected management practices and herd descriptors was also gathered from the DHIA dataset as potential explanatory variables. These included herd milking frequency (i.e. 2 times per day versus 3 or more times per day), herd involvement with bovine somatotropin (bST) (i.e. non-users, having 0% of the herd enrolled; intermediate users, with >0-50% of the herd enrolled; and committed users, with ≥50% of the herd enrolled), individual cow supplementation with bST during a lactation (i.e. yes or no), herd size (expressed on the log base 10 scale and as a deviation from its mean) and herd size expansion (expressed as the percentage change in herd size from the preceding year). In addition, the use of synchronized breeding was considered, being defined on a herd-year basis using the adjusted Chi-square categorization method proposed by Miller et al. (2007), whereby herds were classified as either having synchronized breeding or not.

Deciding which of these factors were to be incorporated as explanatory variables, and at what level of the hierarchical model, was based on a sequential model selection approach described below.

Animal Care and Use Committee approval was not obtained for this study because the data were obtained from an existing performance records database.

Model specification and posterior inference

We specify the statistical model for analysis using the bivariate generalized hierarchical Bayesian approach recently developed by our research group (Chapter 3 in this dissertation). For this study, cow-specific records capture the residual subject level of

the hierarchy, while herd-year clusters or contemporary groups represent the random cluster level. Take y_{1j} to be the observed MY and y_{2j} to be the observed PO on subject j; $j=1,\ldots,n$. As previously proposed (Albert and Chib, 1993, Sorensen and Gianola, 2002), we assume y_{2j} to be determined by an underlying normally distributed latent variable y_{2j}^{*} such that $y_{2j} = I\left(y_{2j}^{*} > 0\right)$ for $I(.)$ representing the indicator function. This is equivalent to specifying a probit link in a generalized linear mixed model (GLMM) for y_{2j}. The underlying bivariate GLMM, as per Chapter 3 of this dissertation, is then

$$\begin{bmatrix} y_{1j} \\ y_{2j}^{*} \end{bmatrix} = \begin{bmatrix} \mu_{1j} + e_{1j} \\ \mu_{2j} + e_{2j} \end{bmatrix} \quad \text{where} \quad \begin{bmatrix} \mu_{1j} \\ \mu_{2j} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{j(1)}'\boldsymbol{\beta}_1 + \mathbf{z}_j'\mathbf{u}_1 \\ \mathbf{x}_{j(2)}'\boldsymbol{\beta}_2 + \mathbf{z}_j'\mathbf{u}_2 \end{bmatrix}. \tag{1}$$

Here, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are p_1 x 1 and p_2 x 1, respectively, vectors of selected classical fixed effects whereas $\mathbf{u}_1$ and $\mathbf{u}_2$ are each q x 1 vectors of classical random effects of herd-year with subscripts denoting the outcome (1 for MY, 2 for PO). The known incidence vectors $\mathbf{x}_{j(1)}'$, $\mathbf{x}_{j(2)}'$ and $\mathbf{z}_j'$ connect these effects to their respective responses and are specific to cow j. Similarly, $e_{1,i}$ and $e_{2,i}$ are residual effects on the corresponding response variables specific to the jth subject (cow). Each pair of random effects specific to the kth herd-year cluster, namely $\mathbf{u}_{.k} = \left[u_{1k} \;\; u_{2k}\right]'$, and each pair of residual effects specific to the jth cow subject, namely $\mathbf{e}_{.j} = \left[e_{1j} \;\; e_{2j}\right]'$, are mutually independent with null means and (co)variances defined by:

$$\mathbf{G}_k = \text{var}\begin{bmatrix} u_{1,k} \\ u_{2,k} \end{bmatrix} = \begin{bmatrix} \sigma^2_{u_1 k} & \sigma_{u_{12}k} \\ \sigma_{u_{12}k} & \sigma^2_{e_2 k} \end{bmatrix} \text{ and } \mathbf{R}_j = \text{var}\begin{bmatrix} e_{1,j} \\ e_{2,j} \end{bmatrix} = \begin{bmatrix} \sigma^2_{e_1 j} & \sigma_{e_{12}j} \\ \sigma_{e_{12}j} & \sigma^2_{e_2 j} \end{bmatrix}. \quad (2)$$

Here $\sigma^2_{u_1 k}$ and $\sigma^2_{u_2 k}$ are the random effects (i.e. herd-level) variances in MY and PO, respectively, and $\sigma_{u_{12}k}$ is the corresponding random effects covariance between the two traits, specific to the kth herd. Similarly, $\sigma^2_{e_1 j}$ and $\sigma^2_{e_2 j}$ represent the residual (i.e. cow-level) variances for MY and PO, respectively, with $\sigma_{e_{12}j}$ being the corresponding residual covariance between the outcomes for the jth cow.

Following the methodological developments proposed in Chapter 3 of this dissertation, the bivariate posterior joint density of the complete data (as implied by the GLMM presented in Equation (1)) is factorized into a marginal component and a conditional component; this is equivalently to specifying a square-root free Cholesky decomposition on $\mathbf{R}_j$ and $\mathbf{G}_k$ (Bello et al., 2010; Chapter 1) that results in the following reparameterization:

$$\mathbf{R}_j = \begin{bmatrix} \sigma^2_{e_1 j} & \varphi^{(e)}_j \sigma^2_{e_1 j} \\ \varphi^{(e)}_j \sigma^2_{e_1 j} & \sigma^2_{e_{2|1}j} + \left(\varphi^{(e)}_j\right)^2 \sigma^2_{e_1 j} \end{bmatrix} \text{ and } \mathbf{G}_k = \begin{bmatrix} \sigma^2_{u_1 k} & \varphi^{(u)}_k \sigma^2_{u_1 k} \\ \varphi^{(u)}_k \sigma^2_{u_1 k} & \sigma^2_{u_{2|1}k} + \left(\varphi^{(u)}_k\right)^2 \sigma^2_{u_1 k} \end{bmatrix}. \quad (3)$$

Here, $\varphi^{(e)}_j = \dfrac{\sigma_{e_{12}j}}{\sigma^2_{e_1 j}}$ represents a residual cow-specific regression coefficient for $e_{2,j}$ on $e_{1,j}$ such that $e_{2j} = \varphi^{(e)}_j e_{1j} + e_{2|1j}$ where $e_{2|1j} \sim N\left(0, \sigma^2_{e_{2|1}j}\right)$ is conditionally

independent of e_{1j}. Then, $\sigma^2_{e_{2|1}j} = \sigma^2_{e_2 j} - \left(\varphi_j^{(e)}\right)^2 \sigma^2_{e_1 j}$ represents the cow-level variance of y^*_{2j} conditional on y_{1j} and is constrained to be equal to 1 for all cows j=1,...n, to ensure identifiability of the remaining GLMM parameters, as described in Chapter 3. Similarly, $\varphi_k^{(u)} = \frac{\sigma_{u_{12}k}}{\sigma^2_{u_1 k}}$ represents a random herd-specific regression coefficient of u_{2k} on u_{1k}, such that $u_{2k} = \varphi_k^{(u)} u_{1k} + u_{2|1k}$ where $u_{2|1k} \sim N\left(0, \sigma^2_{u_{2|1}k}\right)$ is conditionally independent of u_{1k} and $\sigma^2_{u_{2|1}k} = \sigma^2_{u_2 k} - \left(\varphi_k^{(u)}\right)^2 \sigma^2_{u_1 k}$ is the herd-level variance of u_{2k} conditional on u_{1k}. Note that the regression coefficients $\varphi_j^{(e)}$ and $\varphi_k^{(u)}$ describe the cow- and herd-level components of the association between MY and PO to first postpartum insemination. These parameters, along with the logarithms of $\sigma^2_{e_1 j}$, $\sigma^2_{u_1 k}$ and $\sigma^2_{u_{2|1}k}$, are completely unconstrained without compromising positive semi-definiteness of $\mathbf{R}_j$ and $\mathbf{G}_k$. Therefore, we model heterogeneity of associations between MY and PO by specifying a liner mixed model on $\varphi_j^{(e)}$ and $\varphi_k^{(u)}$, as follows:

$$\varphi_j^{(e)} = \mathbf{x}'_{j(3)}\boldsymbol{\gamma}_e + \mathbf{z}_j'\mathbf{m}_e \tag{4}$$

$$\varphi_k^{(u)} = \mathbf{x}'_{k(4)}\boldsymbol{\gamma}_u + \mathbf{w}'_k\mathbf{m}_u \tag{5}$$

Here, $\boldsymbol{\gamma}_e$ and $\boldsymbol{\gamma}_u$ represent p_3 x 1 and p_4 x 1, respectively, vectors of unknown fixed effects with corresponding known incidence row vectors $\mathbf{x}'_{j(3)}$ and $\mathbf{x}'_{k(4)}$; $\mathbf{m}_e$ represents

a q x 1 vector of unknown random herd-year specific effects on the cow-level association such that $\mathbf{m}_e \sim N\left(\mathbf{0}, \mathbf{I}\sigma^2_{m_e}\right)$ with associated known incidence row vector $\mathbf{z}'_j$. In turn, $\mathbf{m}_u$ represents a r x 1 vector of unknown random county-specific effects such that $\mathbf{m}_u \sim N\left(\mathbf{0}, \mathbf{I}\sigma^2_{m_u}\right)$, with $\mathbf{w}'_k$ being the corresponding known row incidence vector for herd k. We borrow the terms "fixed effects" and "random effects" from the classical linear mixed effects model framework, whereby "fixed effects" refer to the effects of systematic management factors that can be subsequently inferred upon in other studies and "random effects" pertains to factors of potentially exchangeable effects that can be characterized by a distribution (Robinson, 1991).

We also model heterogeneity of variances (i.e. $\sigma^2_{e_1 j}$, $\sigma^2_{u_1 k}$ and $\sigma^2_{u_{2|1} k}$) as separate functions of fixed and random effects (Foulley et al., 1990) using the logarithmic link function:

$$\log\left(\sigma^2_{e_1 j}\right) = \mathbf{x}'_{j(5)} \log\left(\boldsymbol{\tau}_{e_1}\right) + \mathbf{z}'_j \log\left(\mathbf{v}_{e_1}\right), \tag{6}$$

$$\log\left(\sigma^2_{u_1 k}\right) = \mathbf{x}'_{k(6)} \log\left(\boldsymbol{\tau}_{u_1}\right) + \mathbf{w}'_k \log\left(\mathbf{v}_{u_1}\right), \quad \text{and} \tag{7}$$

$$\log\left(\sigma^2_{u_{2|1} k}\right) = \mathbf{x}'_{k(7)} \log\left(\boldsymbol{\tau}_{u_{2|1}}\right) + \mathbf{w}'_k \log\left(\mathbf{v}_{u_{2|1}}\right). \tag{8}$$

Here $\boldsymbol{\tau}_{e_1}$, $\boldsymbol{\tau}_{u_1}$ and $\boldsymbol{\tau}_{u_{2|1}}$ represent p_5 x 1, p_6 x 1 and p_7 x 1 vectors of unknown fixed effects with $\mathbf{x}'_{j(5)}$, $\mathbf{x}'_{k(6)}$ and $\mathbf{x}'_{k(7)}$ being the corresponding known incidence vector. Furthermore, $\mathbf{v}_{e_1}$ represents a q x 1 vector of unknown random herd-year-specific effects, each specified by independent inverted gamma priors

$v_{e_1 j} \mid \eta_{e_1} \sim IG\left(\eta_{e_1}, \eta_{e_1} - 1\right)$ as in Chapter 3, with $\mathbf{z}'_j$ being the corresponding known incidence vector. In turn, $\mathbf{v}_{u_1}$ and $\mathbf{v}_{u_{2|1}}$ represent r x 1 vectors of unknown random county-specific effects on the herd-level variances, each specified with independent inverted-gamma priors as $v_{u_1 k} \mid \eta_{u_1} \sim IG\left(\eta_{u_1}, \eta_{u_1} - 1\right)$ and $v_{u_{2|1} k} \mid \eta_{u_{2|1}} \sim IG\left(\eta_{u_{2|1}}, \eta_{u_{2|1}} - 1\right)$, with known incidence vector $\mathbf{w}'_k$. As anticipated, the constraint $\sigma^2_{e_{2|1} j} = 1 \ \forall j$ needs to be imposed to ensure parameter identifiability.

At this point, our proposed bivariate GLMM accommodates 7 submodels that specify heterogeneity at different hierarchical levels; namely Equation (1) defines 2 linear models on location parameters, one per outcome of interest; Equations (4) and (5) each specify a linear model on the cow-level and herd-level associations, respectively, between MY and PO; and Equations (6), (7) and (8) define a total of 3 linear mixed model specifications on conditional variances, one at the cow-level and 2 at the herd-level.

Prior specifications were the same as those used for the simulation study described in Chapter 3 of this dissertation. Briefly, flat unbounded priors were specified on γ_e, $\sigma^2_{m_e}$, γ_u, $\sigma^2_{m_u}$, as well as for $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\tau}_{e_1}$, $\boldsymbol{\tau}_{u_1}$ and $\boldsymbol{\tau}_{u_{2|1}}$. For each herd-specific pair of elements from ($\mathbf{u}_1$, $\mathbf{u}_2$), we specified an independent normal prior density with null mean and variance-covariance matrix $\mathbf{G}_j$. Also, we adopted $\mathbf{m}_e \sim N\left(\mathbf{0}, \mathbf{I}\sigma^2_{m_e}\right)$ and $\mathbf{m}_u \sim N\left(\mathbf{0}, \mathbf{I}\sigma^2_{m_u}\right)$. For $\mathbf{v}_{e_1}$, $\mathbf{v}_{u_1}$ and $\mathbf{v}_{u_{2|1}}$, we specified inverse gamma priors with

shape and scale parameters $\left(\eta_{e_1}, \eta_{e_1} - 1\right)$, $\left(\eta_{u_1}, \eta_{u_1} - 1\right)$ and $\left(\eta_{u_{2|1}}, \eta_{u_{2|1}} - 1\right)$, respectively. In turn, each η_l is characterized by a vague, though proper, prior density given by $\eta_l \propto \left(1+\eta_l\right)^{-2}$; for $\eta_l > 0$ and $l = e_1, u_1, u_{2|1}$ (Kizilkaya and Tempelman, 2005). Furthermore, we imposed standard linear model restrictions on elements of the "fixed" effects (β_1, β_2, γ_e, γ_u, τ_{e_1}, τ_{u_1} and $\tau_{u_{2|1}}$) to ensure identifiability of the parameters, as per Bello et al. (2010; Chapter 1)

For the final selected model (see later), we ran the MCMC chain for 5,000 burn-in cycles followed by 280,000 iterations that were saved for inference. MCMC convergence and sampling diagnostics were monitored graphically using trace plots and autocorrelation plots, and also following Raftery and Lewis (1992). In addition, we report the effective sample size (ESS) as a measure of the number of effectively independent samples or Monte Carlo error amongst the 280,000 dependent MCMC samples (Sorensen et al., 1995).

We summarize posterior densities for each parameter of interest using posterior means, posterior standard deviations and the 95% highest posterior density intervals (HPD). For any comparison of interest between two parameters, say generically θ_1 and θ_2, we also report the Bayesian *P*-value defined as: *P*-value $=2\min\left(\Pr\left(\theta_1 - \theta_2 \geq 0 \mid \mathbf{y}\right), \Pr\left(\theta_1 - \theta_2 < 0 \mid \mathbf{y}\right)\right)$. For parameters γ_e and γ_u, the null hypothesis evaluates the difference between factor levels against a contrast null value that is equal to zero. In turn, for τ_{e_1}, τ_{u_1} and $\tau_{u_{2|1}}$, the null hypothesis evaluates the ratio between factor levels against a contrast null value equal to one.

Model selection

Competing models were compared using the Deviance Information Criteria (DIC) (Spiegelhalter et al., 2002) as an indicator of model fit. Smaller values of DIC are indicative of improved fit, and generally, DIC differences of 7 or greater are believed to indicate a decisive difference in model fit (Spiegelhalter et al., 2002). Table 1 lists the systematic fixed effects factors and covariates (i.e., management practices and herd attributes) that were considered for inclusion into the cow-level and herd-level of the hierarchical linear model. As indicated in Equations (4) and (6), the random effect of herd-year cluster was evaluated as a source of variability in cow-level (co)variances; similarly, random county-specific effects were considered in the modeling of herd-level (co)variances, as per Equations (5), (7) and (8).

The classical fixed effect factors for modeling location parameters β_1 and β_2 always included the effects of parity, calving season and year; in addition, β_1 included Legendre polynomials of order 5 on days in milk and β_2 included linear and quadratic polynomial function of days in milk. These effects were not of primary inferential interest in themselves, thus were not subjected to model selection. That is, we preferred an overspecified model for Equation (1) to provide robust inference for fixed and random effects in the 5 other "mixed effects" submodels. That is, our primary objective was to characterize sources of heterogeneity on the associations between MY and PO (as per Equations (4) and (5)) as well as on their variability (as per Equations (6), (7) and (8)).

We implemented DIC-based model selection in a forward stepwise manner, such that each factor and covariate was evaluated one at a time for model inclusion based on

their contribution to model fit, as described in Chapter 2 of this dissertation. We first selected univariate best-fitting models, one for each of MY and PO. For MY, selection for fixed effects factors influencing the cow-level variance in Equation (6) and the herd-level variance in Equation (7) consisted of four steps, as depicted in Table 2. For PO, selection of fixed and random effects was restricted to the herd-level variance due to identifiability constraints on the cow-level variance, as previously explained (i.e. $\sigma^2_{e_{2|1}j} = 1$ for all j). On the herd-level variance, none of the factors and covariates in Table 1 were identified to improve DIC-based model fit, thereby rendering a final univariate model for PO with homogeneous between-herd variance for all k herds (i.e. $\sigma^2_{u_{2|1}k} = \sigma^2_{u_{2|1}}$).

The univariate models thus selected were then connected as a null bivariate model to further investigate the cow-level and herd-level associations between MY and PO as per Equations (4) and (5). Table 3 provides step-by-step details and final outcomes on the selection of fixed effects and random explanatory factors on the cow-level and herd-level regression coefficients. Our inference on heterogeneity of reparameterized covariances should be strengthened by our sequential approach to model selection since these inferences are already based on important sources of heterogeneity of variances in Equations (6), (7) and (8).

RESULTS AND DISCUSSION

Associations between Milk Yield and Pregnancy Outcome at First Postpartum Insemination.

Steps of model selection on cow-level and herd-level regression coefficients describing the association between MY and PO are shown in Table 3. None of the fixed effects listed in Table 1 were recognized by DIC-based model selection as sources of heterogeneity on the cow-level or on the herd-level association. Therefore, posterior inference is based on overall means for each regression coefficients, namely $\varphi_j^{(e)} = \varphi^{(e)}$ for all j, and $\varphi_k^{(u)} = \varphi^{(u)}$ for all k.

Posterior inference on the cow-level and herd-level associations between MY and PO, expressed in the underlying liability scale, is shown in Table 4. Within herds, cows showed no significant association between MY and PO at first postpartum insemination, as indicated by a 95% HPD on $\varphi^{(e)}$ that overlapped with zero. Thus, there was no evidence that cows with higher daily yields were more or less likely to become pregnant to first insemination than lower yielding herdmates. Conversely, the herd-level regression coefficient indicated evidence for an overall antagonism between herd milk yield and herd fertility, based on negative values of the lower and upper boundaries of the 95% HPD on $\varphi^{(u)}$ (i.e. [-0.017, -0.005]). This suggests that, on average, herds with greater daily milk yields had impaired pregnancy rates to first postpartum insemination. The cow-herd duality of the association between MY and fertility may be indicative of different underlying mechanisms at each dimension of performance, as supported by previous work from our group (Chapter 2 of this dissertation). Indeed, physiological mechanisms in the cow may not necessarily align with the workings of the managerial business unit (i.e. herd) within which cows are handled.

In the previous paragraph, we summarized a conceptual interpretation of $\varphi^{(e)}$ and $\varphi^{(u)}$ in their respective liability scales. In turn, interpreting these parameters in a probability-of-pregnancy scale is likely to be more meaningful from an application perspective. However, the probit link function implemented in Equation (1) hinders a straightforward interpretation of the regression coefficients in the observed probability scale. As proposed in Chapter 3, we translate the herd-level association parameter (i.e. $\varphi^{(u)}$) into an expected herd-level differential in pregnancy rates, namely $\Delta_u = \Phi\left(\mu_2 + \varphi^{(u)} u_1\right) - \Phi(\mu_2)$, whereby Δ_u is dependent upon a baseline herd fertility, expressed as $\Phi(\mu_2)$. Figure 1A illustrates the herd-level differential pregnancy rate Δ_u at the posterior mean for $\varphi^{(u)}$ (as per Table 4) and evaluated over a grid of plausible values of baseline herd fertility $\Phi(\mu_2)$ and plausible relative herd milk yields u_1. Note that these dependencies introduce heterogeneity on the probability-scaled Δ_u despite homogeneous $\varphi^{(u)}$ on the liability scale. Also apparent from Figure 1A is the non-proportional nature of the probit link function, which yields a symmetric curvature of the regression surface along the axis of baseline fertility. As a consequence, Δ_u is maximum (minimum) at a herd baseline fertility of 50% pregnancy rate to first postpartum insemination for any herd with above (below) average daily milk yield and Δ_u decreases (increases) symmetrically as the herd baseline fertility departs from 50% pregnancy rate. Figure 1B illustrates the two-dimensional association between Δ_u and relative herd milk yields for arbitrarily selected values of herd baseline fertility, namely 10, 30 and 50%

pregnancy rate to first postpartum insemination. Note that the association depicted in Figure 1B is also non-linear, though the probit-induced curvature is much alleviated in these circumstances. An examination of Figure 1B reveals that, overall, the herd-level differential pregnancy rate Δ_u decreased with increasing relative herd milk yield, thus supporting an antagonism between MY and PO. However, this antagonism was most pronounced (i.e. steeper rate of change) for herds with a baseline fertility of 50% pregnancy rate and was partially curbed (i.e. relatively more moderate rate of change) as baseline herd fertility departed from 50%. Days in milk at breeding are known to play a role in determining baseline fertility, whereby conception rates rise as the first half of lactation proceeds (Huang et al., 2009, Tsuruta et al., 2009). As a consequence, management strategies that define timing of first postpartum insemination may be partially responsible for the observed differences in intensity of the antagonism between milk yield and fertility at first insemination. Furthermore, the correlation between test-day milk yield and fertility traits is known to fluctuate during lactation (Berry et al., 2003, Tsuruta et al., 2009). Hence, a decision on duration of the voluntary waiting period, combined with synchronization strategies for high insemination rate soon thereafter, are likely to determine herd baseline fertility and milk yield at first postpartum insemination. As a result, the timing of these decisions may help explain heterogeneity in the production-reproduction association as per the observed differences in rates of change in herd pregnancy rate per kg of milk yield at first postpartum insemination.

Interestingly, the sign of the cow-level and herd-level association parameters between milk yield and pregnancy outcome at first insemination indicate a type of relationship that is contrary to that previously estimated between milk production and

reproductive performance on a whole-lactation basis (Chapter 2). In that study, the association between 305-d cumulative milk yield and calving interval was estimated to be favorable at the herd-level but antagonistic at the cow-level. However, the discrepancy of results should not be entirely surprising given that first insemination pregnancy rate and milk yield indicate performance at a discrete and singular point in time, whereas 305-d milk yield and calving interval represent cumulative appraisals over a much longer period of time. Again, the mechanisms underlying the short term relationships between milk yield and pregnancy outcome may not necessarily be the same as those involved in the long term (i.e. entire lactation) association. If anything, these results indicate that the relationship between milk production and reproductive performance of dairy cows is even more complex than anticipated previously. Multiple and varied components appear to contribute to the observed heterogeneity, including the dual cow-herd dimensions and highly diverse and dynamic herd management scenarios evaluated over short or long-term periods of time.

Regional Patterns in the Association between MY and PO.

Differences between Michigan counties were apparent in the magnitude of the herd-level regression coefficient that describes the association between MY and fertility of dairy cows among herds, as indicated in Table 3. The magnitude of this heterogeneity between Michigan counties is quantified by the variance component $\sigma^2_{m_u}$, for which we report a posterior inference summary (Table 4). Based on the Empirical Rule (Ott and Longnecker, 2001) and using the posterior mean $1.3*10^{-4}$ (reproductive liability points

per kg)^2 as a point estimate for $\sigma^2_{m_u}$, one might anticipate a range of $\pm 2 \cdot \sqrt{1.3 * 10^{-4}} = 0.0456$ (reproductive liability points per kg milk yield) between the most extreme county-specific effects. If we assume known $\sigma^2_{m_u}$ and multivariate normal county-specific herd-level associations centered at its posterior mean of -0.011, then counties would be expected to have relationship parameters ranging from -0.034 to 0.012 reproductive liability points per kg increase in milk yield at first postpartum insemination. The upper limit of this range being positive indicates that for a subset of Michigan counties, herds with higher daily milk yields are also expected to have greater pregnancy rates to first insemination, thus effectively counteracting the overall relationship antagonism reported among herds. Specific sources of heterogeneity among counties in the nature of the herd-level relationship between MY and PO are unclear and will require further investigation. In addition to typical environmental factors, such as weather and soil conditions, it may be of interest to evaluate differences between counties in the cultural legacy and profit priority of dairy farming in a given area.

Variability in Milk Yield at First Postpartum Insemination.

Table 5 summarizes the posterior inference on variance components for MY. At the herd-level, the MY variance was 22.6±1.3 kg^2, as estimated by the posterior mean and posterior standard deviation. Within herds, multiple explanatory factors and covariates were identified as potential sources of cow-level heteroskedasticity for MY. In particular, MY at first insemination was ~89% more dispersed in multiparous compared to primiparous herdmates (P-value<0.0001). Cow-level heteroskedasticity was also

associated with calving season (P-value=0.005), whereby cows calved in the Summer had the most homogeneous MY in contrast to cows calved in either Winter or Spring, which had the most unstable yield responses. Fall calving was characterized by intermediate variability in MY, relative to the aforementioned seasons. Furthermore, the consistency between cows in milk yield performance was linked to herd size (P-value<0.0001; Figure 2). As herd size increased by a factor of 10 (i.e. from 25 to 250, from 250 to 2500), the variability in MY between cows increased by ~82%. Finally, the consistency of cow MY at first postpartum insemination differed substantially among herds. This herd-specific heterogeneity was quantified by the coefficient of variation (CV) $\sigma_{v,e_1} = \frac{1}{\sqrt{\eta_{e_1} - 2}}$, the posterior means of which indicated a CV of cow-level variances for MY of approximately 29% (95% HPD = [0.27, 0.32]). Indeed, the posterior means for cow-level variances in MY at first insemination for the most and least variable herds, relative to a typically variable herd (=1), were 2.06 and 0.48, respectively. Hence, the estimated ratio between extreme herds with the largest and the smallest herd-specific relative variances was 4.3. The reasons for such differences between herds in the consistency of their cow performance are not readily apparent and will require investigation of additional features of management and herds that were beyond the scope of the current study. However, the high magnitude of heterogeneity of variances suggests the need to explicitly model heteroskedasticity in the analysis of dairy data, even if not of direct interest to the research, in order to attain relevant inference.

Variability in Pregnancy Outcome to First Postpartum Insemination.

Modeling of the conditional variability in PO was limited to the herd-level dimension due to identifiability constraints on the corresponding cow-level variance (i.e. $\sigma^2_{e_{2|1}j} = 1$ for all j), as required by the probit link implementation in Equation (1). The herd-level conditional variance on pregnancy outcome, expressed as a posterior mean in the liability scale, was 0.047 (liability points)^2 with 95% HPD = [0.039, 0.055]. The DIC did not support any of the herd-level explanatory variables in Table 1 as potential sources of herd-level heteroskedasticity.

CONCLUSIONS

This study provides evidence for strong heterogeneity in the nature and magnitude of the association between milk yield and pregnancy outcome at first postpartum insemination. A dual cow-herd dimension and various management factors are clearly involved in such heterogeneity, thus substantiating the idea that the production-reproduction link is not a one-size-fits-all concept. Indeed, herds do not necessarily mimic the performance capacity of the dairy cow when looking jointly at milk production and reproduction. As a result, it is essential that, in analyzing dairy data, cows and herds are explicitly recognized as separate, though nested, sources of variation. Explicit acknowledgement of the cow-herd duality is a key feature of hierarchical modeling, which we thus recommend. Furthermore, our results indicate that management practices may partially describe heterogeneity in the performance link between first insemination milk yield and pregnancy outcome, such that it may be possible to elicit management scenarios conducive to jointly optimizing milk production and reproductive performance of dairy cows. Ultimately, the goal of this line of research is to elicit guidelines that dairy

managers can implement to jointly optimize milk yield and reproductive efficiency of dairy cows in commercial operations. Further research will be needed to provide further insight on other practices of interest that were not considered in this study.

ACKNOWLEDGMENTS

We thank Dr. John Clay and his staff at the Dairy Records Management Systems, Raleigh, NC for providing the DHIA dataset used in this study. This study was partially funded by the Elwood Kirkpatrick Dairy Research Endowment, the Michigan Milk Producers Association, the College of Agriculture and Natural Resources and the Department of Animal Science at Michigan State University.

Table 4.1. List of fixed effects (classification factors and linear regression on covariates) evaluated as explanatory variables for heterogeneity of cow-level and herd-level (co) variances on milk yield and pregnancy outcome to first postpartum insemination in Michigan dairy cows.

Cow-level (co)variability

- Days in milk at first postpartum insemination (days)
- Parity (Primi- vs. Multiparous)
- Calving season (Winter, Spring, Summer, Fall)
- Year (2005, 2006)
- Milking frequency (2 vs 3+ times per day)
- Individual cow treatment with bovine somatotropin during lactation (Yes/No)
- Level of herd supplementation with bovine somatotropin (0%, >0 to 50% and >50% of the herd)
- Reproductive management practices: Use of synchronization strategies (Yes/No)
- Herd size (number of heads)
- Herd expansion (% change in herd size from preceding year)

Herd-level (co)variability

- Calving season (Winter, Spring, Summer, Fall)
- Year (2005, 2006)
- Milking frequency (2 vs 3+ times per day)
- Level of herd supplementation with bovine somatotropin (0%, >0 to 50% and >50% of the herd)
- Reproductive management practices: Use of synchronization strategies (Yes/No)
- Herd size (number of heads)
- Herd expansion (% change in herd size from preceding year)

Table 4.2. Sequential details of the forward model selection procedure implemented on variance components for a univariate model on test-day milk yield at first postpartum insemination of Michigan dairy cows. Selection of fixed and random effects into the model was based on model fit as determined by Deviance Information Criteria (DIC).

	DIC difference	
Factors and covariates entering the model:	**Relative to Null Model**	**Relative to Model in Preceding Step**
Null Model, consisting of: • Fixed effects on the mean, including parity, calving season, year and Legendre polynomials of order 5 on days in milk, as per Table 1; and • Random clustering effect of herd on the mean.	**0**	.
Step 1: Evaluation of fixed effects on the cow-level variance		
1.1) Parity (Primi- vs. Multiparous)	-3530	-3530
1.2) Calving season (Winter, Spring, Summer, Fall)	-3622	-92
1.3) Herd size (number of heads)	-3678	-56
No additional effects entered the model	.	.
Step 2: Evaluation of random effects on the cow-level variance		
2.1) Clustering effect of herd	-6804	-3126
Step 3: Evaluation of fixed effects on the herd-level variance		
No effects entered the model	.	.
Step 4: Evaluation of random effects on the herd-level variance		
No effects entered the model	.	.

Table 4.3. Sequential details of the forward model selection procedure implemented on the Cholesky-reparameterized covariances (expressed as regression coefficients) between milk yield and pregnancy outcome to first postpartum insemination in Michigan dairy cows. Selection of fixed and random effects into the model was based on model fit as determined by Deviance Information Criteria (DIC).

	DIC difference	
Factors and covariates entering the model:	**Relative to Null Model**	**Relative to Model in Preceding Step**
Null Model, consisting of: • Univariate model on milk yield at first postpartum insemination, as per Table 2. • Univariate model on pregnancy outcome to first postpartum insemination, including systematic effects on the mean (parity, calving season, year and linear and quadratic polynomials on days in milk), a random clustering effect of herd on the mean, a homogeneous *herd*-level variance (as per DIC-based model selection), and a cow-level variance fixed at 1 for reasons of parameter identifiability. • Covariances between traits are modeled as homogeneous and estimated accordingly.	0	.
Step 1: Evaluation of fixed effects on the cow-level regression coefficient		
No effects entered the model	.	.
Step 2: Evaluation of random effects on the cow-level regression coefficient		
No effects entered the model	.	.
Step 3: Evaluation of fixed effects on the herd-level regression coefficient		
No effects entered the model	.	.
Step 4: Evaluation of random effects on the herd-level regression coefficient		
4.1) Clustering effect of county	-7	-7

Table 4.4. Posterior mean (PMEAN), posterior standard deviation (PSD), 95% highest posterior density interval (HPD) and effective sample size (ESS) on *cow*-level and *herd*-level reparameterized covariances (expressed as regression coefficients, namely, $\varphi^{(e)}$, $\varphi^{(u)}$ and $\sigma^2_{m_u}$, respectively) between milk yield and pregnancy outcome at first postpartum insemination in Michigan dairy cows.

Regression coefficients	PMEAN	PSD	HPD	ESS
$\varphi^{(e)}$, liability/ kg	$-8.5*10^{-4}$	$5.3*10^{-4}$	$[-2.1*10^{-4}, 1.9*10^{-3}]$	104,360
$\varphi^{(u)}$, liability/kg	-0.011	0.003	[-0.017, -0.005]	26,710
$\sigma^2_{m_u}$,(liability/kg)^2	$1.3*10^{-4}$	$9.4*10^{-5}$	$[1.0*10^{-8}, 3.1*10^{-4}]$	5,627

$\sigma^2_{m_u}$ defines random county-specific heterogeneity on the *herd*-level regression coefficients.

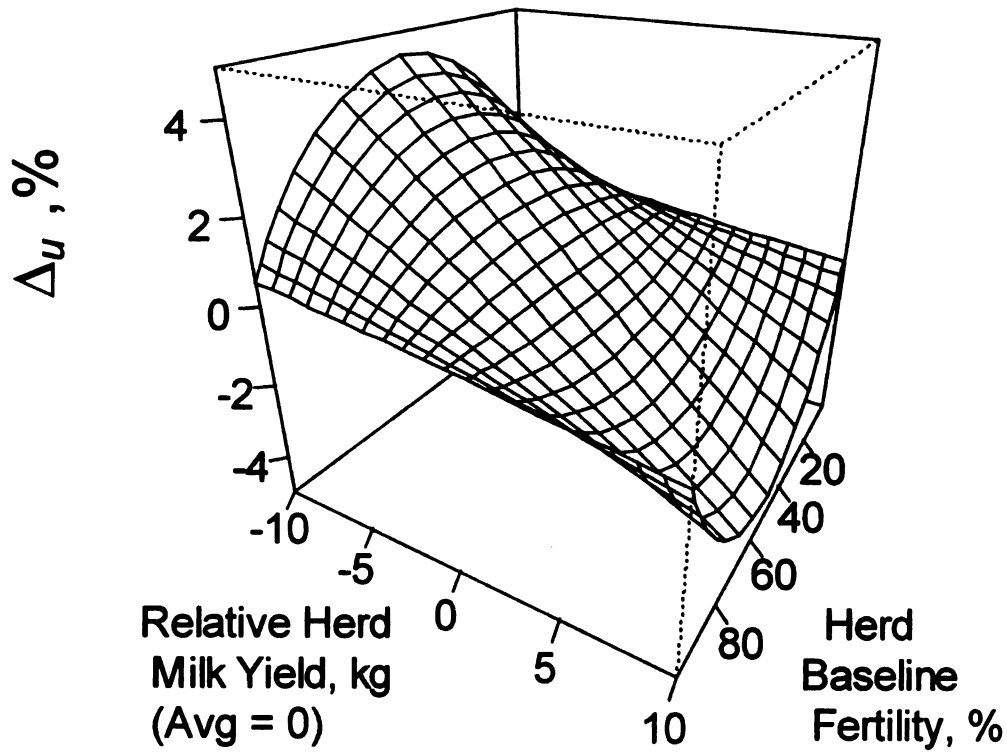
Table 4.5. Posterior means (PMEAN), posterior standard deviations (PSD), 95% highest posterior density intervals (HPD) and effective sample size (ESS) for *cow*-level and *herd*-level variances for milk yield at first postpartum insemination in Michigan dairy cows.

Variance Components	PMEAN	PSD	HPD	ESS
Herd-Level Variance				
Between Herds, kg^2	22.6	1.3	[20.1, 25.1]	241,167
Cow-level variance				
Parity				
Primiparous, kg^2	44.8 [a]	0.8	[43.2, 46.4]	6,290
Multiparous, kg^2	84.8 [b]	1.5	[81.9, 87.7]	5,398
Season				
Winter, kg^2	64.5 [a]	1.2	[62.1, 66.9]	6,940
Spring, kg^2	64.7 [a]	1.2	[62.4, 67.1]	6,809
Summer, kg^2	57.3 [b]	1.1	[55.2, 59.5]	7,386
Fall, kg^2	60.3 [c]	1.2	[58.0, 62.5]	5,864
Herd Size				
10X change in herd size, $(100\ kg)^2$	1.82	0.06	[1.69, 1.94]	4,135
Between Herds				
Coefficient of Variation	0.29	0.01	[0.27, 0.32]	35,216

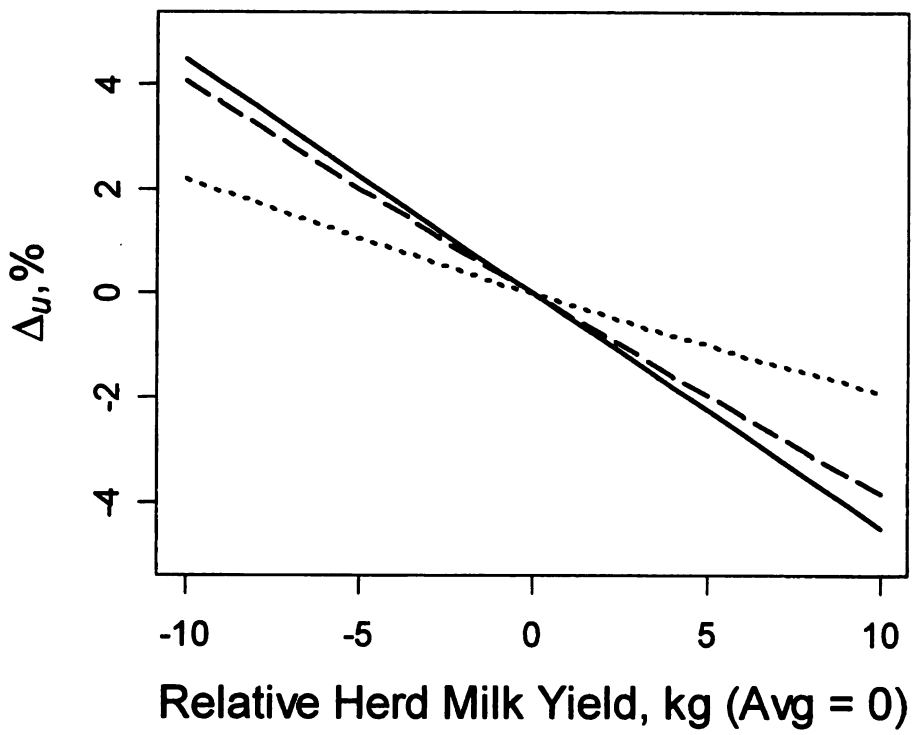
([a, b, c]) Letters indicate significant differences (Bayesian p-value < 0.001) in *cow-to-cow* variation between levels of management factor.

Figure 4.1. (A) Three-dimensional surface regression plot illustrating the *herd*-level differential on the conditional probability of pregnancy to first insemination (i.e. Δ_u) as a function of plausible values of herd baseline fertility, expressed as $\Phi(\mu_2)$, and herd milk yield relative to a typical herd (i.e. u_1), as per $\Delta_u = \Phi\left(\mu_2 + \varphi^{(u)} u_1\right) - \Phi(\mu_2)$. (B) Two-dimensional regression plot illustrating the *herd*-level differential conditional probability of pregnancy (i.e. Δ_u) as a function of herd milk yield relative to a typical herd average (i.e. u_1), at arbitrarily selected values of herd baseline fertility (i.e. $\Phi(\mu_2)$), namely 10% (·····), 30% (– – –) and 50% (——) herd pregnancy rate to first insemination.

A


Δ_u ,%
4
2
0
-2
-4
-10
-5
0
5
10
Relative Herd Milk Yield, kg (Avg = 0)
20
40
60
80
Herd Baseline Fertility, %

B


Δ_u,%
4
2
0
-2
-4
-10
-5
0
5
10
Relative Herd Milk Yield, kg (Avg = 0)

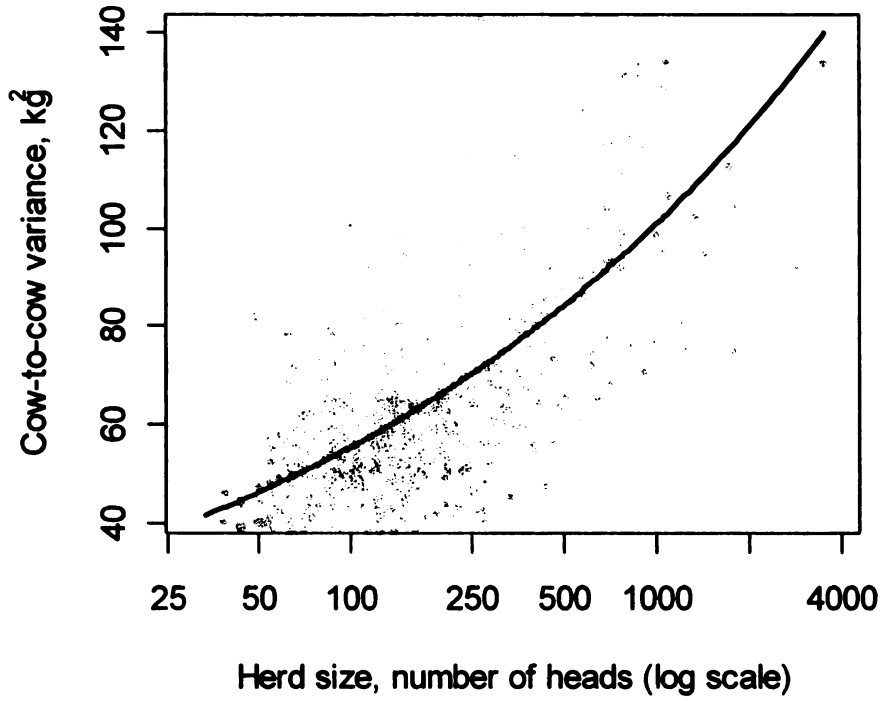


Figure 4.2. *Cow*-level variance estimate (black line) describing the cow-to-cow (i.e. cow-level) variation in milk yield at first postpartum insemination in Michigan dairy cows, expressed as a function of herd size (in log base 10 scale). Dots represent herd-specific posterior means for the *cow*-level variance.

References

Albert, J. H. and S. Chib. 1993. Bayesian Analysis of Binary and Polychotomous Response Data. Journal of the American Statistical Association 88(422):669-679.

Bello, N. M., J. P. Steibel, and R. J. Tempelman. 2010. Hierarchical Modeling of Random and Residual Variance-Covariance Matrices in Bivariate Mixed Effects Models. Biometrical Journal. 52(3):297–313.

Berry, D. P., F. Buckley, P. Dillon, R. D. Evans, M. Rath, and R. F. Veerkamp. 2003. Genetic parameters for body condition score, body weight, milk yield, and fertility estimated using random regression models. Journal Dairy Science 86(11):3704-3717.

Butler, W. R. and R. D. Smith. 1989. Interrelationships between energy balance and postpartum reproductive function in dairy cattle. Journal Dairy Science 72(3):767-783.

Emanuelson, U. and P. A. Oltenacu. 1998. Incidences and effects of diseases on the performance of Swedish dairy herds stratified by production. Journal Dairy Science 81(9):2376-2382.

Foulley, J. L., D. Gianola, M. S. Cristobal, and S. Im. 1990. A Method for Assessing Extent and Sources of Heterogeneity of Residual Variances in Mixed Linear-Models. Journal of Dairy Science 73(6):1612-1624.

Hare, E., H. D. Norman, and J. R. Wright. 2006. Trends in calving ages and calving intervals for dairy cattle breeds in the United States. Journal Dairy Science 89(1):365-370.

Huang, C., S. Tsuruta, J. K. Bertrand, I. Misztal, T. J. Lawlor, and J. S. Clay. 2009. Trends for conception rate of Holsteins over time in the southeastern United States. Journal of Dairy Science 92(9):4641-4647.

Jamrozik, J., J. Fatehi, G. J. Kistemaker, and L. R. Schaeffer. 2005. Estimates of genetic parameters for Canadian Holstein female reproduction traits. Journal of Dairy Science 88(6):2199-2208.

Kizilkaya, K. and R. J. Tempelman. 2005. A general approach to mixed effects modeling of residual variances in generalized linear mixed models. Genetic Selection and Evolution 37(1):31-56.

Leblanc, S. 2010. Assessing the Association of the Level of Milk Production with Reproductive Performance in Dairy Cattle. Journal of Reproduction and Development 56:S1-S7.

Littell, R. C., W. W. Stroup, and R. J. Freund. 2002. SAS for Linear Models. Fourth ed. SAS Institute Inc., Cary, NC.

Lopez-Gatius, F., I. Garcia-Ispierto, P. Santolaria, J. Yaniz, C. Nogareda, and M. Lopez-Bejar. 2006. Screening for high fertility in high-producing dairy cows. Theriogenology 65(8):1678-1689.

Lucy, M. C. 2001. Reproductive loss in high-producing dairy cattle: where will it end? Journal of Dairy Science 84(6):1277-1293.

Miller, R. H., H. D. Norman, M. T. Kuhn, J. S. Clay, and J. L. Hutchison. 2007. Voluntary waiting period and adoption of synchronized breeding in dairy herd improvement herds. Journal Dairy Science 90(3):1594-1606.

Ott, R. L. and M. Longnecker. 2001. An Introduction to Statistical Methods and Data Analysis. Fifth ed. Duxbury, United States of America.

Raftery, A. E. and S. Lewis. 1992. How Many Iterations in the Gibbs Sampler? Pages 763-773 in Bayesian Statistics. J. M. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith, ed. Oxford University Press, Oxford, UK.

Robinson, G. K. 1991. That BLUP is a good thing - the estimation of random effects. Statistical Science 6:15-51.

Sorensen, D. and D. Gianola. 2002. Likelihood, Bayesian and MCMC Methods in Quantitative Genetics. Statistics for Biology and Health. Springer Verlag, New York.

Sorensen, D. A., S. Andersen, D. Gianola, and I. Korsgaard. 1995. Bayesian-Inference in Threshold Models Using Gibbs Sampling. Genetics Selection Evolution 27(3):229-249.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society - Series B 64:1-34.

St-Pierre, N. R. 2001. Invited review: Integrating quantitative findings from multiple studies using mixed model methodology. Journal of Dairy Science 84(4):741-755.

Tempelman, R. J. 2010. Addressing scope of inference for global genetic evaluation of livestock. in Proceedings of the 47th Meeting of the Brazilian Animal Science Society. Salvador, Bahia, Brazil.

Tsuruta, S., I. Misztal, C. Huang, and T. J. Lawlor. 2009. Bivariate analysis of conception rates and test-day milk yields in Holsteins using a threshold-linear model with random regressions. Journal Dairy Science 92(6):2922-2930.

CONCLUSIONS

1. This dissertation in the context of statistical inference on dairy production systems

Milk yield and reproductive performance of dairy cows constitute the foundation of a successful dairy business. Simultaneous optimization of both outcomes is of primordial interest to the profitability and sustainability of the dairy industry. However, despite abundant literature on the subject, the nature of the association between milk yield and fertility remains a largely controversial topic (Butler and Smith, 1989, Laben et al., 1982, Leblanc, 2010, Lopez-Gatius et al., 2006, Lucy, 2001, Norman et al., 2009, among others). We argue that this controversy is partially due to the multiple shortcomings of the statistical methodologies commonly implemented in dairy management studies. The most frequent limitations of previous studies include single-trait analyses that limit specification of explanatory covariates to either milk yield or reproduction (but not both), and an ubiquitous under-appreciation of cows and herds as separate, yet interconnected, units of performance.

Multivariate hierarchical Bayesian models present a general statistical framework to address these methodological shortcomings. The multivariate setting allows for simultaneous investigation of two or more outcomes, whereas the hierarchical realm naturally accounts for cows and herds as different components of the question (Sorensen and Gianola, 2002). In Chapters 1 and 3 of this dissertation, I present and validate methodological extensions to bivariate hierarchical Bayesian models that naturally overcome the aforementioned limitations. These extensions represent an important advance over the current statistical literature due to an explicit linear model specification

on covariance parameters, thereby introducing the opportunity to investigate sources of heterogeneity in the association (i.e. correlation) between multiple outcomes of interest. The novelty of these methodological developments relies on their introducing a whole new dimension to study heterogeneity in complex multivariate biological systems. Explicit design and modeling of heterogeneity has been recently advocated as a robust venue to guarantee reproducibility and external validation of scientific results and to decrease uncertainty due to spurious findings in certain areas of the biological sciences (Richter et al., 2010, Richter et al., 2009). In particular, the proposed methodological developments directly address the motivating question for this dissertation, thereby allowing for explicit modeling of heterogeneity in the association between milk yield and dairy reproductive performance. In Chapters 2 and 4 of this dissertation, I applied this novel methodology to large datasets from commercial dairy farms in Michigan to investigate the nature of the association between milk production and reproductive performance of dairy cows and of commercial dairy herds. In addition, I evaluated management factors and herd attributes as potential determinants of heterogeneity in these associations.

2. Address of Specific Aims

1) To develop and validate a hierarchical Bayesian extension to classical bivariate mixed effects methods to model heterogeneity in residual and random (co)variance matrices for the joint analysis of two Gaussian phenotypes.

The statistical developments presented in Chapter 1 of this dissertation fully address this aim and constitute a methodological advancement in the investigation of a new dimension of heterogeneity in complex systems. Specifically these developments allow for evaluation of heterogeneous covariances (or correlations) between two outcomes of interest. In Chapter 1, the proposed methodology is restricted to outcomes of continuous nature which can be approximated with a normal distribution; this constraint is circumvented in Chapter 3, as described later. It should be noted that the methodological developments associated with Specific Aim 1 are relevant to the scientific literature on multivariate statistics and on hierarchical linear models independently of specific subject-matter applications.

2) To use the methodology developed in Specific Aim 1) to investigate the *within*-herd (cow-level) and *between*-herd (herd-level) associations between indicators of comprehensive (i.e. entire lactation) milk production and reproductive performance of Michigan dairy cows, including the evaluation of various management factors and herd attributes potentially affecting these associations.

This Specific Aim was accomplished through a small data application in Chapter 1, followed by a more comprehensive investigation in Chapter 2. In both chapters, we focused on cumulative milk yield at 305-d in lactation and calving interval as lactation-encompassing indicators of performance. In Chapter 1, the investigation of cow-level and herd-level associations is intended only for the purpose of demonstrating the application of the corresponding methodology and is thus limited to a subset of the population of

interest and to an arbitrarily selected subset of management factors. The investigation of cow-level and herd-level associations between milk production and reproductive performance becomes more comprehensive in Chapter 2. This chapter provides a more extensive coverage of the dairy population in Michigan, coupled with a sequential approach to selecting among the many management factors and herd attributes that potentially contribute to heterogeneity.

3) To develop and validate a hierarchical Bayesian implementation of a bivariate generalized linear mixed-effect model for heterogeneous variance-covariance matrices in the context of a joint analysis of Gaussian and non-Gaussian traits.

Chapter 3 of this dissertation directly addresses Specific Aim 3 by describing and validating generalized linear modeling extensions to the bivariate hierarchical Bayesian method presented in Chapter 1. These developments extend the methodology presented in Chapter 1 by accommodating non-Gaussian responses, including binary, ordered categorical, count and censored traits, to the modeling of heterogeneous covariances. The need to incorporate non-Gaussian traits is driven by the non-continuous nature of performance outcomes related to fertility and fitness. In particular, the motivating question of this dissertation calls for consideration of pregnancy outcome, a binary response that is critical to the investigation of dairy reproductive performance in a way that is meaningful to daily farm management.

4) To implement the methodology developed in Specific Aim 3) to investigate the associations between milk yield at and pregnancy outcome to first postpartum insemination of Michigan dairy cows, accounting for cow and herd as hierarchical units of performance and evaluating the role of management practices and herd attributes as potential sources of heterogeneity.

This Specific Aim was addressed in depth in Chapter 4, where I investigated management factors and herd attributes that are potentially linked to the association between daily milk yield and pregnancy outcome at first postpartum insemination. This investigation was based on a large dataset from commercial dairy herds in Michigan and model selection techniques. In addition, Chapter 3 presents a small data application of the methodology described for Specific Aim 3.

3. Implications for optimization of dairy cow performance

Results from Chapters 1 through 4 in this dissertation provide overwhelming evidence that forsakes the concept of "one-size-fits-all" as an overly simplistic attempt to describe the nature of the association between milk production and reproductive performance of dairy cows. Results from this dissertation indicate that the nature of this association is highly heterogeneous and consists of multiple dimensions, rending the issue substantially more complex than ever anticipated. The statistical methodology specifically developed herein to tailor the intricacies of the production-reproduction relationship provides a formal substantiation of this conclusion. Previous investigations have clearly over-simplified this association. Instead, the evidence presented in this

dissertation urge us to consider a more sophisticated, yet realistic, scenario by which this association is recognized as heterogeneous at multiple levels. The following paragraphs describe the multidimensional/multifactorial components of this heterogeneity.

First, results clearly indicate a strong hierarchical duality of the production-reproduction association comprised by at least two separate, yet interconnected components. Specifically, these components comprise herds as business production units, and cows as physiological units of performance managed within those herds. The outcomes of the delicate intricacies of a cow's physiology do not necessarily mirror those of management mechanisms and the business decisions of the production system (i.e. herd) in which the cow is immersed. I personally believe that failure to recognize the cow-herd duality while investigating the association between milk production and reproduction is probably the main reason for the abundant contradictory evidence on the subject. If studies do not appropriately account for cows and herds as a hierarchical duality, it is likely that the resulting characterization of the production-reproduction association becomes an indiscriminate merging of sources of variation into one large *mélange*. I speculate that the form of this *mélange* may be directly determined by the relative proportion of information contributed by cows and by herds. The relative proportions of the cow-level and herd-level components is highly specific to each individual dataset and thus, innately, not reproducible. This concept can be easily demonstrated using data simulation techniques whereby the cow-level and herd-level hierarchical components can be alternatively accounted for or disregarded in the evaluation of the association between milk production and reproduction. Interestingly, conclusions on the association between outcomes turn out to be mutually contradicting if

cow- and herd- hierarchical levels are or are not explicitly modeled, thereby supporting the argument that blunt disregard of a hierarchical data structure can easily lead to dangerously biased results.

Second, the nature and magnitude of the association between milk yield and reproductive performance differed across management scenarios, indicating that management practices and herd attributes may be potential contributors to the multilayered heterogeneity in this association. In particular, benefits on the production-reproduction association were evident under highly specialized management practices that typically characterize intensive production systems. Nonetheless, I personally find intriguing, to say the least, that intensive production management practices appear as a converging point for the successful future of animal agriculture from perspectives as diverse as business profitability and technical efficiency (Cabrera et al., 2010), environmental sustainability (Capper et al., 2009, Capper et al., 2008) and food supply challenges of the century (Simmons, 2009).

At this point, the following note of caution could not be emphasized enough: the observational nature of the data used in this dissertation precludes any type of cause-and-effect conclusion as inappropriate and over-generalized. That being said, the link identified between intensive management and benefits on the association between milk production and reproduction certainly warrants further investigation. For example, results from this dissertation could be used to identify herds with extreme rankings for their estimates of relative production-reproduction association (i.e. herds with inferences unusually distal from zero, both favorable and unfavorable, say the upper and lower 5th percentiles). These individualized herd entities could then be studied retrospectively to

explore any potentially new important management and environmental factor that may be linked to the production-reproduction association. Such retrospective evaluation could also be used to characterize management scenarios conducive to favorable associations, or just as important, scenarios of antagonistic association that would be recommended against. Again, this evaluation would still not render cause-and-effect conclusions but it will likely identify desirable and undesirable management conditions, from which it might be possible, in some cases, to proceed with randomized experimentation. It should be pointed out that the management practices and herd attributes considered in this dissertation are limited to those available through the information processing facilities that centralize dairy herd management data. At least two key components of dairy cow management were not available in the data and are likely to be relevant in future investigations, namely herd health management and body condition score records. Many herds do indeed keep individual cow records of vaccination schemes, clinical presentations of disease and treatment, as well as accounts of regular evaluations of body condition score throughout the lactation of a cow. Farms normally find this information useful to guide decision making on individual cows or groups within the farm. However, health and body condition information is not routinely recovered by data processing centers, thus creating a breach between data available at the farm and processed data from regional centers. This disconnect may be partially reinforced by the lack of rigorous uniform standards for collecting health data across farms. Therefore, while health data may likely be meaningful within a farm, the lack of standardized practices for record keeping on health events renders health data useless to make comparisons between farms.

A third dimension of heterogeneity in the production-reproduction association pertains to the long- versus short-term coverage of the outcomes used to characterize milk production and reproduction. Results in this dissertation suggest that the mechanisms that underlie performance measures spanning a whole lactation (i.e. cumulative milk yield at 305-d and calving interval) may differ from the mechanisms at work for "snap-shot" performance indicators corresponding to a discrete and singular point in time (i.e. daily milk yield at and pregnancy outcome to first postpartum insemination). Indeed, cows with greater lactation yields had longer calving intervals, but herd calving intervals were either shorter or unaffected among herds with highest cumulative 305-d yields. Conversely, herds with greater milk yields at the time of first insemination had impaired pregnancy rates, but within such herds, cows with higher daily yields were not any more or less likely to become pregnant to first postpartum insemination than lower yielding herdmates. Clearly, the dual cow-herd components of the association between milk production and reproduction behave differently on a whole-lactation basis as compared to a point in time. It can certainly be speculated that a "snapshot" indicator may be more volatile and sensitive to rapid circumstantial changes in management, whereas the summative nature of whole-lactation indicators of performance is likely to make them more stable and reflective of long-term practices. Adjusting management recommendations to each short- and long-term scenario undoubtedly adds an extra layer of complexity to the challenge of jointly optimizing milk yield and reproductive performance of dairy cows.

The question of management-driven heterogeneity in the cow-level and herd-level associations between milk production and reproduction in lactating cows is undoubtedly

a major issue to the dairy industry. Overall, this dissertation supports that a comprehensive assessment of this question implies considerably more than just a simplistic pendulum between "favorable" or "antagonistic" conclusions. It is only through a comprehensive appreciation of the multidimensional sources of heterogeneity in the production-reproduction association that it will be possible to elicit management scenarios conducive to joint optimization of milk yield and reproductive efficiency. By means of this dissertation, it is my humble intention to provide a broader framework for and a novel perspective on the controversy of the association between milk yield and fertility of dairy cows in a way that recognizes many of the complexities of the problem. In so doing, this approach attempts to provide realistic insight from which to design management recommendations targeted to optimize overall performance of dairy cows and commercial dairy herds. Nevertheless, I recognize that, by its own nature, the multidimensional sources of heterogeneity unveiled for the production-reproduction association impair the formulation of comprehensive blanket management recommendations. It certainly appears that there are no recommendations that are suitable for all cows and all herds. Further research will undoubtedly be required to clarify what these recommendations may be in specific circumstances.

4. Opportunities for future studies

The work presented in this dissertation provides a foundation for future studies both in the arena of statistical methodology as well as that of dairy management and production. The opportunity is particularly exciting in that statistical methods and dairy applications can be induced to create a chain reaction of mutual developments in a truly

interdisciplinary approach. Moreover, many of these methodological developments could also be implemented to other subject-matter applications across a wide range of biological sciences.

In the pursuit of understanding the multiple interconnections within a complex biological system, the methodological developments proposed in this dissertation unveiled a whole new dimension of heterogeneity; that of covariances or correlations between outcomes of interest. Possibilities for extending this work in ways that are relevant to dairy management are numerous. The hierarchical Bayesian models presented in Chapters 1 and 3 provide a general framework that can be easily built upon to accommodate, for instance, more than two outcomes of interest in a truly multivariate modeling approach. Such extension would allow incorporation of additional fitness outcomes to the joint analysis of quantitative performance responses and reproductive outcomes. Particularly interesting candidate outcomes are those related to health status (Wu et al., 2007, Wu et al., 2008) and assessments of energy balance (Roche et al., 2009), provided that issues of data collection standards (as discussed in the previous section) can be appropriately addressed. Furthermore, a longitudinal component to multiple outcomes (i.e. repeated measures over time) could be easily accommodated in our model, thereby allowing for evaluation of pregnancy outcomes to consecutive inseminations throughout a lactation cycle. Also, inclusion of interaction terms between the fixed effects modeled on variances and covariances would be desirable to fine tune delicate interdependencies between management factors and their contributions to heterogeneous associations. One might ask, for example, if the observed differences in the production-reproduction association between intensive and traditional management practices depends upon

lactation. In other words, are the management scenarios linked with more favorable associations shared by both primiparous and multiparous cows?

Modeling of additional random components, such as additive genetic effects, is a natural extension to our model that is likely to be of special interest to animal breeders and geneticists in their quest to assess the heritable component of the correlation between performance outcomes, as well as potential sources of heterogeneity (Berry et al., 2003, Tsuruta et al., 2009). The integration of genetics and management (i.e. environment) into a comprehensive plan for simultaneously improving milk yield and dairy fertility looks certainly very promising, especially as the era of genomic selection unveils.

As a more elegant alternative to the discrete model selection approach implemented in Chapters 2 and 4, it will be of interest to investigate Bayesian Model Averaging (BMA) techniques to further enhance model fit to the data by considering a large number of candidate models. Potential ways to incorporate BMA to our proposed multivariate hierarchical model include building upon an attractive proposal by Chen and Dunson (2003), and Kinney and Dunson (2007).

Finally, the statistical methodology developed in this dissertation is general enough that it could be easily implemented in other subject-matter applications in which the joint evaluation of multiple outcomes with potentially heterogeneous correlations may be of interest. Examples include, but are not limited to, other livestock production systems, meta-analyses of multiple related research studies, multicenter studies and longitudinal data pertaining to the biomedical and agricultural sciences, and animal/plant breeding, genetics and genomics.

References

Berry, D. P., F. Buckley, P. Dillon, R. D. Evans, M. Rath, and R. F. Veerkamp. 2003. Genetic relationships among body condition score, body weight, milk yield, and fertility in dairy cows. J Dairy Sci 86(6):2193-2204.

Butler, W. R. and R. D. Smith. 1989. Interrelationships between energy balance and postpartum reproductive function in dairy cattle. Journal Dairy Science 72(3):767-783.

Cabrera, V. E., D. Solis, and J. del Corral. 2010. Determinants of technical efficiency among dairy farms in Wisconsin. Journal of Dairy Science 93(1):387-393.

Capper, J. L., R. A. Cady, and D. E. Bauman. 2009. The environmental impact of dairy production: 1944 compared with 2007. Journal of Animal Science 87(6):2160-2167.

Capper, J. L., E. Castaneda-Gutierrez, R. A. Cady, and D. E. Bauman. 2008. The environmental impact of recombinant bovine somatotropin (rbST) use in dairy production. Proceedings of the National Academy of Sciences of the United States of America 105(28):9668-9673.

Chen, Z. and D. B. Dunson. 2003. Random effects selection in linear mixed models. Biometrics 59(4):762-769.

Kinney, S. K. and D. B. Dunson. 2007. Fixed and random effects selection in linear and logistic models. Biometrics 63(3):690-698.

Laben, R. L., R. Shanks, P. J. Berger, and A. E. Freeman. 1982. Factors Affecting Milk-Yield and Reproductive-Performance. Journal of Dairy Science 65(6):1004-1015.

Leblanc, S. 2010. Assessing the Association of the Level of Milk Production with Reproductive Performance in Dairy Cattle. Journal of Reproduction and Development 56:S1-S7.

Lopez-Gatius, F., I. Garcia-Ispierto, P. Santolaria, J. Yaniz, C. Nogareda, and M. Lopez-Bejar. 2006. Screening for high fertility in high-producing dairy cows. Theriogenology 65(8):1678-1689.

Lucy, M. C. 2001. Reproductive loss in high-producing dairy cattle: where will it end? Journal of Dairy Science 84(6):1277-1293.

Norman, H. D., J. R. Wright, S. M. Hubbard, R. H. Miller, and J. L. Hutchison. 2009. Reproductive status of Holstein and Jersey cows in the United States. Journal of Dairy Science 92(7):3517-3528.

Richter, S. H., J. P. Garner, C. Auer, J. Kunert, and H. Wurbel. 2010. Systematic variation improves reproducibility of animal experiments. Nat Methods 7(3):167-168.

Richter, S. H., J. P. Garner, and H. Wurbel. 2009. Environmental standardization: cure or cause of poor reproducibility in animal experiments? Nat Methods 6(4):257-261.

Roche, J. R., N. C. Friggens, J. K. Kay, M. W. Fisher, K. J. Stafford, and D. P. Berry. 2009. Invited review: Body condition score and its association with dairy cow productivity, health, and welfare. J Dairy Sci 92(12):5769-5801.

Simmons, J. 2009. Food Economics and Consumer Choice. Online. Available: http://www.dairycast.com/food-economics-and-consumer-choice?utm_source=DairyCast+update&utm_campaign=882bf92449-dairycast_2009_05_18AB&utm_medium=email.

Sorensen, D. and D. Gianola. 2002. Likelihood, Bayesian and MCMC Methods in Quantitative Genetics. Statistics for Biology and Health. Springer Verlag, New York.

Tsuruta, S., I. Misztal, C. Huang, and T. J. Lawlor. 2009. Bivariate analysis of conception rates and test-day milk yields in Holsteins using a threshold-linear model with random regressions. Journal Dairy Science 92(6):2922-2930.

Wu, X. L., B. Heringstad, Y. M. Chang, G. de Los Campos, and D. Gianola. 2007. Inferring relationships between somatic cell score and milk yield using simultaneous and recursive models. Journal of Dairy Science 90(7):3508-3521.

Wu, X. L., B. Heringstad, and D. Gianola. 2008. Exploration of lagged relationships between mastitis and milk yield in dairy cows using a Bayesian structural equation Gaussian-threshold model. Genet Sel Evol 40(4):333-357.