4
(1997)

This is to certify that the

dissertation entitled

Spatial and Temporal Comparison of Selected Cancers in Dogs and Humans, Michigan, 1964 - 1994

presented by

Daniel John O'Brien, DVM

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Large Animal Clinical Sciences (Epidemiology)

Major professor

John B. Kaneene, DVM MPH PhD

Date 9/30/98

DATE DUE	DATE DUE	DATE DUE

SPATIAL AND TEMPORAL COMPARISON OF SELECTED CANCERS IN DOGS AND HUMANS, MICHIGAN, 1964-1994

VOLUME I

By

Daniel John O'Brien, DVM

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Large Animal Clinical Sciences
(Epidemiology)

1998

ABSTRACT

SPATIAL AND TEMPORAL COMPARISON OF SELECTED CANCERS IN DOGS AND HUMANS, MICHIGAN, 1964-1994

By

Daniel John O'Brien, DVM

Research into the aggregation or "clustering" of cases of diseases, particularly cancers, remains a common priority for epidemiological research in both veterinary and human medicine. The utility of such research for describing the spatial patterns of cancer cases, for the purpose of generating hypotheses regarding the underlying processes which determine them, has not been fully investigated. Few studies investigating the geographic distribution of cancer in companion animals exist, due to the limitations of methods incorporating rate calculations, and the inability to enumerate the populations-at-risk they require. Even fewer studies have simultaneously studied the geographic and time distributions of biologically similar cancers in humans and companion animals. Thus, the objectives of this study were to demonstrate the utility of rate-independent methods 1) for the quantitative description of cancer patterns in time and space, 2) for generation of hypotheses concerning processes affecting cancer aggregation, and 3) for simultaneous comparison of those patterns in companion animals and humans. Data records for cases of selected cancers in dogs diagnosed at the Michigan State University Veterinary Medical Center between 1964 and 1994 were obtained from the Veterinary Medical Data Base,

and some characteristics described by frequency table analyses. Results were largely consistent with the literature, though notable differences, particularly with respect to breed, were described. Residence addresses of cases of four of those cancers in three Michigan counties were obtained from patient medical records, geocoded, mapped, and subjected to *K* function spatial analysis and one-dimensional nearest-neighbor temporal analysis. Records for human cases of the same cancers diagnosed in the same areas and time period obtained from the Surveillance, Epidemiology and End Results Program and the Michigan Cancer Registry were analyzed by identical methods. Analysis of demographic factors potentially affecting canine cancer distributions was also conducted using Poisson regression. Among the canine cases, significant spatial clustering was found that varied by county and cancer. Temporal patterns concurred minimally with spatial ones, and varied from clustering to dispersion depending on the cancer under study. Demographic factors studied had little influence. Comparative analyses suggested that processes determining spatial aggregation of cases in dogs and humans were not independent, did not act uniformly over different geographic areas, that those processes operated at spatial scales of < 2000 meters, and that they tended to act upon dogs more strongly at shorter distances than upon humans. There was little evidence of interspecies correspondence of case clustering in time, though broad tendencies of particular neoplasms to be clustered or dispersed were similar. Though exploratory, the study demonstrates the value of rate-independent methods for the simultaneous study of disease in humans and their companion animals.

This work is dedicated to the memory of my good friend and colleague Scott Richard Allen, who was murdered in the spring of 1986. Until I see you again, my friend, may you rest in peace.

ACKNOWLEDGMENTS

It is with great honor and humility that I acknowledge those individuals and entities whose contributions have made this work possible:

My best friend and the greatest joy of my life, Susan, whose patience and support through many lonely hours, days and years stand as a living testament to the sacrifices that can be borne in the name of love.

My father Ed and mother Barbara, whose unconditional love and selflessness have made possible every good thing I have been fortunate enough to accomplish in my life. In a life spent constantly reaching for things that once seemed unattainable, it has been on their shoulders that I have stood.

My advisor, John Kaneene, whose friendship, indomitable optimism and steadfast belief in me and my ideas has more than once served as the sole inspiration that caused me to continue my research in the face of seemingly endless setbacks.

Jim Lloyd, who I am both proud and happy to call my counselor and my good friend.

The other members of my guidance committee: Marie Swanson, Mike Rip, Bob Dunstan, Ashley Robinson, and the late Bob Leader who, beyond their responsibilities to my training, have accepted and embraced my unconventional research interests, and treated me more as a respected colleague than as a student.

Art Getis, without whose patient guidance from afar this work would never have been possible.

The many great teachers whose influence it has been my good fortune to fall under in my lifetime: Linda Sisung; Mike VanWashenova; Jim Spooner; Philip Smith; Philip Allessandro; Marius Wittner; Joe Sanderson; Marge Goodell; Garelick; Dick Hill; Chris Brown; John Kaneene; Pat Lederle; Ivan Mao; Carl Ramm; and Jim Lloyd.

And, finally, for financial support of this research: The Population Medicine Center; Bob Poppenga; the Postgraduate Training Grant Program of the Department of Large Animal Clinical Sciences; and the Dissertation Completion Fellowship Program of The Graduate School, Michigan State University.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	*Meaning*
CDC	Centers for Disease Control and Prevention
CSR	Complete Spatial Randomness
d.f.	Degrees of Freedom
ESRI	Environmental Systems Research Institute
GIS	Geographic Information System(s)
ICD	International Classification of Diseases
ICDO-2	International Classification of Diseases for Oncology, Second Revision
I/RA	Intestinal/Rectal Adenocarcinoma
LY	Lymphosarcoma
MA	Mammary Adenocarcinoma
MC	Mammary Carcinoma
ME	Melanoma
MMT	Mixed Mammary Tumor
MSU-VMC	Michigan State University Veterinary Medical Center
NA	Nasal Adenocarcinoma
PA	Prostatic Adenocarcinoma
PCOP	Purdue Comparative Oncology Program
PPA	Point Pattern Analysis
SCS	Spindle Cell Sarcoma

SMR	Standardized Mortality (Morbidity) Ratio
SNVDO	Standard Nomenclature of Veterinary Diseases and Operations
TCC	Transitional Cell Carcinoma of the Urinary Bladder/Urethra
TIGER	Topologically Integrated Geographic Encoding and Reference
VMDB	Veterinary Medical Data Base
VMDP	Veterinary Medical Data Program

INTRODUCTION

The study of the aggregation of cases of disease in space and in time has a long and storied history. Such celebrated investigations as those carried out by Dr. John Snow in London in the nineteenth century[1] are now viewed not only with deserved awe for their insightfulness and innovation, but also with the reverence afforded the central icons upon which scientific disciplines are founded and from which they derive their lasting sense of purpose and identity. In spite of this, studies of case aggregations, or "cluster investigations" as they are commonly called, are an enigma in epidemiology. Nearly 140 years after Dr. Snow's death, cluster investigations continue to generate controversy[2,3] and active research.[4-7] Regardless of whether human or animal diseases are the subject of scrutiny, formidable challenges continue to accompany this most legendary of epidemiologic tools.

Problem Statement

Investigations into the clustering of cases of disease, and particularly of neoplastic disease, have been numerous in human medicine. The appearance of cancer cases occurring in close proximity to one another has invariably caused considerable public concern.[8,9] The public often attaches extreme significance to any perception of clusters of cancer in an area, and this perceived significance is often supported by the mass media, which often publicizes these occurrences and speculates on causes despite limited

understanding of the pathogenesis and dynamics of the disease. Media speculation can magnify this perceived significance until public reaction approaches hysteria. Concern, anger, and fear are common. The Centers for Disease Control and Prevention (CDC)[10] and others have pointed out that "the public perceives the need to investigate clusters and environmental exposures to their satisfaction", and commonly exert extreme pressure on public health authorities to investigate perceived clusters. The public demands an explanation that suits its perceptions, even though such explanations are often quite costly and difficult to come by.

But when a cluster is found, what is its significance, if any? The public often believes that any cluster of cancers must have a biological cause, but the historic experience strongly suggests that they are often impossible to distinguish from random events.[11-13] Moreover, the traditional approaches using incidence rates in small geographic areas are beset with problems which have proven formidable,[14,15] and perhaps intractable, despite the best efforts of researchers and public health authorities alike. Traditional cluster investigations which have concentrated, largely in vain, on finding well-defined causes, have often neglected the opportunity to generate useful descriptions of disease patterns[15] and hypotheses which can be tested using other methods.[16]

From the standpoint of veterinary medicine, the difficulties in investigating cancer clusters are less emotionally charged than in human medicine, but in some respects even more formidable, as they are more fundamental.[17] Chief among these is a lack of good data on the occurrence of cancers in animal

populations. Though great strides have been made in diagnosis and therapy in recent years, cancers in domestic animals are very often terminal illnesses, not necessarily due to the severity of the diseases themselves or the inability to treat them, but due to euthanasia. Owners are often unable or unwilling to pursue diagnosis and treatment because cost is perceived as, or really is, prohibitive, despite the fact that sensitive diagnostics and curative (or at least palliative) therapies are available. As a result, with some notable historical exceptions which will be discussed later, there is very little comprehensive data on animal cancers in the general population. This paucity of data has drastically limited the number and statistical power of cluster investigations of cancer in domestic animals. What data does exist has generally been gathered from cancer cases examined at referral hospitals. The extent to which those cases are representative of the general animal population is subject to considerable uncertainty, as are the epidemiologic conclusions drawn by studying them.

Another huge barrier to cluster investigations in veterinary medicine exists. Many of the traditional approaches used to investigate disease aggregation in human medicine that rely on the comparison of incidence rates or standardized mortality or morbidity ratios (SMRs) amongst geographic areas are practically impossible to employ in most animal populations. To calculate prevalence or incidence, one needs to be able to enumerate a reference population or a population-at-risk; to calculate SMRs, one needs to be able to determine the number of expected cases of disease. While some animal populations subject to confinement or intensive management (*e.g.*, food animals)

have been enumerated adequately, companion animals such as dogs, by and large, have not. Reasonably accurate enumeration has been accomplished in a some localized areas.[18-26] However, more generalized methods based on sample surveys of pet ownership,[27] designed to capture populations across wider geographic areas, while a useful approximation, may overestimate the prevalence of pet ownership,[23] and thus the companion animal population. Moreover, there are substantial numbers of dogs and cats that have no owners and thus would not be enumerated in surveys of pet ownership; one study estimated that 14% of their study population was composed of such dogs.[24] Yet these animals, despite not being counted, are still at risk for the development of disease. Such uncertainties may render estimates of populations-at-risk, and the rates calculated from them, inaccurate. Also, to the extent that risk factors which vary from region to region and over time are important determinants of the occurrence of a disease, incidence rates for that disease can only be considered accurate in the area and time where the population-at-risk was enumerated. As a result, cluster investigations which attempt to extrapolate population estimates or incidence rates from other regions or time periods may well suffer serious problems with internal validity. These factors have combined to drastically limit the number and quality of cluster investigations in the veterinary medical literature.

Rationale

Thus, with respect to investigations of clusters of cancer and other diseases, neither veterinary nor human medicine has, as yet, developed definitive methods with which to address the public's concerns and the medical community's need for information on which to base intervention and control, therapy, and allocation of health care resources. Additionally, in veterinary medicine, so little work has been done in this area (or indeed, in spatial analysis in general[28]) that cancer cluster research in dogs and other companion animals is effectively still in its infancy. While the number and scientific sophistication of existing studies, and the specific methodological problems, differ between veterinary and human medicine, the need for continuing research remains a common priority.

Intuitively, it seems few would argue with the notion that the health of humans and animals are linked. With respect to zoonotic diseases this link is well established.[29-31] Much has been made in the fields of environmental health and risk assessment of the relevance of findings in animal studies to human health. The existence of a connection between the health of animals and humans sharing a common environment has been implied, even assumed, yet studies in which patterns of disease have been studied in humans and animals simultaneously in the same location and time period are uncommon.[32]

Consequently, opportunities exist for studies capable of addressing some of the historical deficiencies of disease aggregation investigations in veterinary medicine, studies that:

- Address the inadequacies of incidence rate-based methods, or alternatively, that attempt to apply methods which do not rely on rates, and so avoid their currently intractable difficulties;
- Allow the distribution of clustered cases of disease to be spatially and temporally described for the generation of hypotheses which may be further investigated elsewhere;
- Apply such techniques to large, animal medical data bases to conduct investigations with sufficient statistical power;
- Compare the spatial and temporal occurrence of neoplastic disease in humans and the companion animals that live closely with them and which may share their exposures, for the purpose of exploring similarities, differences, and the existence of linkages between human and animal health.

The desire to conceive a study capable of addressing these issues is the underlying motivation for the current work.

Objectives

The objectives of this research are four-fold:

1) To describe the frequencies, and some general characteristics (age, gender, weight, vital status at discharge, and breed), of cases of some selected cancers in dogs in the state of Michigan during the thirty year period between 1964 and 1994;

2) To describe and compare the spatial and temporal distributions of some selected canine and human cancers in three Michigan counties during that same period;

3) To determine, using methods novel to veterinary epidemiology, if these spatial distributions are independent between the two species; and

4) To generate some plausible hypotheses for these distributions that may ultimately, through further research, suggest risk factors or etiologies that both dogs and humans hold in common.

Overview

With respect to format, each chapter of this dissertation is written in a format suitable for publication as an individual manuscript. Consequently, with the exception of Chapter 1 (which is essentially a review of the literature), each chapter contains an abstract, introduction, methods, results, and discussion. Chapter 1 reviews the literature of relevance to the current study. This includes the epidemiological and statistical literature regarding various methods for investigating aggregation of disease, previous approaches taken to the study of cancer aggregation in companion animals, and the literature of simultaneous studies of companion animal and human cancer. The chapter also recounts the reasoning behind the methodological approaches chosen for the current study. Chapter 2 describes the case frequencies and some characteristics of some selected canine cancers diagnosed at the Michigan State University Veterinary Medical Center and judged to be biologically similar to comparable cancers in

humans. Chapter 3 recounts the analysis of the spatial and temporal clustering of a subset of these canine neoplasms, while Chapter 4 uses the same analytical methods to compare the spatial and temporal distributions of human cases of these cancers with those seen in the dogs. The final section attempts to synthesize the various chapters into a whole by summarizing findings, drawing conclusions, and speculating on potential directions for future research.

REFERENCES

REFERENCES

1. Snow J. Report on the cholera outbreak in the Parish of St. James, Westminster during the autumn of 1854. Presented to the vestry by the Cholera Inquiry Committee, July 1855. London: Churchill, 1855:175p.

2. Rothman KJ. A sobering start to the cluster busters' conference. Am J Epidemiol 1990;132 Suppl 1:S6-S13.

3. Wartenberg D, Greenberg M. Methodological problems in investigating disease clusters. Sci Tot Environ 1992;127(1,2):173-85.

4. Proceedings of the National Conference on Clustering of Health Events. Am J Epidemiol 1990;132 Suppl 1:S1-S202.

5. Jacquez, GM, editor. Proceedings of the workshop on statistics and computing in disease clustering, Port Washington, New York, 1992. Stat Med 1993;12:1751-968.

6. Lawson A, Waller L, Biggeri A, editors. Spatial disease patterns. Stat Med 1995;14:2289-501.

7. Jacquez GM, editor. Conference on Statistics and Computing in Disease Clustering, British Columbia, Canada, July 21-22, 1994. Stat Med 1996;15:681-952.

8. Bithell JF, Stone RA. On statistical methods for analyzing the geographical distribution of cancer cases near nuclear installations. J Epidemiol Comm Health 1989;43:79-85.

9. Osborne JS III, Shy CM, Kaplan BH. Epidemiologic analysis of a reported cancer cluster in a small rural population. Am J Epidemiol 1990;132 Suppl 1:S87-S95.

10. Caldwell GC. Twenty-two years of cancer custer investigations at the Centers for Disease Control. Am J Epidemiol 1990;132 Suppl 1:S43-S47.

11. Neutra R, Swan S, Mack T. Clusters galore: insights about environmental clusters from probability theory. Sci Tot Environ 1992;127:187-200.

12. Hole DJ, Lamont DW. Problems in the interpretation of small area analysis of epidemiological data: the case of cancer incidence in the West

of Scotland. J Epidemiol Comm Health 1992;46:305-10.

13. Olsen SF, Martuzzi M, Elliott P. Cluster analysis and disease mapping—why, when, and how? A step by step guide. BMJ 1996;313:863-6.

14. Elliott P. Investigation of disease risks in small areas. Occup Environ Med 1995;52:785-9.

15. Elliott P, Martuzzi M, Shaddick G. Spatial statistical methods in environmental epidemiology: a critique. Stat Methods Med Res 1995;4:137-59.

16. Beral V. Childhood leukemia near nuclear plants in the United Kingdom: the evolution of a systematic approach to studying rare disease in small geographic areas. Am J Epidemiol 1990;132 Suppl 1:S63-S68.

17. Tjalma RA. Implications of animal cancers to human neoplasia: epidemiologic considerations. Int J Cancer 1968;3:1-6.

18. Leslie BE, Meek AH, Kawash GF, McKeown DB. An epidemiological investigation of pet ownership in Ontario. Can Vet J 1994;35(4):218-22.

19. Wise JK, Yang JJ. Veterinary service market for companion animals, 1992. Part I: Companion animal ownership and demographics. J Am Vet Med Assoc 1992;201(7):990-2.

20. Franti CE, Kraus JF, Borhani NO, Johnson SL, Tucker SD. Pet ownership in rural Northern California (El Dorado County). J Am Vet Med Assoc 1980;176(2):143-9.

21. Griffiths AO, Brenner A. Survey of cat and dog ownership in Champaign County, Illinois, 1976. J Am Vet Med Assoc 1977;170(11):1333-40.

22. Franti CE, Kraus JF, Borhani NO. Pet ownership in a suburban-rural area of California, 1970. Public Health Rep 1974;89(5):473-84.

23. Patronek GJ, Beck AM, Glickman LT. Dynamics of dog and cat populations in a community. J Am Vet Med Assoc 1997;210(5):637-42.

24. Heussner JC, Flowers AI, Williams JD, Silvy NJ. Estimating dog and cat populations in an urban area. Anim Regulation Studies 1978;1:203-12.

25. Schneider R, Vaida ML. Survey of canine and feline populations:

Alameda and Contra Costa Counties, California, 1970. J Am Vet Med Assoc 1975;166(5):481-6.

26. Lengerich EJ, Teclaw RF, Mendlein JM, Mariolis P, Garbe PL. Pet populations in the catchment area of the Purdue Comparative Oncology Program. J Am Vet Med Assoc 1992;200(1):51-6.

27. Wise JK. U.S pet ownership & demographics sourcebook. Schaumburg, Illinois: Center for Information Management, American Veterinary Medical Association, 1997.

28. Hungerford LL. Use of spatial statistics to identify and test significance in geographic disease patterns. Prev Vet Med 1991;11:237-42.

29. Acha PN, Szyfres B. Zoonoses and communicable diseases common to man and animals, 2nd ed. Washington (DC) : Pan American Health Organization, Pan American Sanitary Bureau, Regional Office of the World Health Organization, 1987; Pan American Health Organization scientific publication; no. 503:963 p.

30. Bell JC, Palmer SR, Payne JM. The zoonoses : infections transmitted from animal to man. Baltimore: Edward Arnold; 1988:241p.

31. Hubbert WT, McCulloch WF, Schnurrenburger PR, editors. Diseases transmitted from animals to man. 6th ed. Springfield (IL): C. C. Thomas, 1975:1206 p.

32. O'Brien DJ, Kaneene JB and Poppenga RH. The use of mammals as sentinels for human exposure to toxic contaminants in the environment. Environ Health Perspect 1993;99:351-68.

Chapter 1

RATE-INDEPENDENT METHODS FOR THE STUDY OF DISEASE AGGREGATION IN EPIDEMIOLOGY

Context For The Current Study

In order to develop a methodological approach that has the greatest opportunity for success within the constraints and objectives of the current effort, it is prudent to examine the strategies taken by others in the investigation of clusters of disease. Cases of disease, in a very broad sense, can be thought of as events characterized by both a place (a geographic reference) and a time (a temporal reference). In this respect, they can be approached analytically in manners similar to other events which have space and time references, through modeling as spatial processes. The literature on spatial statistics is well developed, due in large part to the efforts of researchers in geography and ecology, and a number of excellent general references exist which can provide valuable background to the epidemiologist.[1-6] More specialized references dealing with applications in environmental science,[7] and population biology[8] may also be of utility. In addition, Cliff and Haggett[9] have published a useful atlas which extensively addresses geographic approaches to the analysis of disease distribution. This impressive and profusely illustrated work provides introductory material on cartographic methods, as well as spatial and time series methodology as they apply to epidemiology. In general, contagious diseases are emphasized; examples for animal diseases include fowl pest in England and rabies in continental Europe.

As one narrows this broad focus specifically to the analysis of disease aggregation, one finds that a trio of recent monographs[10-12] and a number of reviews of cluster evaluation methods already exist.[13-16] These articles have focused primarily on statistical issues. In addition, a fine review which places its emphasis on the epidemiological issues of cluster investigation and geographic epidemiology has been published by Paul Elliott and his coworkers at the London School of Hygiene.[17] More recently, Waller and Jacquez[18] have reviewed the issues surrounding statistical tests of hypotheses as they relate to disease aggregation. Their work is notable in that it explores the way that various cluster analysis methods imply the existence of particular underlying models of disease. Consequently, the designation of a statistically significant cluster of disease by one test may not easily be compared with one designated by a different test. From a practical standpoint, these differences can be used to help guide the choice of a test, depending on the underlying disease model one expects in a particular situation.

In examining these works then, it becomes clear that the theoretical aspects of the analysis of disease clustering have been, and continue to be, well described and scrutinized by the scientific community. Since that task has already been accomplished, it is not the intention here to provide a comprehensive and critical review of cluster assessment methods, nor to mount an in-depth mathematical or statistical evaluation. Rather, the discussion that follows is intended to provide an overview of some general methodological approaches that have been developed to investigate disease aggregation, with

perspective offered on how some of these approaches might find application in the investigation of animal diseases, particularly, chronic ones such as cancer. More broadly, the discussion will examine how they may enable an approach to the simultaneous assessment of disease aggregation in both animals and the humans with whom they share their environment.

Some Methodological Approaches to the Study of Disease Aggregation

Methods which are designed to analyze clustering of disease events have been categorized in a number of different ways. First, particular techniques are designed to assess clustering in space, clustering in time, or the interaction between the two. Second, techniques can also be subdivided based on whether they are *general* tests or *focused* tests :[19] These terms were originally developed by Besag and Newell.[20] General tests determine whether a pattern of disease cases in an area exhibits clustering anywhere within that area, and assesses its statistical significance, whereas focused tests determine whether or not cases of disease are significantly clustered around a particular predetermined point, often some putative source of increased disease risk. Third, clustering methods may be considered *proactive* or *reactive*.[21] Proactive methods are essentially a form of surveillance, monitoring over time for evidence of increases in the occurrence of a disease in a specified area. These may be contrasted with reactive methods, which search for evidence of clustering in an area where an increased occurrence of disease has already been reported. These latter methods have been the subject of considerable controversy with

respect to the validity of testing their statistical significance. In greatly simplified terms, the problem lies in the fact that the conventional tool used for judging statistical significance, the probability of a Type I error (or *p*-value) can only be interpreted in relation to the testing of a hypothesis posed *a priori*, *i.e.*, one formulated ahead of time without knowledge of the outcome. Testing whether the number of cases of disease is statistically significantly greater than what would be expected due to random variation in an area that was specifically chosen for testing because many cases were known to occur there is not valid. It may be thought of as a form of preselection bias,[22] since the same group of cases that were used to specify the hypothesis are used to test its significance as well. A rather nice discussion of the issues surrounding hypothesis testing in reactive studies for a general medical audience has been published by Olsen *et al.*,[23] while more in depth treatments intended for epidemiologists and statisticians are also available.[22,24-26]

For organizational purposes here, the methods will be grouped rather loosely according to whether they have been used primarily to study clustering in space, in time, or space/time interaction. However, it should also be noted conspicuously that this categorization is somewhat arbitrary. Many of the techniques can be used with equal utility for a geographic metric, a temporal one, or both.

Space/Time Methods

Knox's Method and Its Extensions

It is fitting to begin with a discussion of these methods because, from a historical standpoint, the development of tests of disease clustering (with the notable exception of Karl Pearson's purely spatial technique[27]) began with this group. This set of tests, nicely reviewed by Williams,[28] is intended to determine whether cases which are relatively close in time are also relatively close in space. The foundation method is that of Knox,[29] a method he used initially to assess the aggregation of cases of childhood leukemia in Northern England.[30] Knox's test requires a defined study area and time period be defined *a priori*, and requires as data the geographic coordinates (latitude/longitude or some other vertical/horizontal coordinates) and time coordinates of each case. All $n(n-1)/2$ possible pairs of cases are evaluated with respect to each other in terms of their distances apart in both space and time. The number of pairs that are "close" in space, in time, in both space and time, or in neither are determined and entered into the cells of a 2 x 2 contingency table. Statistical significance is determined by testing the observed frequency of case pairs that are considered close in both time and space against the expected frequency, which is considered to be a Poisson random variable. Note that the criteria for what constitutes "closeness" ("termed critical distances") are determined by the investigator, and should be based on known characteristics of the disease being studied and the at-risk population. In other words, one has to specify, say, that cases occurring within 500 meters and 1 week of each other are considered

"close". This arbitrariness is the principal drawback of Knox's test. If the critical distances are not known with reasonable certainty and so are chosen incorrectly, even strong clustering can be overlooked. Alternatively, if the critical distances are chosen after the data have been inspected, picking those critical distances which yield a significant test statistic, then comparison with the expected Poisson distribution is no longer valid, and so the test is essentially inoperative. It can be argued that performing the test iteratively with a range of critical distances for time and space may yield some additional clues to the etiology of clustering, but such an approach risks becoming a "fishing" expedition, with resulting concern for the effect that the multiple comparisons may have on the test of significance. In such circumstances, it seems prudent that the test results be stated as exploratory. Critical distances are also dependent on the metrics used for distance and time unless they are chosen as percentiles. An additional weakness of Knox's approach is that it is not sensitive to detection of disease processes that result only in spatial clustering, or only in temporal clustering. In spite of these limitations, Knox's approach has been one of the most widely used clustering techniques, and has the important advantages of not requiring knowledge of the size or demographic characteristics of the population-at-risk, or of the background frequency of the disease under study.

Using Knox's method as a point of departure, other authors developed alternative tests of space-time clustering. One such test which, like Knox, made use of critical distances, was developed by Pinkel and coworkers.[31] They

defined a critical spatial distance and assumed that cases further than that apart from one another were unrelated. The time distances between cases in that group were then used as a comparison distribution against which the time distances between the spatially close cases (*i.e.*, cases less than the critical distance apart) were compared. The distribution of time between cases was partitioned into percentiles for both the spatially close and not close groups, and expressed as "relative to an identified distribution" or "ridits". While no formal test of statistical significance was provided by the authors, confidence intervals for the ridits were calculated and compared between the two groups to get a idea of whether they overlapped, suggesting a lack of space-time clustering. The test could be extended to a two sample case to compare the distributions of more than one disease as well.

Barton and David[32] generally confirmed Knox's assumption that his test statistic had a Poisson distribution, deriving its exact mean and variance for large samples. They also originated a method to detect changes in the spatial distribution of cases over time,[32,33] somewhat analogous to an analysis of variance. They assessed temporal clustering by comparing the time between successive pairs of cases with the average time interval between all the cases within one year time cells. When the observed interval was less than average, clustering in time was said to occur. Spatial coordinates were obtained for the cases. Centroids for the point cloud of cases occurring in each of the time cells, and the overall centroid of all the cases during the entire study period, were determined. The mean squared distance between the two should theoretically

be similar to the mean squared distance between all of the cases and the overall centroid if spatial patterns tend not to vary over time, that is, if there is no space-time interaction. The test statistic will approximate 1 in that case, with values less than 1 indicating that cases close in time are also close spatially. The authors evaluated it against an *F* distribution or a normal distribution (depending on the number of temporal clusters) to determine statistical significance. Barton and David's technique, like Knox's has the advantages of not requiring knowledge of the size or make up of the population-at-risk, or of baseline disease rates. What is more, it does not require the subjective choice of critical distances as Knox did. However, the power of the test is limited for detecting clusters of a few cases within short spatial/temporal proximity but otherwise surrounded by widely dispersed cases.

Pike and Smith[34] noted that Knox's approach would be unlikely to detect space-time clustering for diseases for which the latency period was long, and so developed an extension. It is a graphical approach, rooted in the work of David and Barton.[35] Broadly speaking, the test defines two time periods for each case, one of susceptibility to infection, and one of infectivity to others. Similarly, the geographic area occupied by each case is also divided into areas of susceptibility and infectivity. A test statistic *X* is generated which represents the intersections of the areas and periods of various cases' infectivities with other cases' susceptibilities, such that *X* is the number of pairs of cases where one was in the right place at the right time to be infected by the other. The greater the magnitude of the test statistic, the greater the evidence for space-time

interaction attributable to infectivity. The statistical significance of X can be assessed roughly by use of another statistic *t* derived by Pike and Smith, or by a Poisson approximation, but others[28] recommend testing by Monte Carlo methods as being more defensible. Pike and Smith[36] also later developed a separate method incorporating controls matched to the cases, which tried to determine whether cases as a group had more pertinent contacts with each other than did their matched controls.

Another approach generalizes Knox's binary (close versus not-close) measures of space and time proximity by making them continuous.[37] Mantel defined the statistic *Z* as equal to the product of the spatial measure between two cases and the temporal measure between two cases, summed over all pairs of cases. In order to improve the power of the test to detect the short space and time distances characteristic of disease clusters, and minimize the effect of large measures on the statistic, reciprocal transformations were applied to the space and time measures. Constants were added to each to prevent the problem of division by zero in situations where the two cases being measured occupy the same spot in space or time. The statistical significance of *Z* can be determined by randomization, or using formulas the author derived for its expectation and variance. In addition, for large sample sizes, the distribution is approximately normal, although "large" in this case must be more than a few hundred.[28] Although Mantel's procedure avoids the problem of arbitrarily defining critical distances, subjectivity still enters in when defining the constants which are added to the space and time distance measures. Testing many such constants

also has the potential to degrade into a "fishing expedition" as discussed earlier, invalidating the test of significance, although a later paper by the same author suggested that a wide range of arbitrary constants did not affect the statistic greatly.[38] Mantel's method also assumes that the population-at-risk is fairly uniform across the study area, which may or may not hold in specific instances.

Two extensions of Mantel's method, itself an extension of Knox's test, which may be particularly relevant to the discussion here are those of Klauber. He was interested in the situation where one suspects that cases of a similar disease in two[39] or more[40] species have a common etiology. Indeed, the examples the authors used to demonstrate the methods were of the potential interspecies spread of leukemias in cats, dogs and humans. The tests operate on at least two separate sets of points, each cases of disease in separate species, for which the geographic and time coordinates of occurrence are known. The test statistic, like Mantel's, is the product of the measure of spatial distance and the measure of time distance between pairs of cases, summed over all possible pairs of cases. Here, however, one case of each pair is taken from each species. The test provides some flexibility with respect to the distance measures one uses. Binary measures of "closeness" specified by critical distances (as in Knox), or reciprocal distances (as in Mantel) can both be used, each with its attendant problems. Statistical significance is determined by randomization, or, subject to some assumptions, approximate comparisons can be made with a normal distribution. Williams[28] has noted that the test has a high degree of sensitivity; even when clustering involves only a few cases, high levels

of significance can be attained. It should be borne in mind that these tests were designed to detect a tendency for case pairs (one from each species) that are close in space to be close in time as well, and, consequently, are insensitive to clustering only in space or only in time.

Other Space-Time Methods

A space-time clustering method published the same year as Knox's method by Ederer, Myers and Mantel[41] has also been responsible for considerable debate and application in epidemiology. Unlike Knox's method however, it does not rely on analysis of distances between all possible pairs of points in space and time, but rather on the pattern of distribution of cases over groups of time periods within some geographic unit. Their method was based around the analysis of childhood leukemia cases in Connecticut. To fulfill the need to set analysis units *a priori*, the authors used an arbitrary spatial unit of a town (corresponding to a population of 3000 or less), and a temporal unit of 1 year, but grouped the data into five year blocks to minimize the effect of increasing population trend. They go on to examine the theoretical distribution of cases into "five-year-town units", based on Feller's [42, pp.36-38] probability theory of allocating *r* objects ("balls") into *n* cells. They considered only those five-year-town units with two or more cases. Using this approach, the authors determined the expected probability of having, say, all the cases end up in one year of the five year time block, or only some fraction of them. For example, if there were three cases in a five-year-town unit, there is a certain probability that all three would occur in the same year, a certain probability that two would occur in the

same year and one in another, and a certain probability that each of the three would occur in a different year. Ederer *et al.*, simply compared the observed distribution of cases in five-year-town units with these expectations *via* contingency tables and determined significance using a continuity corrected χ^2 statistic. The analysis was somewhat limited to small numbers of cases and time units by the arduous nature of the probability calculations, but two of the authors later published tabulations of the expectation and variance of the test statistic for larger numbers of cases and/or time periods.[43] This test was unique at the time in that it could detect either space-time clustering, or clustering in time alone.

Temporal Methods

In contrast to both space-time and purely spatial methods, there are comparatively few tests developed for the primary purpose of detecting temporal clustering alone. Three temporal clustering tests are described here, one of which was developed first as a spatial method, yet was versatile enough to permit extension by the authors for use with a time metric.

The first of the three is known as the Scan method, and was first proposed by Naus.[44] For a fixed area of study, a time line is created by plotting out in sequence the times of occurrence of all cases of the disease under study. One defines a critical time period or "window" based on the expected duration of the disease outbreak. The window must be a fixed time period, and set before the cases have been examined for evidence of clustering, in order for the test of statistical significance to be valid. The window is moved along the time line to

find the maximum number of cases *n* that fell within that fixed length of time. That maximum, along with the maximum number of windows that can fit within the period specified by the time line, and the total number of disease cases are the basis for the test statistic. It takes the form of a probability, $P(n,N,t/T)$, that of observing *n* or more cases in a time window of length *t*, where *N* is the total number of cases and *T* is the length of the time line. Given the arduous nature the probability calculations necessary for large disease outbreaks, both Naus[45] and other authors[46,47] have published extended tables and approximation formulae to aid in application of the method. Some points of guidance for application of the test warrant comment. First, the Scan statistic makes the assumption that the population-at-risk stays relatively stable over the time of the study. Second, the technique was not intended to detect spatial clustering, and so it is insensitive to purely spatial aggregation of cases. Third, like the critical distances in Knox's test, the time window in the Scan test should be set *a priori* based on known characteristics of the disease. When the pathogenesis of the disease is insufficiently characterized to facilitate a meaningful choice, setting of the time window becomes arbitrary. In order to better describe the disease under study, it may be useful to calculate the statistic using a variety of time lengths for the window. In that case, however, since the statistical significance of the test is directly affected by the window length, the *p* value thus generated has no meaningful interpretation. Issues of multiple comparisons also become relevant. Finally, the statistic may have limited power to detect clustering for extremely rare outcomes.[48] Yet despite these limitations, the technique has

enjoyed wide application in the study of disease outbreaks, including those in animals.[49]

Tango[50] formulated a method in which case frequencies are recorded in a series of successive time intervals. The time distances between these intervals are measured by an arbitrary metric of closeness or association, whose values a_{ij} are inversely related to the difference between any two intervals *i* and *j* in time. Matrix multiplication is used to combine the vector of case frequencies with the matrix of associations to generate an index of association *C*, whose value reaches a maximum value of 1 only in the case where all the cases fall in a single time interval. The test is applicable to grouped data with the assumption that the population-at-risk remains fairly uniform over the study period. The author later developed a method to standardize the index so that statistical significance could be more easily assessed,[51] and extended his temporal method for use in detecting spatial clustering.[52] The extension, according to Tango is "essentially identical in form" to Whittemore's spatial method[53] (to be discussed below), with the exception that Tango considers Whittemore's test "to be less adequate" than the method he proposes.

The third temporal clustering technique, that of Selkirk and Neave, is an extension of a method originally formulated to detect clustering of points on a line, *i.e.*, in one dimension.[54] It is one of a class of methods, originally introduced by Clark and Evans[55,56] for use in ecology, based on the distance of a point to its nearest neighboring point, known, not surprisingly, as "nearest-neighbor" methods. Those authors later noted that the technique could be

generalized to any number of dimensions.[57] Selkirk and Neave distinguish three separate cases: where points are distributed along a closed curve; along a line that extends beyond the section containing the points in both directions; and where the ends of the line are marked by points under study. They note that the two latter cases can be used to test whether events can be assumed to be randomly distributed in time, with time assumed to be linear and measured in some continuous metric (days, years, *etc.*) The test statistic is simply the sum of the distances from each point on the line (for our purposes, a case of disease in time) to the point (case) nearest to it. In a separate publication,[58] the authors used a binomial model, reportedly more accurate than methods used previously, to derive formulas for the exact expectation and variances of the test statistic under each of the three cases. The formulas themselves are presented by Selkirk and Neave,[54] as well as tabulations of percentage points of the test statistic for use in assessing significance for small samples ($n < 20$). The method for testing significance is also lucidly presented by Boots and Getis.[1, pp. 45-49] This general method for assessing temporal clustering has the advantages of not requiring distributional assumptions other than those inherent to the null hypothesis, and being straightforward to calculate even without the aid of a computer.

Spatial Methods

Statistical assessment of clustering of disease cases in space appears to have been initiated by Pearson.[27] He developed a method to investigate whether or not multiple cancer cases were significantly clustered within the same household by comparing the number of cases in occupied households in a particular community with an expected random distribution over those households. The test of significance then simply became a test of χ^2 goodness-of-fit. Mention is made of Pearson's method here in deference to its precedence and historical importance. However, household aggregations of disease constitute a specialized type of spatial clustering, one in which genetic factors, as well as environmental ones, play a significant role. Since issues associated with familial genetic disease risk factors are beyond the focus of the current study, tests of household disease aggregation will not be treated further here. Tests of household clustering other than Pearson's groundbreaking work have been published by Mathen and Chakraborty,[59] and Walter,[60] both of which have been expanded upon by Smith and Pike.[61]

For the present purposes, the remaining spatial clustering methods will be divided into subgroups based on their principal analytic approach. Note at the outset that these divisions are somewhat arbitrary, and because of the diversity of methods, particular tests could reasonably be placed in more than one category. In general, only rate-independent methods will be described, although some of the methods operating on disease rates will be itemized as well in the interest of completeness. Counted among the latter will be those methods which

require knowledge of an expected number of cases in a population-at-risk, since having these data available implies the ability to calculate the expected disease rate. Clearly, in the study of companion animal diseases, neither the expected number of cases nor the size of the population-at-risk are generally known, and knowledge of the former without knowledge of latter would be of limited use for describing the force of the disease in the population.

Moran's *I* and Measures of Spatial Autocorrelation

One way to assess spatial clustering of processes that can be characterized by continuous variables is through measures of spatial autocorrelation. Autocorrelation, put simply, is the tendency for regions which are close to each other spatially to have similar attributes, or, as Kitron and Kazmierczak[62] have described it, the degree of interdependence between values of a variable at different geographic locations. Various authors have developed statistical techniques to quantify and account for this tendency in the analysis of spatial data. The first appears to have been Moran.[63,64] His statistic, *I*, is essentially a measure of covariation of a variable across adjacent regions. It is similar in form to the Pearson correlation coefficient. One can use weights (say, the inverse of distance between the centroids of regions) to give the greatest influence to areas adjacent to another. *I* can range from - 1 to 1, with zero denoting a random distribution and 1 a strong tendency for clustering. It tends to be sensitive to extreme values. Another, the *c* statistic, was proposed by Geary.[65] It is calculated as a weighted comparison of values for pairs of regions, summed over all possible pairs. Positive autocorrelation drives *c* toward an

extreme limit of zero; a value of 1 denotes a random spatial distribution. This statistic is particularly sensitive to the influence of disparate values in adjacent regions. Both of these methods have been used widely and are described in detail elsewhere.[6,66] A third measure of autocorrelation is the rank adjacency statistic, *D*,[67] which quantifies the absolute difference in ranks for the data in adjacent regions, averaged over all regions. It has been used less than *c* and *I*, primarily in atlases of cancer, and is non-parametric. *D* approaches zero as autocorrelation becomes stronger. Statistical significance for *c* and *I* can be assessed by comparisons with a normal distribution, while Monte Carlo simulations are necessary to determine the significance of *D* statistics. Simulation work has also been used to refine the random distribution of *D*.[68] A fourth and more recent measure of autocorrelation has been proposed by Getis and Ord.[69] They introduce a family of statistics *G*, which measure the degree of association between all weighted points within a radius of distance *d* from some weighted reference point. The weighted points can also be used to represent areas, *e.g.*, where the points are centroids of areas such as counties. The *G* statistics are different from other measures of autocorrelation in that they measure autocorrelation at a more local scale, whereas others measure it over the entire area under study. Given this unique property, the authors suggest that *G* statistics can be used in conjunction with more global measures such as Moran's *I* to more completely characterize autocorrelation at different spatial scales. How one determines the statistical significance of *G* can depend on how the statistic distributes for a particular data set, although the authors propose a

normal approximation. In a subsequent paper,[70] the authors reconfigure the statistics to follow a normal distribution with zero mean and unit variance to remove the influence of individual data sets and ease determinations of significance.

With respect to application of these measures to actual disease data, Kitron and Kazmierczak[62] used Moran's *I* to relate numbers of cases of Lyme Disease in humans in Wisconsin to environmental characteristics such as vegetation and surveys of the tick vector of the disease. All three were correlated. This allowed them to produce a map delineating high risk counties. Though he used the autocorrelation on disease rates, Walter's[71,72] pair of papers investigating the spatial patterns of cancer incidence in Ontario are noted here because of the fine job they do pointing out the importance of taking patterns of regional population structure into account when using measures of autocorrelation. The papers also test the power of *c*, *D* and *I* to detect different regional patterns of disease distribution. Moran's *I* had the highest power most often, and *D* the least. Walter has also presented more general commentary on the role of spatial autocorrelation in assessing spatial clustering of data appearing on maps.[73]

Quadrat Methods

A second approach to analysis of spatial aggregation of disease involves analysis of quadrats. Quadrats, originally introduced by plant ecologists,[74] are simply sampling areas within a larger study area that are of any consistent shape and size. Quadrats can also be defined on a temporal scale, as time

periods of predetermined length, if one wishes to test for temporal clustering. A straightforward and detailed discussion of methods and issues involved in quadrat analysis is presented by Boots and Getis,[1, pp. 17-35] while a brief presentation oriented toward epidemiology has been published by Skelton.[75] Briefly, the technique in its simplest form involves recording the numbers of cases that occur in quadrats spread over the study area, tabulating the number of quadrats containing various case frequencies, and comparing those frequencies to the numbers of cases expected under an expected null hypothesis of a random distribution, which is assumed to be Poisson. The observed and expected case frequencies are tested for a statistically significant difference by χ^2. A test statistic following a Student's *t* distribution can be calculated using the mean and variance of the distribution of the observed case frequencies to judge whether the observed pattern is significantly clustered or dispersed. When variance exceeds the mean, clustering is suggested; when it is less than the mean, a regular pattern is indicated. When the two quantities are equal, as is the case under a Poisson distribution, randomness obtains.

A number of the more sophisticated and recent tests of spatial clustering are linked to simpler quadrat methods in that they operate on counts of cases in geographic cells which are generated to cover the study area. They are grouped with quadrat methods here for that reason (although they vary with respect to the uniformity of the size and shape of the geographic cells they employ, and so, strictly speaking, one could also argue that some are not in fact quadrat methods). The general clustering tests of Openshaw *et al.*,[76,77] Turnbull

et al. [78] and Kuldorff,[22,79-81] as well as the focused tests of Stone[82,83] and Lawson/Waller *et al.*,[25,26,84] and Besag and Newell's test (which can be used as either a general or a focused test[20]) all fall into this category. Many of these methods have received detailed and extensive scrutiny elsewhere.[19,21,85,86] All of them require knowledge of either the expected number of cases of disease or the baseline disease rate in each geographic cell, and so will not be treated further here. That being said, these methods appear to be at the cutting edge of cluster assessment in human medicine, and to the extent that problems with accurately enumerating populations-at-risk may at some point be overcome, they have the potential to be of great use in future studies of disease aggregation in companion animals.

Distance Methods

Another group of methods, one used perhaps more than any other to investigate spatial clustering, are those operating on distances measured between pairs of cases. For these methods to be applicable, one must generally have knowledge of the precise geographic location of each case. The advent of Geographic Information Systems (GIS) have greatly increased the ease with which these methods can be applied. For purposes of this discussion, some studies will be described that have investigated disease clustering using four different approaches: spatial intensity, nearest-neighbor techniques, Whittemore's approach, and second-order analysis.

The spatial intensity approach is a type of focused clustering test incorporating the theory of spatial point processes,[5] developed by Diggle and his

associates.[87,88] Spatial intensity refers to the frequency of health events per unit area over a variety of locations. Risk of disease is assumed to vary as some prespecified function of distance from a putative focus of increased risk. For the risk model, one must also specify the background risk λ_0, which reflects the underlying population structure of the study area. λ_0 can be estimated by selecting controls and conditioning on the case and control locations. Alternatively, one can use the spatial distribution of some common disease, the risk of which is assumed to be unaffected by the focus, and use it as an estimate of the background spatial variation in risk, *via* kernel estimation.[89-91] If there is no increased risk of becoming a case associated with distance from the focus, then the spatial intensities of the cases should be the same as those of the control disease. Statistical significance is assessed by goodness-of-fit tests. A very attractive property of using a case-control approach is that it allows for control of covariates using data gathered for individuals, and so avoids the many problems associated with assigning values for covariates based on grouped data (say, county averages). An example of the use of spatial intensity methods is provided by Biggeri *et al.*,[92] who used them in a case-control design to explore the relationship between some sources of air pollution and the risk of lung cancer in Italian men.

Mention has already been made of the nearest-neighbor approach in the context of the assessment of temporal clustering. It is one of the older distance-based methods, the basic principles of which have been well-documented in standard texts,[1, pp.36-45] and applied to epidemiological data sets.[93] However, it

has also been significantly modified in the past eight years by investigators seeking to adapt it to particular disease investigations. Ross and Davis[94] used two dimensional nearest-neighbor analysis to investigate childhood residences of Hodgkin's Disease patients in northwestern Washington. Their application was unique in that they used a permutational approach to derive a study-specific distribution of expected nearest-neighbor distances. This was a response to their observation that the population was not homogeneous across the study area, and consequently, the homogeneity assumption of a null Poisson distribution did not hold. Cuzick and Edwards[95] proposed a method where controls are chosen for each case, so as to resemble the cases with respect to disease risk factors. One then identifies *c* nearest neighbors for each case of disease. The test statistic T_c is then simply the sum over all the cases of the number of *c* nearest neighbors for each case that are also cases. In more general terms, under the null hypothesis of random distribution of cases and controls, one would expect the *c* nearest neighbors of a case to be about equally split between cases and controls; if the cases' nearest neighbors are composed predominantly of other cases, spatial clustering is suggested. The significance of the test statistic can be determined either by randomization, or by using the expectation and variance provide by the authors. This test, due to the incorporation of controls, also has the advantage of being able to adjust for confounders and other covariates at the individual, rather than the group, level. In an interesting extension of Cuzick and Edwards' test, Jacquez[96] has published a generalized method to apply in situations where the exact locations of the

cases and controls are not known with certainty (Cuzick and Edwards' original test requires exact locations). He later developed the concept of using uncertain space or time locations into a generalized matrix method[97] that could be used in conjunction with other nearest-neighbor tests, as well as space/time clustering methods like Knox's[29] and Mantel's.[37] Finally, Jacquez[98] has also proposed a nearest-neighbor method for detection of space/time clustering, based around a statistic measuring the number of pairs of cases that are nearest neighbors in time as well as in space. The author includes treatment of how the test can be extended to case-control data and to uncertain space and time locations, and found it to be more powerful than other commonly used tests of space/time clustering.[29,37]

A third distance-based approach to spatial clustering involves a group of methods having their origin in the work of Mantel and Bailar,[99] Whittemore and coworkers[53] and the time clustering test of Tango.[50] Whittemore's test is notable because it is essentially a hybrid test, using not only case counts in geographic cells, but also paired distance measurements. As background for her approach, Whittemore points out that space/time clustering techniques like those of Knox[29] are generally not suitable for study of chronic diseases, since cases caused by a common etiology might be spatially close, but were unlikely to be close in time due to the length and variability of latent periods. In order to overcome the artifactual appearance of clustering that could occur as a result of heterogeneous population distribution over the study area, Whittemore stratifies the cases into census tracts and into age groups, and assumes that the number

of cases in each stratum are independent Poisson processes with means proportional to the number of individuals in the stratum. The test statistic *U* is simply the mean distance δ between all pairs of cases, with the centroid of the census tract acting as the spatial reference for all the cases within it. The expectation and variance of *U* are derived and are approximately normal for large samples. The authors note that the technique could be adapted to detect either spatial or temporal clustering, depending on the metric chosen. Unfortunately, the technique's requirement of knowledge of population numbers in census tracts and age strata limit its usefulness for investigating companion animal diseases. As mentioned previously, Tango[52] has published a similar method which he found to be more powerful. An example of Whittemore's test applied to Hodgkin's Disease data in the San Francisco area was provided by Glaser.[100] In a recent paper, Ranta *et al.*[101] extended Whittemore's approach to incorporate temporal data, making her two dimensional census tracts into three dimensional space/time "cubes", with time as the third dimension. They also introduced the unique feature of weights on the space and time distance measures, such that by varying the weights, a purely spatial test, a purely temporal one, or a space/time interaction test resulted.

Second-order analysis (also called *K* function analysis) is yet another type of distance technique, one in which the theoretical focus is on the variance (or second moment, hence the name) of the distances between all possible pairwise combinations of cases, rather than the mean pairwise distance used for many other tests. It tests hypotheses of randomness by testing the proportion of

the total possible pairs of cases for which both members of the pair are within a specified distance *d* of one another.[102] The technique defines a function *k* which, for a prespecified distance *d*, consists of the reciprocal of the intensity (the mean number of cases per unit area) multiplied by the expected number of additional cases within distance *d* of a case chosen at random.[5, p. 47] The *k* function can be incorporated into the calculation of a test statistic *L*(*d*) which is equal to *d* when the case pattern is random, but which is greater than *d* when cases are clustered or less than *d* when they form a more regular pattern than one would expect at random. Statistical significance can be determined either by randomization tests, or by approximations published by Ripley.[103,104] By calculating the test statistic for a series of distances *d* defined in regular steps of increasing size (say, 100 meters, 200 meters, 300 meters, *etc.*), one can get not only a qualitative idea of whether the cases being studied are significantly clustered, but the quantitative distance scale at which their tendency to cluster is the strongest. That, in turn, can provide insight into the scale at which the disease process that is causing the cases to be clustered is operating. Boots and Getis[1, p. 56] have pointed out a number of advantages of the approach over more traditional ones, among them the fact that it potentially provides more information about spatial pattern than any other existing technique. An additional advantage lies in that fact that it can be used to assess the dependence of patterns of cases from two separate distributions on each other.[5, Chap. 7] For example, one could use second-order analysis to assess whether the geographic pattern of cases of a disease in one species is independent of the

pattern of a similar disease in a second species. Failure to find independence would suggest that the processes which caused the case pattern in each species were dependent on each other. The theory of second-order analysis is described by Diggle[5, p. 47] and Ripley,[105,106] with a somewhat more accessible treatment available as well.[1, p. 56] Although second-order analysis as described here would generally be categorized as a general test of clustering, methods for focused testing have also been published.[102] There is a conspicuous lack of epidemiological studies of disease aggregation employing second-order analysis, such that examples of application of the technique are to be found largely in the geography and ecology literature. The work of Diggle and Chetwynd[107] appears to be the single notable exception. They develop a "random labelling" method for application to inhomogeneous Poisson processes that more accurately describe spatial variations in density of the population-at-risk, applying the method to Cuzick and Edwards' data[95] on childhood hematopoietic malignancies in North Humberside, England.

Cartograms

A final method of analyzing spatial clustering of disease which will be mentioned here is the use of cartograms (maps which distort location and geographic area so as to equalize the density of some other variable, such as population-at-risk).[108] The transformation of map features removes the variability of population from place to place, allowing for the analysis of other factors that influence disease aggregation without population's confounding influence. However, it retains complete geographic detail without the necessity of

combining geographic subdivisions in order to equalize population, as well as preserving adjacency of the subdivisions and the cases that occur there. The technique is computationally intensive, and a visual examination of the map transformations is really necessary in order to appreciate the method. Obviously, such a presentation is beyond the capabilities of the present discussion, so the reader is referred to Selvin *et al.*[108] for a lucid example. Treatment of the theory of the technique has been published elsewhere.[109,110] Selvin and coworkers also illustrate the use of a statistical test of the randomness of the spatial pattern of cases plotted on the density-equalized cartogram. The statistic is based on the mean squared distance among all the cases;[111,112] the authors derive the expectation and variance of their proposed statistic under the null hypothesis of random case distribution. Statistical significance is determined by comparison with a normal distribution. The authors illustrate their technique by analyzing the aggregation of cases of 26 kinds of human cancer in white residents of San Francisco from 1978-1981. Other examples of the use of cartograms for epidemiological data are also available.[113,114]

Methods Incorporating Rates

At the outset of the discussion of spatial methods, it was noted that methods operating on disease rates or requiring knowledge of expected numbers of cases would be mentioned in this discussion only cursorily. Some of the prominent methods are noted here purely in the interest of completeness. Methods designed to detect purely spatial clustering have been described by

Grimson[115,116] and Ohno.[117] A temporal method has been proposed by Chen *et al.*,[118] adapted from a previously developed spatial surveillance method.[119] A technique developed by Aldrich[120,121] known as REMSA can be used to assess clustering according to any characteristic of the cases that can be expressed as a nominal variable (gender, age, place or time of diagnosis, *etc.*). Methods intended specifically for public health surveillance,[122] including the surveillance of congenital anomalies,[119,123-128] are also documented in the biomedical literature.

Previous Approaches Taken in the Epidemiological Study of Cancer Aggregation in Companion Animals, and the Simultaneous Study of Companion Animal and Human Cancers

With the exception of a single recent study[49] using time or space-time clustering methods that do not require rate calculations, all previous study of the geographic distribution of cancer cases in companion animals has thus far roughly fallen into two categories: animal cancer registries and pet censuses. Both have been aimed at sufficiently characterizing an animal population-at-risk or reference population so that disease rates could be calculated. From these have also sprung the majority of the few studies which have attempted to simultaneously investigate cancer occurrence in humans and animals occupying the same geographic area during the same time period.

Rate-Based Methods: Cancer Registries

At the outset of this chapter, it was noted that the assessment of disease aggregation in veterinary medicine has been extremely limited because of limitations in data availability and methodology. Nonetheless, a number of authors have managed to execute studies which have circumvented these difficulties to greater or lesser extents, and it is worthwhile to examine their efforts for the context they provide for the current study. Scrutiny is turned first to animal cancer registries. An overview of existing registries current to 1991, along with some discussion of their limitations and potential, has been published previously.[129]

A seminal work is Priester and McKay's monograph The Occurrence of Tumors in the Domestic Animals.[130] Notably, the foreword for that work recounts that the project arose from "the expectation that research into neoplasia among domestic animals would provide new insight into the origins of human cancer". The monograph is a greatly expanded effort based methodologically on Priester's earlier work with Nathan Mantel.[131] It reports on data collected for over 41,500 cases of neoplasia in eight species at 14 veterinary teaching hospitals in the U.S. and Canada between 1964 and 1977. These data were derived from the Veterinary Medical Data Program (VMDP), initiated at Michigan State University in 1964. The establishment of the VMDP and the coding system which made it possible, the Standard Nomenclature of Veterinary Diseases and Operations (SNVDO)[132] are explained at length by the authors. Noting the impracticality of enumerating census-type reference populations for their cases,

a functional reference population was created by including all records in the VMDP, essentially all animals seen for all reasons during the study period. While admitting that confounding was likely, they noted that such a reference population was a reasonable means of "estimating characteristics of animal populations when no other way of doing so is available." Similarly, a reference statistic called animal-years-at-risk was calculated to provide "a reasonable reflection of qualitative composition of the actual population of each of the species categories". Each animal that visited one of the reporting institutions in a given year contributed a year at-risk to the statistic. The majority of the work consists of tabulations reporting cancer incidence broken down by species, breed, sex, age, tumor behavior (malignancy), geographic region and institution. For all neoplasms that made up more than 5% of the total in a species, additional tabulations of relative risk (*R*) for various histologic types and sites by breed and age are presented, with risk calculations based on the method of Mantel and Haenszel:[133]

$$R = \frac{\sum[\frac{a \cdot d}{N}]}{\sum[\frac{b \cdot c}{N}]} \qquad [1]$$

where *a* and *c* are the numbers of cases with and without the breed/age factor being tested, *b* and *d* represent the numbers of control animals with and without the factor, and *N* is the total number of cases and controls. Animals of all breeds or all ages combined were considered control animals for the purposes of risk

calculation. The monograph also provides plots of relative risk by age for each of these more common histologic types, with relative risk of benign neoplasia in that species plotted on the same axes for comparison.

Another approach that has been carried out with success is the establishment of regional animal cancer registries in defined geographic areas. The groundbreaking work in this area was that of Dorn and coworkers,[134,135] establishing the Animal Neoplasm Registry in Alameda and Contra Costa Counties, California. Those study areas were chosen because of large human and animal populations and the existence of complementary data from other studies. Using data obtained from a preliminary study of characteristics of veterinary practices in the area,[136] the authors enrolled all 65 veterinary practices in the two counties and 11 practices in adjacent counties; formal reporting of cases to the registry began in July of 1963. Practitioners filled out case reports they had been provided by the project, and mailed biopsy or necropsy samples to the registry for histopathological examination. In return for their cooperation in reporting, the participating practitioners received histopathology reports, valuable diagnostic confirmation which very few veterinary practices received at that time. Tissue submissions were classified both according to the SNVDO[132] and the seventh revision of the International Classification of Diseases (ICD);[137] only malignant tumors were included in the study. The authors note that animals in households which did not use veterinary services were not captured by their methods and so were not considered included in the population-at-risk. These animals constituted 13% of the dogs

and 25% of the cats in the baseline animal population recorded in a probability sample survey of households.[138] Estimates of the population-at-risk within various breed-, age- and gender-specific categories were obtained by multiplying by a population sampling factor, and relative risks calculated based on the Mantel-Haenszel method,[133] as described above. Here, controls were chosen from all animals not reported in the Animal Tumor Registry that were identified in Alameda County households using veterinary services.[138] Significance of R was determined by χ^2 tests; a 97.5% level was considered significant to adjust for multiple comparisons. Over the three year course of the study, 5647 incident cases were reported, of which 4842 occurred in dogs. The authors estimated annual incidence rates for cancer of all sites to be 381.2/100,000 in dogs and 20.4/100,000 in cats, and noted that these rates were elevated compared to the corresponding all sites rate in humans in the same county.

It is notable that despite the comprehensive scope and impressive effort made by the study, it was still not able to enumerate a true population-at-risk, only to estimate one, and animals not seen at veterinary clinics were not captured in the risk estimates. Nevertheless, this study and its estimates remain unsurpassed in the veterinary epidemiological literature despite thirty years having past and the substantially greater technological resources now available. This fact speaks to the intractable nature of the problem of calculating accurate incidence rates for cancers in companion animals. More extensive background on the organizational aspects of the survey has also been published.[139,140] A

related study that is also of potential interest is a survey of dogs admitted to the animal control facility in Alameda County,[141] in which a prevalence rate of 70 neoplasms per 1000 dogs was recorded. Some 1096 dogs were examined in that study.

A second tumor registry, the Tulsa Registry of Canine and Feline Neoplasms, was established in 1972.[142] The registry received submissions obtained at necropsy or surgical biopsy from all veterinary practices in Tulsa County and three practices from an adjacent county. While similar in many respects to the structure of the California registry, there were some differences that bear mention. Less than half as many veterinary practices (35) reported cases to the registry during the first year. Classification of submissions was based on SNVDO[132] also, with the exception of mammary tumors, for which a new classification scheme, intended to be more clinically useful, was developed.[143] Rates for benign as well as malignant tumors were calculated. For the purposes of calculating incidence rates, the study considered the sum of all animals visiting a participating hospital during a year to be the population-at-risk. Of particular interest was the fact the status of each case, once diagnosed initially, was also followed up with re-checks at 2 and 6 months post-diagnosis and every 6 months thereafter. The authors note that their registry had the advantage of a more stable dog and cat population that was relatively geographically concentrated and more isolated than most cities of the same size. This was considered important so that the effects of migration could be minimized, and the occurrence of tumors could be attributed to a specific

geographic region, a point the authors considered a relative strength compared to Dorn *et al.*'s studies. Rates of incidence for all site cancers were 1126 cases per 100,000 dogs, with a rate of 507 cases per 100,000 for malignant tumors only. A total of 63,504 dogs were seen during the first year of the registry. The authors point out that because the census did not include owner's names and addresses or other means of uniquely identifying the animals, the number of individuals counted more than once because of crossover from one hospital to another could not be estimated. However, they concluded that the effect of these potential double countings on the denominators of their incidence rates was "probably negligible". Notably, MacVean *et al.* caution against comparing incidence rates of different animal tumor registries, because of the different methods used to derive reference populations. They also discuss the advantages and disadvantages of their denominators at some length, and conclude that ascertainment of a true reference population was essentially impossible, and that the use of a population-at-risk composed of all animals seen at all the participating hospitals (what they term the "veterinarian-using" population) is a reasonable and efficient alternative. Indeed, this seems a sensible, if unfulfilling, resolution of the problem. It must be mentioned, however, that the degree to which the population of dogs visiting veterinary hospitals is representative of the general population of dogs remains unknown, and there is evidence to suggest that spurious inferences regarding the relative risk of cancer in the general population of dogs can result from using hospital populations as populations-at-risk.[144-146]

Rate-Based Methods: Pet Censuses

Another approach that has been taken to characterize companion animal populations for subsequent use in population-based studies is the establishment of pet censuses. While much of the content of these studies focuses on demographic correlates of pet ownership, those issues are not strictly relevant to the present discussion and will not be covered here. One such census has already been described in conjunction with Dorn's work.[138]

The foundation work among these censuses is, arguably, that of Franti and colleagues.[147] It is notable in that it was part of a larger health survey of the human population in Yolo County, California (one county removed from Contra Costa and Alameda Counties), and, consequently, can be viewed as an approach to the simultaneous study of human and animal health in the same geographic area and time period. The study consisted of a two-stage stratified random sample of private households in the county, and gathered demographic and socioeconomic data on the households, as well as health information on individuals within the households. Information concerning health complaints, use of and expenditures for medical services and broader issues such as nutrition, perceptions of environmental quality and family planning was also gathered. Data were collected from 3638 people in 1091 households, of which 67 percent reported owning one or more pets. With respect to the human health issues investigated, pet owners tended to have better nutrition, a greater sensitivity to air pollution, greater access to private physicians and clinics, and a more favorable opinion of family planning. Interestingly, pet owners and non-

owners reported about equal incidents of annoyance or illness associated with uncontrolled dogs and cats. These included such things as allergies, bites, annoyance with barking or feces deposition, fear of attack, etc. More detailed characterization of pets from this study has been presented in a separate report.[148] In a later study, this same group of authors also report some community health findings in humans from El Dorado County, also in northern California.[149] The rationale for investigating this county was that it was largely rural and removed from the influence of an urban area, a characteristic unique among the counties in which pet ownership demographics had been studied thus far. The study was similar to the one carried out in Yolo County, and used the same design and instruments, to which were added some questions intended to gather additional health information. No statistically meaningful association was found between pet ownership and the number of persons reporting having cancer, hypertension, "a heart condition", being mentally or emotionally ill, or having asthma or hay fever. There was a significant positive association between pet ownership and frequent headaches in adults and frequent sore throats in school-aged children, and a significant negative association between pet ownership and frequent diarrhea in preschool children. Franti *et al.* interpreted these findings to mean that pet owners were more aware of chronic health conditions such as hypertension, and that there was little support for an relationship between pet ownership and owner's health. They also concluded that apparent associations between pet ownership and frequent diarrhea and sore throats were more likely due to economic and lifestyle factors than animals

per se. No explanation for the relationship with frequent headaches was hypothesized. As in Yolo, pet owners in El Dorado County were slightly more concerned about air pollution than non-owners.

Other reports derived from regional pet censuses have not directly addressed human health issues as did the California research, but may be of interest for those interested in companion animal population dynamics,[150-153] estimation/management of urban populations,[154-157] or simply to examine the consistency of findings across diverse geographic areas.[158-160]

Rate-Based Methods: Hybrids

A hybrid of the cancer registry and pet census approaches worthy of specific attention here has been carried out by investigators at the Purdue Comparative Oncology Program (PCOP).[161,162] While previous pet censuses executed sample surveys in which households were identified based on demographic characteristics/regions, and a percentage contacted for their responses, the PCOP group used a single stage random-digit dialing telephone survey to estimate populations of pet dogs and cats, as well as cancer case ascertainment. The PCOP itself is a animal tumor registry for two central Indiana counties, Marion and Tippecanoe, which registers cancer cases based on submissions from more than 100 participating private veterinary hospitals. The registry provides histopathologic diagnosis free of charge, and in turn requests submissions from all the suspected cancer patients the participating practitioners examine. Registrations began in 1979, and as with MacVean *et al.*'s study, once diagnosed, cases are followed up biannually until the patient's

death. Between initiation of the study and the time of publication, > 19,000 primary and metastatic tumors were registered in more than 13,000 animals.

Random digit dialing within three digit telephone prefixes serving all or part of the study counties was used to select households for interview. An adult was administered a ten minute interview covering the number and characteristics of human and animal residents, number of unique phone lines, use of veterinary services and household demographics. In addition, interviewees were also asked if they would take their dog or cat to a veterinarian for evaluation if it had "a lump you thought might be cancer", as well as how much they would be willing to pay for cancer treatment. Canine and feline populations were estimated by multiplying the number of dogs and cats reported in interviews by the inverse of the sampling fraction, which, in turn, was calculated as the sum of the weighted number of people residing in the interview households divided by the estimated human population of the county. The weight was the inverse of the number of unique phone lines in each household. Seven hundred thirty-one eligible residences were identified from 2540 telephone numbers, of which > 77% responded. Dog populations estimated in Tippecanoe and Marion Counties (95% confidence limits) were 18,000 (14,445, 21,555) and 144,039 (121,555, 166,523) respectively, while cat populations were 17,165 (12,569, 21,761) and 94,998 (74,348, 115,648), respectively. The proportion of animals that owners reported would be evaluated for cancer (*i.e.*, case ascertainment for the PCOP registry) ranged from 76.1% to 89.7% for cats and from 92.0% to 93.3% for dogs; this variation was not significantly different between either counties or

species. The percentage of owners willing to pay at least $200 for cancer workup differed significantly by county, ranging from 21.7% to 42.7% for cats and 38.9% to 56.7% for dogs. For both species, pet owners in the more urban Marion County were more willing to pay for cancer treatment, although there was no significant difference by county or species in the willingness to owners to have a suspect mass evaluated by a veterinarian.

While the authors were pleased with the general concordance of their demographic findings with those of the 1980 U.S. Census, they did note that their survey had unintentionally oversampled older persons, and that because pet ownership rates tend to be higher in younger persons, their pet population estimates may have been underestimates. They also carefully point out that "estimates from this study may not be generalizable to other periods, geographic areas, cancer registries, or disease registries". Also, to the extent that the interviews overestimated the owners' willingness to seek oncological evaluation for their animals, or the participating veterinarians did not report all suspected tumors to the registry, Lengerich *et al.* point out that their estimates of case ascertainment by the PCOP may be too high. Nonetheless, the study is valuable for its use of random digit dialing, and particularly for its estimates of case ascertainment for a cancer registry. It is arguably the best effort recorded in the literature since Dorn *et al.*

Some points already touched upon are sufficiently important to this discussion to revisit and summarize. First, despite the impressive and labor intensive efforts of all these studies, none was able to enumerate a true

population-at-risk for companion animals, only to estimate one. The PCOP study probably comes closest to accomplishing this goal, as it was the only study which captured estimates of pet animals not seen at veterinary clinics. No study has been able to include ownerless animals in the estimates of population-at-risk, and so any incidence rates calculated will be unavoidably inaccurate. It could be argued that the degree of inaccuracy is likely to be small, but such a conclusion is subject to a great deal of uncertainty. Thus, even under the best of circumstances, the approach of calculating accurate incidence rates for companion animal diseases faces serious problems. Second, even when incidence rates can be calculated and are assumed accurate, their validity is limited to the population from which they were calculated, because of differences in methods used to derive reference populations, not to mention differences in disease risk which vary geographically and over time. This effectively limits the incidence rates derived from studies such as those of MacVean *et al.*[142] and Dorn *et al.*[134,135] to the areas where the studies took place. Moreover, comparison of disease rates between species (say, dogs and humans) even within a common geographic area and time period is of questionable validity unless it can be demonstrated that the methods used to derive the numerators and denominators of the rates are comparable. Third, inferences concerning cancer and other diseases in the general population of dogs from findings employing hospital populations as populations-at-risk are another source of uncertainty that may render incidence rates inaccurate.

While none of these factors dictate that epidemiological studies of

disease in companion animals should forever abandon the use of rates, it does suggest that there remains a need for more widely applicable methods of analysis which can investigate disease aggregation within and between species without reliance on rates.

Rate-Independent Methods

To investigate the possibility of spatial and temporal clustering of horses shedding *Salmonella krefeld* in their feces in an Intensive Care Unit, Knox's test[29] and the Scan test[44] have been employed.[49] The study was intended to evaluate the role of stall cleaning as a possible means of transmission of nosocomial salmonellosis from a horse shedding organisms to others housed nearby at a veterinary medical teaching hospital. Detection of clustering with respect to the index case would suggest cleaning personnel as a means of transmission, whereas random distributions would suggest that horses acquired salmonellosis prior to admission. The investigators considered multiple time windows of 2 to 14 days for the Scan and Knox tests, and multiple distance windows of 1 to 5 hospital stalls. Significant temporal clustering was reported for time windows of 5, 6, 7 and 8 days, but no significant spatial/temporal clustering was detected among any combination of time or distance windows for any of the 190 pairs of case horses examined. It was concluded that the results did not support increased risk for horses housed adjacent to horses shedding *S. krefeld*, but rather that fecal shedding was a result of infection acquired prior to admission to the hospital. Although this appears to be the only study in which

rate-independent clustering methods have been used in the study of disease aggregation in companion animals, several published reports have investigated disease clustering in food animals.[163-166] Note that thus far, no studies have used rate-independent methods to investigate cancer or other chronic diseases in companion animals.

Approaches to the Simultaneous Study of Human and Companion Animal Cancer Aggregation

Only a handful of other published studies have attempted to investigate the simultaneous occurrence of neoplastic disease in humans and animals. Three quarters of these are now nearly three decades old, and half grew out of the previously described Animal Neoplasm Registry in California.

The chronologically first study is mentioned only anecdotally by Tjalma.[167] He notes a preliminary study conducted by the Epizootiology Section of the U.S. National Cancer Institute in which family members exposed to 55 pet dogs diagnosed with leukemia or lymphoma were compared with families exposed to the same number of control dogs free of cancer. One human cancer case was reported among the families of dogs with cancer, while three cases were reported among family members of the controls. Tjalma notes that while sample sizes were too small to enable statistical inference, the results of the study suggest any association between the occurrence of cancer in the two species "is of a low order of magnitude".

Further expanding on their already pioneering work, Dorn, Schneider and

Klauber,[168] using a retrospective design, identified residence households of all canine cancer patients reported to the Animal Neoplasm Registry between July 1, 1963 and June 30, 1965. This amounted to 524 case households. Seven hundred five control households were identified based on a probability sample of dog-owning, veterinarian-using households in Alameda County, California. Owners in both series of households were interviewed to determine the number and ages of all dogs and persons that had lived in the household for 5 years prior to interview; those that had lived in the household for ≥ 6 months were considered household members. Human malignancies reported in interviews were confirmed by checking against registration in the California Tumor Registry or by death certificates. To control for confounding by age, a "household human-cancer risk" index was calculated for each household based on Alameda County sex- and age-specific human cancer incidence rates in 10 year age groups; households were also stratified into three year categories based on the age of the oldest dog in the household. Statistical significance was determined by the Mantel-Haenszel *R* procedure.[133] Four hundred forty-seven different primary canine neoplasms were recorded in the 427 case households used in the analysis. There were 54 human neoplasms reported in 52 of the 427 case households, of which 26 (48%) were verified. Thirty-nine human neoplasms were reported in 38 of the 520 control households included in the analysis, of which 21 (54%) were verified. No significant association of human and canine cancer in the same household was observed, even among the highest household human-cancer risk group, corresponding to persons ≥60 years of

age. Relative risk values for human cancers in households with dogs afflicted with malignancies versus controls were 1.34 for the verified human cases and 1.59 for verified and unverified human cases combined. No confidence intervals for the relative risks were calculated. Sample sizes were insufficient to statistically analyze interspecies relationships between specific cancer types and sites, but interestingly, four verified human digestive neoplasms were all associated with dogs in the same household which had malignant melanoma of the eyelids or lips. The authors noted that the face of the dog was a site with which humans would frequently have close contact, and so concluded further study was warranted to investigate this preliminary association.

Again using the data resources of the Alameda County Tumor Registry and Animal Neoplasm Registry, Schneider[169] compared cases of human breast cancer and canine breast cancer over the five year period spanning 1960 through 1964. The human study population consisted of 1623 female and 8 male cases, all Caucasian. Comparisons were made with histologically-confirmed first occurrences of malignant canine breast neoplasia, of which there were 460 female and 12 male cases. Ages of the dogs at diagnosis were converted to human equivalents,[170] and age-specific incidence rates calculated at four year intervals (to coincide with one year of dog's age, considered to be equivalent to four human years after a dog reaches two years of age). Schneider found that once these age conversions were made, the frequency distribution of female human and canine breast cancer cases by age at diagnosis was similar throughout their lifespans. Median age of the female

cases was 58 years in humans and 10.5 years in bitches. When dog incidence rates were age-adjusted to the human female population standard, the rate in bitches was a bit more than 3 times the rate in women. The magnitude of increase in the age-adjusted rates was equivalent for both species until the age of natural menopause in women, after which the rate of increase in humans subsided, while the canine rate continued to increase at approximately the same exponential rate. With respect to male breast tumors, the frequency distribution of cases by age at diagnosis was also similar across species, with median age among the men of 62 years and 11.5 years in dogs. When age-adjusted to the human population, the rate of breast cancer in male dogs was slightly less than 16 times higher than the rate in men. In both sexes of both species, adenocarcinomas were the predominant histologic type, comprising about 75% of the cases.

Schneider goes on to discuss at length the similarities between breast cancer rates in the two species, with emphasis on the role of endocrine factors in the etiology of the diseases. He notes particularly the sparing effect on risk afforded by ovariohysterectomy in dogs and artificial menopause in women, as well as the hypothesis of a "critical cutoff age", after which the spaying of bitches no longer reduces risk, since carcinogenesis has already been initiated during the rapid breast growth that occurs between puberty and maturity. Based on these data, that critical cutoff age was considered to be about 2.5 years in the bitch, equivalent to about 26 years in women. Interestingly, he goes on to note research in women[171] which suggested that pregnancies after the age of 25 had

no protective effect, while first pregnancies prior to age 20 halved a woman's risk of breast cancer. Schneider considered this supportive evidence for the existence of a critical cutoff age in women, although the author was careful to note that "whether menopausal changes are directly related to breast cancer risk in women...remains to be seen".

A unique brief report expounding the value of "medical elementology" has been published by Dobrowolski and Smyk.[172] The authors describe this approach as "consisting of a multidisciplinary investigation of the role of selected elements of the natural environment in health hazard, including 'cancer clusters'". Various environmental samples (soil, water, plants, food, cattle, milk and blood) were collected from areas where clusters of unspecified human cancers and "cattle leukemia" (presumably, bovine leukosis) were apparently documented previously, as well as "in control areas". Fungal cultures and trace element analyses were then performed. Increased concentrations of lead, mercury, nickel, rubidium, potassium, manganese, copper and zinc, and decreased amounts of magnesium, calcium, iron, cobalt and selenium were reported "in the soil and food chain in 'cancer clusters'. Three mold species *Aspergillus flavus*, *Penicillium meleagrinum* and *Cladosporium herbarium* were found statistically more frequently in the homes and farm buildings of cancer patients and leukemic cattle than in control sites. The authors implicate selenium deficiency and the presence of aflatoxins as etiologic agents responsible for neoplasms in both humans and cattle. While an interesting approach, the paucity of methodological details and failure to account in any way

for confounding factors suggests the conclusions of this study be viewed with prudent skepticism.

Finally, in the interest of completeness, it should be noted that a number of studies[173-183] have used case-control methods to investigate cancer in pet populations, generally dogs, and subsequently speculated on the implications of the results for the health of the dogs' human owners. These studies, and the potential role of pet epidemiological studies in assessing environmental cancer risks for humans, have been the subject of a recent review.[184] However, since these studies did not assess geographic aggregation of cancer or simultaneous disease occurrence in both animals and humans, they will not be discussed further here. This is not to say, by any means, that they are irrelevant though, since risk factors identified by them might figure into the design of future studies of cancer aggregation in either, or both, species. In addition, the case-control design offers a powerful means of investigating causality, as a follow-up to studies of clustering that describe case distributions and generate causal hypotheses.

An Approach For the Present Study

In reflecting on the bodies of work reviewed here concerning the study of disease aggregation and the simultaneous study of cancer in humans and companion animals, two points seem to stand out. First, while a multitude of diverse methods exist for assessing disease clustering, there is as yet no test generally recognized as definitive. While the merits and weaknesses of these

various methods have been and will continue to be debated at length, it seems clear that until such a time as a definitive approach is identified, the choice of an approach to the study of disease clustering will continue to be driven by the integral, if mundane, considerations of data and resource availability. This may be particularly true with respect to the study of companion animal diseases. Much attention has already been paid in this paper to the limitations placed on investigations of companion animal diseases by the inability to accurately enumerate populations-at-risk. This precludes the calculation of disease rates, and eliminates clustering methods which use them from consideration for use in the present study. The potential approaches are further constrained by the related fact that expected numbers of cancers in companion animals cannot be defined with accuracy. Even if one attempts to extrapolate from case frequencies found in other studies, the unknown size and characteristics of the underlying dog population in Michigan would render such expected numbers essentially meaningless for describing the force of disease in the population. Thus, the choice of a methodological approach for the present study is already substantially limited.

Second, though thirty years have lapsed since the groundbreaking work of Dorn, Schneider, Franti, Klauber and their colleagues, there has been essentially no new or substantial research which has studied the distribution of biologically similar neoplasms in companion animals and humans living in the same defined geographic area during the same time period. This fact speaks not only to formidable nature of conducting such research, but also to the

scientific vision of those innovators and the considerable significance of their contribution. Notwithstanding its obvious scientific merit, it appears highly questionable whether the approach of comparing disease rates between geographically co-located human and animal cancer registries will ever be applied on anything but an infrequent and isolated basis. When considered in the context of the monumental and rapid pace of innovation that has taken place in other aspects of medical and epidemiological research during the same time period, it seems clear that there is a fundamental need for innovative methodological approaches that can be more generally applied to explore the potential epidemiological linkages between cancer in animals and humans.

It is in this context that a simultaneous investigation of the spatial and temporal distribution of selected cancers in dogs and humans in Michigan from 1964 to 1994 is undertaken. The intention to simultaneously assess the two populations further limits the possible methodological approaches to clustering that can be employed here. Only Klauber[39,40] has attempted such a study using rate-independent methods. However, Klauber's technique was intended to assess space/time interaction for diseases where an infectious etiology was considered likely. It, like other extensions of Knox's[29] test, may have insufficient power to detect clustering of neoplasms which often have long and variable latencies.[53] Moreover, Klauber's methods are insensitive to clustering only in space or only in time. Finally, Klauber's tests, like Knox's, require the *a priori* designation of critical space and time distances, or, like Mantel's,[37] require *a priori* designation of time and distance constants which define what should or

should not be considered a cluster. In the present study, knowledge of the etiology and pathogenesis of canine cancers is sufficiently limited so as to render any such choice of critical distances or constants little more than a guess. While the possibility exists that other cluster assessment methods might be adapted to simultaneous study of disease in two or more species, such adaptations have not as yet been proposed, with the notable exception of second-order (*k* function) analysis. That technique allows such an assessment.

Finally, the quantitative sophistication of the various clustering methodologies nearly necessitates the availability of computer software in order for them to be implemented by anyone other than a statistician. While a number of software implementations of techniques discussed in this paper have recently become available,[75,185-188] at the time this study was initiated (1992), only software to implement *K* function analysis was available (indeed, a glance at the dates of publication of the clustering papers reviewed here reveals that a substantial number have come into being only since 1992). Thus, *K* function analysis[1, pp. 56-68] will be used in this study to assess spatial clustering. To assess time clustering, the one dimensional nearest-neighbor technique of Selkirk and Neave[54,58] will be used due to its ease of calculation even without the aid of computer software, and because it can be applied without the designation of the critical time windows that are required by the Scan method.[44-47] As with the setting of critical spatial distances, the current knowledge of companion animal cancer etiology and pathogenesis is insufficient to provide for a rigorously defensible choice of critical time windows that should be used to define what is

and what is not considered a cluster. For comparison to humans, the present study will limit its scope to the study of tumors in dogs, in part because of the availability of sufficient numbers of cases for study and to maximize statistical power, and in part because of the author's affection for the species.

A number of significant contributions are anticipated from this study. Chief among these is the simultaneous description of the distributions of canine and human neoplasms in common geographic and time periods, utilizing a large data base, a task which to the authors' knowledge has not yet been reported. First, if concurrence of patterns of a particular tumor in the two species is discovered, and it is known *a priori* that the disease is similar in dogs and humans, it may be possible to hypothesize risk factors for the disease that are common to both species. While the testing of such hypotheses is beyond the scope of this study, they can be investigated by others in future studies. Second, the study will investigate the utility of a new methodology, *K* function analysis, for investigations of disease aggregation in veterinary epidemiology. Third, the study will be a first step in evaluating the use of dogs as sentinels[184,189,190] for some specific human cancers. If cases of a particular type of tumor in both man and dog are determined to be spatially dependent, future occurrences of a cluster of that type of tumor in dogs may potentially act as an early warning of human cases to follow. Such information could be invaluable for investigating environmental risk factors, allocating health care resources, and possibly intervening to prevent exposure of other at-risk individuals, both animal and human. Fourth, if it is possible to identify particular types of tumors that occur in

similar space and time circumstances in both humans and dogs, dogs may then be prospective animal models for future laboratory and clinical research investigating that particular neoplastic disease in humans. An animal model identified in such a manner would have an advantage over strictly laboratory species, in that it would be known to develop the disease of interest under real-world exposure conditions, thus avoiding some of the difficulties encountered when extrapolating results to genetically heterogeneous humans from inbred laboratory animal strains. Moreover, while the identification of the dog as an animal model for human cancer could benefit the health of humans, it would most assuredly benefit the health and welfare of dogs as well. While much is obviously to be gained by research which investigates disease in animals exclusively, research investigating the health of animals in relation to that of humans can not only elucidate the pathogenesis and prevention of disease in animals, but also focus the attention and comprehensive resources of the human medical research community on companion animals, and emphasize the importance of an interdisciplinary approach to animal health research. Finally, a less tangible, though no less significant, contribution which may arise from this research relates to the human-animal bond. The complexities of the relationships between humans and companion animals are only beginning to be understood. Associations between the occurrence of neoplastic disease between dogs and humans, if found, could potentially focus attention on the biological links between the health and well being of companion animals and humans, as well as the emotional and psychological links. Such recognition

could serve to broaden the traditional definition of the human-animal bond, and increase the credibility of the view that human and animal health are necessarily interdependent components within the overall context of environmental health.

REFERENCES

REFERENCES

1. Boots BN, Getis A. Point Pattern Analysis. Newbury Park (CA):Sage Publications; 1988:93p.

2. Cressie NAC. Statistics for spatial data. New York: John Wiley and Sons; 1993:900p.

3. Ripley BD. Spatial statistics. New York: John Wiley and Sons; 1981:252p.

4. Ripley BD. Statistical inference for spatial processes. Cambridge: Cambridge University Press; 1988:148p.

5. Diggle PJ. Statistical analysis of spatial point patterns. London: Academic Press; 1983:148p.

6. Cliff AD, Ord JD. Spatial processes: models and applications. London: Pion; 1981:266p.

7. Haining, RP. Spatial data analysis in the social and environmental sciences. Cambridge: Cambridge University Press; 1990:409p.

8. Renshaw, E. Modelling biological populations in space and time. Cambridge: Cambridge University Press; 1993:403p.

9. Cliff AD, Haggett P. Atlas of disease distributions: analytic approaches to epidemiological data. Oxford: Basil Blackwell; 1988:300p.

10. Elliott P, Cuzick J, English D, Stern R, editors. Geographical and environmental epidemiology: methods for small-area studies. Oxford: World Health Organization Regional Office for Europe by Oxford University Press; 1992:382 p.

11. Elliott P. Methodology of enquiries into disease clustering. London: Small Area Health Statistics Unit; 1989.

12. Alexander FE, Boyle P. Methods for investigating localized clustering of disease. Lyon: International Agency for Research on Cancer; 1996:247p (IARC scientific publication series; no 135).

13. Marshall RJ. A review of methods for the statistical analysis of spatial patterns of disease. J R Stat Soc A 1991;154:421-41.

14. Hills M, Alexander F. Statistical methods used in assessing the risk of disease near a source of possible environmental pollution: a review. J R Stat Soc A 1989;152:353-63.

15. Alexander F, Cuzick J. Methods for the assessment of disease clusters. In: Elliott P, Cuzick J, English D, Stern R, editors. Geographical and environmental epidemiology: methods for small area studies. Oxford: World Health Organization Regional Office for Europe by Oxford University Press; 1992:221-30.

16. Wartenberg D, Greenberg M. Solving the cluster puzzle: clues to follow and pitfalls to avoid. Stat Med 1993;12(19-20):1763-70.

17. Elliott P, Martuzzi M, Shaddick G. Spatial statistical methods in environmental epidemiology: a critique. Stat Methods Med Res 1995;4:137-59.

18. Waller LA, Jacquez GM. Disease models implicit in statistical tests of disease clustering. Epidemiology 1995;6(6):584-90.

19. Waller LA, Turnbull BW, Gustafsson G, Hjalmars U, Andersson B. Detection and assessment of clusters of disease: an application to nuclear power plant facilities and childhood leukaemia in Sweden. Stat Med 1995;14:3-16.

20. Besag J, Newell J. The detection of clusters in rare diseases. J R Stat Soc A 1991;154:143-55.

21. Waller LA. Statistical power and design of focused clustering studies. Stat Med 1996;15:765-82.

22. Kulldorff M, Feuer EJ, Miller BA, Freedman LS. Breast cancer clusters in the northeast United States: a geographic analysis. Am J Epidemiol 1997;146(2):161-70.

23. Olsen SF, Martuzzi M, Elliott P. Cluster analysis and disease mapping—why, when, and how? A step by step guide. BMJ 1996;313:863-6.

24. Rothman KJ. A sobering start to the cluster busters' conference. Am J Epidemiol 1990;132 Suppl 1:S6-S13.

25. Lawson AB. On the analysis of mortality events associated with a

prespecified fixed point. J R Stat Soc A 1993;156:363-77.

26. Waller LA, Turnbull BW, Clark LC, Nasca P. Spatial pattern analyses to detect rare disease clusters. In: Lange N, editor. Case studies in biometry. New York: John Wiley and Sons; 1994:3-23.

27. Pearson K. On the appearance of multiple cases of disease in the same house. Biometrika 1912;8:404-35.

28. Williams GW. Time-space clustering of disease. In: Cornell RG, editor. Statistical methods for cancer studies. New York: Marcel Dekker; 1984. p. 167-227.

29. Knox G. The detection of space-time interactions. Appl Stat 1964;13:25-9.

30. Knox G. Epidemiology of childhood leukemia in Northumberland and Durham. Br J Prev Soc Med 1964;18:17-24.

31. Pinkel K, Dawd JE, Bross IDJ. Some epidemiological features of malignant solid tumors of children in the Buffalo, NY area. Cancer 1963;16:28-33.

32. David FN, Barton DE. Two space-time interaction tests for epidemicity. Br J Prev Soc Med 1966;20:44-8.

33. Barton DE, David FN, Merrington M. A criterion for testing contagion in time and space. Ann Hum Genet 1965;29:97-102.

34. Pike MC, Smith PG. Disease clustering: a generalization of Knox's approach to the detection of space-time interactions. Biometrics 1968;24:541-56.

35. Barton DE, David FN. The random intersection of two graphs. In: David FN, editor. Research papers in statistics. New York: Wiley; 1966. p. 455-9.

36. Pike MC, Smith PG. A case-control approach to examine diseases for evidence of contagion, including diseases with long latent periods. Biometrics 1974;30:263-79.

37. Mantel N. The detection of disease clustering and a generalized regression approach. Cancer Res 1967;27:209-20.

38. Glass AG, Mantel N. Lack of space-time clustering of childhood leukemia,

Los Angeles County, 1960-64. Cancer Res 1969;29:1995.

39. Klauber MR. Two-sample randomization tests for space-time clustering. Biometrics 1971;27:129-42.

40. Klauber MR. Space-time clustering tests for more than two samples. Biometrics 1975;314:719-26.

41. Ederer F, Meyers MH, Mantel N. A statistical problem in space and time: do leukemia cases come in clusters? Biometrics 1964;20:626-38.

42. Feller W. An introduction to probability theory and its application, vol I. 2nd ed. New York: Wiley; 1957.

43. Mantel N, Kryscio RJ, Myers MH. Tables and formulas for extended use of the Ederer-Myers-Mantel disease-clustering procedure. Am J Epidemiol 1976;104:576-84.

44. Naus JI. The distribution of the size of the maximum cluster of points on a line. J Am Stat Assoc 1965;60:532-8.

45. Naus JI. Approximations for distributions of scan statistics. J Am Stat Assoc 1982;77:177-83.

46. Wallenstein S. A test for detection of clustering over time. Am J Epidemiol 1980;111:367-71.

47. Wallenstein S, Neff N. An approximation for the distribution of the SCAN statistic. Stat Med 1987;6:197-207.

48. Aldrich TE, Drane JW. Cluster 3.1: Instruction manual, system for epidemiologic analysis. Atlanta (GA): Agency for Toxic Substances and Disease Registry (US); 1993:30; USDHHS publication PB93-155083INC.

49. Paré J, Carpenter TE, Thurmond MC. Analysis of spatial and temporal clustering of horses with *Salmonella krefeld* in a intensive care unit of a veterinary hospital. J Am Vet Med Assoc 1996;209(3):626-8.

50. Tango T. The detection of disease clustering in time. Biometrics 1984;40:15-26.

51. Tango T. Asymptotic distribution of an index for disease clustering. Biometrics 1990;46(2):351-7.

52. Tango T. An index for cancer clustering. Environ Health Perspect 1990;87:157-62.

53. Whittemore AS, Friend N, Brown BW Jr, Holly EA. A test to detect clusters of disease. Biometrika 1987;74(3):631-5.

54. Selkirk KE, Neave HR. Nearest-neighbour analysis of one-dimensional distributions of points. Tijdschr Econ Soc Geogr 1984;75(5):356-62.

55. Clark PJ, Evans EC. Distance to nearest neighbour as a measure of spatial relationships in populations. Ecology 1954;35(4):445-53.

56. Clark PJ, Evans EC. On some aspects of spatial patterns in biological populations. Science 1955;121:397-8.

57. Clark PJ, Evans EC. Generalisation of a nearest neighbour measure of dispersion for use in K dimensions. Ecology 1979;60(2):316-7.

58. Neave HR, Selkirk KE. Nearest neighbour analysis of the distribution of points on a line. Nottingham (UK): Nottingham University Department of Mathematics; 1983; Nottingham Statistics Group Research Report no. 08-83.

59. Mathen KK, Chakraborty PN. A statistical study on multiple cases of disease in households. Sankhya 1950;10:387-92.

60. Walter SD. On the detection of household aggregation of disease. Biometrics 1974;30:525-38.

61. Smith PG, Pike MC. Generalizations of two tests for the detection of household aggregation of disease. Biometrics 1976;32:817-28.

62. Kitron U, Kazmierczak JJ. Spatial analysis of the distribution of Lyme Disease in Wisconsin. Am J Epidemiol 1997;145(6):558-66.

63. Moran PAP. The interpretation of statistical maps. J R Stat Soc B 1948;10:243-51.

64. Moran PAP. Notes on continuous stochastic phenomena. Biometrika 1950;37:17-23.

65. Geary R. The contiguity ratio and statistical mapping. Incorporated Statistician 1954;5:115-45.

66. Upton GJG, Fingleton B. Spatial data analysis by example: Vol. 1: point pattern and quantitative data. New York: John Wiley and Sons; 1985:394 p.

67. Kemp I, Boyle P, Smans M, Muir CS. Atlas of cancer in Scotland: 1975-1980, incidence and epidemiological perspective. Lyon: International Agency for Research on Cancer; 1985. IARC scientific publication no. 72.

68. Smans, M. Analysis of spatial aggregation. Recent Results Cancer Res 1989;114:83-6.

69. Getis A, Ord JK. The analysis of spatial association by use of distance statistics. Geogr Anaylsis 1992;24(3):189-206.

70. Ord JK, Getis A. Distributional issues concerning distance statistics. Proceedings of the annual meeting of the North American Regional Science Association International. Houston (TX): North American Regional Science Association International; 1993.

71. Walter SD. The analysis of regional patterns in health data. I. Distributional considerations. Am J Epidemiol 1992;136(6):730-41.

72. Walter SD. The analysis of regional patterns in health data. II.The power to detect environmental effects. Am J Epidemiol 1992;136(6):742-59.

73. Walter SD. Visual and statistical assessment of spatial clustering in mapped data. Stat Med 1993;12:1275-91.

74. Greig-Smith P. Quantitative plant ecology. 2nd ed. Washington (DC): Butterworths; 1964:256 p.

75. Skelton AG. Quadrat analysis software for the detection of spatial or temporal clustering. Stat Med 1996;15:939-41.

76. Openshaw S, Charlton M, Wymer C, Craft AW. A mark 1 analysis machine for the automated analysis of point data sets. Int J Geogr Info Sys 1987;1:335-58.

77. Openshaw S, Craft AW, Charlton M, Birch JM. Investigation of leukemia clusters by use of a geographical analysis machine. Lancet 1988;1:272-3.

78. Turnbull BW, Iwano EJ, Burnett WS, Howe HL, Clark LC. Monitoring for clusters of disease: application to leukemia incidence in upstate New York. Am J Epidemiol 1990;132 Suppl 1:S136-S143.

79. Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. Stat Med 1995;14:799-810.

80. Hjalmars U, Kulldorff M, Gustafsson G, Nagarwalla N. Childhood leukaemia in Sweden: using GIS and a spatial scan statistic for cluster detection. Stat Med 1996;15:707-15.

81. Kulldorff M. A spatial scan statistic. Communicat Stat Theory Methods 1997;26:1481-96.

82. Bithell JF, Stone RA. On statistical methods for analyzing the geographical distribution of cancer cases near nuclear installations. J Epidemiol Comm Health 1989;43:79-85.

83. Stone RA. Investigations of excess environmental risks around putative sources: statistical problems and a proposed test. Stat Med 1988;7:649-60.

84. Waller LA, Turnbull BW, Clark LC, Nasca P. Chronic disease surveillance and testing of clustering of disease and exposure: application to leukemia incidence and TCE-contaminated dumpsites in upstate New York. Environmetrics 1992;3:281-300.

85. Waller LA, Turnbull BW. The effect of scale on tests of disease clustering. Stat Med 1993;12:1869-84.

86. Waller LA, Lawson AB. The power of focused tests to detect disease clustering. Stat Med 1995;14:2291-308.

87. Diggle P. A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point. J R Stat Soc A 1990;153:340-62.

88. Diggle P, Rowlingson B. A conditional approach to point process modelling of elevated risk. J R Stat Soc A 1994;157:433-40.

89. Diggle P. A kernel method for smoothing point process data. Appl Statist 1985;34:138-47.

90. Silverman BW. Density estimation for statistics and data analysis. London: Chapman and Hall; 1986.

91. Berman M, Diggle PJ. Estimating weighted integrals of the second-order

intensity of a spatial point process. J R Stat Soc B 1989;51:81-92.

92. Biggeri A, Barbone F, Lagazio C, Bovenzi M, Stanta G. Air pollution and lung cancer in Trieste, Italy: spatial analysis of risk as a function of distance from sources. Environ Health Perspect 1996;104(7):750-4.

93. Lewis MS. Spatial clustering in childhood leukaemia. J Chronic Dis 1980;33:703-12.

94. Ross A, Davis S. Point pattern analysis of the spatial proximity of residences prior to diagnosis of persons with Hodgkin's Disease. Am J Epidemiol 1990;132 Suppl 1:S53-S62.

95. Cuzick J, Edwards R. Spatial clustering for inhomogeneous populations (with discussion). J R Stat Soc B 1990;73-104.

96. Jacquez GM. Cuzick and Edward's test when exact locations are unknown. Am J Epidemiol 1994;140:58-64.

97. Jacquez GM. Disease cluster statistics for imprecise space-time locations. Stat Med 1996;15:873-85.

98. Jacquez GM. A *k* nearest neighbour test for space-time interaction. Stat Med 1996;15:1935-49.

99. Mantel N, Bailar JC. A class of permutational and multinomial tests arising in epidemiological research. Biometrics 1970; 26:687-700.

100. Glaser SL. Spatial clustering of Hodgkin's Disease in the San Francisco Bay Area. Am J Epidemiol 1990;132 Suppl 1:S167-S177.

101. Ranta J, Pitkäniemi J, Karvonen M, Virtala E, Rusanen J, Colpaert A, et al. Detection of overall space-time clustering in a non-uniformly distributed population. Stat Med 1996;15:2561-72.

102. Getis A, Franklin J. Second-order neighborhood analysis of mapped point patterns. Ecology 1987;68(3):473-7.

103. Ripley BD. Spectral analysis and the analysis of pattern in plant communities. J Ecol 1978;66:965-81.

104. Ripley BD. Tests of 'randomness' for spatial point patterns. J R Stat Soc B 1979;41:368-74.

105. Ripley BD. Modelling spatial patterns (with discussion). J R Stat Soc B 1977;39:172-212.

106. Ripley BD. The analysis of geographical maps. In: Bartels CPA, Ketellapper RH, editors. Exploratory and explanatory statistical analysis of spatial data. Boston: Martinus Nijhoff; 1979:53-72.

107. Diggle P, Chetwynd A. Second-order analysis of spatial clustering for inhomogeneous populations. Biometrics 1991;47:1155-63.

108. Selvin S, Merrill D, Schulman J, Sacks S, Bedell L, Wong L. Transformations of maps to investigate clusters of disease. Soc Sci Med 1988;26(2):215-21.

109. Tobler WR. Geographic area and map projections. Geogr Rev 1973;53(1):59-78.

110. Tobler WR. Geographic area and map projections. Ann N Y Acad Sci 1973;219:215-20.

111. Schulman S, Selvin S, Merrill DW. Density equalized map projections: a method for analysing clustering around a fixed point. Stat Med 1988;7(4):491-505.

112. Selvin S, Merrill DW, Sacks S, Wong L, Bedell L, Schulman J. Transformation of maps to investigate clusters of disease. Berkeley (CA): Lawrence Berkeley Laboratory; 1984. Report no. LBL-18550.

113. Howe MG. Statistical and cartographic methods of analysing spatial patterns of mortality. Geoforum 1979;10:311-22.

114. Howe MG. Mortality from selected malignant neoplasms in the British Isles: the spatial perspective. Soc Sci Med 1981;15D:199-211.

115. Grimson RC, Wang KD, Johnson PWC. Searching for hierarchical clusters of disease: spatial patterns of sudden infant death syndrome. Soc Sci Med 1981;15:287-93.

116. Grimson RC, Rose RD. A versatile test for clustering and proximity analysis of neurons. Meth Inform Med 1991;30:299-303.

117. Ohno Y, Aoki K, Aoki N. A test of significance for geographic clusters of disease. Int J Epidemiol 1979;8:273-81.

118. Chen R, Mantel N, Connelly RR, Isaacson P. A monitoring system for chronic diseases. Meth Inform Med 1982;21:86-90.

119. Chen R. Statistical techniques in birth defects surveillance systems. Contr Epidemiol Biostat 1979;1:184-89.

120. Aldrich TE. Detecting space-time aggregationof rare events [abstract]. Am J Epidemiol 1984;120:464.

121. Aldrich TE, Wilson CC, Easterly CE. Population surveillance for rare health events. In: Proceedings of the 1985 Public Health Conference on Records and Statistics. Hyattsville (MD) : National Center for Health Statistics; 1986. DHHS publication no. (PHS) 86-1214, 525 p.

122. Hardy RJ, Schroeder GD, Cooper SP, Buffler PA, Prichard HM, Crane M. A surveillance system for assessing health effects from hazardous exposures. Am J Epidemiol 1990;132(1):532-42.

123. Weatherall JAC, Haskey JC. Surveillance of malformations. Brit Med Bull 1976;32:39-44.

124. Grimson RC. The clustering of disease. Math Biosci 1979;46:257-78.

125. Edmonds LD, Layde PM, James LM, Flynt JW, Erickson JD, Oakley GP. Congenital malformations surveillance: two American systems. Int J Epidemiol 1981;247-52.

126. Chen R, McDowell M, Terzian E, Weatherall J. Eurocat guide to monitoring methods for malformation registers. Bruxelles: European Economic Community Concerted Action Project, 1983.

127. Chen R. Revised values for the parameters of the sets technique for monitoring the incidence rate of a rare disease. Meth Inform Med 1986;25:47-9.

128. Hill GB, Spicer CC, Weatherall JAC. The computer surveillance of congenital malformations. Brit Med Bull 1968;24:215-18.

129. Monsein DL. An overview of animal cancer registries in the United States and suggestions for improved applications. Compend Contin Educ Pract Vet 1991;13(7):1139-46.

130. Priester WA, McKay FW. The occurrence of tumors in domestic animals. Ziegler JL, editor. Bethesda (MD): National Cancer Institute (US);

1980:210 p. (National Cancer Institute Monographs; vol 54).

131. Priester WA, Mantel N. Occurrence of tumors in domestic animals. Data from 12 United States and Canadian colleges of veterinary medicine. J Natl Cancer Inst 1971;47:1333-44.

132. Priester WA, ed. Standard nomenclature of veterinary diseases and operations. 2nd (abridged) ed. Bethesda (MD): National Cancer Institute (US); 1975. DHEW publication no. (NIH)76-1028.

133. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959;22:719-48.

134. Dorn CR, Taylor ON, Frye FL, Hibbard HH. Survey of animal neoplasms in Alameda and Contra Costa Counties, California. I. Methodology and description of cases. J Natl Cancer Inst 1968;40:295-305.

135. Dorn CR, Taylor ON, Schneider R, Hibbard HH, Klauber MR. Survey of animal neoplasms in Alameda and Contra Costa Counties, California. II. Cancer morbidity in dogs and cats from Alameda County. J Natl Cancer Inst 1968;40:307-18.

136. Dorn CR, Chaulk LE. A survey of practice characteristics and animal admissions in two California counties. J Am Vet Med Assoc 1964;145:358-60.

137. World Health Organization. Manual of the international statistical classification of diseases, injuries and causes of death, 1955 revision. Geneva: World Health Organization, 1957.

138. Human Population Laboratory: Alameda County population, 1965. Berkeley (CA): California State Department of Public Health; 1966. Series A, no 7.

139. Taylor, DON. California's four-phase cancer field research program. J Am Vet Med Assoc 1965;147:1492-97.

140. Dorn CR. An animal tumor registry as a source of morbidity information. Proc US Livestock Sanit Assoc 1967;443-54.

141. Dorn CR, Taylor DON, Chaulk LE, Hibbard HH. The prevalence of spontaneous neoplasms in a defined canine population. Am J Pub Health 1966;56:254-65.

142. MacVean DW, Monlux AW, Anderson PS Jr, Silberg SL, Roszel JF. Frequency of canine and feline tumors in a defined population. Vet Pathol 1978;15(6):700-15.

143. Monlux AW, Roszel JF, MacVean DW, Palmer TE. Classification of epithelial canine mammary tumors in a defined population. Vet Pathol 1977;14:194-217.

144. Robinson GW. A comparison of licensed and hospital dog populations. J Am Vet Med Assoc 1968;152(9):1383-89.

145. Masi, AT. Potential uses and limitations of hospital data in epidemiologic research. Am J Pub Health 1965;55:658-67.

146. Wagstaff DJ, Goyings LS, Langham RF. Canine cancer distribution as related to data source. Am J Vet Res 1967;28:1479-82.

147. Franti CE, Kraus JF, Borhani NO. Pet ownership in a suburban-rural area of California, 1970. Public Health Rep 1974;89(5):473-84.

148. Franti CE, Kraus JF. Aspects of pet ownership in Yolo County, California. J Am Vet Med Assoc 1974;164:166-71.

149. Franti CE, Kraus JF, Borhani NO, Johnson SL, Tucker SD. Pet ownership in rural Northern California (El Dorado County). J Am Vet Med Assoc 1980;176(2):143-9.

150. Nassar R, Mossier JE. Canine population dynamics: a study of the Manhattan, Kansas, canine population. Am J Vet Res 1980;41:1798-803.

151. Nassar R, Mossier JE. Feline population dynamics: a study of the Manhattan, Kansas, feline population. Am J Vet Res 1982;43:167-70.

152. Nassar R, Mossier JE. Understanding the dynamics of your community's pet population. Vet Med 1986;81:1120-6.

153. Patronek GJ, Glickman LT. Development of a model for estimating the size and dynamics of the pet dog population. Anthrozoos 1994;7:25-42.

154. Heussner JC, Flowers AI, Williams JD, Silvy NJ. Estimating dog and cat populations in an urban area. Anim Regulation Studies 1978;1:203-12.

155. Anvik JO, Hague AE, Rahaman A. A method of estimating urban dog populations and its application to the assessment of canine fecal pollution

and endoparasitism in Saskatchewan. Can Vet J 1974;15:219-23.

156. Heussner JC, Grant WE. Ecological aspects of urban dog management: a simulation model. Anim Regulation Studies 1978;1:355-74.

157. Rangel MCF, Lara JC, Aluja AS. The canine population of Mexico City: an estimation study. Anim Regulation Studies 1980;3:281-90.

158. Leslie BE, Meek AH, Kawash GF, McKeown DB. An epidemiological investigation of pet ownership in Ontario. Can Vet J 1994;35(4):218-22.

159. Griffiths AO, Brenner A. Survey of cat and dog ownership in Champaign County, Illinois, 1976. J Am Vet Med Assoc 1977;170(11):1333-40.

160. Nassar R, Mosier JE, Williams LW. Study of the feline and canine populations in the greater Las Vegas area. Am J Vet Res 1984;45:282-7.

161. Lengerich EJ, Teclaw RF, Mendlein JM, Mariolis P, Garbe PL. Pet populations in the catchment area of the Purdue Comparative Oncology Program. J Am Vet Med Assoc 1992;200(1):51-6.

162. Teclaw R, Mendlein J, Garbe P, Mariolis P. Characteristics of pet populations and households in the Purdue Comparative Oncology Program catchment area. J Am Vet Med Assoc 1992;201:1725-9.

163. White ME, Schukken YH, Tanksley B. Space-time clustering of, and risk factors for, farmer-diagnosed winter dysentery in dairy cattle. Can Vet J 1989;30:948-51.

164. Carpenter TE, Snipes KP, Hird DW. A spatial epidemiologic investigation of fowl cholera. Kenya Vet 1994;18:487-9.

165. Ward M, Carpenter TE, Johnson SJ. Spatial analysis of seroconversion of sentinel cattle to bluetongue viruses in Queensland. Aust Vet J 1996;74(2):128-31.

166. Singer RS, Case JT, Carpenter TE, Walker RL, Hirsh DC. Assessment of spatial and temporal clustering of ampicillin-resistant and tetracycline-resistant strains of Pasteurella multocida and P. haemolytica isolated from cattle in California. J Am Vet Med Assoc 1998;212(7):1001-5.

167. Tjalma RA. Implications of animal cancers to human neoplasia: epidemiologic considerations. Int J Cancer 1968;3:1-6.

168. Schneider R, Dorn CR, Klauber MR. Cancer in households: a human-canine retrospective study. J Natl Cancer Inst 1968;41(6):1285-92.

169. Schneider R. Comparison of age, sex, and incidence rates in human and canine breast cancer. Cancer 1970;26(2):419-26.

170. Lebeau A. L'age du chien et celui de l'homme, essai de statistique sur la mortalite canine. Bull Acad Vet France 1953;26:229-32.

171. Cole P, MacMahon B. Oestrogen fractions during early reproductive life in the etiology of breast cancer. Lancet 1969;1:604-6.

172. Dobrowolski JW, Smyk B. Environmental risk factors of cancer and their primary prevention. J Environ Pathol Toxicol Oncol 1993;12(1):55-7.

173. Ragland WL, Gorham JR. Tonsilar carcinoma in rural dogs. Nature 1967;214:925-6.

174. Reif JS, Cohen D. Retrospective radiographic analysis of pulmonary disease in rural and urban dogs. Arch Environ Health 1970;20:684-9.

175. Reif JS, Cohen D. The environmental distribution of canine respiratory tract neoplasms. Arch Environ Health 1971;22:136-40.

176. Hayes HM Jr, Hoover R, Tarone RE. Bladder cancer in pet dogs: a sentinel for environmental cancer? Am J Epidemiol 1981;114(2):229-33.

177. Harbison ML, Godleski JJ. Malignant mesothelioma in urban dogs. Vet Pathol 1983;20:531-40.

178. Glickman LT, Domanski LM, Maguire TG, Dubielzig RR, Churg A. Mesothelioma in pet dogs associated with exposure of their owners to asbestos. Environ Res 1983;32:305-13.

179. Glickman LT, Shofer FS, McKee LJ, Reif JS, Goldschmidt MH. Epidemiologic study of insecticide exposures, obesity, and risk of bladder cancer in household dogs. J Toxicol Environ Health 1989;28:407-14.

180. Hayes HM, Tarone RE, Casey HW, Huxsoll DL. Excess of seminomas observed in Vietnam service U.S. military working dogs. J Natl Cancer Inst 1990;82:1042-6.

181. Hayes HM, Tarone RE, Cantor KP, Jessen CR, McCurnin DM, Richardson RC. Case-control study of canine malignant lymphoma:

positive association with dog owner's use of 2,4-dichlorophenoxyacetic acid herbicides. J Natl Cancer Inst 1991;83:1226-31.

182. Reif JS, Dunn K, Oglivie GK, Harris CK. Passive smoking and canine lung cancer risk. Am J Epidemiol 1992;135(3):234-39.

183. Reif JS, Lower KS, Oglivie GK. Residential exposure to magnetic fields and risk of canine lymphoma. Am J Epidemiol 1995;141(4):352-9.

184. Bukowski JA, Wartenberg D. An alternative approach for investigating the carcinogencity of indoor air pollution: pets as sentinels of environmental cancer risk. Environ Health Perspect 1997;105(12):1312-9.

185. Schlattmann P. The computer package Dismapwin. Stat Med 1996;15:931.

186. Hall HI, Lee CV, Kaye WE. Cluster: a software system for epidemiologic cluster analysis. Stat Med 1996;15:943-50.

187. Jacquez GM. Statistical software for the clustering of health events. Stat Med 1996;951-2.

188. Kulldorff M, Rand K, Williams G. SaTScan, version 1.0, program for the space and time scan statistic [computer program]. Bethesda (MD): National Cancer Institute (US); 1996.

189. O'Brien DJ, Kaneene JB and Poppenga RH. The use of mammals as sentinels for human exposure to toxic contaminants in the environment. Environ Health Perspect 1993;99:351-68.

190. National Research Council. Animals as sentinels of environmental health hazards. Committee on Animals as Monitors of Environmental Hazards, Board on Environmental Studies and Toxicology, Commission on Life Sciences, National Research Council (US). Washington (DC): National Academy Press; 1991:160 p.

Chapter 2

FREQUENCIES AND CHARACTERISTICS OF SOME SELECTED CANINE CANCERS, MICHIGAN, 1964-1994

Abstract

The objective of this study was to report frequencies and characteristics for some selected canine cancers occurring in Michigan between 1964 and 1994. The sample population consisted of thirteen hundred forty-six dogs diagnosed with one of ten types of malignant neoplasia judged to be of relevance to similar neoplasms in humans. Case records of all histologically-confirmed cases of neoplasia diagnosed in dogs at the Michigan State University Veterinary Medical Center (MSU-VMC) between March 1, 1964 and March 31, 1994 were obtained from the Veterinary Medical Data Base. From these, cases of a subgroup of ten types of neoplasia (intestinal/rectal adenocarcinomas, lymphosarcomas, mammary adenocarcinomas, mammary carcinomas, melanomas, mixed mammary tumors, nasal adenocarcinomas, prostatic adenocarcinomas, spindle cell sarcomas and transitional cell carcinomas of the urinary bladder and urethra) were chosen for further study based on selection criteria relevant to a larger environmental epidemiological study. Frequency tables were generated for sex, breed, vital status at discharge, weight and age within each type of neoplasia, and differences in frequencies of these variables between tumor bearing dogs and comparable controls were evaluated using χ^2 and Fisher's Exact tests. Lymphosarcomas comprised the single most commonly diagnosed malignant tumor in our thirty year study period. Compared

to dogs presenting to MSU-VMC for all reasons, those dogs with neoplastic disease encompassed greater proportions of spayed females and neutered males. Dogs afflicted with cancer were older, and considerably less likely than the reference dogs to be discharged alive . Breed distribution differed little between cancer cases and the surrogate reference population when ranked by case frequency. In contrast, ranking by percent of individuals of a breed afflicted with the tumor under study often differed markedly from rankings based on frequency of cases. Readily accessed sources of disease data accumulated over decades may provide a strong foundation for formulating hypotheses regarding the etiology of neoplasia in companion animals, despite the limitations in scope imposed by the inability to enumerate true reference populations and populations-at-risk, and may eventually lead to an improved understanding of the environmental determinants of cancer in other species, including humans.

Introduction

The proliferation of computing technology during the past thirty years has had an enormous impact on a number of scientific disciplines by enabling the routine storage of large amounts of data. Moreover, access to and summarization of data have been facilitated, allowing data bases to be exploited for hypothesis generation to an extent that would not have been possible previously. Veterinary medical data have been encompassed by this progressing technology as well. Information of relevance to epidemiologists, researchers in the basic sciences and clinicians alike can be accessed with

relative ease, yet the available data base resources seem surprisingly underutilized. Thus, there is a need for studies which extract and summarize descriptive information from these data bases and present it in forums that are more widely read by both basic researchers in veterinary medicine and veterinary practitioners. Dissemination of such information to those who can interpret it in the context of their own particular areas of scientific expertise and integrate it into their own research programs is an essential step in the scientific process. The extent to which this occurs may ultimately be the best measure of the usefulness of these data base resources.

Data concerning risk factors for chronic diseases (*e.g.*, breed, age, reproductive status) can be used to formulate testable research hypotheses, while also accounting for confounders during analysis of causal factors. Summaries of large numbers of case records can confirm or call into question patterns reported in smaller case series. Evidence of geographic and temporal variation can be manifest. Information on vital status at discharge from hospital can be used to help evaluate the success of factors under the control of the clinician, such as diagnostic and therapeutic regimens, as well as prognosis for future cases. In addition, the specific ongoing need of veterinary oncologists for epidemiologic data has been pointed out,[1] and data on spontaneously occurring tumors in animals may be of particular interest for comparative research into similar disease processes in humans.[2-4]

To help disseminate information that may find application in situations such as these, we briefly report here descriptive data for some selected canine

cancers occurring in Michigan during the thirty year period between 1964 and 1994.

Materials and Methods

The present work was part of a larger exploratory study intended to simultaneously compare the distributions of some selected cancers in dogs and humans living in the same geographic region during the same time period. Canine case records were obtained from the Veterinary Medical Data Base (VMDB). The VMDB, housed at Purdue University, is a computerized system used to archive information on patients examined at the majority of veterinary schools in North America. The system was used to retrieve the records of all histologically-confirmed cases of neoplasia (n = 6,520) diagnosed in dogs at the Michigan State University Veterinary Medical Center (MSU-VMC) between March 1, 1964 and March 31, 1994.

Because of the wide variability in the recorded frequency of individual diagnoses (1-167) for different tumors, a small group of histologic types was chosen for further study. These ten histologic types, the SNVDO codes considered to comprise each tumor type (*i.e.*, the case definition) and the number of cases available for analysis during the study period are presented in Table 2.1. Criteria used for neoplasm selection included the following: 1) Medical importance in veterinary, as well as human, medicine; 2) Similarity of pathogenesis or biological mechanisms between dogs and humans; 3) Major causes of morbidity and mortality in both species; 4) Potential for the dog as an

animal model of the human disease, 5) Possibility of plausible environmental risk factors contributing to disease distribution and 6) frequency of occurrence. For each record retrieved in this final subset, data fields (coding) available for analysis included unique patient identification number, discharge status (nominal, 5 levels) and date of discharge, length of stay (continuous), sex/reproductive status (nominal, 7 levels), breed (nominal, 200 levels), age (ordinal, 10 levels), weight (ordinal, 9 levels), up to five diagnoses and up to three operations, coded according to the Standard Nomenclature of Veterinary Diseases and Operations (SNVDO)[5], diagnostic procedures used (twelve categories, each coded binary), and zip code of residence at diagnosis. SNVDO diagnosis codes take the form of an 8 digit number, with the first four digits designating topography and the second four designating etiology (for these cancer records, the first four digits denote anatomic site and the second four denote histologic type). These digits are followed by a malignancy (behavior) code character. For the selected subgroup of ten histologic types, codes were filtered so that records for benign tumors and metastatic sites were excluded. For dogs examined more than once, only the record documenting the first diagnosis of a tumor was included.

Records were selected based on primary diagnosis code only. To judge whether a significant number of cases were being missed by failing to consider diagnoses two through five, a random sample of fifty records was chosen from the full data set of 6,520 histologically confirmed neoplasia cases. Of those fifty records, the proportion recording any data value at all for diagnoses 2 through 5

were 27/50, 9/50, 3/50 and 1/50, respectively; those recording neoplasms were 12/50, 5/50, 1/50 and 0/50, respectively. Only one of the fifty records recorded a diagnosis code corresponding to one of the histologic types selected for this study, and upon examination, that record had already been captured in our study population based on primary diagnosis code. Thus, it was concluded that case selection based solely on primary diagnosis was unlikely to miss an appreciable number of cases.

Frequency tables were generated for sex, breed, vital status at discharge, weight and age using the SAS System (Proc Freq)[6]. Because the distribution of study variables in the general population of dogs was unknown, the frequencies of study variables in dogs diagnosed with neoplasia during the study period were compared with distributions of the same variables among all dogs seen at the MSU-VMC for all reasons during the same time period. This latter group of dogs were effectively considered a surrogate reference population for the purpose of this study. While not sufficient for extrapolation to the general dog population, the MSU-VMC reference population nonetheless offers some perspective into the similarities and differences between distributions of the study variables amongst tumor-bearing dogs versus the population of dogs which gave rise to them. Distributions of the five study variables in dogs diagnosed with neoplastic disease versus those in the reference population were compared statistically by simple contingency table analysis using Stata software[7]. The distributions of these variables in dogs diagnosed with each specific malignant cancer under study versus those in the reference population

were analyzed similarly.

Data concerning body weight are difficult to interpret in a purely descriptive work such as this. There are few literature references to body weight as a risk factor for the particular cancer types in this study. Moreover, it is impossible within the design limitations of this study to meaningfully assess the independent effect of weight as a risk factor, given the confounding effects of age, gender and breed data with which it is likely to be highly correlated. It is recognized here that confounding variables could potentially be accounted for analytically using multivariable methods, through stratified analysis, or by developing standards from purebred association data to judge when a cancer case of given age/breed/sex was obese. Though desirable, such an analysis was beyond the designated scope and resources available for this study. Consequently, information on body weights in this study were tabulated in the interest of completeness and internal consistency, but not treated further except where the availability of specific literature references allowed descriptive comparisons.

Results

The frequencies and percentages of cases sorted by sex/reproductive status, age at diagnosis, body weight at diagnosis, and vital status at discharge are shown in Tables 2.2 through 2.5, respectively. A ranking of breeds by percent of all cases within each tumor type and by percent of total individuals within each breed is presented in Table 2.6. Tables 2.7-2.10 compare the

distributions of animals by sex/reproductive status, age, weight and vital status at discharge, respectively, among all histologically-confirmed cases of neoplasia and among all dogs seen for all reasons during the 30 year study period. Note that Tables 2.7-2.10 do not duplicate the data from Tables 2.2-2.6, since the latter summarize cases of the ten selected histologic types, which are a subset [*n* = 1346] of the population comprised by all dogs with histologically-confirmed cancers [*n* = 6520]. Table 2.11 compares the nine most frequently reported breeds among the studied tumors with the percentage each comprises of 1) all histologically-confirmed neoplasms and 2) all dogs seen for all reasons during the 30 year study period. The percent of total breed individuals composed of cases of neoplasia is reported as well. A comparison of vital status at discharge for different age categories of the five most commonly reported breeds is displayed in Figure 2.1.

The proportionate morbidity due to each type of neoplasm (as a percentage of the total sample of 6,520 histologically-confirmed neoplasms recorded during the period) was as follows: intestinal/rectal adenocarcinomas, 1.2% (75 cases); lymphosarcomas, 6.9% (453 cases); mammary adenocarcinomas, 2.6% (172 cases); mammary carcinomas, 0.2% (10 cases); melanomas, 2.1% (136 cases); mixed mammary tumors, 0.8% (52 cases); nasal adenocarcinomas, 1.5% (97 cases); prostatic adenocarcinomas, 0.4% (26 cases); spindle cell sarcomas, 3.8% (250 cases); and transitional cell carcinomas of the urinary bladder and urethra, 1.2% (75 cases).

Discussion

Tumor-bearing dogs versus reference population

Compared to dogs presenting to MSU-VMC for all reasons, those dogs with neoplastic disease encompassed substantially smaller proportions of intact females (21.3 vs. 33.9%) and males (33.2 vs. 40.8%), and substantially greater percentages of spayed females (31.9 vs. 16.6%) and neutered males (13.5 vs. 7.1%) (Table 2.7). Not surprisingly, substantially fewer litters were reported (<0.1 vs. 1.3%). It seems likely that these differences can be accounted for at least to some extent by differences in the age distributions of the two groups (Table 2.8). Neoplasia is predominantly a disease of older animals,[8-11] and consistent with this, the age distribution of tumor-bearing animals as a whole was shifted strongly toward animals greater than four years of age. While 89% of the dogs afflicted with cancer were older than 4, only 39% of the reference population was. Differences in age distribution may well explain the apparent shift of frequencies of tumor-bearing dogs toward the heavier weight classes as well (Table 2.9), with relatively heavier adult animals predominating among the dogs with cancer, and 30% of the reference population comprised of pups.

Dogs with neoplasia were considerably less likely than the reference dogs to be discharged alive (71.8 vs. 93%), yet interestingly, nearly three-quarters of the tumor-bearing animals were discharged alive (Table 2.10). We did not analyze separately the discharge status for dogs bearing benign tumors compared to malignant ones, but the inclusion of dogs with benign disease among the overall number of dogs diagnosed with neoplasia may have inflated

this percentage somewhat. Breed distribution (Table 2.11) differed little between cases with cancer and all dogs seen at MSU-VMC during the study period. Among the nine breeds noted most frequently in our study, the proportion of all breed individuals diagnosed with neoplasia was greater than the corresponding proportion in the study population as a whole for Dobermans and mongrels, while this proportion was less than the study population for the German shepherd, American cocker spaniel, Labrador retriever and miniature poodle breeds. This is an interesting observation given the seemingly widely-held belief that mixed breed dogs are less likely to develop cancer on average, and German shepherds more so. When analyzed by case frequencies alone (Table 2.11), breed distributions were significantly different in dogs diagnosed with neoplasia compared to all individuals seen for all reasons ($\chi^2 = 80.39$, d.f. = 8, $p < 0.001$). However, when analyses were performed on our rankings based on percent of breed individuals affected (Table 2.6), no significant differences were detected between breeds within any of the studied cancers (Fisher's Exact Test; p value range: 0.28 - 1.0).

Regardless of which of the remaining study variables (sex, age, weight or vital status) was analyzed, its distribution was significantly different by χ^2 contingency table analysis in dogs with neoplasia versus its distribution in all individuals seen for all reasons.

Intestinal/Rectal Adenocarcinomas

There was little evidence of any sex predilection for this histologic type in our study. This is consistent with most previous reports,[12] although Klausner and Hardy report males as more frequently affected by intestinal neoplasms in general.[13] The disease was somewhat less common in neutered males in the present study, but this distribution agreed well with the percentage of castrates in the reference population as a whole. Patients with these tumors were diagnosed in middle to old age; only about 5% of the cases in our study were less than four years of age, and over 70% were seven years or older. This agrees well with historical findings.[12-15] A large majority (>70%) were discharged alive. Of those that were not, less than a third died of their disease. Intestinal/rectal adenocarcinomas were one of only two neoplasms studied for which a purebred (German shepherd) reported more cases than mixed breeds. This is consistent with some reports which suggest that this breed is predisposed,[13,14] but others found no such association.[12] Straw[14] reports the boxer and collie breeds may also be predisposed. While the collie was the fourth most common breed recorded in our study, the boxer was not among the top ten. The Airedale, old English sheepdog, Great Dane, and standard and miniature poodle ranked highest with respect to the proportion of all breed individuals diagnosed with this neoplasm in our study. Distributions by sex\reproductive status, age, body weight, and vital status at discharge all differed significantly in dogs with this cancer versus all individuals seen for all reasons.

Lymphosarcomas

Lymphosarcomas comprised the single most commonly diagnosed malignant tumor in our thirty year study period. There was little evidence of gender predilection, though nearly half of all cases of the combined tumors under study that were diagnosed in neutered males were lymphosarcomas. Over 60% of the cases were diagnosed in animals four to ten years of age. Our findings with regard to sex and age distribution are largely the same as those reported in the literature,[16-20] with two points worthy of elaboration. First, it has been noted by Priester and McKay[21] that intact females were less at risk for the disease. In our data, lymphosarcomas occurred more frequently in spayed females as well, and this difference is not explained by an over representation of spayed females in the reference population. Second, only a few cases as young as 4 months of age have apparently been reported,[19] yet the current data recorded two affected animals $\leq$ 2 months old. One, a 15-30 pound mixed breed female from the 2-8 week age class was discharged alive. The other, a spayed female mixed breed weighing 50 to 75 pounds (at < 2 weeks of age!) clearly was miscoded. Slightly over half of the cases were discharged alive. Less than 15% died of their disease while in hospital, but when all the tumors under study are considered, two thirds of the animals that died in hospital were lymphosarcoma cases. With respect to breed, boxers, basset hounds, St. Bernards, Scottish terriers and Airedale terriers have been reported to be at increased risk of developing lymphosarcoma,[16-18,21] while two other authors[20,22] note high risk in the cocker spaniel breed. Of these breeds, only the boxer and

St. Bernard were found among the ten most frequently recorded breeds in the current study, however. Yet another reference[23] referred to anecdotal evidence of higher incidence in the German shepherd and golden retriever, and interestingly, both of these breeds were among the five most frequently reported breeds in our study. Within breeds, the boxer was second only to the Irish wolfhound breed with respect to the percentage of total individuals by breed diagnosed with lymphosarcoma. The Kerry Blue terrier, Pembroke Welsh corgi and Chesapeake Bay retriever also ranked high by this criteria. Distributions by sex\reproductive status, age, body weight, and vital status at discharge all differed significantly in dogs diagnosed with lymphosarcomas versus all individuals seen for all reasons.

Mammary Neoplasms

The literature documents mammary tumors in general as being a disease of intact bitches, with spayed bitches at from three to seven times less risk,[24] and cases in males occurring only rarely.[25-27] By and large, the records in our study fit this pattern as well. The number of intact females diagnosed with mammary neoplasia during our study period was greater than that of spayed females for all three histologic types. Two cases in males were recorded. Historically, many of the male cases recounted in the literature have been associated with hormonal abnormalities, such as Sertoli cell tumors.[27] The first male case in our study was a 7-10 year old German shepherd discharged alive with mammary adenocarcinoma. The second, a 10-15 year old miniature poodle that was euthanized and necropsied, was afflicted with malignant mixed mammary tumor.

While this second case was also diagnosed with metastatic disease in the lung and kidney, neither animal was diagnosed with Sertoli cell tumor or other endocrine disease. The ages at diagnosis for these two dogs were consistent with the average age published in the literature (10.7 years).[26] There were very few cases in females less than four years of age, and regardless of histologic type, the majority of the cases (83% of the adenocarcinomas, 60% of the carcinomas and 79% of the mixed mammary tumors) were diagnosed between the ages of 7 and 15 years. This matches the age distributions noted in other studies quite closely.[24-28] The age distribution for carcinomas appears to peak somewhat earlier than the other to histologic types, but with so few cases (10 in 30 years), it is not appropriate to draw any definitive conclusions. With reference to body weight, the results of one case-control study[29] of conformation, diet and risk of mammary cancer in dogs found that the risk of mammary cancer among spayed dogs was significantly reduced in animals that had been thin at nine to twelve months of age. Risk was also reduced in intact bitches that were thin at the same age, but not significantly so, suggesting the possibility of an interactive effect of body weight and reproductive status on the risk of developing these cancers. Our study did not characterize the weights of the case dogs at times prior to diagnosis, and so it is not possible to address such an effect within the context of this study. Greater than 88% of animals diagnosed with adenocarcinomas and mixed mammary tumors were discharged alive. In contrast, less than half of dogs diagnosed with carcinoma were. Of those that were not, half of the dogs affected with mixed mammary tumors died

in hospital of their disease, whereas none of the dogs with adenocarcinoma did. Mammary neoplasms are reported to be more likely to occur in purebred dogs, with the sporting breeds (pointers, retrievers, setters, spaniels *et al.*) particularly predisposed.[21,24,26,27] Other breeds thought to be at high risk include the Samoyed, Great Pyrenees, toy and miniature poodle, and Boston, Airedale and fox terriers.[25] One study noted a strikingly high ratio of the puli breed among cases of mammary cancer, suggesting a predilection.[28] There seems to be general agreement that mongrels are at decreased risk,[21,24-27] although it has been suggested that this conclusion needs to be validated statistically in larger populations.[30] Consistent with their distribution among the reference animals, mixed breed dogs comprised the largest single breed group reporting cases of adenocarcinomas and mixed tumors in the present work. In contrast, none of the ten dogs diagnosed with carcinoma was a mongrel. Miniature poodles were also ranked highly by number of cases recorded in all three histologic types studied, and the over representation of sporting breeds was quite obvious in our data as well, especially among dogs diagnosed with mixed mammary tumors. The German shepherd and Doberman pinscher breeds frequently recorded cases, despite previous work suggesting only a modestly increased risk for these breeds,[21] but like the mixed, these breeds were popular in the reference population as well, arguing against their predisposition. Three of the hounds were among the top five breeds afflicted with adenocarcinomas when ranked according to percent of breed individuals diagnosed, though none of them was noted in the summary literature as being a breed at increased risk. Distributions

by sex\reproductive status, age and vital status at discharge all differed significantly in dogs diagnosed with any of the three mammary neoplasms under study versus all individuals seen for all reasons, while distributions by body weight were not significantly different from the reference population.

Melanomas

The literature describes these neoplasms as being most common in dogs older than 7 years of age[31-33] with an average age of 9 years at diagnosis.[34,35] The disease is reported to be more common in males,[31-34] although this has been disputed.[35] Darkly pigmented breeds (Scottish, Boston and Airedale terriers, cocker and springer spaniels),[31-33] as well as boxers, Irish setters and terriers, chows and Chihuahuas[34] are recorded as being predilected. Among our cases, males were only slightly more commonly affected (55%) than females, with intact males and spayed females outnumbering castrates and intact bitches. Consistent with previous reports, >90% of these cases were diagnosed in animals older than 7 years of age, but in our study, two thirds of the diagnosed cases were 10 years of age or older, placing the average age at diagnosis somewhat older than previously reported. Nearly all (93%) of our cases were discharged alive; only a single patient died while in hospital. Only three of the top thirteen breeds ranked by case frequency in our study correspond to those noted above as being predisposed. Interestingly, a number of breeds listed as being at higher risk by Priester[21] (German short haired pointer, miniature schnauzer, standard poodle, Labrador retriever, Pekingese) but not other authors[31-35] were also among those most frequently recording cases in our study.

However, by percent of breed individuals comprised by melanoma cases, the Briard, Scottish terrier, standard schnauzer, miniature dachshund and Gordon setter ranked highest. Given reports of its predisposition in the literature, consistent with the results of the current study, the Scottish terrier appears to be at genuinely increased risk of melanoma. Distributions by sex\reproductive status and age differed significantly in dogs diagnosed with malignant melanoma as compared to all individuals seen for all reasons, while distributions for body weight and vital status did not differ significantly from the reference population.

Nasal Adenocarcinomas

Males have been noted as being predisposed to the development of nasal tumors by several authors,[36-40] while another has suggested that this predilection is inconsistent.[41] In variance with the majority of these reports, bitches were more frequently affected than male dogs in our study; the disparity (59 vs. 41%) was substantial, and it was not explained by an over representation of females in the reference population. Only 14% of cases in the present study were diagnosed in dogs less than 7 years of age, which is uniformly consistent with the published literature.[36-42] The youngest diagnosed case noted in the summary references consulted was 1 year of age.[41] Thus, it is noteworthy that the present study recorded one case less than 12 months of age, and another four between the ages of 12 and 24 months. The youngest case was a male Maltese pup weighing 5-15 pounds and discharged alive. Nasal neoplasms as a group were one of the few tumors studied for which specific data on predilection by body size were available in the literature; dogs of medium to large breeds are

reportedly predisposed.[38-41] We noted no marked distribution of cases in the larger breeds; the number of cases weighing 15-50 pounds (44) was about the same as the number weighing 50-100 pounds (43). There were only two cases weighing >100 pounds at diagnosis, while 7 cases weighed between 5 and 15 pounds. About three quarters of affected dogs were discharged alive. Twenty one percent were euthanized after diagnosis. With respect to breed, the largest number of cases were of mixed breed, and the proportion of cases (36%) was substantially higher than the proportion of the reference population accounted for by this breed (21%). Eight of the top eleven breeds ranked by frequency of cases, and all five ranked by percent of breed individuals affected, were either large, dolichocephalic, or both, which is consistent with some previous reports that suggest these breeds are predisposed.[36-41] Whether or not specific breed predispositions for nasal tumors really exist in dogs seems to be somewhat controversial, however. In any case, eight of these eleven breeds have been mentioned by at least one reference as having a high risk of nasal tumors. The Irish setter has been characterized variously as being both at increased[21] and decreased[41] risk, but based upon rank by percent of all Irish setters afflicted with this tumor, it appeared to be at increased risk over the course of this study. Our findings are in agreement with Priester and McKay[21] in finding the Airedale as the most highly predisposed breed. While we found no previous reports addressing breed predisposition for the Samoyed, it appeared among the top eleven breeds as ranked by case frequency in the present study, as well as among the top five breeds as ranked by percentage of breed individuals affected

by nasal adenocarcinoma. These ranks are not explained by the popularity of the breed in the reference population. Distributions by sex\reproductive status, age, body weight, and vital status at discharge all differed significantly in dogs diagnosed with nasal adenocarcinomas versus all individuals seen for all reasons.

Prostatic Adenocarcinomas

The literature reports these rare neoplasms to effect older dogs almost exclusively, with median age at diagnosis of 9-10 years.[43-48] In general, differences in tumor frequency among intact versus castrated dogs have not been noted,[43,44,48] although one case series found a higher incidence in neutered animals.[45] Medium to large breeds of dogs have been reported to have higher case frequencies.[46] Other accounts have not recorded a breed predilection.[43,47,48] Among our 26 cases, intact dogs outnumbered castrates approximately 2:1. This ratio was somewhat less than the nearly 6:1 ratio of intact dogs to castrates in the reference population. Cases were overwhelmingly older animals, with 69% between the ages of 10 and 15 and 92% older than 7 years of age. In terms of weight, the heavier weight classes were not over represented; about a third of the cases weighed >50 pounds at diagnosis. In contrast to most of the other histologic types under study, only 19% of the dogs diagnosed with prostatic adenocarcinomas were discharged alive; a majority of nearly three quarters were euthanized. The most frequently affected breed was the mongrel, followed by the Doberman pinscher. No other breed recorded more than one case. Interestingly, in another recent study[45] the Doberman breed was

most frequently identified to have prostatic disease in general, although breed incidence for adenocarcinomas specifically was not reported. Distributions by reproductive status, age, body weight, and vital status at discharge all differed significantly in dogs diagnosed with adenocarcinomas of the prostate when compared with the population of reference dogs.

Transitional Cell Carcinomas of the Urinary Bladder and Urethra

Although some sources have suggested that these neoplasms may occur more frequently in males than females,[49] others suggest females predominate among affected animals.[46,50-53] It is largely a disease of older dogs.[49-51,53] Some authors report no consistent breed predilection,[49] while others note increased occurrence in terriers (Scottish, Cairn, West Highland White and Airedale), Shetland sheepdogs, beagles and collies.[46] Our study noted females constituting a substantial majority (68%) of the afflicted animals, with spayed bitches more frequently affected by a margin of better than 3:1. Among the males, intact dogs were diagnosed with the disease more than twice as often as castrates, but this ratio agrees with the ratio noted among the reference dogs. Consistent with the literature, over 90% of the cases were diagnosed in animals older than 7 years of age, with more than 70% >10 years old. Fewer than half of the affected animals were discharged alive; 12% died in hospital of their disease. While cases in mixed breeds were most frequently recorded for this histologic type as well, the present study concurred rather well with the breed predispositions reported by Madewell and Theilen.[46] Five of seven breeds noted by those authors also appeared in the top twelve here (as ranked by frequency

of cases), with three of those among the six breeds most frequently recording cases in our study. When ranked according to percentage of their respective breed individuals afflicted with this cancer, Keeshonds, Scottish, wirehaired fox and Airedale terriers, and beagles comprised those with the highest percentages. None of the summary references consulted noted a predilection for the Keeshond breed. Distributions of all five variables under study differed significantly in dogs diagnosed with this neoplasm versus the reference population.

Spindle Cell Sarcomas

It has been recognized that tumors originating in the soft tissues present a formidable diagnostic challenge, with histologic characteristics often overlapping between tumor types,[54] making final histopathologic diagnosis more often dependent on convention than on objective criteria.[33] Given this ambiguity, we decided to consider five arguably similar neoplasms (fibrosarcomas, neurofibrosarcomas, hemangiopericytomas, malignant fibrous histiocytomas and Schwannomas) as a single group for the purposes of our study. Since descriptive information on sex and breed predilection for these tumors has been presented separately in several oncology references, the literature is reported for each subtype separately. However, we consider it more appropriate to summarize our findings for these five histologic types grouped as a whole. Fibrosarcomas are reported to affect primarily adult and aged dogs with no apparent sex or breed predilection,[31,33,55,56] although other authors note a possible predilection for females and cocker spaniels.[57] Neurofibrosarcomas

and Schwannomas have been noted as often referring to the same tumor,[54,57,58] though they are considered a distinct entity by others.[59] They are recorded as occurring primarily in older dogs with no sex predilection, but with fox terriers constituting a predisposed breed.[50] A recent case series[58] includes these two tumors in the larger histologic category of peripheral nerve sheath tumors. In that study, male cases outnumbered females 1.5:1, but no breed predilections were reported. Over seventy percent of the cases followed to death in that study either died or were euthanized due to their cancers. Females older than 5 to 6 years of age and of the German shepherd, boxer, cocker and springer spaniel, and fox terrier breeds are reported as more commonly affected by hemangiopericytomas.[31-33,57] Others report no sex predilection.[54] Malignant fibrous histiocytomas are noted as rare tumors occurring in middle-aged to older dogs, with no apparent breed or sex predilection.[54,57] Other sources report too few cases in dogs to make meaningful conclusions.[33] Records for cases of spindle cell sarcomas in the present study found essentially equal numbers of males and females affected. A consistent trend of increasing frequency of diagnosis with increasing age was prominent, although two animals were diagnosed at less than 1 year of age. Both were males, the younger a golden retriever pup, weighing 50-75 pounds and discharged alive with a diagnosis of fibrosarcoma, and the older a 15-30 pound mongrel discharged alive with malignant fibrous histiocytoma. About three quarters of the cases in this study were discharged alive, with 22% euthanized. Golden retrievers comprised the largest proportion of cases in one recent case series of fibrosarcomas,[56] and that

breed was one of the five most frequently recording cases in the current study as well. Of the few breeds mentioned as possibly predisposed for the five histologic types comprising the spindle cell sarcomas, only the German shepherd and boxer were ranked among the ten breeds most frequently recording cases in this study, and only the boxer appeared at a frequency which differs substantially from its proportional representation in the reference population. When ranked according to percentage of breed individuals afflicted with this tumor, the Briard breed ranked highest (as it did for melanomas), followed by pointers, boxers, pit bull terriers and English setters. As was the case with the majority of the study cancers, distributions by sex\reproductive status, age, body weight, and vital status at discharge all differed significantly in dogs diagnosed with the spindle cell neoplasms as compared to the reference population.

General Comments

It is important to emphasize that univariate frequency tables such as those reported in this study cannot fully describe the potentially complex interrelationships that may exist among multiple study variables, relationships which may confound each other. Although we have presented some simple statistical results here in the interest of completeness, care should be taken in their interpretation. Rigorous multivariable modeling approaches are clearly indicated where potential confounding by various factors could be controlled. These modeling approaches, however, were outside the scope and objectives of the present study, but will be part of a subsequent study already underway.

Returning to an earlier example, one could formulate a number of interpretations for the occurrence of a majority (55%) of our cases of lymphosarcoma at body weights > 50 pounds (Table 2.4). It could be argued that obesity plays a role in the occurrence of this neoplasm, or, alternatively, that a predominance of cases in large breeds such as the boxer is a confounding factor responsible for the distribution of cases to the heavier weight classes. Because body weight may also be related to age or gender, those variables could also conceivably exert a confounding effect. Without a multivariable analysis where the independent effects of individual study variables on tumor occurrence can be assessed while simultaneously controlling for the effects of potential confounding variables, it is not possible to definitively address the roles played by each as risk factors for cancer in dogs. In recognition of these issues, our conclusions here have intentionally avoided definitive statements that are not justified by a mainly descriptive treatment such as ours. Nevertheless, there are significant findings, particularly the role of breed, worth noting.

Interpretation of the role of breed as a risk factor in tumor occurrence is not straightforward. This study has chosen to present rankings based on both overall frequency of reported cases and on the percentage of all individuals of a given breed that is accounted for by cases of a particular neoplasm. Both rankings convey useful information if taken in context. Ranking by overall frequency answers the question of what dogs present most often with a given tumor, while percentage of breed individuals affected may better address the issue of differential breed susceptibility. Moreover, it is interesting to compare

rankings by the two methods, and to note how often they differ. This suggests that the overall frequency of diagnosis of a tumor, broken down by breed, may say little about actual breed predispositions for that tumor. While this may seem an obvious statement, many of the breed predispositions recorded in the literature seemed to correspond more closely to our rankings based on case frequency than to those based on percent of breed individuals affected. Since the latter should theoretically be more representative of actual breed predisposition, some caution may be in order when drawing conclusions about breed susceptibilities based on the literature. The mammary neoplasms were a notable exception to this, with the literature consistently reporting high incidence in sporting and hunting dogs, which coincided quite closely with our findings. It must also be mentioned that these rankings by percent of all individuals of a breed diagnosed with a particular neoplasm may be somewhat unstable. Among the less common breeds, even a small number of cases may lead, perhaps artifactually, to a high ranking for that breed. Animals of a particular familial genetic line could be contributing those few cases, and their increased risk of disease may not be indicative of the breed as a whole. Future studies using large sample sizes should be encouraged to further address the breed issue while controlling for other factors *via* multivariable modeling.

While this study has defined a functional comparison population for the tumor cases from the population of all the MSU-VMC cases that gave rise to it, uncertainty remains regarding the specifics of occurrence of these neoplasms in the general population of dogs. Experience in both veterinary and human

medicine has shown that practitioners are more likely to refer cancer patients for specialized care than those patients afflicted with more easily treated conditions. Consequently, the differences between the characteristics of the general population and those of the referred population may be less pronounced for cancer patients than they would be for patients with other diseases. However, the possibility that the referral cases that comprise studies such as this one may differ in substantial ways from dogs not referred to teaching hospitals should in no way be excluded. Indeed, previous evidence suggests that differences between hospital and general populations do exist.[60] As was noted in great detail in Chapter 1, incidence rates for some canine cancers have been reported, but extrapolation of rates derived from one region and time period to others may underestimate the variability of local environmental factors that could influence the occurrence of cancer. Thus, some caution in extrapolation of results seems warranted.

Still, as Miller has pointed out,[61] "unavoidable imperfections in scientific design should not discourage the search for new clues to the causes of disease". It is hoped that projects such as this one may stimulate further research into the etiology and environmental determinants of cancer in companion animals as well as other species, including humans. Some specific research directions seem indicated. A multivariable assessment of this same data set, perhaps employing a logistic model with an ordinal outcome, would help assess the simultaneous influence of such factors as sex/reproductive status, age, weight and breed on the occurrence of the cancers studied. In

addition to the cases studied here and their referents, such a study could also incorporate cases of these neoplasms with benign disease, to get an idea of how these factors may vary with clinical stage of the cancer, as well as a second referent group comprised of dogs diagnosed at hospitals other than MSU-VMC, to address the possibility of referral bias. Another piece of research which could prove both fascinating and valuable would be a systematic review of the canine cancer literature, focusing on citations which mention particular breeds as being predisposed to particular cancers. The review would trace back those citations to their original sources, comparing the characteristics of the original study populations from which the breed predispositions were derived with those of the study populations in which they were subsequently cited. It would be interesting to see if the methods originally used to determine breed predispositions were scientifically defensible. A similar analysis approach might be used to compare study populations to which cancer incidence rates have been extrapolated with those from which they were derived. These types of research approaches might provide some insight on whether the faith placed in references to breed predispositions and incidence rates that appear in the veterinary literature is in fact justified. Finally, as was pointed out here by the record for the lymphosarcoma patient weighing 50 to 75 pounds at < 2 weeks of age, some data records in the VMDB are miscoded despite the best efforts of clinicians and medical records personnel alike. A study quantifying the frequency and characteristics of such miscoded records could also prove useful for future research. While such a study has been reported by Pollari and coworkers for

case records at one veterinary teaching hospital,[62] no such data for the VMDB as a whole apparently exist at this time.

REFERENCES

REFERENCES

1. Reeves MJ, Postorino Reeves N. Epidemiology and the veterinary oncologist—evaluation and critical appraisal of the scientific oncology literature: How to read the clinical literature. In: Gilson SD, editor. Surgical Oncology. *Vet Clin North Am Small Anim Pract* 1995;25(1):1-18.

2. Madewell BR. *Etiology of neoplastic disease*. In: Clinical oncology (cancer-diagnosis, treatment & prognosis). Refresher course for veterinarians. Proceedings #69. Sydney: Post-Graduate Committee in Veterinary Science, University of Sydney; 1984:176-82.

3. Tjalma RA. Cancer epizootiology. In: Meier H, editor. Epizootiology of cancer in animals. *Ann NY Acad Sci* 1963;108(3):849-54.

4. Misdorp, W. Veterinary cancer epidemiology. *Vet Q* 1996;18(1):32-6.

5. Priester WA, editor. Standard nomenclature of veterinary diseases and operations. 2nd (abridged) ed. Bethesda (MD): National Cancer Institute (US); 1975. DHEW publication no. (NIH)76-1028.

6. SAS Institute Inc. The FREQ Procedure. In: SAS/STAT User's Guide, Version 6, 4th ed., Vol. 1. Cary (NC): SAS Institute, Inc.; 1989. p. 851-89.

7. Stata Corp. Stata Statistical Software: Release 5.0. College Station (TX): Stata Corporation; 1997.

8. Moulton JE, editor. Tumors in domestic animals. 3rd ed. Berkeley (CA): University of California Press; 1990:672 p.

9. Theilen GH, Madewell BR, editors. Veterinary cancer medicine. 2nd ed. Philadelphia: Lea & Febiger; 1987:676 p.

10. Withrow SJ, MacEwen EG, editors. Clinical veterinary oncology. Philadelphia : Lippincott; 1989:507 p.

11. Withrow SJ, MacEwen EG, editors. Small animal clinical oncology. 2nd ed. Philadelphia: WB Saunders; 1996:589 p.

12. Head KW. Tumors of the alimentary tract. In: Moulton JE, editor. Tumors in domestic animals. 3rd ed. Berkeley (CA): University of California Press; 1990. p. 398-412.

13. Klausner JS, Hardy RM. Alimentary tract, liver, and pancreas. In: Slatter DH, editor. Textbook of small animal surgery. Philadelphia: WB Saunders; 1985. p. 2458-61.

14. Straw RC. Tumors of the intestinal tract. In: Small animal clinical oncology. 2nd ed. Philadelphia: WB Saunders; 1996. p. 252-61.

15. Valerius, KD, Powers, BE, McPherron, MA, Hutchison, JM, Mann, FA, Withrow, SJ. Adenomatous polyps and carcinoma *in situ* of the canine colon and rectum: 34 cases. J Am Anim Hosp Assoc 1997;33(2):156-60.

16. Davies AP, Klausner JS, Grindem CB. Hematopoetic system. In: Slatter DH, editor. Textbook of small animal surgery. Philadelphia: WB Saunders; 1985. p. 2478-82.

17. MacEwen EG, Young KM. Canine lymphoma and lymphoid leukemias. In: Small animal clinical oncology. 2nd ed. Philadelphia: WB Saunders; 1996. p. 451-79.

18. Madewell BR, Theilen GH. Hematopoetic neoplasms, sarcomas and related conditions. Part IV, Canine. In: Veterinary cancer medicine. 2nd ed. Philadelphia: Lea & Febiger; 1987. p. 392-406.

19. Moulton JE, Harvey JW. Tumors of the lymphoid and hematopoietic tissues. Lymphoma of the Dog. In: Moulton JE, editor. Tumors in domestic animals. 3rd ed. Berkeley (CA): University of California Press; 1990. p. 240-4.

20. Squire RA. Spontaneous hematopoietic tumors of dogs. In: Lingeman CH, Garner FM, editors. Symposium on comparative morphology of hematopoietic neoplasms. Washington (DC): Armed Forces Institute of Pathology (US) and Bethesda (MD): National Cancer Institute (US), 1969. p. 97-109. (National Cancer Institute Monographs; vol 32).

21. Priester WA, McKay FW. The occurrence of tumors in domestic animals. Ziegler JL, editor. Bethesda (MD): National Cancer Institute (US); 1980:210 p. (National Cancer Institute Monographs; vol 54).

22. Smith HA. Leukemic neoplasia in the dog. In: Meier H, editor. Epizootiology of cancer in animals. *Ann NY Acad Sci* 1963;108(3):633-41.

23. Moore AS. Canine lymphoma. In: Treatment of cancer in companion animals. Refresher course for veterinarians. Proceedings #224. Sydney:

Post-Graduate Committee in Veterinary Science, University of Sydney; 1994. p. 37-9.

24. Johnston SD. Female reproductive system. In: Slatter DH, editor. Textbook of small animal surgery. Philadelphia: WB Saunders; 1985. p. 2545-56.

25. MacEwen EG, Withrow KM. Tumors of the mammary gland. In: Small animal clinical oncology. 2nd ed. Philadelphia: WB Saunders; 1996. p. 356-72.

26. Madewell BR, Theilen GH. Tumors of the mammary gland. Canine mammary neoplasms. In: Veterinary cancer medicine. 2nd ed. Philadelphia: Lea & Febiger; 1987.p. 328-37.

27. Moulton JE. Tumors of the mammary gland. Mammary tumors of the dog. In: Moulton JE, editor. Tumors in domestic animals. 3rd ed. Berkeley(CA): University of California Press; 1990. p. 518-46.

28. Boldizsar H, Szenci O, Muray T, Csenki J. Studies on canine mammary tumours. I. Age, seasonal and breed distribution. Acta Vet Hung 1992;40(1-2):75-87.

29. Sonnenschein EG, Glickman LT, Goldschmidt MH, McKee LJ. Body conformation, diet, and risk of breast cancer in pet dogs: a case-control study. Am J Epidemiol 1991;133(7):694-703.

30. Ferguson HR. Canine mammary gland tumors. In: Brown NO, editor. Symposium on clinical veterinary oncology. Vet Clin North Am Small Anim Pract 1985;15(3):501-11.

31. Henderson RA, Core DM. Skin. In: Slatter DH, editor. Textbook of small animal surgery. Philadelphia: WB Saunders; 1985. p. 2441-52.

32. Madewell BR, Theilen GH. Tumors of the skin and subcutaneous tissues. In: Veterinary cancer medicine. 2nd ed. Philadelphia: Lea & Febiger; 1987. p. 233-325.

33. Pulley LT, Stannard AA. Tumors of the skin and soft tissues. In: Moulton JE, editor. Tumors in domestic animals. 3rd ed. Berkeley (CA): University of California Press; 1990. p. 23-87.

34. Strafuss AC. Skin tumors. In: Brown NO, editor. Symposium on clinical veterinary oncology. Vet Clin North Am Small Anim Pract

1985;15(3):473-92.

35. Vail DM, Withrow SJ. Tumors of the skin and subcutaneous tissues. In: Small animal clinical oncology. 2nd ed. Philadelphia: WB Saunders; 1996. p. 167-91.

36. Hayes HM, Wilson GP. Comparative aspects of nasal passage carcinoma in dogs with man. In: Reznik G, Stinson SF, editors. Nasal tumors in animals and man. Vol. II. Tumor pathology. Boca Raton (FL): CRC Press, Inc.; 1983. p. 229-38.

37. Madewell BR, Theilen GH. Tumors of the respiratory tract and thorax. Neoplasms of the nasal passages and paranasal sinuses. In: Veterinary cancer medicine. 2nd ed. Philadelphia: Lea & Febiger; 1987. p. 536-43.

38. Moulton JE. Tumors of the respiratory system. Nasal cavity and paranasal sinus tumors of the dog. In: Moulton JE, editor. Tumors in domestic animals. 3rd ed. Berkeley (CA): University of California Press; 1990. p. 308-13.

39. Norris AM, Henderson RA, Withrow SJ. Respiratory system. Neoplasms of the nasal and paranasal sinuses. In: Slatter DH, editor. Textbook of small animal surgery. Philadelphia: WB Saunders; 1985. p. 2583-7.

40. Withrow SJ. Tumors of the respiratory system. Nasal tumors. In: Small animal clinical oncology. 2nd ed. Philadelphia: WB Saunders; 1996. p. 281-6.

41. Beck ER, Withrow SJ. Tumors of the canine nasal cavity. In: Brown NO, editor. Symposium on clinical veterinary oncology. Vet Clin North Am Small Anim Pract 1985;15(3):521-33.

42. Ogilvie GK, LaRue SM. Canine and feline nasal and paranasal sinus tumors. Vet Clin North Am Small Anim Pract 1992;22(5):1133-44.

43. Bell FW, Klausner JS, Johnston SD, Feeney DA, Hayden DW. Clinical and pathologic features of spontaneous prostatic adenocarcinoma in 31 dogs. In: Proceedings of the eighth annual conference of the Veterinary Cancer Society; 1988 Oct 5-8; Estes Park, Colorado. Veterinary Cancer Society, 1988.

44. Johnston SD. Male reproductive system. In: Slatter DH, editor. Textbook of small animal surgery. Philadelphia: WB Saunders; 1985. p. 2535-45.

45. Krawiec DR, Heflin D. Study of prostatic disease in dogs: 177 cases (1981-1986). J Am Vet Med Assoc 1992;200(8):1119-22.

46. Madewell BR, Theilen GH. Tumors of the urogenital tract. In: Veterinary cancer medicine. 2nd ed. Philadelphia: Lea & Febiger; 1987. p. 567-600.

47. Nielsen SW, Kennedy PC. Tumors of the genital systems. Carcinoma of the prostate. In: Moulton JE, editor. Tumors in domestic animals. 3rd ed. Berkeley (CA): University of California Press; 1990. p. 493-5.

48. Withrow SJ, Reeves NP. Tumors of the male reproductive tract. Canine prostatic tumors. In: Small animal clinical oncology. 2nd ed. Philadelphia: WB Saunders; 1996. p. 376-8.

49. Nielsen SW, Moulton JE. Tumors of the urinary system. Tumors of the urinary bladder and urethra. In: Moulton JE, editor. Tumors in domestic animals. 3rd ed. Berkeley (CA): University of California Press; 1990. p. 465-72.

50. Caywood DD, Osborne CA. Urinary system. In: Slatter DH, editor. Textbook of small animal surgery. Philadelphia: WB Saunders; 1985. p. 2561-74.

51. Norris AM, Laing E, Withrow SJ, Oglivie GL, Tomlinson J, Pidgeon GL, et al. A clinical pathologic study of cancer of the canine urinary bladder. In: Proceedings of the seventh annual conference of the Veterinary Cancer Society; Oct 26-28; Madison, Wisconsin. Veterinary Cancer Society; 1987. p. 28.

52. Stone EA. Urogenital tumors. In: Brown NO, editor. Symposium on clinical veterinary oncology. Vet Clin North Am Small Anim Pract 1985;15(3):597-608.

53. Withrow SJ. Tumors of the urinary system. Urinary bladder cancer. In: Small animal clinical oncology. 2nd ed. Philadelphia: WB Saunders; 1996. p. 385-90.

54. MacEwen EG, Withrow SJ. Soft tissue sarcomas. In: Small animal clinical oncology. 2nd ed. Philadelphia: WB Saunders; 1996. p. 211-26.

55. Richardson RC. Solid tumors. In: Brown NO, editor. Symposium on clinical veterinary oncology. Vet Clin North Am Small Anim Pract 1985;15(3):557-67.

56. Ciekot PA, Powers BE, Withrow SJ, Straw RC, Oglivie GK, LaRue SM. Histologically low-grade, yet biologically high-grade, fibrosarcomas of the mandible and maxilla in dogs: 25 cases (1982-1991). J Am Vet Med Assoc 1994;204(4):610-15.

57. Muller GH, Kirk RW, Scott DW. Neoplastic diseases. Mesenchymal neoplasms. In: Small animal dermatology. 3rd ed. Philadelphia: WB Saunders; 1983. p. 738-71.

58. Brehm DM, Vite CH, Steinberg HS, Haviland J, vanWinkle T. A retrospective evaluation of 51 cases of peripheral nerve sheath tumors in the dog. J Am Anim Hosp Assoc 1995;31(4):349-59.

59. Cordy DR. Tumors of the nervous system and eye. Schwannoma. In: Moulton JE, ed. Tumors in domestic animals. 3rd ed. Berkeley (CA): University of California Press; 1990. p. 652-4.

60. Robinson GW. A comparison of licensed and hospital dog populations. J Am Vet Med Assoc 1968;152(9):1383-9.

61. Miller RW. Foreword. In: Priester WA, McKay FW. The occurrence of tumors in domestic animals. Ziegler JL, editor. Bethesda (MD): National Cancer Institute (US); 1980. p. ix. (National Cancer Institute Monographs; vol 54).

62. Pollari FL, Bonnett BN, Allen DG, Bamsey SC, Martin SW. Quality of computerized medical record abstract data at a veterinary teaching hospital. Prev Vet Med 1996;27(3-4):141-154.

Table 2.1. Case definitions, selected canine neoplasms, Michigan, 1964-1994

Neoplasm	SNVDO Code(s)	Number of cases
Intestinal/Rectal Adenocarcinomas (I/RA)	60408091, 65008091, 65108091,66008091,66808091	75
Lymphosarcomas (LY)	____8310	453
Mammary Adenocarcinomas (MA)	19008091	172
Mammary Carcinomas (MC)	19008191	10
Melanomas (ME)	____8174	136
Mixed Mammary Tumors (MMT)	19008080	52
Nasal Adenocarcinomas (NA)	31008091,31308091	97
Prostatic Adenocarcinomas (PA)	76408091	26
Spindle Cell Sarcomas* (SCS)	____8451,____8452,____8531,____8700F,____8960	250
Transitional Cell Carcinomas of the Urinary Bladder and Urethra (TCC)	73008101,74008101	75
Total		*1346*

*Includes fibrosarcomas, neurofibrosarcomas, hemangiopericytomas, malignant fibrous histiocytomas and Schwannomas.

Table 2.2. Frequency (%) of cases by sex/reproductive status
(Square brackets [ ] denote % of all dogs of that sex seen for all reasons. For example, 14.7% all intestinal/rectal adenocarcinomas were reported in females, while cases of intestinal/rectal adenocarcinoma comprised 0.025% of all female individuals seen for all reasons. Neoplasm abbreviations follow Table 2.1.)

Table 2.2

Neoplasm	Female	Spayed Female	Male	Neutered Male	*Total (Cum. %)*
I/RA*	11 (14.7) [0.025]	26 (34.7) [0.12]	31 (41.3) [0.059]	7 (9.3) [0.075]	*75* *(100)*
LY*	84 (18.5) [0.19]	131 (28.9) [0.61]	175 (38.6) [0.33]	63 (13.9) [0.68]	*453* *(100)*
MA†	93 (54.1) [0.21]	78 (45.3) [0.36]	1 (0.6) [0.002]	0 (0.0) [0.0]	*172* *(100)*
MC†	7 (70.0) [0.016]	3 (30.0) [0.014]	0 (0.0) [0.0]	0 (0.0) [0.0]	*10* *(100)*
ME*	18 (13.2) [0.041]	44 (32.4) [0.2]	64 (47.1) [0.12]	10 (7.4) [0.1]	*136* *(100)*
MMT†	33 (63.5) [0.075]	18 (34.6) [0.083]	1 (1.9) [0.002]	0 (0.0) [0.0]	*52* *(100)*
NA*	22 (22.7) [0.05]	35 (36.1) [0.16]	29 (29.9) [0.055]	11 (11.3) [0.12]	*97* *(100)*
PA*	-	-	17 (65.4) [0.032]	9 (34.6) [0.1]	*26* *(100)*
SCS*	54 (21.6) [0.12]	69 (27.6) [0.32]	106 (42.4) [0.2]	21 (8.4) [0.23]	*250* *(100)*
TCC*	12 (16.0) [0.027]	39 (52.0) [0.18]	17 (22.7) [0.032]	7 (9.3) [0.075]	*75* *(100)*
Total *[Cum. %]*	*334* *[0.75]*	*443* *[2.0]*	*441* *[0.83]*	*128* *[1.4]*	*1346* *(100)*

Distribution by sex/repro. status significantly different in dogs with this cancer vs. reference group by: *χ^2 ($p \leq 0.005$) or †Fisher's Exact Test ($p \leq 0.01$).

Table 2.3. Frequency (%) of cases by age at diagnosis
(Square brackets [ ] denote % of all dogs of that age seen for all reasons.
Neoplasm abbreviations follow Table 2.1)

Neoplasm	<2 yrs	2-<4 yrs	4-<7 yrs	7-<10 yrs	≥10 yrs	*Total (Cum. %)*
I/RA*	2 (2.7) [0.0036]	2 (2.7) [0.0086]	17 (22.7) [0.078]	26 (34.7) [0.16]	28 (37.3) [0.22]	*75* *(100)*
LY†	21 (4.6) [0.038]	61 (13.5) [0.26]	142 (31.3) [0.65]	141 (31.1) [0.86]	88 (19.4) [0.7]	*453* *(100)*
MA*	2 (1.2) [0.0036]	3 (1.7) [0.013]	24 (14.0) [0.11]	63 (36.6) [0.39]	80 (46.5) [0.64]	*172* *(100)*
MC*	0 (0.0) [0.0]	1 (10.0) [0.0043]	3 (30.0) [0.014]	4 (40.0) [0.025]	2 (20.0) [0.016]	*10* *(100)*
ME*	1 (0.7) [0.0018]	3 (2.2) [0.013]	8 (5.9) [0.037]	35 (25.7) [0.21]	89 (65.4) [0.71]	*136* *(100)*
MMT*	0 (0.0) [0.0]	0 (0.0) [0.0]	8 (15.4) [0.037]	16 (30.8) [0.098]	28 (53.8) [0.22]	*52* *(100)*
NA*	5 (5.2) [0.009]	1 (1.0) [0.0043]	8 (8.2) [0.037]	36 (37.1) [0.22]	47 (48.5) [0.38]	*97* *(100)*
PA*	0 (0.0) [0.0]	0 (0.0) [0.0]	1 (3.9) [0.0046]	6 (23.1) [0.037]	18 (69.2) [0.14]	*25* *(96.2)‡*
SCS†	13 (5.2) [0.024]	28 (11.2) [0.12]	42 (16.8) [0.19]	74 (29.6) [0.45]	92 (36.8) [0.73]	*249* *(99.6)‡*

Table 2.3 (cont'd)

Neoplasm	<2 yrs	2-<4 yrs	4-<7 yrs	7-<10 yrs	≥10 yrs	*Total (Cum. %)*
TCC*	1 (1.3) [0.0018]	2 (2.7) [0.0086]	3 (4.0) [0.014]	15 (20.0) [0.092]	54 (72.0) [0.43]	*75* *(100)*
Total *[Cum. %]*	*45* *[0.082]*	*101* *[0.43]*	*256* *[1.2]*	*416* *[2.5]*	*526* *[4.2]*	*1344* *(99.6)*

Distribution by age significantly different in dogs with this cancer vs. all individuals seen for all reasons by: *Fisher's Exact Test ($p \leq 0.002$) or $^{\dagger}\chi^2$ ($p < 0.001$).

‡No age was recorded for one patient for each of these histologic types.

Table 2.4. Frequency (%) of cases by body weight at diagnosis
(Square brackets [ ] denote % of all dogs of that weight seen for all reasons.
Neoplasm abbreviations follow Table 2.1)

Neoplasm	<15 Lbs.	15-<30 Lbs.	30-<50 Lbs.	50-<75 Lbs.	75-<100 Lbs.	≥100 Lbs.	Not Known	*Total (Cum. %)*
I/RA*	9 (12.0) [0.037]	15 (20.0) [0.056]	8 (10.7) [0.038]	27 (36.0) [0.1]	10 (13.3) [0.1]	3 (4.0) [0.074]	2 (2.7) [0.011]	*74* *(98.7)*§
LY†	25 (5.5) [0.1]	67 (14.8) [0.25]	90 (19.9) [0.43]	153 (33.8) [0.58]	65 (14.3) [0.68]	29 (6.4) [0.72]	23 (5.1) [0.12]	*452* *(99.8)*§
MA‡	24 (14.0) [0.1]	45 (26.2) [0.17]	36 (20.9) [0.17]	41 (23.8) [0.16]	15 (8.7) [0.16]	2 (1.2) [0.049]	9 (5.2) [0.049]	*172* *(100)*
MC‡	1 (10.0) [0.004]	3 (30.0) [0.011]	0 (0.0) [0.0]	3 (30.0) [0.011]	1 (10.0) [0.01]	0 (0.0) [0.0]	2 (20.0) [0.011]	*10* *(100)*
ME‡	26 (19.1) [0.11]	38 (27.9) [0.14]	19 (14.0) [0.092]	37 (27.2) [0.14]	8 (5.9) [0.083]	3 (2.2) [0.074]	5 (3.7) [0.027]	*136* *(100)*
MMT‡	9 (17.3) [0.037]	17 (32.7) [0.064]	11 (21.2) [0.053]	10 (19.2) [0.037]	5 (9.6) [0.052]	0 (0.0) [0.0]	0 (0.0) [0.0]	*52* *(100)*

Table 2.4 (cont'd)

Neoplasm	<15 Lbs.	15-<30 Lbs.	30-<50 Lbs.	50-<75 Lbs.	75-<100 Lbs.	≥100 Lbs.	Not Known	*Total (Cum. %)*
NA*	7 (7.2) [0.029]	23 (23.7) [0.087]	21 (21.6) [0.1]	30 (30.9) [0.11]	13 (13.4) [0.14]	2 (2.1) [0.049]	0 (0.0) [0.0]	*96* *(98.9)*§
PA*	0 (0.0) [0.0]	8 (30.8) [0.03]	8 (30.8) [0.039]	6 (23.1) [0.023]	2 (7.7) [0.021]	0 (0.0) [0.0]	1 (3.8) [0.005]	*25* *(96.2)*§
SCS†	14 (5.6) [0.058]	42 (16.8) [0.16]	30 (12.0) [0.14]	102 (40.8) [0.39]	36 (14.4) [0.38]	19 (7.6) [0.47]	7 (2.8) [0.038]	*250* *(100)*
TCC*	5 (6.7) [0.021]	23 (30.7) [0.087]	23 (30.7) [0.11]	14 (18.7) [0.053]	2 (2.7) [0.021]	3 (4.0) [0.074]	3 (4.0) [0.016]	*73* *(97.5)*ǀ
Total *[Cum. %]*	*120* *[0.5]*	*281* *[1.1]*	*246* *[1.2]*	*423* *[1.6]*	*157* *[1.7]*	*61* *[1.5]*	*52* *[0.28]*	*1340* *(99.1)*

Distribution by weight significantly different in dogs with this cancer vs. all individuals seen for all reasons by: *Fisher's Exact Test ($p \leq 0.03$) or †χ^2 ($p < 0.001$). Superscript ‡ indicates non-significant comparisons ($p > 0.05$) using Fisher's Exact Test. "Not Known" category not included in calculations.

§No weight class value was recorded for one patient of each of these histologic types.

ǀNo weight class value was recorded for two patients of this histologic type.

Table 2.5. Frequency (%) of cases by vital status at discharge
(Square brackets [ ] denote % of all dogs with that discharge status seen for all reasons.
Neoplasm abbreviations follow Table 2.1)

Neoplasm	Alive	Died-Necropsy	Died-No Necropsy	Euthanasia-Necropsy	Euthanasia-No Necropsy	*Total (Cum. %)*
I/RA*	53 (70.7) [0.044]	5 (6.7) [0.3]	1 (1.3) [0.089]	12 (16.0) [0.48]	4 (5.3) [0.11]	*75* *(100)*
LY†	235 (51.9) [0.19]	52 (11.5) [3.1]	14 (3.1) [1.3]	124 (27.4) [5.0]	28 (6.2) [0.75]	*453* *(100)*
MA*	152 (88.4) [0.13]	0 (0.0) [0.0]	0 (0.0) [0.0]	16 (9.3) [0.64]	4 (2.3) [0.11]	*172* *(100)*
MC*	4 (40.0) [0.003]	1 (10.0) [0.06]	0 (0.0) [0.0]	5 (50.0) [0.2]	0 (0.0) [0.0]	*10* *(100)*
ME‡	126 (92.6) [0.1]	1 (0.7) [0.06]	0 (0.0) [0.0]	6 (4.4) [0.24]	3 (2.2) [0.08]	*136* *(100)*
MMT*	46 (88.5) [0.038]	3 (5.8) [0.18]	0 (0.0) [0.0]	3 (5.8) [0.12]	0 (0.0) [0.0]	*52* *(100)*

Table 2.5 (cont'd)

Neoplasm	Alive	Died-Necropsy	Died-No Necropsy	Euthanasia-Necropsy	Euthanasia-No Necropsy	*Total (Cum. %)*
I/RA*	53 (70.7) [0.044]	5 (6.7) [0.3]	1 (1.3) [0.089]	12 (16.0) [0.48]	4 (5.3) [0.11]	*75* *(100)*
NA*	74 (76.3) [0.061]	2 (2.1) [0.12]	1 (1.0) [0.089]	18 (18.6) [0.72]	2 (2.1) [0.053]	*97* *(100)*
PA*	5 (19.2) [0.004]	2 (7.7) [0.12]	0 (0.0) [0.0]	16 (61.5) [0.64]	3 (11.5) [0.08]	*26* *(100)*
SCS*	186 (74.4) [0.15]	8 (3.2) [0.48]	1 (0.4) [0.089]	46 (18.4) [1.8]	9 (3.6) [0.24]	*250* *(100)*
TCC*	33 (44.0) [0.027]	7 (9.3) [0.42]	2 (2.7) [0.18]	27 (36.0) [1.1]	6 (8.0) [0.16]	*75* *(100)*
Total *[Cum. %]*	*914* *[0.75]*	*81* *[4.8]*	*19* *[0.58]*	*273* *[11.0]*	*59* *[1.6]*	*1346* *(100)*

Distribution by weight significantly different in dogs with this cancer vs. all individuals seen for all reasons by: *Fisher's Exact Test ($p \leq 0.02$) or $^{\dagger}\chi^2$ ($p < 0.001$). Superscript ‡ indicates non-significant comparisons ($p > 0.05$) using Fisher's Exact Test.

Table 2.6. Tumor occurrence: Ranking of top five breeds

Neoplasm	By % of all cases of this tumor accounted for	*n*	%	By % of all individuals of this breed accounted for by cases of this tumor*	*n*	%
Intestinal/Rectal Adenocarcinomas						
	German Shepherd	12	16.0	Airedale Terrier	2	0.27
	Mixed Breed	11	14.7	Old English Sheepdog	3	0.23
	Miniature Poodle	7	9.3	Great Dane	4	0.21
	Collie	5	6.7	Standard Poodle	3	0.20
	Great Dane	4	5.3	Miniature Poodle	7	0.19
Lymphosarcomas						
	Mixed Breed	100	22.1	Irish Wolfhound	5	1.81
	Golden Retriever	36	7.9	Boxer	27	1.73
	Boxer	27	6.0	Kerry Blue Terrier	2	1.69
	German Shepherd	25	5.5	Pembroke Welsh Corgi	2	1.64
	Doberman Pinscher	18	4.0	Chesapeake Bay Retriever	5	1.35

Table 2.6 (cont'd)

Neoplasm	By % of all cases of this tumor accounted for	*n*	%	By % of all individuals of this breed accounted for by cases of this tumor*	*n*	%
Mammary Adenocarcinomas						
	Mixed Breed	27	15.7	Pointer	4	1.27
	Miniature Poodle	13	7.6	Norwegian Elkhound	4	0.79
	German Shepherd	10	5.8	Treeing-Walker Coonhound	2	0.78
	American Cocker Spaniel	8	4.7	English Setter	4	0.39
	Doberman Pinscher	6	3.5	Afghan Hound	3	0.38
Mammary Carcinomas						
	Doberman Pinscher	2	20.0	Doberman Pinscher‡	2	0.06
	Miniature Poodle	2	20.0	Miniature Poodle‡	2	0.05
	Boxer†	1	10.0			
	German Shepherd†	1	10.0			
	Labrador Retriever†	1	10.0			

Table 2.6 (cont'd)

Neoplasm	By % of all cases of this tumor accounted for	*n*	%	By % of all individuals of this breed accounted for by cases of this tumor*	*n*	%
Melanomas						
	Mixed Breed	31	22.8	Briard	2	1.32
	Miniature Poodle	12	8.8	Scottish Terrier	5	0.69
	Miniature Schnauzer	6	4.4	Standard Schnauzer	2	0.53
	Dachshund§	5	3.7	Miniature Dachshund	2	0.48
	German Shepherd§	5	3.7	Gordon Setter	2	0.42
Mixed Mammary Tumors						
	Mixed Breed	9	17.3	Weimaraner	2	0.31
	Miniature Poodle	5	9.6	German Short-haired Pointer	3	0.24
	Beagle‡	3	5.8	English Setter	2	0.19
	German Shorthaired Pointer‡	3	5.8	English Springer Spaniel	2	0.16
	American Cocker Spaniel‡	3	5.8	Brittany Spaniel	2	0.14

Table 2.6 (cont'd)

Neoplasm	By % of all cases of this tumor accounted for	*n*	%	By % of all individuals of this breed accounted for by cases of this tumor*	*n*	%
Nasal Adenocarcinomas						
	Mixed Breed	35	36.1	Airedale Terrier	4	0.55
	Collie	5	5.2	Shetland Sheepdog	4	0.24
	Golden Retriever	5	5.2	Samoyed	3	0.23
	Irish Setter	5	5.2	Irish Setter	5	0.20
	Airedale Terrier¶	4	4.1	Collie	5	0.17
Prostatic Adenocarcinomas						
	Mixed Breed	13	50.0	Doberman Pinscher‡	2	0.06
	Doberman Pinscher	2	7.7	Mixed Breed‡	13	0.05
	Beagle**	1	3.8			
	Bouvier Des Flandres**	1	3.8			
	Dachshund**	1	3.8			

Table 2.6 (cont'd)

Neoplasm	By % of all cases of this tumor accounted for	*n*	%	By % of all individuals of this breed accounted for by cases of this tumor*	*n*	%
Spindle Cell Sarcomas						
	Mixed Breed	67	26.8	Briard	2	1.32
	Golden Retriever	16	6.4	Pointer	4	1.27
	German Shepherd	15	6.0	Boxer	11	0.70
	Miniature Poodle	13	5.2	Pit Bull	2	0.56
	Doberman Pinscher	12	4.8	English Setter	5	0.48
Transitional Cell Carcinomas of the Urinary Bladder and Urethra						
	Mixed Breed	29	38.7	Keeshond	2	0.61
	Beagle	6	8.0	Scottish Terrier	3	0.42
	Collie††	3	4.0	Wirehaired Fox Terrier	3	0.27
	Wirehaired Fox Terrier††	3	4.0	Airedale Terrier	2	0.27
	Miniature Poodle††	3	4.0	Beagle	6	0.20

Table 2.6 (cont'd)

Neoplasm	By % of all cases of this tumor accounted for	*n*	%	By % of all individuals of this breed accounted for by cases of this tumor*	*n*	%
*Breed rankings based on this method were tested for significant differences by Fisher's Exact Test within each cancer type. No significant differences were found between breeds for any of the studied cancers (*p* values ranged from 0.277 [mammary adenocarcinomas] to 1.0 [mammary carcinomas]). †Based on percentage of cases, Skye Terrier, American Cocker Spaniel and Other Purebred (Specific Breed Unknown) also have this rank. ‡Since no other breeds recorded more than one case for this tumor, no other breeds were ranked. §Based on percentage of cases, Labrador Retriever and Scottish Terrier also have this rank. ǁBased on percentage of cases, English Springer Spaniel also has this rank. ¶Based on percentage of cases, German Shepherd, Labrador Retriever and Shetland Sheepdog also have this rank. **Based on percentage of cases, German Shepherd, Miniature Poodle, Samoyed, Miniature Schnauzer, American Cocker Spaniel, English Springer Spaniel, Finnish Spitz and Other Purebred (Specific Breed Unknown) also have this rank. ††Based on percentage of cases, Scottish Terrier also has this rank.						

Table 2.7. Proportional allocation by VMDB sex/reproductive status category of total cases of histologically-confirmed neoplasia, and total individuals seen by MSU-VMC, 3/1/64-3/31/94*

VMDB Sex/reproductive status category	Number of individuals/litters of pups with neoplasia (%)	Number of individuals/litters of pups seen for all reasons (%)
Litter	2 (<0.1)	1740 (1.3)
Female	1388 (21.3)	43,959 (33.9)
Spayed female	2079 (31.9)	21,592 (16.6)
Male	2167 (33.2)	52,945 (40.8)
Neutered male	880 (13.5)	9282 (7.1)
Female, unknown reproductive status	*1 (<0.1)*	*21 (0.2)*
Other/unknown	*3 (<0.1)*	*320 (0.2)*
Total animals, all breeds	6520	129,859

*Distribution by sex/reproductive status significantly different in dogs with neoplasia vs. all individuals seen for all reasons (χ^2 = 1848.38, d.f. = 4, p < 0.001). Italicized categories were not included in χ^2 calculation.

Table 2.8. Proportional allocation by age of total cases of histologically-confirmed neoplasia, and total individuals seen by MSU-VMC, 3/1/64-3/31/94*

Age	Number of dogs with neoplasia (%)	Number of dogs seen for all reasons (%)
<8 weeks	12 (0.2)	4426 (3.4)
2 - <6 months	30 (0.5)	18,005 (13.9)
6 - <12 months	82 (1.3)	16,372 (12.6)
1 - <2 years	182 (2.8)	16,584 (12.8)
2 - <4 years	399 (6.1)	23,262 (17.9)
4 - <7 years	1164 (17.9)	21,846 (16.8)
7 - <10 years	2102 (32.2)	16,335 (12.6)
10 - <15 years	2393 (36.7)	11,796 (9.1)
≥15 years	147 (2.3)	739 (0.6)
Unknown	*3 (<0.1)*	*494 (0.4)*
Total animals, all breeds	6514†	129,859

*Distribution by age significantly different in dogs with neoplasia vs. all individuals seen for all reasons (χ^2 = 11064.17, d.f. = 8, $p < 0.001$). Italicized category was not included in χ^2 calculation.

†No age class value was recorded for six animals in this group.

Table 2.9. Proportional allocation by body weight of total cases of histologically-confirmed neoplasia, and total individuals seen by MSU-VMC, 3/1/64-3/31/94*

Weight (lbs.)	Number of dogs with neoplasia (%)	Number of dogs seen for all reasons (%)
<5	9 (0.1)	3710 (2.9)
5 - <15	409 (6.3)	20,355 (15.7)
15 - <30	1231 (19.0)	26,546 (20.7)
30 - <50	1315 (20.2)	20,721 (16.0)
50 - <75	2057 (31.7)	26,450 (20.4)
75 - <100	861 (13.3)	9597 (7.4)
≥100	341 (5.3)	4036 (3.1)
Unknown	*272 (4.2)*	*18,444 (14.2)*
Total animals, all breeds	6495†	129,859

*Distribution by body weight significantly different in dogs with neoplasia vs. all individuals seen for all reasons (χ^2 = 1274.76, d.f. = 6, $p < 0.001$). Italicized category was not included in χ^2 calculation.

†No weight class value was recorded for twenty five animals in this group.

Table 2.10. Proportional allocation by vital status at discharge of total cases of histologically-confirmed neoplasia, and total individuals seen by MSU-VMC, 3/1/64-3/31/94*

Vital status at discharge	Number of dogs with neoplasia (%)	Number of dogs seen for all reasons (%)
Alive	4684 (71.8)	120,820 (93.0)
Died-necropsy	377 (5.8)	1672 (1.3)
Died-no necropsy	83 (1.3)	1118 (0.9)
Euthanasia-necropsy	1125 (17.3)	2494 (1.9)
Euthanasia-no necropsy	251 (3.8)	3755 (2.9)
Total animals, all breeds	6520	129,859

*Distribution by vital status at discharge significantly different in dogs with neoplasia vs. all individuals seen for all reasons (χ^2 = 9848.43, d.f. = 4, p < 0.001).

Table 2.11. Proportional allocation by breed of total cases of histologically-confirmed neoplasia, and total individuals seen by MSU-VMC, 3/1/64-3/31/94*

Breed	Rank for study tumors†	Number of dogs with neoplasia (%)	Number of dogs seen for all reasons (%)	Percent of total individuals of that breed diagnosed with neoplasia
Mixed	1	1566 (24.0)	27,540 (21.2)	5.7
German Shepherd	2	391 (6.0)	9573 (7.4)	4.1
Miniature Poodle	2	302 (4.6)	6815 (5.2)	4.4
Doberman Pinscher	4	240 (3.7)	3619 (2.8)	6.6
Labrador Retriever	5	314 (4.8)	7079 (5.5)	4.4
Collie	6	152 (2.3)	3028 (2.3)	5.0
Beagle	6	146 (2.2)	2964 (2.3)	4.9
Miniature Schnauzer	8	135 (2.1)	2877 (2.2)	4.7
American Cocker Spaniel	8	174 (2.7)	4231 (3.3)	4.1
Total animals, all breeds		6520	129,859	5.0

*Distribution by breed significantly different in dogs with neoplasia vs. all individuals seen for all reasons (χ^2 = 80.39, d.f. = 8, $p < 0.001$).

†Based on number of cases recorded for the ten study cancers combined. A breed was assigned a score of two if it was among the top five breeds (by case frequency) for a particular tumor, and one if it was sixth through tenth. Scores were then summed over all the ten tumor types.

Mixed			German Shepherd			Miniature Poodle			Neoplasm	Doberman Pinscher			Labrador Retriever		
<7	7-10	>10	<7	7-10	>10	<7	7-10	>10		<7	7-10	>10	<7	7-10	>10
○	O	○	○	○	○	○	○O	○	I/RA	-	○O	-	○O	○O	-
○	○	○	O	O	○	○	-	○O	LY	○	○	O	O	○	-
○	○	○	○	○	○	○	○	○	MA	○	○	O	○	○	○
-	-	-	-	●	-	○	-	O	MC	○	O	-	○	-	-
○	○	○	-	○	○	○	○	○	ME	-	○	○	-	○	○
○	○	○	○	-	○	-	○	○	MMT	-	○	-	-	-	-
○	○	○	○	○	●	-	○	○	NA	-	-	-	-	○	○
-	O	O	-	-	●	-	-	O	PA	O	-	O	-	-	-
○	○	○	○	○	○	○O	○	○	SCS	○	○	○O	○	○	○
○	○●	O	-	-	O	-	-	O	TCC	O	-	-	-	O	○

Figure 2.1. Comparison of age (in years), breed and vital status at discharge, selected cancers, Michigan, 1964-1994 (Plotting symbol reflects the mode of observations for vital status at discharge for cases in that category. Where two symbols appear in one cell, a tie occurred. ○ = Alive; O = Euthanasia; ● = Dead; - = No data. Neoplasm abbreviations follow Table 2.1)

Chapter 3

SPATIAL AND TEMPORAL DISTRIBUTION OF SELECTED CANINE CANCERS IN INGHAM, OAKLAND AND WAYNE COUNTIES, MICHIGAN, 1964-1994

Abstract

While rates are commonly used to compare regional disease occurrence, rate-independent methods may also prove useful in circumstances where geographic occurrence of a disease is known, but calculation of disease rates is not feasible. This is frequently the case for diseases in companion animals, where accurate enumeration of populations-at-risk is often arduous. This study had two objectives: to assess the usefulness of a rate-independent method for investigating disease aggregation in companion animals; and, to assess the spatial and temporal clustering of canine cases of some neoplasms that are biologically similar in dogs and humans. Geographic Information Systems and Point Pattern Analysis were used to assess the spatial and temporal clustering of incident cases of four types of canine cancer in three counties in Michigan between 1964 and 1994, and to generate hypotheses concerning disease aggregation. Significant ($p < 0.01$) spatial clustering was found that varied by county and cancer type. Temporal patterns generally did not concur with spatial ones, and varied from significant ($p = 0.02$) clustering to dispersion. These results demonstrate the utility of distance based methods for assessing clustering of disease, and suggest that processes determining the aggregation of canine cancer cases do not act uniformly over space or time.

Introduction

Over the last two decades, considerable attention has been directed toward clusters of disease in space and time. Epidemiological investigations of suspected disease clusters have often been initiated in hopes of allaying public fears of environmental causes for a wide variety of diseases, particularly cancers. Yet the limitations of the available investigative techniques to point out specific etiologies have led to extensive discussion of the utility of cluster investigations and the circumstances under which they should be conducted, if at all.[1-4] Nonetheless, other authors have also pointed out that cluster investigations may have value for descriptive purposes and for hypothesis generation.[5-7]

The statistical methods available for cluster investigations are numerous, and while many employ morbidity and mortality rates to compare regional occurrences of disease, rate-independent methods operating on the distances between cases in space and time may also prove useful. This may be particularly true in circumstances where the geographic occurrence of a disease can be ascertained, but the calculation of disease rates is not feasible. This is frequently the case in the study of diseases in companion animals, where accurate enumeration of a population-at-risk is often impossible. Similar difficulties may arise in human communities in many parts of the world. Thus, rate-independent methods may have important applications in the study of both animal and human populations.

Reported here are the results of an investigation conducted under such

circumstances, where Point Pattern Analysis (PPA) techniques[8,9] were employed to describe the spatial and temporal distribution of some canine cancers in Michigan during the thirty year period between 1964 and 1994. The study objectives were twofold: 1) to demonstrate the usefulness of a rate-independent method for investigating disease aggregation in companion animals; and, 2) to assess the spatial and temporal clustering of canine cases of some neoplasms that are biologically similar in both dogs and humans. The usefulness of this methodology for generating hypotheses about factors which may help to explain the occurrence of companion animal diseases is also noted.

Materials and Methods

Case Description

The present work was part of a larger exploratory study intended to simultaneously compare the distributions of some selected cancers in dogs and humans living in the same geographic region during the same time period. The canine case records were obtained from the Veterinary Medical Data Base (VMDB). The VMDB, housed at Purdue University, is a computerized system used to archive information on patients examined at the majority of the veterinary colleges in Canada and the United States. The system was used to retrieve the records of all histologically-confirmed cases of neoplasia diagnosed in dogs at the Michigan State University Veterinary Medical Center (MSU-VMC) between March 1, 1964 and March 31, 1994. From those, records pertaining to ten cancers judged to have similar characteristics in both dogs and humans were

chosen for further study. Case records were filtered so that records for benign tumors and metastatic sites were excluded.

The sole geographic reference available for cases through VMDB was zip code of residence at diagnosis. While readily available, zip codes are less than desirable as geographic references,[10,11] since they change over time, and their population sizes are often quite variable.[12] The boundaries of zip codes areas are often not known with accuracy, and changes are not well documented.[12] Moreover, distances computed based on the centroids of such areas are subject to error.[13] To circumvent these difficulties, street addresses were obtained for each case by linking the patient identification number from VMDB with MSU-VMC medical records. Case frequencies by county for each of the ten cancers were mapped (MapViewer, Golden Software, Golden, CO). Due to uneven distribution of (and, in some instances, lack of) cases of the study cancers across the state, spatial analysis was limited to four neoplasms (lymphosarcomas, mammary adenocarcinomas, melanomas and spindle cell sarcomas) occurring in three counties (Ingham, Oakland and Wayne). These three counties encompass a wide range of demographics broadly representative of southern Lower Michigan. Ingham County contains Lansing, a small city which serves as the state capitol, in its northwest corner. The city's economic base has historically been divided between heavy manufacturing associated with the automobile industry, state government offices, and Michigan State University, which lies on the eastern edge of the city. The remainder of Ingham County is rural and agricultural. Oakland County contains Pontiac, an older

industrial city, also associated with the automobile industry. Much of Oakland County is comprised of affluent suburbs of the City of Detroit in its southern reaches, with some rural agricultural areas in the northern part of the county. Most of Wayne County is composed of the City of Detroit, historically a large heavy industrial city that has suffered substantial loss of human population and decline of socioeconomic status in recent years. The western and southern portions of the county are close-in working class suburban areas, which have gradually overtaken the few remaining agricultural parts of the county. The only persistant rural areas are in the extreme southwest. The most affluent areas in the county occupy a narrow strip of land in the northeast portion along the Detroit River and Lake St. Clair.

Figures 3.1-3.4 depict the statewide case distributions of the study cancers, while the study counties are identified in Figure 3.5. In general, the statewide case distributions of the cancers not chosen for spatial analysis closely coincided with those that were. Statewide case distribution maps for the six cancers not subjected to spatial analysis are presented in Appendix A (Figures 5.1-5.6).

Street addresses for cases belonging to this subset of neoplasms and counties were standardized using AccuMail software (Group 1 Software, Inc., Lanham, MD) and geocoded using AtlasGIS software (Environmental Systems Research Institute [ESRI] Inc., Redlands, CA) and the Topologically Integrated Geographic Encoding and Referencing (TIGER) files. Detailed description[14] and examples[12,15] of address geocoding methods have been published previously. In

this way, latitude and longitude coordinates for each dog's residence at the time of diagnosis were obtained. Such coordinates obtained from the TIGER files are accurate to approximately 25 meters of a street address' true location.[13] Residence locations of the cases at the time of diagnosis were plotted on maps of the study counties (MapViewer, Golden Software, Golden, CO).

Previous work suggests that demographic factors influence dog ownership.[16-21] To explore the possible influence of demographic factors on the geographic occurrence of the cases, overlay maps were created to show the relationship between case location and population density, median household income, gender (percent female population), median age and percent homeownership within census designated places. These data were abstracted from the 1970, 1980 and 1990 U.S. Censuses, and divided into and plotted by quartiles on map areas corresponding to those defined by the respective census. Cases were subdivided into three data sets per tumor type prior to plotting, so that the temporal midpoint of the data set into which the cases fell approximately corresponded to a census year. Thus, cases diagnosed between 1964 and 1974 were compared to data from the 1970 census, cases from 1975-1984 were compared to 1980 census data, and so on.

Statistical Analysis

To assess the effects of population, county, and the demographic variables on the number of canine cancers per human population, Poisson regression was conducted (Proc Genmod, SAS Institute, Cary, NC). Pearson

coefficients were generated to assess correlation among the demographic predictors. Data for Wayne and Oakland Counties were grouped together for these analyses since they are both encompassed by the Detroit Metropolitan Area. The log of the human population of each census designated place was used as the offset variable, since accurate dog populations were not available. Analyses were performed only for cases occurring after 1974, due to the unavailability of demographic variables for most of the census designated places for the 1970 census (and the resulting large number of missing data points). Income was scaled as thousands of dollars. Two modeling approaches were taken. First, models were fitted for each tumor type, with county and the four demographic factors included as predictors. Then, each of five first-order interaction terms (county x income, county x percent female population, county x median age, county x percent homeownership, and percent female population x median age) considered to be plausibly important predictors *a priori* were included in these main effects models one at a time. Those interactions that were statistically significant at $p \leq 0.2$ were noted and included in the final models for each tumor type. The intent of this approach was to assess the simultaneous effect of all four demographic factors on the number of cases per human population. Moreover, since it was anticipated that the four demographic factors would be highly correlated, retaining all of them in a single model regardless of their statistical significance was expected to help minimize the possibility of mistakenly dropping important predictors from the models due to the effects of multicollinearity. Second, best-fit models were constructed, which

started with all main effects and the five aforementioned first-order interaction terms included in the models. Type 3 analysis was used to determine which terms would be retained in the final models. The intent of this approach was to produce models that best described the data set under study. It could be argued that this approach is more parsimonious, given that the four demographic factors are known to be associated with pet ownership, but not necessarily with the occurrence of the cancers under study. Thus, including all of them in our models might not be justified and, in fact, introduce collinearity that might increase the possibility of Type II errors. Taken together, the two modeling approaches were expected to provide a more comprehensive view of the significance of the demographic factors in predicting occurrences of the cancers under study.

PPA for Ingham and Oakland Counties was carried out by a Fortran software implementation of second-order analysis. Also referred to as *K* function analysis,[8,9] the technique operates on the interevent distances between all combinations of pairs of mapped points *i,j* within a predesignated search radius of distance *d*. The analysis is termed second-order because of its focus on the variance of the test statistic across a series of progressively larger stepped *d* up to some maximum radius. The size of the steps is set to reveal the interevent distances at which clustering, if present, is strongest, *i.e.*, the scale at which the process determining the pattern of events is operating most strongly. In the present work, the "events" were the geographic locations of the case dogs' street addresses; the cases were analyzed at fifty 100 meter steps *d*, with a maximum

search radius of 5000 meters. For PPA, longitude coordinates were adjusted downward by a factor of 50/69 to reflect the distance encompassed by a degree of longitude/latitude at the latitude of the study counties (42-43°). The *K* function[9] is defined:

$$K(d) = \lambda^{-1} E[\text{number of further events within distance } d \text{ of an arbitrary event}] \quad [1]$$

where λ is the intensity (here, the mean number of cases per unit area) and *E* is the expectation operator. The Fortran implementation tests the null hypothesis that the spatial point process giving rise to the observed geographic pattern of cases is consistent with complete spatial randomness (CSR). The null hypothesis of CSR asserts that the number of events in a planar region *A* with area $|A|$ is Poisson with mean $\lambda|A|$, and given *N* events $\mathbf{x}_i$ in *A*, the $\mathbf{x}_i$ form an independent random sample from the uniform distribution on *A*. Diggle[9] has noted that while CSR may, strictly speaking, be untenable for description of natural phenomena, tests of CSR are used as a means of exploring a data set and as an aid to formulation of hypotheses concerning the pattern of points and the process that generated it. The *K* function test model, based on the CSR formula of Ripley,[22] is:

$$L(d) = \left[\frac{A \sum_{i=1}^{N} \sum_{j=1}^{N} k_{ij}}{\pi N(N-1)} \right]^{1/2} \qquad \text{for } i \neq j \quad [2]$$

where *L(d)* is a linear expression of the expected number of cases *J* occurring within distance *d* of all *I* cases, *A* is the area of the study region, *k* is the *K*

function, $\pi = 3.1415926$ and N is the total number of cases. The expectation of $L(d)$ is d when the null hypothesis of CSR holds. Upon rejection of the null hypothesis, one of two alternatives obtains: 1) $L(d) > d$, indicating clustering or 2) $L(d) < d$, indicating dispersion, *i.e.*, cases demonstrating a more regular pattern than would be expected at random. The statistical significance of the CSR hypothesis can be tested by randomization, as it was in the present study. We generated 99 Poisson process point patterns of size N for each search radius d within each cancer by county analysis combination. When the observed $L(d)$ was more extreme than any of the simulated values at any d, the null hypothesis was considered rejected. Alternatively, Ripley's T statistic[23], can be used to determine significance. Results of K function analysis can be presented graphically, as plots of $L(d)$ vs. d, which show whether the cases are clustered, random or dispersed, as well as the strength of the clustering/dispersion. Plots of the change in $L(d)$ over the range of search radii d show the relative tendency toward clustering/dispersion.

Dates of discharge from hospital were converted to integer format (number of days between 1/1/60 and the discharge date) using the SAS System.[24] Assessment of the temporal distribution of the cases in all three study counties was accomplished using a one-dimensional nearest-neighbor technique.[25] Although the technique was originally applied to evaluate aggregation of points on a line, Selkirk and Neave[25] point out it can also be used to test whether events in time can be assumed to be randomly distributed. As with the spatial analysis, one dimensional analysis of cases in time determines

which of three possible conditions holds: either the cases occur randomly in time, are clustered, or are dispersed. Statistical significance is determined by comparison of the test statistic with percentage points of a normal distribution.

Results

Spatial analysis

Success of address assignment is depicted in Table 3.1. It was possible to obtain geographic coordinates of residence at the time of diagnosis for approximately 97% of all the cases, regardless of tumor type.

Examination of the mapped point patterns of the case dogs' residences suggests two somewhat different underlying processes (Figure 3.6). There was strong aggregation of cases in the northwest portion of Ingham County in the vicinity of the City of Lansing and its suburban areas, corresponding closely with areas of higher population. In the other two study counties, by contrast, the distribution of cases was considerably more diffuse than that in Ingham County. These geographic patterns were, in general, consistent across all four of the cancers under study.

The results of Poisson regression analyses suggest that the ratio of cases to human population varies significantly from county to county. In models which retained all four demographic factors *a priori* (Table 3.2), county was a significant ($p \leq 0.05$) predictor of case frequency in four of the eight tumor by decade models. In all of those instances, the cases were significantly more likely to be diagnosed in Ingham County than in Wayne and Oakland Counties.

No other main effect term was statistically significant in any of the models. A single interaction term, percent female population x county, was significant in predicting the occurrence of lymphosarcomas between 1985 and 1994. Results of the best fit models (Table 3.3) confirm the significance of county in predicting the occurrence of cases. County was a highly significant ($p < 0.01$) predictor for all eight tumor by decade models. In addition, median income was either significant or approached significance at $p = 0.05$ in half of the models, although the direction of the effect was inconsistent. In two of those models, the effect of income on case occurrence varied significantly by county ($p = 0.03$) as well. Percent female population was a significant ($p = 0.02$) negative predictor of the number of canine mammary adenocarcinomas diagnosed between 1985 and 1994. None of the other demographic factors provided significantly improved fit and so were not retained.

As suspected, the demographic predictors were significantly correlated, sometimes highly so. For 1980 census data, Pearson correlations between the four demographic variables ranged between -0.18 and 0.62, while for 1990 census data, correlations varied from -0.29 to 0.72 (data not shown).

The results of the *K* function analyses are presented in Table 3.4. Five of the eight cancer by county analyses found the cases to be significantly clustered geographically at the 99% level of significance. All four cancers were significantly clustered in Ingham County, while only lymphosarcomas were significantly clustered in Oakland County. Two of the five significant analyses suggested clustering that, while statistically significant, was slight.

Lymphosarcomas in Oakland County were clustered significantly only at a *d* of 900 meters, while melanomas in Ingham County displayed significance only at *d* of 200 and 1600 meters. Graphical output of the *K* function analysis confirmed the relative strength of case aggregation suggested by the mapped case patterns (Figure 3.7). Plots of the change in *L(d)* with respect to *d* across the range of search radii suggested that where significant clustering was observed, it took place at distances less than 2000 meters, with the strongest clustering at a scale of 200 meters or less (Figure 3.8).

Temporal analysis

Case frequencies by year of diagnosis for the cancers under study are presented in Figure 3.9. Frequency of diagnosis for lymphosarcomas, mammary adenocarcinomas and melanomas tended to parallel one another over the period of the study, with highs and lows in the number of diagnosed cases occurring during approximately the same years. Peaks of diagnosis for all three occurred in the late 1970s and mid 1980s. In contrast, the diagnosis of cases of spindle cell sarcomas did not parallel the other three cancers as closely, with peak numbers of cases preceding them in 1974 and 1983. Regardless of tumor type, there was a common decline from the peak numbers of cases registered in the eighties to the low numbers recorded in the early 1990s.

Summary results of one-dimensional nearest neighbor temporal analysis are presented in Table 3.5. In contrast to the spatial analysis, it was possible to reject the null hypothesis of a random distribution in only two of the twelve cancer by county analyses. Cases of lymphosarcomas were significantly

clustered in time (p = 0.04) for dogs in Oakland County, as were mammary adenocarcinomas in Wayne County dogs (p = 0.02). Cases of spindle cell sarcoma in Ingham County dogs approached significant temporal dispersion (p = 0.08). When all individual test statistics were considered, lymphosarcomas and mammary adenocarcinomas showed a tendency toward clustering in time, whereas melanomas and spindle cell sarcomas tended toward a pattern of dispersion in time in Ingham and Wayne Counties, but toward temporal clustering in Oakland County.

Discussion

Although the general epidemiological approach to the study of diseases in companion animals has a great deal in common with the approaches used to study human diseases, in some areas of inquiry, unique problems have thus far limited knowledge of factors underlying the occurrence of disease. Tools as fundamental as incidence rates are quite often impossible to apply to populations of cats or dogs, simply because the denominator necessary for calculation of the rate is not known with accuracy. In an attempt to circumvent this, general methods have been developed to estimate the numbers of pets in a community,[16] based on sample surveys of pet ownership. While a useful approximation, these assume that rates of pet ownership vary little within geographic areas as large as states. It is not known whether that assumption is valid. Moreover, evidence suggests that such surveys may overestimate the prevalence of pet ownership.[26] Also, substantial numbers of dogs and cats have

no owners and thus would not be enumerated in surveys of pet ownership; one group of authors estimated such dogs to comprise 14% of their study population.[27] Yet these animals are nonetheless at risk for the development of disease, and so using populations counted in this way to calculate incidence rates seems likely to underestimate the actual number of cases. Another approach that has been successfully used to obtain the denominator necessary for rate calculations has been the development of local pet censuses[17,19,20,26-28] and cancer registries.[29-34] These methods are particularly useful for long term surveillance in well-defined geographic areas. Yet they also necessitate considerable effort and cooperation from owners and/or veterinarians to establish and maintain. In addition, to the extent that local rates of disease are influenced by factors that exhibit substantial geographic or temporal variation, extrapolation of rates derived from these studies to other areas or time periods may be a questionable practice.

Substantial uncertainty also remains with respect to the etiologies of common canine cancers. Of the four types of neoplasia examined in this study, a well-established etiology has been described only for mammary adenocarcinomas, whose occurrence is known to be strongly related to endocrine factors.[35-37] While nutrition,[38] genetic factors,[31,39] radiation,[40-43] immunologic factors,[44,45] viruses,[46-52] parasites,[53] surgical implants[54] and pesticide exposure[55] have all been investigated as causes for the various cancers under study here, many questions remain with respect to the pathogenesis of these diseases. Thus, there remains a need for techniques

which can be generally applied to determine aggregation of disease in companion animals without the necessity of calculating rates, and can be easily used to generate hypotheses concerning the distribution and causes of disease at the population level. As demonstrated here, PPA techniques can be applied to accomplish both of these tasks. Moreover, these techniques are not limited to the study of cancer in dogs; they can be used with equal utility to study non-neoplastic diseases, or human diseases for which a population-at-risk, and so disease rates, cannot be determined.

In this study, spatial patterns of canine cancer showed variation by county, by tumor, and over time. The most remarkable positive finding was the highly significant ($p < 0.01$) clustering of all four study cancers in Ingham County. Of equal interest was the general lack of any clustering in Oakland County that paralleled concentrations of human population. Overlay maps of demographic factors suggest that the significant aggregation in Ingham County could be described as an urban versus rural phenomenon, with cases occurring in areas of relatively high human population (Figure 3.10) and largely invariant to the influence of income, age, gender or homeownership. Previous studies in humans have also noted associations between cancer occurrence and population density.[56-58] This case concentration occurred despite the fact that people living in rural areas are more likely to own dogs than their urban counterparts.[16,17,20,21] In contrast, three of four of the cancers under study were distributed randomly in Oakland County. Overlay maps suggested that, of the demographic factors considered, the Oakland cases corresponded most closely

to areas of higher median household income, especially as time progressed through the study period. This geographic pattern (Figure 3.11) approximates a "concentric ring" configuration, consistent with other geographic phenomena occurring around large industrial cities. The Ingham cases did not exhibit a similar relationship with income. Regression analyses largely confirmed the relationships suggested from mapping. Irrespective of whether models containing all demographic factors or best-fit models were considered, there was little consistent evidence that any of the demographic factors had a significant effect on the number of these cancers per human population. County, in contrast, was either a significant predictor or approached significance in most or all of the models.

In light of these findings, it appears that the spatial distribution of factors associated with pet ownership does not necessarily explain the spatial occurrence of these cancers. Moreover, the expectation that the number of dog cancer cases will always be higher in areas of high human population appears not to hold consistently either. Thus, while the cancers displaying the strongest clustering in this study did coincide with areas of higher human population, the statistically significant clustering of lymphosarcomas in Oakland County dogs suggests that significant clusters of canine cancers cannot always be assumed to be an artifact of high human population. Most importantly, it is clear that the underlying processes that determined the aggregation of cases in this study were not acting uniformly across counties or over time.

What processes, then, determine the spatial distribution of these

diseases in those circumstances where they do not coincide with human demographics or high human population? While the ability to determine what factors are causal is beyond the scope of this study, some hypotheses can be formulated. First, geographic distance from the referral institution may play a role. If travel distance is a significant determinant of whether or not case workup is pursued by the dog's owner, then one might expect more cases to be diagnosed in areas close to the referral institution. This may partially explain the apparent differential effects of human population and household income by county in this study. If a client lives close to the referral institution, she may be more willing to pursue a definitive diagnosis regardless of income, whereas those of limited means living further away may perceive the expense of a referral to be prohibitive. The use of study populations of dogs diagnosed at a larger number of hospitals (*e.g.*, biopsy submissions from private practitioners to pathology laboratories) might help to address the plausibility of this hypothesis. Second, the population density of dogs may differ sufficiently from that of humans such that even a uniform geographic incidence of cases in dogs would not be identified by comparing their locations with human density. Without means of accurately determining dog population numbers, this hypothesis could prove difficult to investigate. Third, there may be substantial geographic heterogeneity with respect to the age of the dog population. As in humans, neoplasia in dogs is generally a disease of older individuals.[59-62] To the extent that the canine population in particular areas is composed predominantly of older dogs, cancer cases could appear clustered in those areas. Studies of

interarea variation in age distribution of the dog population might test this hypothesis. Finally, it is possible that the spatial aggregation of cancer cases is not entirely an artifact of any population or demographic factor, but rather is determined by local environmental risk factors which play significant causal roles in the pathogenesis of the diseases. The fact that clustering was strongest at short distances (< 200 meters) in this study is not inconsistent with such a hypothesis. While this is perhaps the most fascinating hypothesis of all, it seems likely to be the most elusive to investigate as well.

Examining the distribution of the study cancers in time, it is interesting that the highs and lows in the number of cases diagnosed were roughly parallel for some, but not all, of them. Whatever factors influence the diagnosis of these diseases in time apparently do not act uniformly across all canine neoplasms, but neither are they entirely independent. The steep decline in the number of cases diagnosed in the early 1990s was a consistent and prominent characteristic of all cancers under study. Such a dramatic drop in incident cases seems unlikely to have actually occurred, and may be an artifact of frequency of diagnosis. The decline roughly coincided with opening of private referral hospitals in the Detroit Metropolitan Area, and one can hypothesize that some of the cases that would otherwise have been diagnosed at MSU-VMC may have gone elsewhere. Again, studies of populations diagnosed at a larger number of institutions might shed light on this.

The results of the analysis of temporal aggregation challenge interpretation. In the case of lymphosarcomas and mammary adenocarcinomas,

which, in general, showed a tendency to cluster in time, one can hypothesize a number of causes. A common exposure to an infectious agent or to a point source of a carcinogenic agent could be consistent with temporal clustering, (and indeed, with the strong short distance spatial clustering noted for those cancers as well). Non-etiological factors bearing on the likelihood of diagnosis could also elicit such a pattern. For example, arrival of a practitioner with expertise in oncology, or availability of a specific diagnostic technique, may make it more probable that previously unknown cases of disease will be diagnosed. With respect to melanomas and spindle cell sarcomas however, which showed a general tendency toward a dispersed pattern, explanations are less obvious. Dispersion implies that when one case occurs, it inhibits the occurrence of other cases so that a regular pattern results over time. Neither etiological nor other factors come to mind that provide a straightforward explanation for such a tendency. Interestingly, only one cancer (lymphosarcomas in Oakland County) which exhibited significant spatial clustering was also significantly clustered in time (Tables 3.3 & 3.4). Given the variability of latency periods associated with cancer, this may not be unexpected.

Some limitations inherent to these techniques must be kept in mind, limitations which dictate that these results be considered exploratory. For example, it was mentioned earlier that two of the five *K* function analyses suggested clustering that while statistically significant at the 99% level, should be considered slight. This may appear counterintuitive. However, it has been

argued that using a number of search radii *d*, rather than specifying a single *d* based on some known characteristic of the spatial process under study, amounts to testing multiple dependent hypotheses. Accepting that, one would expect a certain small percentage of *L(d)* to show rejection of the null hypothesis even when CSR holds. Discussion of this topic is beyond the scope of this paper, but is treated elsewhere.[9, p. 12] It will suffice for purposes of this study to reiterate that rejection of CSR at multiple (rather than one or two) *d* presents a much stronger case for clustering (Figure 3.7).

The present study found no apparent relationship between the spatial occurrence of the cases and demographic factors such as the age or gender of humans in the same areas. However, because the age and gender of the cases' owners were unavailable, these variables were expressed as median age and proportion of female population for relatively large places defined by the U.S. Census. It is possible that a relationship which was not detected could become apparent with the use of less aggregated variables. Issues concerning aggregation bias have been discussed in detail by Diggle and Elliot.[63] Moreover, the present regression results used human population as an offset variable for the canine cases, so to the extent that dog population density does not parallel human population density, the validity of those results could be affected. It is also important to note that this analysis does not have the ability to distinguish between etiological factors that may determine the spatial occurrence of cancer and other factors that may simply be affecting the likelihood of diagnosis. In this study, all cases were diagnosed at a teaching

hospital, and so factors which may have affected the likelihood of referral from a general practitioner to MSU-VMC may have been the forces actually driving the spatial pattern. Nonetheless, even non-etiological information gained from this work could prove to be of value, for instance, for studying the utilization of veterinary health care resources and access to care. This demonstrates the broad utility of *K* function analysis to generate hypotheses, biological or otherwise.

In summary, this study used GIS and rate-independent, distance-based PPA methods to assess the clustering of incident cases of four types of canine cancer in three counties in Michigan between 1964 and 1994. The chief purpose of the study was to demonstrate the utility of these techniques to assess and generate hypotheses about disease aggregation in situations where a population-at-risk cannot be enumerated. Significant spatial clustering was found that varied by county and type of cancer. Temporal patterns concurred only minimally with spatial ones, and varied from clustering to dispersion depending on the cancer under study.

With an eye toward future investigations, it may be worthwhile here to restate briefly the hypotheses generated from this approach, and to provide some comments on the utility of the analysis. First, these results suggest human population concentration may be an inadequate surrogate measure of the geographic distribution of cancers in dogs, depending on the area under study. This runs counter to what has traditionally been assumed. Second, within the limitations imposed by the study design, demographic factors associated with

pet ownership appear to be poor predictors of which geographic areas will yield aggregations of cancer cases. Third, as a corollary hypothesis of these first two, it appears factors other than high human population and pet ownership demographics play a significant role in determining the spatial distribution of canine cancers. Some of these factors worthy of investigative research include: 1) distance of the case's residence from a veterinarian capable of oncological diagnosis and case workup, 2) variation in canine population density as compared to human population density, 3) local variations in the age distribution of the dog population and 4) local environmental factors. The first of these factors might be investigated by repeating this study design using a population of dogs diagnosed at many hospitals (for example, using a data base of histopathologic accessions mailed to a central diagnostic laboratory by private practitioners). Alternatively, one could survey or interview owners of canine oncology patients to ascertain what factors figured into their decisions to pursue diagnosis for their dogs. The second and third factors could be approached with geographically-referent sample surveys of the characteristics of owners and their dogs which recorded data at the household level (rather than pooling data to the county or state level as has been done previously). With respect to factor number 4, studies of the influence of local environmental factors on cancer in dogs might best be accomplished with case-control studies carefully designed to minimize recall biases. The case-control method is arguably the best epidemiological tool currently available for identifying specific cancer risk factors following the identification of significant case clustering. The work of Hayes and

coworkers[55] provides an excellent example of how such a study can be successfully approached. Such studies would need to account in some way for the mobility of dogs and their owners. That is, residence addresses of the cases would need to be followed throughout the animals' lives, not just at the time of diagnosis (as measured here), in order to attribute the causal exposure for each case to a specific study area. Finally, ideally, it might be of interest, funding permitting, to attempt sample surveys to identify canine populations-at-risk in Ingham and Oakland Counties, so that incidence rates could be calculated for the cancers under study here and compared with the results obtained using these rate-independent methods.

Regarding the temporal analysis, there was little evidence in this study to support a hypothesis of significant case clustering in time, or the coincidence of spatial and temporal clustering of the same cancer in the same county. This, coupled with the finding that fluctuations in the number of cases diagnosed over time tended to coincide for three of the four cancers studied suggests the hypothesis that aggregation of cases in time may be due to non-etiological factors. Finally, this author cannot propose a hypothesis for why melanomas and spindle cell sarcomas tended toward dispersion over time, but the concept of some process causing cases to occur only at regular time intervals certainly warrants further work.

Overall, there is, in the author's view, considerable reason to be satisfied with the demostrated success of the analysis approach taken in this study. Without question, uncertainties remain. Yet, they are arguably of no greater

magnitude than the substantial uncertainties accompanying rate-based methods (as discussed in Chapter 1), and the present study was carried out at considerably less cost in terms of both research support and labor expended. Moreover, these methods are easily applicable over diverse areas. These methods have proven capable of providing accurate quantitative spatial description of cancer diagnoses in companion animals without enumeration of populations-at-risk. Experience gained in the present study suggests that they may be most suited to initial exploratory inquiries aimed at examining the general spatial distributions of diseases in populations of animals or humans where populations-at-risk cannot be accurately or easily ascertained. The descriptive information and hypotheses they generate can then be further investigated using other research designs intended to identify causal factors for the diseases under study.

REFERENCES

REFERENCES

1. Elliott P, Martuzzi M, Shaddick G. Spatial statistical methods in environmental epidemiology: a critique. Stat Meth Med Res 1995;4:137-59.

2. Hole DJ, Lamont DW. Problems in the interpretation of small area analysis of epidemiological data: the case of cancer incidence in the West of Scotland. J Epidemiol Comm Health 1992;46:305-10.

3. Neutra R, Swan S, Mack T. Clusters galore: insight about environmental clusters from probability theory. Sci Tot Environ 1992;127:187-200.

4. Rothman KJ. A sobering start to the cluster busters' conference. Am J Epidemiol 1990;132 Suppl 1:S6-S13.

5. Beral V. Childhood leukemia near nuclear plants in the United Kingdom: the evolution of a systematic approach to studying rare disease in small geographic areas. Am J Epidemiol 1990;132 Suppl 1:S63-S68.

6. Olsen SF, Martuzzi M, Elliott P. Cluster analysis and disease mapping—why, when, and how? A step by step guide. BMJ 1996;313:863-6.

7. Waller LA, Jacquez GM. Disease models implicit in statistical tests of disease clustering. Epidemiology 1995;6:584-90.

8. Boots BN, Getis A. Point Pattern Analysis. Newbury Park (CA):Sage Publications; 1988.

9. Diggle PJ. Statistical Analysis of Spatial Point Patterns. New York: Academic Press; 1983.

10. Croner CM, Sperling J, Broome FR. Geographic Information Systems (GIS): New perspectives in understanding human health and environmental relationships. Stat Med 1996;15:1961-77.

11. Sperling J. Considerations for linking ZIP codes and other small areas using TIGER. Hyattsville (MD): National Center for Health Statistics (US) Cartography Guest Lecture Series, 1993.

12. Rushton G, Lolonis P. Exploratory spatial analysis of birth defect rates in

an urban population. Stat Med 1996;15(7-9):717-26.

13. Rushton G. Describing and demonstrating the analytical applications of GIS. Lansing (MI): Michigan Department of Public Health Workshop; 1996.

14. Vine MF, Degnan D, Hanchette C. Geographic Information Systems: Their use in environmental epidemiologic research. Environ Health Perspect 1997;105(6):598-605.

15. Tobias RA, Roy R, Alo CJ, Howe HL. Tracking human health statistics in "Radium City". Geo Info Systems 1996;(July):50-3.

16. Wise JK. U.S pet ownership & demographics sourcebook. Schaumburg (IL): Center for Information Management, American Veterinary Medical Association; 1997.

17. Leslie BE, Meek AH, Kawash GF, McKeown DB. An epidemiological investigation of pet ownership in Ontario. Can Vet J 1994;35(4):218-22.

18. Wise JK, Yang JJ. Veterinary service market for companion animals, 1992. Part I: Companion animal ownership and demographics. J Am Vet Med Assoc 1992;201(7):990-2.

19. Franti CE, Kraus JF, Borhani NO, Johnson SL, Tucker SD. Pet ownership in rural Northern California (El Dorado County). J Am Vet Med Assoc 1980;176(2):143-9.

20. Griffiths AO, Brenner A. Survey of cat and dog ownership in Champaign County, Illinois, 1976. J Am Vet Med Assoc 1977;170(11):1333-40.

21. Franti CE, Kraus JF, Borhani NO. Pet ownership in a suburban-rural area of California, 1970. Public Health Rep 1974;89(5):473-84.

22. Ripley BD. Spatial Statistics. New York: John Wiley & Sons; 1981. p. 159-60.

23. Ripley BD. The analysis of geographical maps. In: Bartels CPA, Ketellapper RH, editors. Exploratory and explanatory statistical analysis of spatial data. Boston: Martinus Nijhoff; 1979. p. 53-72.

24. SAS Institute Inc. SAS date and time values. In: SAS Language: Reference, Version 6, 1st ed. Cary (NC): SAS Institute Inc.; 1990. p. 128-31.

25. Selkirk KE, Neave HR. Nearest-neighbour analysis of one-dimensional distributions of points. Tijdschr Econ Soc Geogr 75(5):356-62.

26. Patronek GJ, Beck AM, Glickman LT. Dynamics of dog and cat populations in a community. J Am Vet Med Assoc 1997;210(5):637-42.

27. Heussner JC, Flowers AI, Williams JD, Silvy NJ. Estimating dog and cat populations in an urban area. Anim Regulation Studies 1978;1:203-12.

28. Schneider R, Vaida ML. Survey of canine and feline populations: Alameda and Contra Costa Counties, California, 1970. J Am Vet Med Assoc 1975;166(5):481-6.

29. Lengerich EJ, Teclaw RF, Mendlein JM, Mariolis P, Garbe PL. Pet populations in the catchment area of the Purdue Comparative Oncology Program. J Am Vet Med Assoc 1992;200(1):51-6.

30. Schneider R. Comparison of age- and sex-specific incidence rate patterns of the leukemia complex in the cat and the dog. J Natl Cancer Inst 1983;70(5):971-7.

31. Dorn CR, Schneider R. Inbreeding and canine mammary cancer: a retrospective study. J Natl Cancer Inst 1976;57(3):545-8.

32. Dorn CR, Taylor DO, Frye FL, Hibbard HH. Survey of animal neoplasms in Alameda and Contra Costa Counties, California. I. Methodology and description of cases. J Natl Cancer Inst 1968;40(2):295-305.

33. Dorn CR, Taylor DO, Schneider R, Hibbard HH, Klauber MR. Survey of animal neoplasms in Alameda and Contra Costa Counties, California. II. Cancer morbidity in dogs and cats from Alameda County. J Natl Cancer Inst 1968;40(2):307-18.

34. MacVean DW, Monlux AW, Anderson PS Jr, Silberg SL, Roszel JF. Frequency of canine and feline tumors in a defined population. Vet Pathol 1978;15(6):700-15.

35. MacEwen EG, Withrow KM, editors. Tumors of the mammary gland. In: Small animal clinical oncology. 2nd ed. Philadelphia: WB Saunders; 1996. p. 356-72.

36. Madewell BR, Theilen GH, editors. Tumors of the mammary gland. Canine mammary neoplasms. In: Veterinary cancer medicine. 2nd ed.

Philadelphia: Lea & Febiger; 1987. p. 328-37.

37. Moulton JE, editor. Tumors of the mammary gland. Mammary tumors of the dog. In: Tumors in domestic animals. 3rd ed. Berkeley (CA): University of California Press; 1990. p. 518-46.

38. Sonnenschein EG, Glickman LT, Goldschmidt MH, McKee LJ. Body conformation, diet, and risk of breast cancer in pet dogs: a case-control study. Am J Epidemiol 1991;133(7):694-703.

39. Onions DE. A prospective survey of familial canine lymphosarcoma. J Natl Cancer Inst 1984;72(4):909-12.

40. Hoekstra HJ, Sindelar WF, Kinsella TJ, Mehta DM. Intraoperative radiation therapy-induced sarcomas in dogs. Radiat Res 1989;120(3):508-15.

41. White RA, Jefferies AR, Gorman NT. Sarcoma development following irradiation of acanthomatous epulis in two dogs. Vet Rec 1986;118(24):668.

42. Bruenger FW, Lloyd RD, Miller SC, Taylor GN, Angus W, Huth DA. Occurrence of mammary tumors in beagles given radium-226. Radiat Res 1994;138(3):423-34.

43. Moulton JE, Rosenblatt LS, Goldman M. Mammary tumors in a colony of beagle dogs. Vet Pathol 1986;23(6):741-9.

44. Gordon BR, Moroff S, Hurvitz AL, Matus RE, MacEwen EG, Good RA, et al. Circulating immune complexes in sera of dogs with benign and malignant breast disease. Cancer Res 1980;40(10):3627-31.

45. Keller ET. Immune-mediated disease as a risk factor for canine lymphoma. Cancer 1992;70(9):2334-7.

46. Hardy WD Jr. Naturally occurring retroviruses (RNA tumor viruses). II. Cancer Invest 1983;1(2):163-74.

47. Albert DM. The association of viruses with uveal melanoma. Trans Am Opthalmol Soc 1979;77:367-421.

48. Frese K. Virus-like particles in metastases of malignant oral melanomas in dogs. Naturwissenschaften 1973;60(1):55.

49. Kasza L, Hayward AH, Betts AO. Isolation of a virus from a cat sarcoma in an established canine melanoma cell line. Res Vet Sci 1969;10(2):216-8.

50. Winters WD. Human adenovirus antibody in sera of normal and tumour-bearing dogs. Vet Rec 1979;105(10):216-20.

51. Kakuk TJ, Conner GH, Langham RF, Moore JA, Mitchell JR. Isolation of a canine herpesvirus from a dog with malignant lymphoma. Am J Vet Res 1969;30(11):1951-60.

52. Teske E, de Vos JP, Egberink HF, Vos JH. Clustering in canine malignant lymphoma. Vet Q 1994;119(23):705-17.

53. Stephens LC, Gleiser CA, Jardine JH. Primary pulmonary fibrosarcoma associated with Spirocerca lupi infection in a dog with hypertrophic pulmonary osterarthropathy. J Am Vet Med Assoc 1983;182(5):496-8.

54. Sinibaldi K, Rosen H, Liu SK, DeAngelis M. Tumors associated with metallic implants in animals. Clin Orthop 1976;118(Jul-Aug):257-66.

55. Hayes HM, Tarone RE, Cantor KP, Jessen CR, McCurnin DM, Richardson RC. Case-control study of canine malignant lymphoma: positive association with dog owner's use of 2,4-dichlorophenoxyacetic acid herbicides. J Natl Cancer Inst 1991;83(17):1226-31.

56. Nasca PC, Burnett WS, Greenwald P, Brennan K, Wolfgang P, Carlton K. Population density as an indicator of urban-rural difference in cancer incidence, upstate New York, 1968-1972. Am J Epidemiol 1980;112(3):362-75.

57. Bako G, Dewar R, Hanson J, Hill G. Population density as an indicator of urban-rural differences in cancer incidence, Alberta, Canada, 1969-73. Can J Public Health 1984;75(2):152-6.

58. Mahoney MC, LaBrie DS, Nasca PC, Wolfgang PE, Burnett WS. Population density and cancer mortality differentials in New York State, 1978-1982. Int J Epidemiol 1990;19(3):483-90.

59. Moulton JE, editor. Tumors in domestic animals. 3rd ed. Berkeley (CA): University of California Press; 1990:672 pp.

60. Theilen GH, Madewell BR, editors. Veterinary cancer medicine. 2nd ed. Philadelphia: Lea & Febiger; 1987:676 pp.

61. Withrow SJ, MacEwen EG, editors. Clinical veterinary oncology. Philadelphia : Lippincott; 1989:507 pp.

62. Withrow SJ, MacEwen EG; editors. Small animal clinical oncology. 2nd ed. Philadelphia: WB Saunders; 1996:589 pp.

63. Diggle P, Elliott P. Disease risk near point sources: statistical issues for analyses using individual or spatially aggregated data. J Epidemiol Comm Health 1995;49 Suppl 2:S20-S27.

Table 3.1. Outcome of address assignment/geocoding by case frequency (%)

Cancer	Unsuccessful		Successful	Total
	No/Incomplete Street Address	Address Could Not Be Coded		
Lymphosarcoma	4 (1.8)	2 (0.9)	221 (97.3)	227 (100)
Mammary Adenocarcinoma		2 (2.1)	96 (97.9)	98 (100)
Melanoma	2 (2.7)		72 (97.3)	74 (100)
Spindle Cell Sarcoma	2 (1.6)	2 (1.6)	125 (96.8)	129 (100)

Table 3.2. Results of Poisson regression analysis of demographic factors on case frequencies*

Parameter	Parameter estimate (*p*-value)							
	Lymphosarcoma		Mammary Adenocarcinoma		Melanoma		Spindle Cell Sarcoma	
	1975-1984	1985-1994	1975-1984	1985-1994	1975-1984	1985-1994	1975-1984	1985-1994
Intercept	11.48	60.53	0.57	8.83	-10.34	2.46	-7.65	-108.50
County (Wayne/Oakland relative to Ingham)	-10.42 (0.51)	21.40 (0.12)	-8.16 (0.02)	0.42 (0.42)	-8.43 (0.05)	-5.53 (0.02)	-7.43 (0.09)	-1.50 (<0.01)
Median Income	0.12 (0.16)	-0.12 (0.06)	0.009 (0.81)	-0.04 (0.11)	0.05 (0.12)	-0.02 (0.51)	-0.29 (0.3)	-0.005 (0.81)
% Female Population	-0.45 (0.08)	-1.31 (0.20)	-0.17 (0.46)	-0.40 (0.22)	0.09 (0.73)	-0.23 (0.38)	-0.02 (0.93)	1.87 (0.28)
Median Age	0.03 (0.57)	-2.34 (0.09)	-0.02 (0.90)	0.21 (0.08)	-0.22 (0.24)	0.09 (0.44)	0.26 (0.35)	3.50 (0.18)
% Homeownership	-0.02 (0.08)	0.03 (0.34)	-0.01 (0.51)	-0.06 (0.13)	-0.002 (0.92)	-0.05 (0.17)	-0.05 (0.33)	-0.004 (0.89)
Income x County	-0.06 (0.45)	0.10 (0.15)					0.31 (0.27)	
% Female Population x County	0.20 (0.49)	-0.59 (0.05)		-0.45 (0.22)				

Table 3.2 (cont'd)

Parameter	Parameter estimate (*p*-value)			
	Lymphosarcoma	Mammary Adenocarcinoma	Melanoma	Spindle Cell Sarcoma
Median Age x County	0.20 (0.22)	0.18 (0.13)	0.25 (0.12)	-0.15 (0.56)
% Homeownership x County	-0.03 (0.43)	0.07 (0.10)	0.06 (0.10)	0.06 (0.26)
% Female Population x Median Age	0.05 (0.10)			-0.07 (0.19)

*Empty cells denote interaction terms for which the parameter estimates were not statistically significant at $p \leq 0.20$ in a model fitting the main effects and that lone interaction term.

Table 3.3. Results of Poisson regression analysis of demographic factors on case frequencies, best fit models

Parameter	Parameter estimate (*p*-value)							
	Lymphosarcoma		Mammary Adenocarcinoma		Melanoma		Spindle Cell Sarcoma	
	1975-1984	1985-1994	1975-1984	1985-1994	1975-1984	1985-1994	1975-1984	1985-1994
Intercept	-10.85	-7.73	-9.46	8.17	-11.17	-10.57	-5.22	-10.69
County (Wayne/Oakland relative to Ingham)	-1.47 (<0.01)	-4.24 (<0.01)	-2.65 (<0.01)	-2.24 (<0.01)	-1.84 (<0.01)	-1.65 (<0.01)	-7.84 (<0.01)	-1.32 (<0.01)
Median Income	0.06 (<0.01)	-0.07 (0.08)			0.05 (0.01)		-0.28 (0.07)	
% Female Population				-0.36 (0.02)				
Income x County		0.083 (0.03)					0.34 (0.03)	

Table 3.4. Results of *K* function spatial analyses

Neoplasm	Ingham County	Oakland County
Lymphosarcoma	Clustered*	Clustered*
Mammary Adenocarcinoma	Clustered*	Random
Melanoma	Clustered*	Random
Spindle Cell Sarcoma	Clustered*	Random
		*$p \leq 0.01$

Table 3.5. Results of one-dimensional nearest-neighbor temporal analyses

Neoplasm	Ingham County	Oakland County	Wayne County
Lymphosarcoma	Random ($p = 0.26$)	Clustered ($p = 0.04$)	Random ($p = 0.92$)
Mammary Adenocarcinoma	Random ($p = 0.88$)	Random ($p \approx 0.64$)*	Clustered ($p = 0.02$)
Melanoma	Random ($p = 0.78$)	Random ($p = 0.54$)	Random ($p = 0.22$)
Spindle Cell Sarcoma	Random ($p = 0.08$)	Random ($p = 0.44$)	Random ($p = 0.48$)

*p-value is approximate only, due to small number of cases (16).

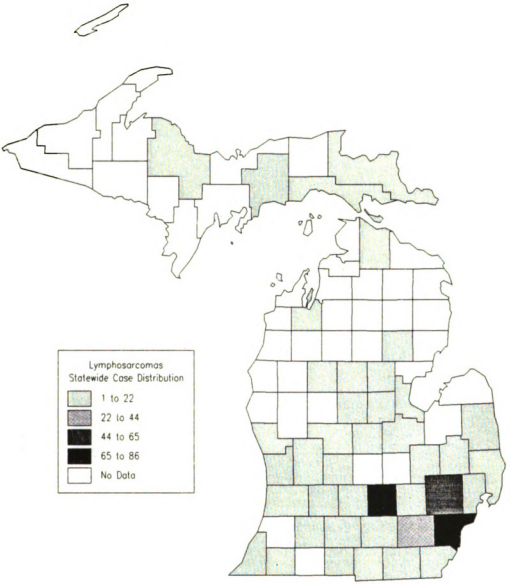


Figure 3.1. Case frequencies by county, lymphosarcoma, Michigan, 3/1/64 through 3/31/94

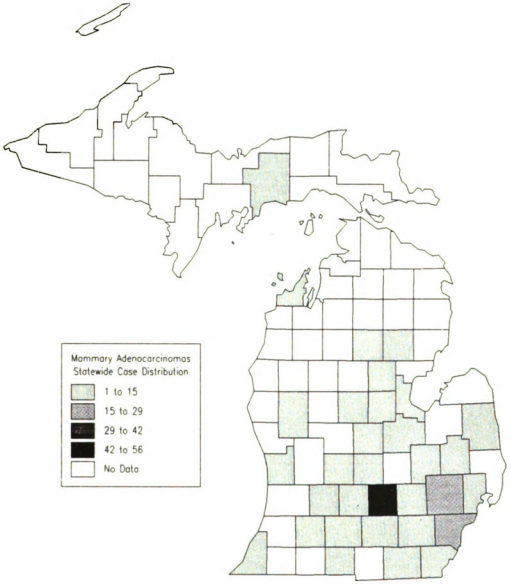


Figure 3.2. Case frequencies by county, mammary adenocarcinoma, Michigan, 3/1/64 through 3/31/94

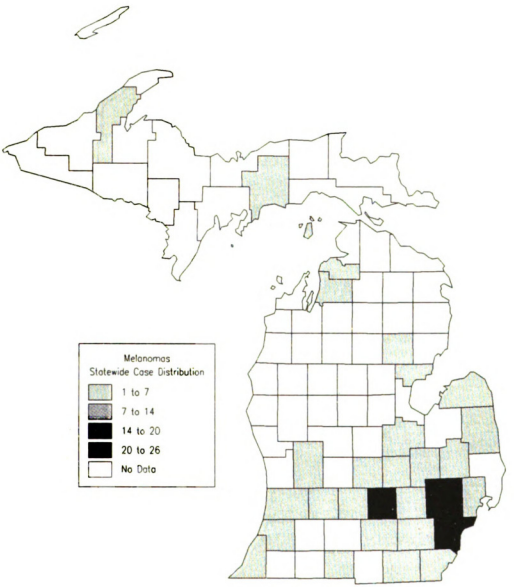


Figure 3.3. Case frequencies by county, melanoma, Michigan, 3/1/64 through 3/31/94

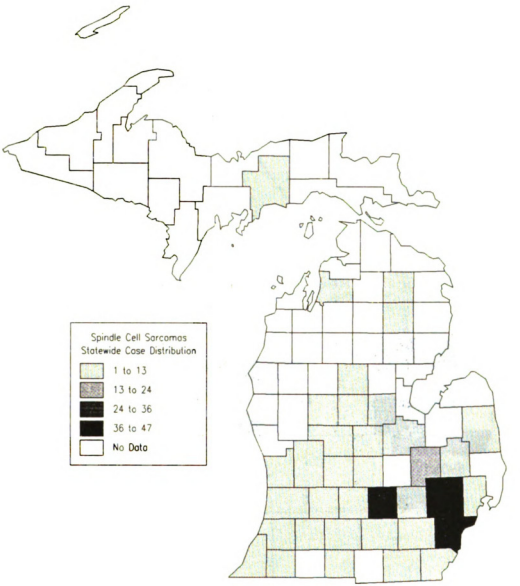


Figure 3.4. Case frequencies by county, spindle cell sarcoma, Michigan, 3/1/64 through 3/31/94

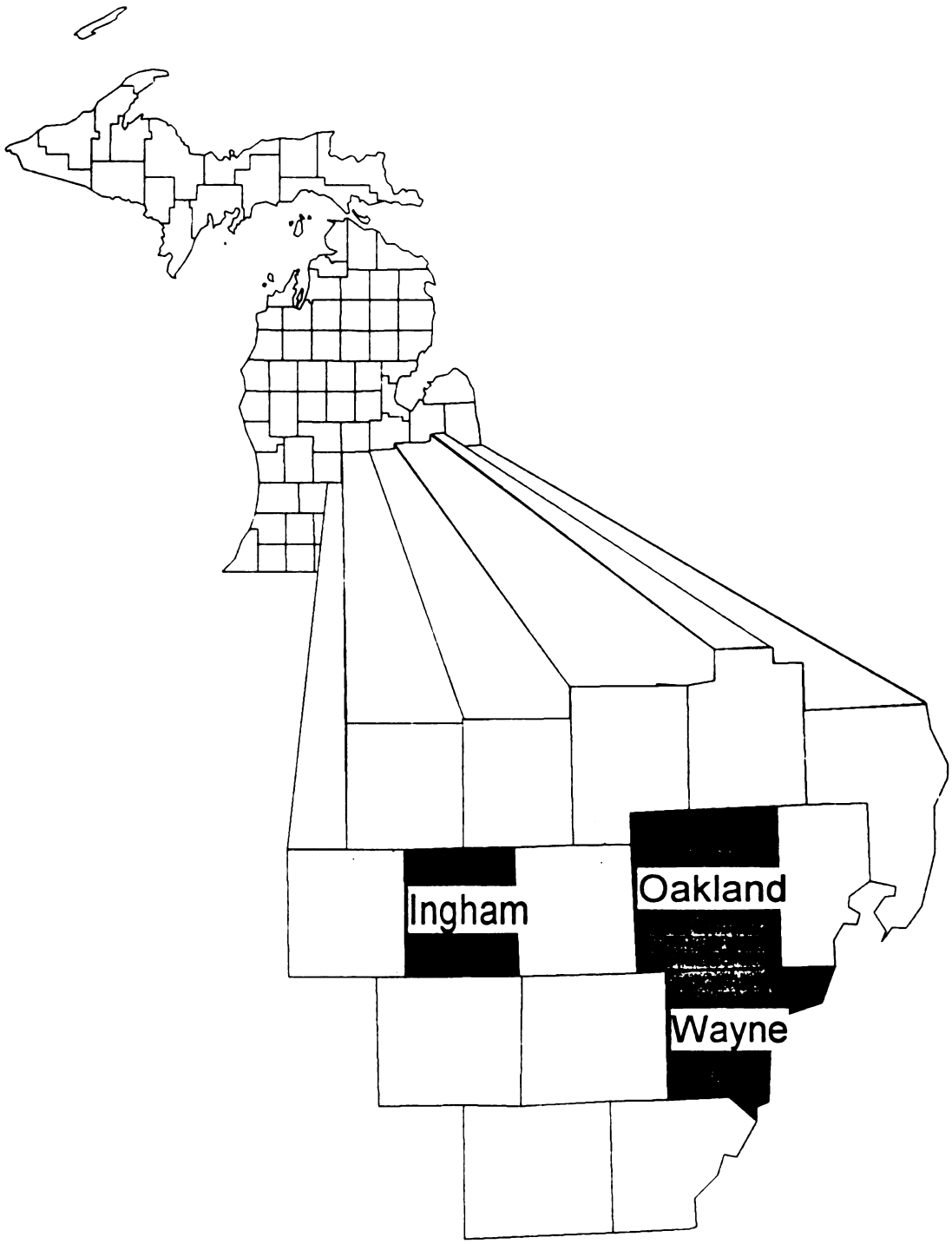


Figure 3.5. Area of study: Ingham, Oakland and Wayne Counties, Michigan, U.S.A.

Figure 3.6. Distribution of canine lymphosarcoma in Ingham, Oakland and Wayne Counties, Michigan, 1964 - 1994. Open circles mark locations of cases' residences at the time of diagnosis (Analogous maps for the other three cancers under study appear as Figures 5.7-5.9 in Appendix A).

Figure 3.6.

Figure 3.7. *K* function analysis graphical output, lymphosarcomas, Ingham (top) and Oakland Counties. Results for case data are depicted by closed circles joined by the solid line. Dotted lines delineate the 99% confidence bounds on the null hypothesis of CSR. Any case data plotting outside the confidence bounds denote rejection of the null hypothesis; results plotting above/left denote significant clustering of cases, while results plotting below/right denote significant dispersion of cases. The further case data plot away from the confidence bounds, the stronger the clustering/dispersion of the cases. Note the substantially weaker clustering of cases in Oakland County, despite statistically significant clustering at 900 meters (Analogous plots for the other three cancers under study appear as Figures 5.10-5.15 in Appendix A).

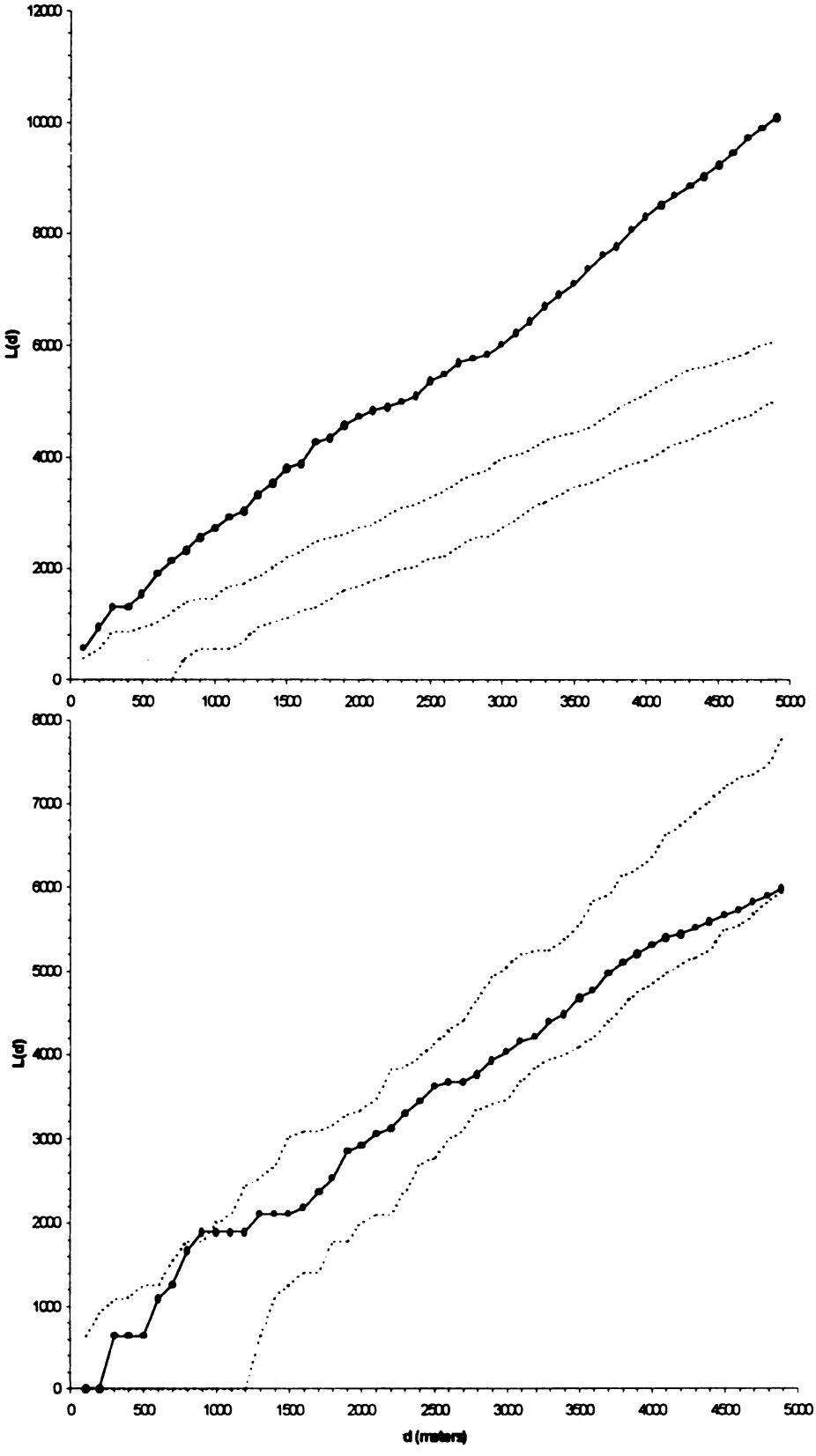
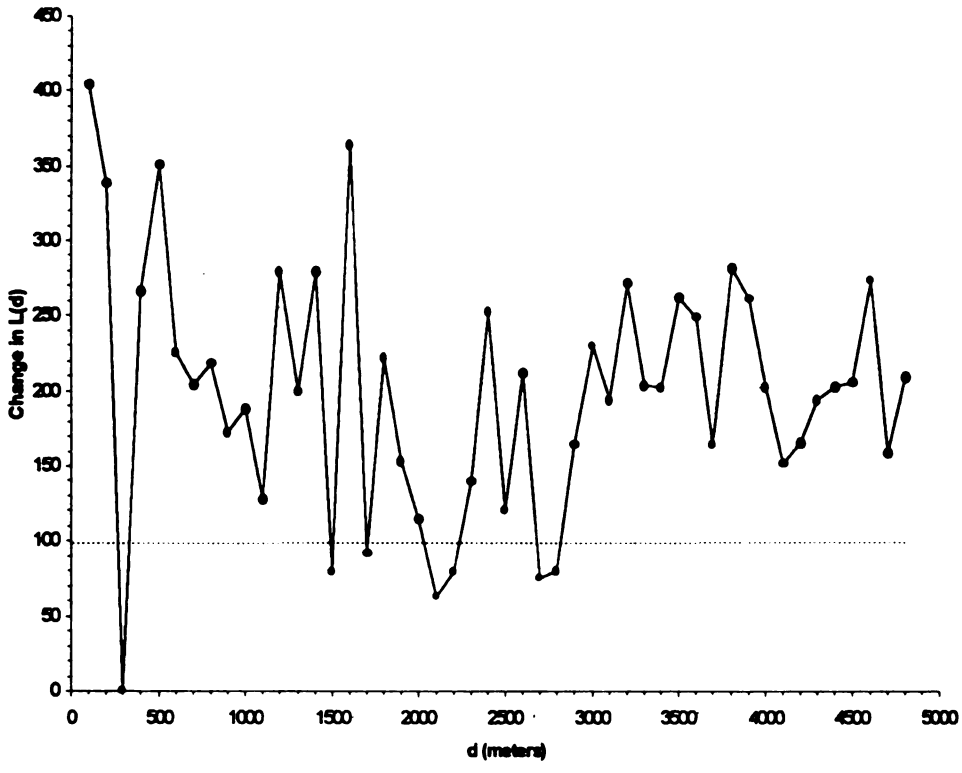


Figure 3.7.

Figure 3.8. Plots of $\Delta L(d)$, lymphosarcomas, Ingham (top) and Oakland Counties. Results for case data are depicted by closed circles joined by the solid line. Dotted lines denote the expected change in $L(d)$ under the null hypothesis of CSR. Changes in $L(d)$ in excess of 100 as the step distance d increases by 100 suggest a tendency toward clustering, with larger changes signifying stronger clustering. Note the strongest tendency for clustering of cases was at distances less than 200 meters (Analogous plots for the other three cancers under study appear as Figures 5.16-5.21 in Appendix A).


Change in L(d)
d (meters)
0
50
100
150
200
250
300
350
400
450
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000

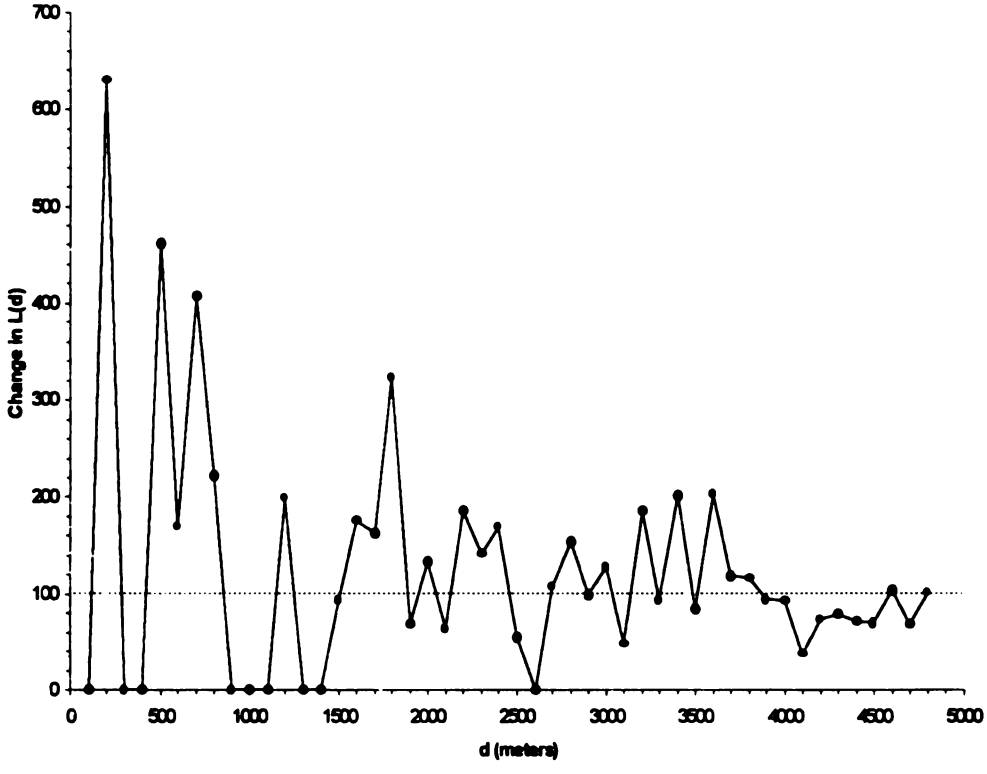

Change in L(d)
d (meters)
0
100
200
300
400
500
600
700
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000

Figure 3.8.

Figure 3.9. Case frequencies by year. In the top plot, open bars represent lymphosarcomas and black bars represent mammary adenocarcinomas, while in the bottom plot, open bars represent melanomas and black bars represent spindle cell sarcomas.

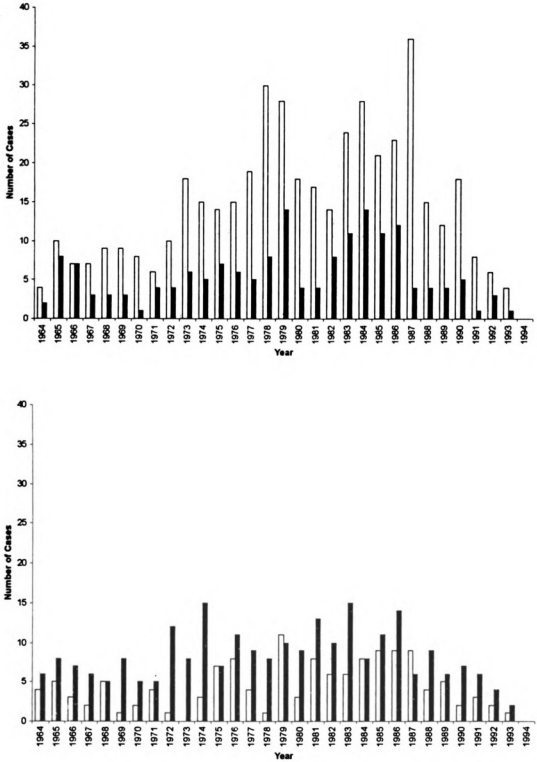

Number of Cases
0
5
10
15
20
25
30
35
40
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
Year
Number of Cases
0
5
10
15
20
25
30
35
40
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
Year

Figure 3.9.

Figure 3.10. Cases of lymphosarcoma (open circles) in Ingham, Oakland and Wayne Counties, Michigan, 1985 - 1994, and distribution by 1990 human population density. More dense stippling represents higher density.

Figure 3.10

Figure 3.11. Cases of lymphosarcoma (open circles) in Ingham, Oakland and Wayne Counties, Michigan, 1975 - 1984, and distribution by 1980 median household income. Shadings depict quartiles of income, with darker shading representing higher median household income.

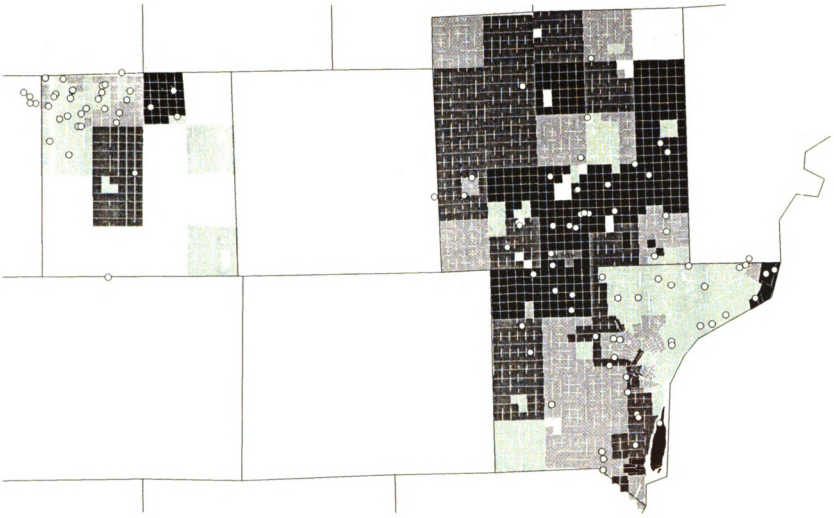


Figure 3.11.

SPATIAL AND TEMPORAL COMPARISON OF SELECTED CANCERS IN DOGS AND HUMANS, MICHIGAN, 1964-1994

VOLUME II

By

Daniel John O'Brien, DVM

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Large Animal Clinical Sciences
(Epidemiology)

1998

Chapter 4

SPATIAL AND TEMPORAL COMPARISON OF SELECTED CANCERS IN DOGS AND HUMANS, INGHAM, OAKLAND AND WAYNE COUNTIES, MICHIGAN, 1964-1994

Abstract

Concentration on identification of causal factors has often meant that the potential of investigations of cancer clustering to explore disease patterns and generate hypotheses has not been fully exploited. Simultaneous studies of morbidity in humans and other species in particular have not been fully investigated. Thus, a study was undertaken to investigate the geographic and time distributions of some biologically similar neoplasms in dogs and humans living in Michigan between 1964 and 1994. The study objective was to describe and compare the patterns of cancer in the two species while assessing the strength and dependence of those patterns. Histologically-confirmed incident human and canine cases of four cancers in three counties were mapped; *K* function spatial analysis and one-dimensional nearest-neighbor temporal analysis performed on residence addresses and dates of hospital discharge or diagnosis, respectively. Results suggest that processes determining spatial aggregation of cases in dogs and humans were not independent ($p \leq 0.01$) of each other and did not act uniformly over different geographic areas. Results suggest those processes operate at spatial scales < 2000 meters regardless of species, and that they tend to act upon dogs more strongly at shorter distances than on humans. Little evidence of interspecies concurrence of temporal

clustering was found, though general tendencies of particular neoplasms to be clustered or dispersed over time were similar. While the results of this study should properly be viewed as exploratory, they suggest that the underlying processes determining the geographic distribution of cancer in dogs and humans do not act independently, do not effect the two species equally, and vary by area and type of neoplasia.

Introduction

Aggregations (or "clusters") of disease in space and/or time have generated substantial debate[1-3] as well as research.[4-7] Most of the attention from both of these spheres has focused on identifying causes of disease aggregation, and making inferences concerning the role of putative environmental risk factors in influencing the spatial and temporal patterns of disease. Causality is unquestionably a worthy subject of inquiry. Yet the concentration of intellectual resources on this enigmatic area has often meant that the potential for investigations of disease aggregation to explore disease patterns and generate hypotheses has not been fully exploited.[8-10] In particular, Marshall[11] has noted that a broad ecological view of the spectrum of morbidity in humans and other species has been neglected.

In their ground breaking epidemiological work, Dorn, Schneider, Klauber and their colleagues initiated a series of investigations designed to describe the incidence of common cancers in dogs living in selected Northern California counties, and to compare cancer burdens in the dog population with those in

humans living in the same counties during the same period.[12-15] In the thirty years that have elapsed since those studies were published, there has been a great deal of speculation concerning the interrelation between the health of humans and the health of dogs and other animals that closely share their environment, and presumably their exposures.[16-18] In spite of this interest in the concept of sentinel animals, and the monumental advances made in epidemiological methods over the past thirty years, no new epidemiological research has emerged to surpass the work of Dorn's group. No new studies simultaneously investigating disease occurrence in companion animals and humans have been published since. This is likely due in large part to the paucity of good quality data on companion animal diseases, and the inability to accurately characterize the populations-at-risk necessary for the calculation of incidence rates for companion animal diseases. The difficulties in mounting laborious and expensive animal population census studies may also be a contributing factor. Furthermore, no research comparing the simultaneous spatial and temporal aggregation of cancer in humans and companion animals appears ever to have been reported.

In an attempt to help address these apparent gaps, a study was undertaken to investigate the geographic and time distributions of some biologically similar neoplasms in dogs and humans living in the State of Michigan during the thirty year period between 1964 and 1994. The objective of the study was to describe and compare the patterns of cancer in the two species using easily and inexpensively applicable quantitative methods, methods

capable of generating hypotheses concerning the processes which produce patterns of cancer in space and time, and capable of assessing the strength and dependence of those patterns between dogs and humans.

Methods

Data Acquisition

Cases for this retrospective registry-based study were gathered from three separate computerized data bases. Canine case records were obtained from the Veterinary Medical Data Base at Purdue University, a central repository for case records of patients examined at the majority of the veterinary medical colleges in North America. Records were obtained for all incident, histologically-confirmed cases of neoplasia in dogs diagnosed at the Michigan State University Veterinary Medical Center between March 1, 1964 and March 31, 1994. Histologic classification was based on the Standard Nomenclature of Veterinary Diseases and Operations.[19] Only records for first diagnoses were retained, eliminating duplicate records for the same patient. Records for benign neoplasms and metastatic sites were excluded. From this pool, cases of ten types of neoplasia, judged by a panel of board-certified specialists in veterinary pathology and internal medicine to be biologically similar to neoplasms in humans, were selected for further study. These case records were linked to medical records maintained at the Michigan State University Veterinary Medical Center *via* unique patient identification numbers to ascertain the residence street address of each case at the time of diagnosis. The majority of the cases lived in

only a few Michigan counties, and were of a small number of histologic types. To maximize the numbers of cases for quantitative analysis while concentrating efficiently on areas recording most of the cases, the study population of dogs was further restricted to cases of four cancers (lymphosarcoma, mammary adenocarcinoma, melanoma and spindle cell sarcoma) occurring in three counties (Ingham, Oakland and Wayne). Histologic classifications included in the case definitions of each of the cancers under study are presented in Table 4.1.

Human case records were obtained from separate sources depending on county of residence. Histologic classification of human cancers was based on the International Classification of Diseases for Oncology, Second Edition.[20] Histologic codes judged to be comparable to those of the canine cases were selected by a board-certified veterinary pathologist; these case definitions are presented in Table 4.1. Records for tumors with behavior codes corresponding to benign disease were excluded for lymphomas and spindle cell sarcomas, but some codes indicating benign behavior were included for mammary tumors and melanomas based on the professional judgement of the pathologist. Records for metastatic sites were excluded, in order to avoid including the same patient in the study population more than once. Based on those criteria, human case records were obtained from the Michigan Cancer Registry for patients residing in Ingham County, and from the Surveillance, Epidemiology and End Results Program of Metropolitan Detroit[21] for patients residing in Oakland and Wayne Counties. Human case records were not available for the entire thirty year time

period encompassed by the canine records. Data for human cases from Oakland and Wayne Counties were obtained for the period from 1/1/73 to 3/31/94, while data for human cases from Ingham County were obtained for the period from 1/1/85 to 3/31/94. Only records corresponding to first diagnoses of the study neoplasms were included to avoid duplicate records for a single patient. Study design, all procedures, and measures to ensure anonymity of study subjects were approved by the Michigan State University Committee on Research Involving Human Subjects and by the scientific advisory panel of the Michigan Cancer Registry prior to acquisition of any human case records.

Since available resources did not permit analysis of all of the human case records available, a stratified random sample of cases was selected for further analysis. Cases were stratified by date of diagnosis within cancer type within county. The sample was selected in such a way as to preserve the proportional representation of each histologic type among the four study tumors in the data set as a whole (*e.g.*, lymphomas comprised *x*% of the cases of the four study tumors residing in Oakland County, and so *x*% of the Oakland County sample was also made up of that histologic type). Similarly, sample cases were selected from two time strata (1973-1983, 1984-1994) so as to preserve the proportional incidence of cases over time (*e.g.*, *y*% of lymphomas in Oakland County were diagnosed between 1973 and 1983, so *y*% of the sample of lymphomas from Oakland County was also comprised of cases from that time period). Since case records from Ingham County were only available from 1985 onwards, the entire sample for that county fell within the second time stratum. Because of their

small proportional incidence, spindle cell sarcomas were oversampled by a factor of five. Random selections were made using a commercially available random number generator (Microsoft Excel 5.0, Microsoft Corporation, Redmond, Washington, USA) operating on random seeds.

Mapping and Spatial Analysis

Detailed discussion of the spatial and temporal analysis of the canine cases has been presented in Chapter 3. Methods used for analysis of the human cases in this study were identical. Briefly, street addresses of cases' residences at the time of diagnosis were standardized (AccuMail, Group 1 Software, Inc., Lanham, Maryland, USA) and geocoded (AtlasGIS, Environmental Systems Research Institute Inc., Redlands, California, USA) to obtain latitude and longitude (Universal Transverse Mercator) coordinates. Maps of residences of cases of both species were created to display geographic distributions at the county level (MapViewer 2.13, Golden Software, Inc., Golden, Colorado, USA); no maps were made at finer scales of resolution in order to protect the anonymity of the cases.

Latitude/longitude coordinates were subjected to second-order (*K* function) statistical analysis[22, pp. 56-68; 23, Chap. 4-7] (PPA, DongMei Chen and Arthur Getis, Department of Geography, San Diego State University, San Diego, California, USA). Originally employed for analyses in geography and ecology, second-order analysis has also been adapted for study of disease clustering.[24] Its theoretical focus is on the variance (or second moment, hence the name) of the distances between all possible pairwise combinations of cases. The

technique defines a function *k* which, for a prespecified distance *d*, equals the reciprocal of the intensity (the mean number of cases per unit area) multiplied by the expected number of additional cases within distance *d* of a case chosen at random.[23, p. 47] The *k* function can be used to calculate a test statistic *L*(*d*) which equals *d* when the pattern of cases is random, is greater than *d* when the cases cluster in space or less than *d* when they form a more regular pattern than one would expect at random. By calculating *L*(*d*) for a series of regular stepped distances *d* of increasing size (say, 100 meters, 200 meters, 300 meters, *etc.*), it is possible to determine both whether the cases are significantly clustered, and the distance(s) at which their tendency to cluster is strongest. The latter can provide insight into the geographic scale at which the process causing the cases to be clustered is operating. Results are depicted graphically, as plots of *L(d)* vs. *d*, or of the change in *L(d)* over the range of stepped distances *d*, the former showing the presence, statistical significance and overall strength of clustering or dispersion,[22, pp. 65-66] and the latter showing at which particular distances *d* the tendency toward clustering or dispersion is operating most strongly.[22, p. 67] In addition, *K* function analysis can be used to assess the statistical dependence of case patterns from two separate distributions on each other,[23, Chap. 7] *e.g.*, to assess whether the geographic pattern of cases of a disease in one species is independent of the pattern of a similar disease in a second species. A finding of independence would suggest that the processes which caused the case pattern in each species were acting independently of each other.

In the present work, the *K* function analysis calculated *L*(*d*) at fifty 100

meter stepped distances *d*, from 100 to 5000 meters. The first statistical test carried out tested the null hypothesis that the observed spatial pattern of human cases was consistent with complete spatial randomness, *i.e.*, consistent with a homogeneous planar Poisson process. The statistical significance of the randomness hypothesis was tested by Monte Carlo simulation, in which ninety-nine homogeneous Poisson process point patterns of size *N* were generated for each stepped distance *d* within each cancer by county analysis combination. When the *L(d)* calculated for the observed cancer case data was more extreme than any of the simulated values at any *d*, the null hypothesis was considered rejected. The null hypothesis of independence of the observed case patterns formed by the dog cancers from those formed by the comparable human cancers was statistically tested (PPA, DongMei Chen and Arthur Getis, Department of Geography, San Diego State University, San Diego, California, USA). The *K* function analysis was repeated, this time measuring all possible pairwise distances between case pairs where one member of the pair was a dog and the other a human case of the same cancer. The test statistic *L(d)* was calculated again for the observed data at each distance *d*. To test statistical significance, the coordinates of the human cases were randomized while keeping the coordinates of canine cases constant, and *L(d)* recalculated. The process was repeated ninety-nine times, and if the *L(d)* of the observed case data was more extreme than any of the simulated values at any *d*, the hypothesis of independence of the human cases from the dog cases was considered rejected. Because the software implementations used were constrained to analysis of

geographic areas with square or rectangular dimensions, spatial analysis was only conducted for cases residing in Ingham and Oakland Counties.

It is openly recognized here that the assumption of an underlying geographic pattern of complete spatial randomness for the cancers under study is, strictly speaking, difficult to defend. A number of authors[24-26] have pointed out that because of natural geographic variations in population density, cases of disease in humans are more likely to occur in urban areas than in rural ones, simply because the number of people available to become diseased is greater. Thus, rejection of a null hypothesis of complete spatial randomness, when it occurs, could simply be a result of clustering of the population itself, rather than clustering of disease over and above the spatial aggregation of the population.

The dog population, by comparison, is less well characterized. Although one might expect dog population density to mirror that of humans, there is also data to suggest that rural areas have higher rates of dog ownership.[27-30] Adding to the uncertainty is the fact that even the relatively limited dog population data available suggest that a substantial percentage of the dog population (13-15%) is missed by pet censuses.[31,32] The geographic distribution of those dogs is unknown. Another potential source of uncertainty might be the tendency for dogs living close to the Michigan State University Veterinary Medical Center to be diagnosed there more frequently than dogs from other areas. The existence and magnitude of such a referral bias have not, as yet, been characterized. Thus, assumption of complete spatial randomness as a null distribution for the dog population, while perhaps implausible, has not been definitively refuted

either.

Diggle[23, p. 5] has pointed out that while complete spatial randomness may be unlikely in real world situations, tests of it may nonetheless be defensible and useful for exploring a data set and helping to generate hypotheses concerning spatial patterns and what causes them. Rejection of a null hypothesis of complete spatial randomness may be of little interest in and of itself, although it is useful as a screening method, since patterns for which it cannot be rejected may not be worthy of further investigative efforts. In addition, it can be used to roughly distinguish between spatial patterns that tend toward clustering versus those that tend toward dispersion. It has also been pointed out[1] that a test of significance of clustering may offer little insight unless it also helps describe the disease pattern itself. Thus the test of statistical significance and its underlying assumptions are not necessarily the most important concern in a spatial analysis. Such is arguably the case in the present study, where the simultaneous comparison of the geographic patterns of cancers in dogs and humans is the salient issue.

Temporal Analysis

Dates of hospital discharge (for dogs, all counties, and humans, Oakland and Wayne Counties) or diagnosis (for humans, Ingham County) were converted to continuous, integer format (days since January 1, 1960) (Statistical Analysis System 6.10, SAS Institute Inc., Cary, North Carolina, USA), and analyzed using one-dimensional nearest-neighbor analysis.[22, pp. 45-49] Originally developed to test for clustering of points on a line, the method can also be used to test for

clustering of events in time.[33] It computes a test statistic which is the sum of the distances from each case to the case nearest it in time. This observed statistic is used along with the exact expectation and variance to calculate a *z* score which is compared with a normal distribution to determine two-tailed statistical significance. As with the *K* function analysis, one-dimensional nearest neighbor analysis allows determination of whether a group of cases is clustered, random, or dispersed in time.

Results

Success of geocoding of the human cases' residences is depicted in Table 4.2. In all cancer by county combinations, geographic coordinates were obtainable for greater than 89% of the cases; for nine of the twelve cancer by county combinations, success was greater than 95%. This proportion is slightly lower than the 97% success recorded previously for the canine cases.

Visual inspection of maps of the geographic distributions of the human and canine cancer cases' residences at diagnosis suggest disease patterns that vary by county and species, and to a lesser extent, by type of neoplasm. As expected, locations of human cases closely followed human population density, with the possible exception of melanomas in Wayne County. Geographic locations of canine cases, by comparison, closely followed gradients of human population density in Ingham County for all four cancers under study, but did not in Metropolitan Detroit (Oakland and Wayne Counties). There, cases in dogs appeared to be more widely dispersed through the counties than human cases

of the same cancer. Notably, this tendency displayed variation depending on the cancer being studied, with the effect being particularly prominent for canine lymphosarcomas and spindle cell sarcomas (Figure 4.1), while canine melanomas came closest to following human population density even in Oakland and Wayne Counties.

Results of *K* function spatial analysis and nearest-neighbor temporal analysis for both dogs and humans are summarized in Table 4.3. For the spatial analysis of the human cases' residences, the null hypothesis of a random distribution was rejected for all but one of the eight cancer by county combinations tested. Examination of the plots of $L(d)$ vs. d confirmed the subjective impressions noted on visual examination of the mapped patterns. In Ingham County, residences of human lymphoma cases showed significant clustering. Compared with canine cases, which were also found to be significantly clustered, the human cases exhibited comparatively weaker clustering tendencies than dogs at short (100 to 1000 meters) and long (3700 to 5000) distances d. Over the intervening distances, the strength of clustering was essentially identical in both dogs and humans. In Oakland County, the case pattern was significantly clustered in both species as well, but the tendency towards clustering was much stronger in humans. The null hypothesis of a random pattern of canine lymphomas was rejected only at one distance d, 900 meters, whereas in humans, clustering was evident at d from 1000 to 5000 meters. Interestingly, the distances at which both species' spatial patterns first showed significant clustering were within 100 meters of each other (900 and

1000 meters in dogs and humans, respectively).

With respect to mammary tumors, human cases' residences showed significant clustering in Ingham County across the entire range of tested distances *d*. This was quite similar to the pattern observed for dogs, which were also significantly clustered across all *d*. The case pattern in dogs was characterized by stronger clustering than the human pattern at all distances. In contrast, in Oakland County, human cases' residences were significantly clustered at *d* from 900 to 1600 meters, and *d* greater than 1800 meters. By way of comparison, the canine pattern was random.

Residences of human melanoma cases in Ingham County were significantly clustered at all distances $d > 900$ meters. While clustering among dogs attained statistical significance at two distances *d* (200 and 1500 meters), over the rest of the tested range, a random case pattern could not be rejected. Thus, the comparative strength of clustering was considerably stronger in humans except at very short ($\leq$ 200 meters) distances. In Oakland County, the pattern of human cases' residences attained statistically significant clustering over several distance ranges (1100-1700, 3800-3900, and 4500-4900 meters), although the overall strength of clustering was weak. The comparable pattern of dog cases' residences was random.

For spindle cell sarcomas, the human case pattern was random in Ingham County, while canine cases' residences showed significant spatial clustering at distances $d \geq 600$ meters. The opposite tendency occurred in Oakland County, where the case pattern in humans was significantly clustered at distances

between 1700 and 2300 meters, while the case pattern in dogs was random.

Recall that plots of the change in $L(d)$ per 100 meter change in d show the approximate distances at which the strength of spatial clustering is strongest, since under the null hypothesis of a random pattern, one would expect $L(d)$ to change the same amount as d.[22, p. 67] In the present study, these plots for the human cases' residence data suggest that where significant clustering was present, it was operating at distances less than 2000 meters. The distances d at which the tendency to cluster was strongest exhibitted some variation, with the largest number of cancer by county analysis combinations (lymphomas and melanomas in both counties) showing strongest clustering at $d < 300$ meters. Mammary neoplasms clustered most strongly at a distances of 500 and 900 meters in Ingham and Oakland Counties, respectively, while spindle cell sarcomas in Oakland County clustered most strongly at a d of 700 meters. Comparing these findings with those in dogs, clustering among the canine case residences, where present, was also operating at distances less than 2000 meters, with strongest clustering tendencies at distances $d < 200$ meters, regardless of tumor or county. Among the four cancer by county analysis combinations for which significant clustering was found in both species, the distance range over which clustering was operating tended to be at shorter distances among the canine cases than among the humans. Clustering of lymphomas in Ingham County occurred at $d < 1500$ meters among dogs vs. $d <$ 2000 meters among humans; clustering of lymphomas in Oakland County was operating at $d < 900$ meters among dogs vs. $d < 1700$ meters among humans;

and clustering of mammary neoplasms in Ingham County occurred at $d < 1100$ meters among dogs vs. $d < 1500$ meters in humans. In contrast, clustering of melanomas in Ingham County was operating at $d < 1400$ meters regardless of species.

K function analysis testing the independence of the spatial patterns formed by the dog cases' and human cases' residences rejected the null hypothesis of independent patterns in all eight cancer by county combinations tested (Table 4.3). Significant ($p \leq 0.01$) clustering was found in all cases, with human cases clustered in the direction of the canine cases. Strength of clustering (as measured by the number of *d* at which independence was rejected) varied considerably by both county and cancer, with independence of the spatial patterns rejected more strongly for cases residing in Ingham County than Oakland County. Strength of rejection by type of neoplasia was greatest for lymphomas in Ingham County and mammary tumors and weakest for melanomas and lymphomas in Oakland County, with spindle cell sarcomas intermediate.

Results of one-dimensional nearest-neighbor analysis (Table 4.3) suggest significant clustering of human cases over time in three of the twelve cancer by county analysis combinations studied: mammary neoplasms and melanomas in Oakland County, and mammary neoplasms in Wayne County. By way of comparison, only two of the twelve cancer by county analysis combinations showed significant temporal clustering among the canine cases. The only neoplasm/county where significant temporal clustering was found in both humans and dogs was mammary neoplasms in Wayne County. However,

interestingly, notwithstanding statistical significance, broad tendencies of particular cancers to be clustered or dispersed in time corresponded fairly well between species. In general, lymphomas and mammary neoplasms in both species tended towards clustering in time, whereas cases of melanoma and spindle cell sarcoma tended to be dispersed (*i.e.*, to occur at regular intervals) over time.

Discussion

The study of spontaneously occurring cancers in domesticated animals to gain insight into the pathogenesis of human cancer has been the subject of considerable scientific discussion.[34-39] All four of the canine neoplasms involved in the current study have previously been concluded to be sufficiently similar biologically to the corresponding diseases in humans to be considered animal models of the human diseases.[15,36,40-56] Thus, pathogenetic characteristics of these diseases described in dogs in clinical and laboratory settings are considered important for understanding the human disease. Using a similar line of reasoning, one might expect the epidemiologic characteristics of these diseases in dog populations to be of potential value for understanding the dynamics of the diseases in humans, or vice versa. The intensive scrutiny of these cancers in human patients has led to considerable understanding of their causes and pathogenesis in people. In contrast, with the exception of mammary neoplasms, the specific etiological factors which give rise to these cancers in dogs are poorly understood by comparison. Yet even without certain knowledge

of common etiologies, the biological similarity of these cancers in dogs and humans is, in and of itself, sufficient justification for comparing their spatial patterns to discover what those patterns may convey about the underlying processes which help determine the geographic occurrence of the diseases over time.

Some explanation is warranted here to clarify a distinction between findings of statistically significant clustering which are strong and obvious and those that might well be considered weak or equivocal. For the methods used here, statistically significant clustering was defined to have occurred when the $L(d)$ value calculated for the observed data was more extreme than the ninety-nine simulated values at any of the distances d tested. Consequently, formal statistical rejection of a null hypothesis of randomness could have occurred at any one distance d or at many. In both situations, the outcome of the test formally designates the spatial pattern of cases to be significantly clustered, yet those situations where rejection of randomness occurred at only one or a few d should be considered evidence of clustering that is more equivocal than those where rejection occurred at many d (Figure 4.2). Accordingly, in the present study, the spatial patterns in a number of the cancer by county analysis combinations (dog lymphomas in Oakland County, dog melanomas in Ingham County, and human melanomas, Oakland County), though formally considered clustered, should be recognized as showing weaker or more equivocal evidence of spatial clustering.

Perhaps the most prominent finding in the current study was the variation

in spatial pattern by geographic area. If the underlying processes determining the geographic distribution of these cancers are sufficiently similar in both dogs and humans, one might hypothesize that their spatial patterns would tend to be similar regardless of geographic setting. Such might be the case if, for example, the occurrence of cancer in dogs inevitably followed dog population density, which in turn mimicked human population density. Alternatively, such a hypothesis might also hold if the processes determining the spatial distributions were unaffected by local environmental factors that vary from place to place. The results of the present study do not appear to be consistent with either scenario. By assuming identical random underlying spatial distributions for both dogs and humans and analyzing them using identical methodology, the effect of human population density on interspecies comparisons of the spatial patterns is neutralized. For some of the cancers studied here (lymphomas and mammary tumors in Ingham County), the spatial patterns were similar enough between the species to suggest that human population density could be the simple underlying process determining the geographic distribution of the cases. Yet examination of the patterns formed by cases of those same cancers in another county (Oakland) suggests that human population density cannot be a universal determinant of spatial distribution for canine cancer cases (Figure 4.3). Examination of cancers for which the spatial patterns of cases in dogs and humans are random, or the strength and statistical significance of their clustering dissimilar (melanomas, spindle cell sarcomas), suggest a similar conclusion. Taken together, these findings allow one to hypothesize that either the spatial

pattern of these cancers in dogs is not primarily determined by dog population density (if dog density follows human density), or alternatively, that dog population density is relatively independent of human population density. Whichever of these hypotheses actually holds, it suggests the simplistic assumption that the geographic distributions of dog and human cancers are necessarily determined by a common process (*e.g.*, human population density) may be flawed.

Similarly, the assumption of etiological factors acting as uniformly potent causes of human and canine cancer unaffected by local environmental factors may also have limited validity. The design of this study does not allow firm conclusions about the actual causes of the study cancers to be drawn. However, these results do suggest the hypothesis that whatever factors are determining the spatial distribution of cases can vary substantially from place to place. Local environmental factors acting as risk factors for these cancers are not inconsistent with such an hypothesis, though clearly, they are only one of potentially numerous and speculative explanations. Differential sensitivity to a common etiological exposure might also explain why cases of a cancer in one species tend to be aggregated geographically while those in another are not.

The other important outcome which warrants discussion is interspecies comparison of the distances at which clustering is occurring, and occurring most strongly. Among the four tumor by county analysis combinations where both canine and human cases were significantly clustered, three of four showed clustering to be operating at shorter distances among dogs than among humans.

Melanomas in Ingham County, which closely paralleled human population density, were the lone exception. This suggests that the processes associated with aggregation, whatever they are, were acting over a shorter distance range on dogs. In addition, the strength of the clustering caused by those processes, in general, was greatest at shorter distances from the dog cases' residences than from the human cases' residences. Given this, one might hypothesize that whatever processes are determining the clustering of the cases have an upper limit of influence at relatively longer distances for humans than dogs, and exert their maximal influence closer to dogs than to humans living in the same area. Notably, this maximum distance at which the clustering processes were potent appears to have been < 2000 meters regardless of species.

It is also interesting to note that particular distances *d* seem to recur in these analyses, suggesting the hypothesis that they may be of some special importance to the processes which determine the spatial patterns. For example, the distance at which the clustering of several types of cancer (human lymphomas and melanomas in Ingham and Oakland Counties, canine lymphomas in Oakland County, and human breast neoplasms in Oakland County) attained statistical significance was from 800 to 1000 meters, a comparatively short distance range of only 200 meters. Also, distances *d* of 300 meters or less were associated with the strongest tendency towards clustering for all five of the canine and two of seven human cancer by county combinations where significant clustering was identified. It is fascinating to speculate abstractly on what sorts of processes might be acting at these distance scales,

but precise explanations are more elusive. Yet such enigmatic questions are exactly the outcome anticipated and desired from a study such as this one.

Results of tests of the independence of the spatial patterns formed by the human and dog cases overwhelmingly suggest that the underlying processes determining the geographic distribution of cases in the two species, whatever they are, are not acting independently. Rather, clustering of the human cases in the direction of the canine cases implies a spatial relationship. However, given the variability in the strength and significance of the results of the within-species randomness tests already discussed, it is also clear that while the processes are not independent, they do not effect the two species equally either. From a methodological standpoint, this illustrates the importance of performing *K* function analyses on the case patterns both within and between each species being studied.

With respect to aggregation in time, the methods used in the present study give little support for concordance of clustering of these cancers between species. In only one instance (breast tumors in Wayne County) did the findings in dogs concur with those in people. This lack of temporal clustering is not altogether unexpected, however, since the long and often variable latencies associated with neoplasia may make it unlikely that even cases due to a common etiological exposure would develop clinical disease at the same time,[57] especially in different species.

While it is interesting to speculate on these results and the value they may hold as points of departure for new research into the processes bearing on

disease aggregation across species, it is nonetheless imperative that the limitations of this study be recognized and kept in mind. That this study is exploratory (rather than explanatory) by design cannot be overemphasized. The techniques used here do not have the ability to distinguish between clustering that is due to etiologic factors for the cancers in question and that due to factors which might affect the likelihood of diagnosis in a particular area. This is particularly relevant with respect to the canine spatial patterns, where the study population was drawn exclusively from those dogs diagnosed at a veterinary teaching hospital. To the extent that the rate of referral of these cancers in the general dog population is low, the population of cases studied here may not be representative of the canine population as a whole, and thus generalizability of the findings may be limited. Other authors have noted the pitfalls of using hospital populations to draw inferences about the general population for both dogs[58,59] and humans.[60] Since the canine patients were generally referred to the Michigan State University Veterinary Medical Center by other veterinary practitioners, factors causing differential likelihood of referral by geographic area could be the processes actually being measured here. Still, even knowledge gained about non-etiological processes causing clustering could potentially be of value, for example, for studying factors influencing utilization of veterinary health care resources and access to care, and how they compare with humans.

Other unavoidable limitations in the data available for analysis could potentially constrain the current findings. The time periods over which cases were registered differed between the canine and human cases, and between

counties for the humans, with the registered human cases covering only a portion of the time spanned by the dogs. To the extent that spatial patterns formed by human cases diagnosed during time periods captured by the available registry data (1973-1994 for Oakland and Wayne Counties, and 1985-1994 for Ingham County) are not representative of the period from 1964-1972 (or 1964-1984), the results presented here could be affected. Eighty percent of the canine cases were diagnosed after 1973, but only 29% after 1984, so the effect of any such bias, if it exists, would be substantially stronger for the Ingham County data. It could also be argued that differences in sample size (Table 3) could account for disparities in the statistical significance of spatial patterns between species for the same cancer (*e.g.*, for mammary tumors in Oakland County). However, this is clearly not the case overall, since in some cases (*e.g.*, spindle cell sarcomas in Oakland County), significant clustering was found for the species with the smaller number of cases, while the one with the larger sample exhibited a random pattern. Finally, as is the case in any situation where novel techniques are applied in a research setting where little previous work exists, the results of the present study must be considered exploratory until replicated by other workers with other data sets.

In spite of these limitations, the present study demonstrates the feasibility of investigating and comparing patterns of cancer occurrence in man and domestic animals without reliance on disease rates and the arduous methods necessary to estimate the populations-at-risk upon which they depend. More importantly, it adds to both the human and veterinary medical literature

descriptive information concerning the simultaneous spatial and temporal distributions of biologically similar cancers in dogs and humans living in the same areas during the same time period. Counted among that information were the suggestive findings that processes determining spatial aggregation of cases in dogs and humans are not independent of each other, do not act uniformly over different geographic areas, that those processes operate at spatial scales of less than 2000 meters regardless of species, and that they tend to act upon dogs more strongly at shorter distances than they do upon humans. There was little evidence of interspecies correspondence of cancer case clustering in time, though broad tendencies of particular neoplasms to be clustered or dispersed over time were similar.

REFERENCES

REFERENCES

1. Rothman KJ. A sobering start to the cluster busters' conference. Am J Epidemiol 1990;132 Suppl 1:S6-13.

2. Neutra R, Swan S, Mack T. Clusters galore: insights about environmental clusters from probability theory. Sci Tot Environ 1992;127:187-200.

3. Wartenberg D, Greenberg M. Methodological problems in investigating disease clusters. Sci Tot Environ 1992;127(1,2):173-85.

4. Proceedings of the National Conference on Clustering of Health Events. Am J Epidemiol 1990;132 Suppl 1:S1-S202.

5. Jacquez, GM, editor. Proceedings of the workshop on statistics and computing in disease clustering, Port Washington, New York, 1992. Stat Med 1993;12:1751-1968.

6. Lawson A, Waller L, Biggeri A, editors. Spatial disease patterns. Stat Med 1995;14:2289-501.

7. Jacquez GM, editor. Conference on Statistics and Computing in Disease Clustering, British Columbia, Canada, July 21-22, 1994. Stat Med 1996;15:681-952.

8. Beral V. Childhood leukemia near nuclear plants in the United Kingdom: the evolution of a systematic approach to studying rare disease in small geographic areas. Am J Epidemiol 1990;132 Suppl 1:S63-S68.

9. Olsen SF, Martuzzi M, Elliott P. Cluster analysis and disease mapping—why, when, and how? A step by step guide. BMJ 1996;313:863-6.

10. Elliott P, Martuzzi M, Shaddick G. Spatial statistical methods in environmental epidemiology: a critique. Stat Methods Med Res 1995;4:137-59.

11. Marshall RJ. A review of methods for the statistical analysis of spatial patterns of disease. J R Stat Soc A 1991;154:421-41.

12. Dorn CR, Taylor ON, Frye FL, Hibbard HH. Survey of animal neoplasms in Alameda and Contra Costa Counties, California. I. Methodology and

description of cases. J Natl Cancer Inst 1968;40:295-305.

13. Dorn CR, Taylor ON, Schneider R, Hibbard HH, Klauber MR. Survey of animal neoplasms in Alameda and Contra Costa Counties, California. II. Cancer morbidity in dogs and cats from Alameda County. J Natl Cancer Inst 1968;40:307-18.

14. Schneider R, Dorn CR, Klauber MR. Cancer in households: a human-canine retrospective study. J Natl Cancer Inst 1968;41(6):1285-92.

15. Schneider R. Comparison of age, sex, and incidence rates in human and canine breast cancer. Cancer 1970;26(2):419-26.

16. O'Brien DJ, Kaneene JB and Poppenga RH. The use of mammals as sentinels for human exposure to toxic contaminants in the environment. Environ Health Perspect 1993;99:351-68.

17. Bukowski JA, Wartenberg D. An alternative approach for investigating the carcinogenicity of indoor air pollution: pets as sentinels of environmental cancer risk. Environ Health Perspect 1997;105(12):1312-19.

18. National Research Council (US). Animals as sentinels of environmental health hazards. Committee on Animals as Monitors of Environmental Hazards, Board on Environmental Studies and Toxicology, Commission on Life Sciences, National Research Council. Washington ,(DC): National Academy Press (US); 1991.

19. Priester WA, editor. Standard nomenclature of veterinary diseases and operations. 2nd (abridged) ed. Bethesda (MD): National Cancer Institute (US); 1975; Dept. Health Education and Welfare Publication No. (NIH)76-1028.

20. Percy C, Van Holten V, Muir C, editors. International Classification of Diseases for Oncology. 2nd ed. Geneva: World Health Organization; 1990.

21. Gloeckler Ries LA, Miller BA, Hankey BF, Kosary CL,Harrass A, Edwards BK, editors. SEER cancer statistics review, 1973-1991. Tables and graphs. Bethesda (MD): Dept. of Health and Human Services (US), National Institutes of Health, National Cancer Institute; 1994. NIH Publication No. 94-2789.

22. Boots BN, Getis A. Point Pattern Analysis. Sage University Paper series on Scientific Geography, Vol. 8. Newbury Park (CA): Sage Publications;

1988.

23. Diggle PJ. Statistical Analysis of Spatial Point Patterns. New York: Academic Press; 1983.

24. Diggle P, Chetwynd A. Second-order analysis of spatial clustering for inhomogeneous populations. Biometrics 1991;47:1155-63.

25. Turnbull BW, Iwano EJ, Burnett WS, Howe HL, Clark LC. Monitoring for clusters of disease: application to leukemia incidence in upstate New York. Am J Epidemiol 1990;132 Suppl 1:S136-S143.

26. Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. Stat Med 1995;14:799-810.

27. Wise JK. U.S pet ownership & demographics sourcebook. Schaumburg (IL): Center for Information Management, American Veterinary Medical Association; 1997.

28. Leslie BE, Meek AH, Kawash GF, McKeown DB. An epidemiological investigation of pet ownership in Ontario. Can Vet J 1994;35(4):218-22.

29. Griffiths AO, Brenner A. Survey of cat and dog ownership in Champaign County, Illinois, 1976. J Am Vet Med Assoc 1977;170(11):1333-40.

30. Franti CE, Kraus JF, Borhani NO. Pet ownership in a suburban-rural area of California, 1970. Public Health Rep 1974;89(5):473-84.

31. Human Population Laboratory: Alameda County population, 1965. Berkeley (CA): California State Department of Public Health; 1966. Series A; No. 7.

32. Heussner JC, Flowers AI, Williams JD, Silvy NJ. Estimating dog and cat populations in an urban area. Anim Regulation Studies 1978;1:203-12.

33. Selkirk KE, Neave HR. Nearest-neighbour analysis of one-dimensional distributions of points. Tijdschr Econ Soc Geogr 1984;75(5):356-62.

34. Mulligan RM. Comparative pathology of human and canine cancer. Ann N Y Acad Sci 1963;108(3):642-90.

35. Immunity to cancer: naturally occurring tumours in domestic animals as models for research. 1. Bull World Health Organ 1973;49(1):81-91.

36. Immunity to cancer: naturally occurring tumours in domestic animals as models for research. 2. Bull World Health Organ 1973;49(2):205-13.

37. Owen LM, Bostock DE, Betton GR, Onions DE, Holmes J, Yoxall A et al. The role of spontaneous canine tumours in the evaluation of the aetiology and therapy of human cancer. J Small Anim Pract 1975;16(3):155-62.

38. MacEwen EG. Spontaneous tumors in dogs and cats: models for the study of cancer biology and treatment. Cancer Metastasis Rev 1990;9(2):125-36.

39. Hahn KA, Bravo L, Adams WH, Frazier DL. Naturally occurring tumors in dogs as comparative models for cancer therapy research. In Vivo 1994;8(1):133-43.

40. Johnson RE, O'Conor GT, Titus W, Hafermann MD. Comparative clinical, histologic, and radiotherapeutic aspects of canine and human malignant lymphoma. Radiology 1969;93(2):395-9.

41. Heath H 3d, Weller RE, Mundy GR. Canine lymphosarcoma: a model for study of the hypercalcemia of cancer. Calcif Tissue Int 1980;30(2):127-33.

42. Carter RF, Valli VE, Lumsden JH. The cytology, histology and prevalence of cell types in canine lymphoma classified according to the National Cancer Institute Working Formulation. Can J Vet Res 1986;50(2):154-64.

43. Fivenson DP, Beck ER, Dunstan RW, Nickoloff BJ, Moore PF. Dermal dendrocytes and T-cells in canine mycosis fungoides. Support for an animal model of human cutaneous T-cell lymphoma. Cancer 1992;70(8):2091-8.

44. Teske E. Canine malignant lymphoma: a review and comparison with human non-Hodgkin's lymphoma. Vet Q 1994;16(4):209-19.

45. Mulligan RM. Mammary cancer in the dog: a study of 120 cases. Am J Vet Res 1975;36(9):1391-6.

46. Priester WA. Multiple primary tumors in domestic animals: a preliminary view with particular emphasis on tumors in dogs. Cancer 1977;40(4 Suppl):1845-8.

47. Owen LN. A comparative study of canine and human breast cancer. Invest Cell Pathol 1979;2(4):257-75.

48. MacEwen EG, Patnaik AK, Harvey HJ, Panko WB. Estrogen receptors in canine mammary tumors. Cancer Res 1982;42(6):2255-9.

49. Gilbertson SR, Kurzman ID, Zachrau RE, Hurvitz AI, Black MM. Canine mammary epithelial neoplasms: biologic implications of morphologic characteristics assessed in 232 dogs. Vet Pathol 1983;20(2):127-142.

50. Mottolese M, Morelli L, Agrimi U, Benevolo M, Sciarretta F, Antonucci G, et al. Spontaneous canine mammary tumors. A model for monoclonal antibody diagnosis and treatment of human breast cancer. Lab Invest 1994;71(2):182-7.

51. Dunstan RW, Rosser EJ Jr. Comparative aspects of cutaneous melanocytic neoplasms in the dog. Am J Dermatopathol 1985;7 Suppl:197-201.

52. Goldschmidt MH. Benign and malignant melanocytic neoplasms of domestic animals. Am J Dermatopathol 1985;7 Suppl:203-212.

53. Gleiser CA, Raulston GL, Jardine JH, Gray KN. Malignant fibrous histiocytoma in dogs and cats. Vet Pathol 1979;16(2):199-208.

54. Richardson RC, Anderson VL, Voorhees WD 3d, Blevins WE, Inskeep TK, Janas W, et al. Irradiation-hyperthermia in canine hemangiopericytomas: large-animal model for therapeutic response. J Natl Cancer Inst 1984;73(5):1187-94.

55. Madewell BR, Griffey SM, Munn RJ. Ultrastructure of canine vasoformative tumors. J Vasc Res 1992;29(2):50-5.

56. Pumarola M, Anor S, Borras D, Ferrer I. Malignant epithelioid schwannoma affecting the trigeminal nerve of a dog. Vet Pathol 1996;33(4):434-6.

57. Whittemore AS, Friend N, Brown BW Jr, Holly EA. A test to detect clusters of disease. Biometrika 1987;74(3):631-5.

58. Wagstaff DJ, Goyings LS, Langham RF. Canine cancer distribution as related to data source. Am J Vet Res 1967;28:1479-82.

59. Robinson GW. A comparison of licensed and hospital dog populations. J Am Vet Med Assoc 1968;152(9):1383-9.

60. Masi, AT. Potential uses and limitations of hospital data in epidemiologic research. Am J Pub Health 1965;55:658-67.

Table 4.1. Case definitions, selected canine and human neoplasms, Michigan, 1964-1994

Neoplasm	Histologic Classifications Included*	
	Dog	Human
Lymphoma	8310	9590-9592, 9670-9673, 9675, 9676, 9680-9686, 9693, 9700-9702, 9709, 9720, 9723, 9731, 9732
Mammary Neoplasia	19008091	8140 (including 8140/0, but not 8140/1), 8141, 8260 (including 8260/0), 8323 (including 8323/0), 8470-8473 (including 8470/0 & 8471/0), 8480 (including 8480/0), 8481, 8490, 8500, 8503-8506 (including 8503/0, 8504/0, 8505/0 & 8506/0, but not 8503/2), 8520-8522, 8550 (including 8550/0), 8560, 8562, 8570-8572 and sites 500-509
Melanoma	8174	8720, 8721, 8725/0, 8730, 8740/0, 8743, 8745, 8750/0, 8760/0, 8771-8774 (including 8771/0 & 8772/0), 8780 (including 8780/0), 8790/0
Spindle Cell Sarcoma	8700F, 8451, 8531, 8702, 8452	8801, 8810-8814, 8830, 9150, 9330, 9540, 9560-9561

*Histologic classification of canine cancers based on the Standard Nomenclature of Veterinary Diseases and Operations, while human cancer classification based on the International Classification of Diseases for Oncology, Second Edition.

Table 4.2. Outcome of address assignment/geocoding, selected human cancers, Michigan, 1973-1994, by number of cases (%)

Neoplasm	County	Unsuccessful		Successful	Total
		No/ Incomplete Street Address	Address Could Not Be Coded		
Lymphoma	Ingham		1 (2.1)	46 (97.9)	47 (100)
	Oakland	1 (1.9)		51 (98.1)	52 (100)
	Wayne	2 (3.4)		56 (96.6)	58 (100)
Mammary Neoplasia	Ingham	3 (1.5)	1 (0.5)	195 (98.0)	199 (100)
	Oakland	14 (7.5)	5 (2.7)	168 (89.8)	187 (100)
	Wayne	5 (2.6)	4 (2.1)	182 (95.3)	191 (100)
Melanoma	Ingham	1 (2.2)	1 (2.2)	43 (95.6)	45 (100)
	Oakland	2 (5.0)	1 (2.5)	37 (92.5)	40 (100)
	Wayne	1 (4.2)		23 (95.8)	24 (100)
Spindle Cell Sarcoma	Ingham		1 (4.0)	24 (96.0)	25 (100)
	Oakland	2 (9.5)		19 (90.5)	21 (100)
	Wayne	1 (4.2)		23 (95.8)	24 (100)

Table 4.3. Summary of results of spatial (*K* function)* and temporal (nearest-neighbor) analyses, selected canine and human neoplasms, Michigan, 1964-1994

Neoplasm	County	*n*†	Spatial Analysis		Temporal Analysis	
			Dog	Human	Dog	Human
Lymphoma	Ingham‡	78 *46*	Clustered ($p \leq 0.01$)	Clustered ($p \leq 0.01$)	Random ($p = 0.26$)	Random ($p = 0.10$)
	Oakland‡	59 *51*	Clustered ($p \leq 0.01$)	Clustered ($p \leq 0.01$)	Clustered ($p = 0.04$)	Random ($p = 0.22$)
	Wayne	84 *56*	—	—	Random ($p = 0.92$)	Random ($p = 0.50$)
Mammary Neoplasia	Ingham‡	57 *195*	Clustered ($p \leq 0.01$)	Clustered ($p \leq 0.01$)	Random ($p = 0.88$)	Random ($p = 0.58$)
	Oakland‡	16 *168*	Random	Clustered ($p \leq 0.01$)	Random ($p \approx 0.64$)§	Clustered ($p = 0.007$)
	Wayne	23 *182*	—	—	Clustered ($p = 0.02$)	Clustered ($p = 0.003$)
Melanoma	Ingham‡	25 *43*	Clustered ($p \leq 0.01$)	Clustered ($p \leq 0.01$)	Random ($p = 0.78$)	Random ($p = 0.58$)
	Oakland‡	23 *37*	Random	Clustered ($p \leq 0.01$)	Random ($p = 0.54$)	Clustered ($p = 0.018$)
	Wayne	24 *23*	—	—	Random ($p = 0.22$)	Random ($p = 0.32$)

Table 4.3 (cont'd)

Neoplasm	County	*n*[†]	Spatial Analysis		Temporal Analysis	
			Dog	Human	Dog	Human
Spindle Cell Sarcoma	Ingham[‡]	45 *24*	Clustered ($p \leq 0.01$)	Random	Random ($p = 0.08$)	Random ($p = 0.12$)
	Oakland[‡]	38 *19*	Random	Clustered ($p \leq 0.01$)	Random ($p = 0.44$)	Random ($p \approx 0.46$)[§]
	Wayne	42 *23*	—	—	Random ($p = 0.48$)	Random ($p = 0.32$)

*Spatial analysis was limited to Ingham and Oakland Counties.
[†]Canine sample size listed in plain typeface; human sample size listed in italics.
[‡]Spatial patterns of canine and human cases of this cancer in this county were not independent of each other ($p \leq 0.01$)
[§]*p* value considered approximate due to small sample size (< 20).

Figure 4.1. Distribution of canine and human spindle cell sarcomas in Ingham (left), Oakland (top right) and Wayne Counties, Michigan, 1964-1994. Open and shaded diamonds mark locations of canine and human cases' residences, respectively, at the time of diagnosis (Analogous maps for the other three cancers under study appear as Figures 6.1-6.3 in Appendix B).

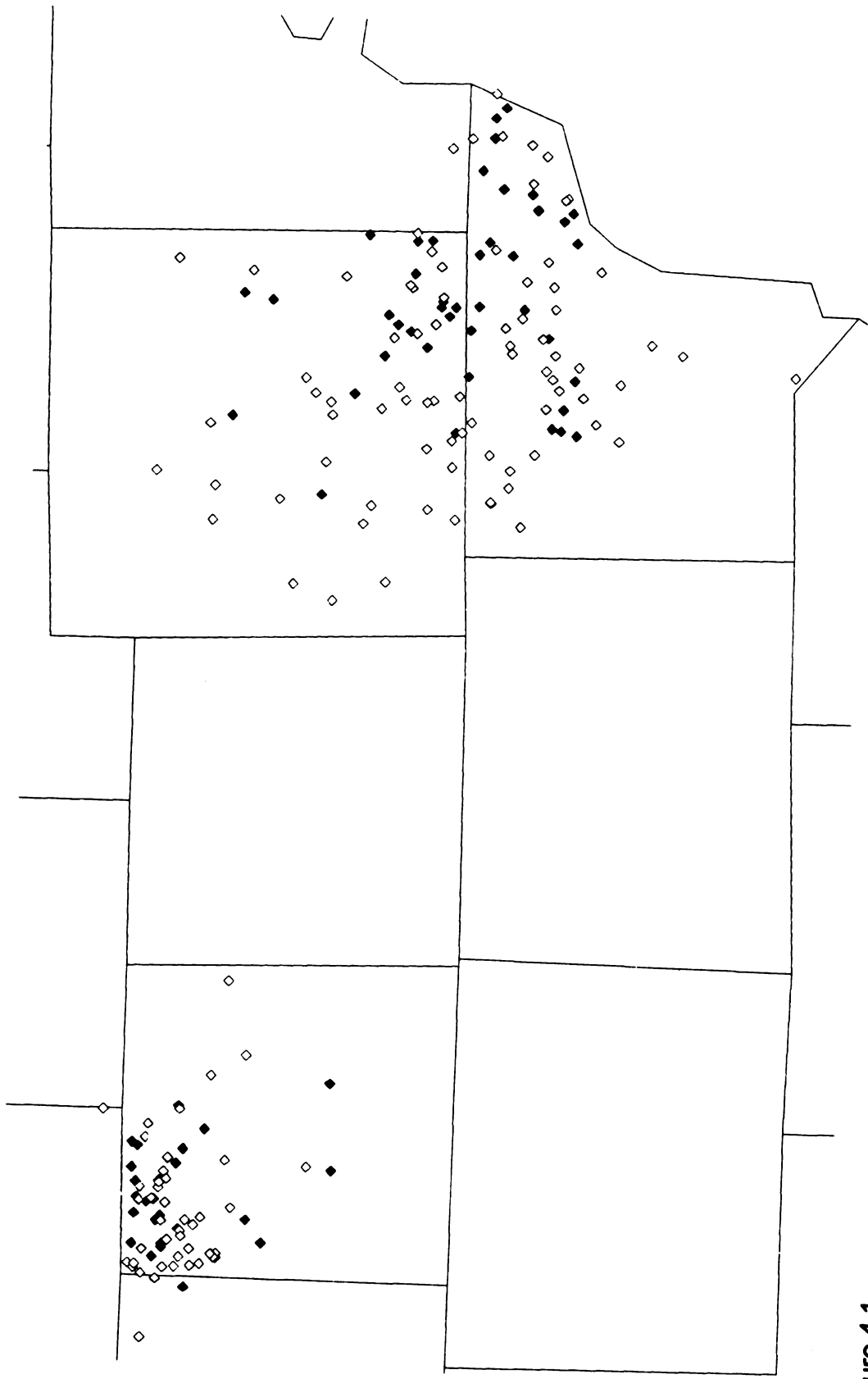


Figure 4.1.

Figure 4.2. Graphical results of second-order (*K* function) spatial analysis, melanomas, Ingham County, Michigan 1964-1994. The observed value of the test statistic $L(d)$ at a given distance d is marked by the solid black triangles and line (canine cases' residences) and the open red triangles and solid red line (human cases' residences). The boundaries of the 99% confidence band on the null hypothesis of complete spatial randomness are denoted by the black (dog) and red (human) dotted lines. Note that the plot of observed $L(d)_{human}$ breaches the upper confidence limit at all $d \geq 900$ meters, while the plot of observed $L(d)_{dog}$ does so only at two distances, 200 and 1600 meters. Since both the human and canine spatial patterns had observed $L(d)$ values which exceeded the upper 99% confidence limit determined by simulation, both formally reject the null hypothesis of randomness, and consequently, both are considered significantly clustered at $p \leq 0.01$. Yet due to the number of d at which the observed $L(d)_{human}$ exceeded the upper confidence limit, and the magnitude of those exceedances, the spatial pattern of human cases' residences is interpreted as showing strong evidence of spatial clustering, while that of the canine cases' residences may be considered weak to equivocal, in spite of its high degree of statistical significance.

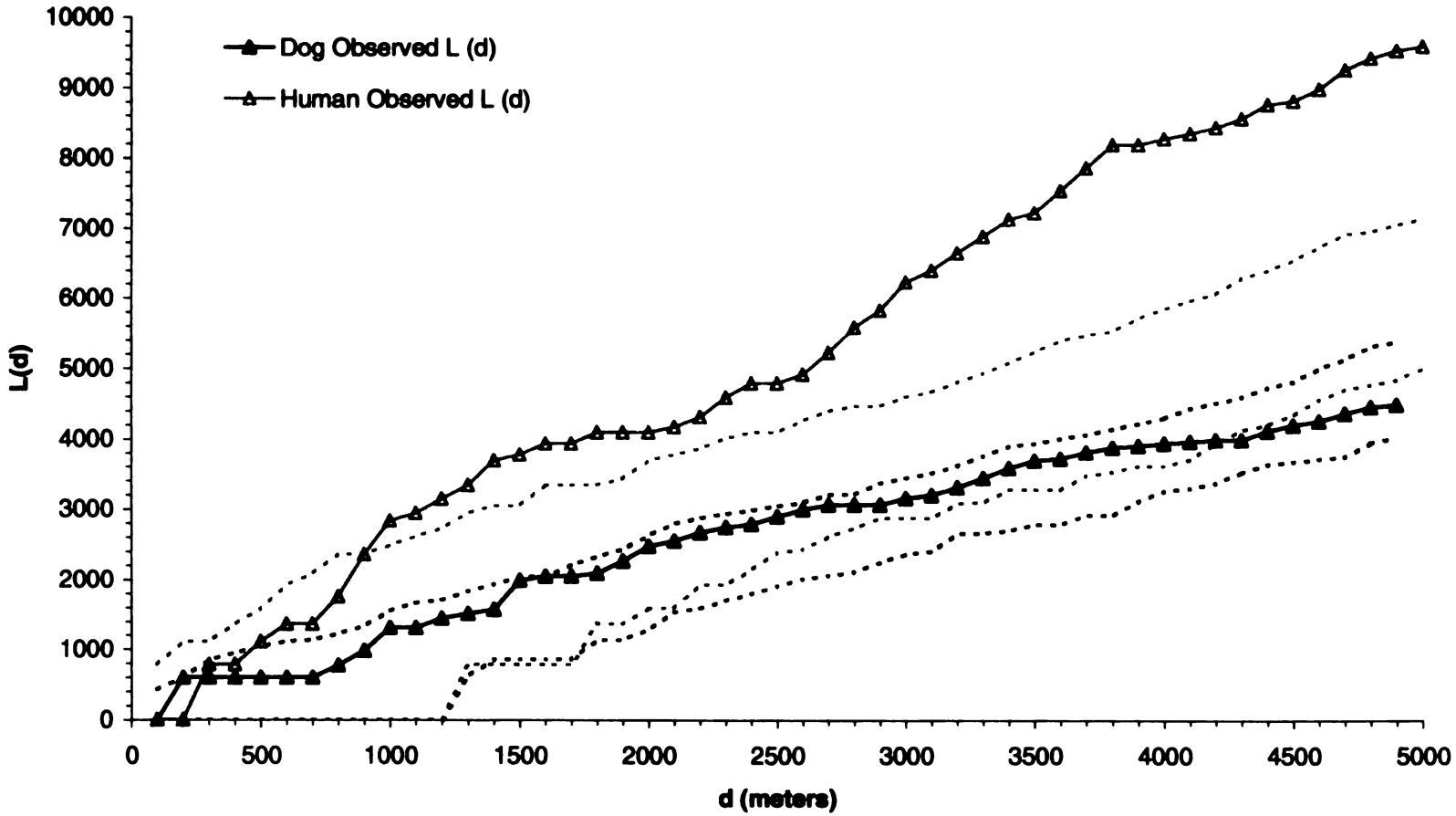
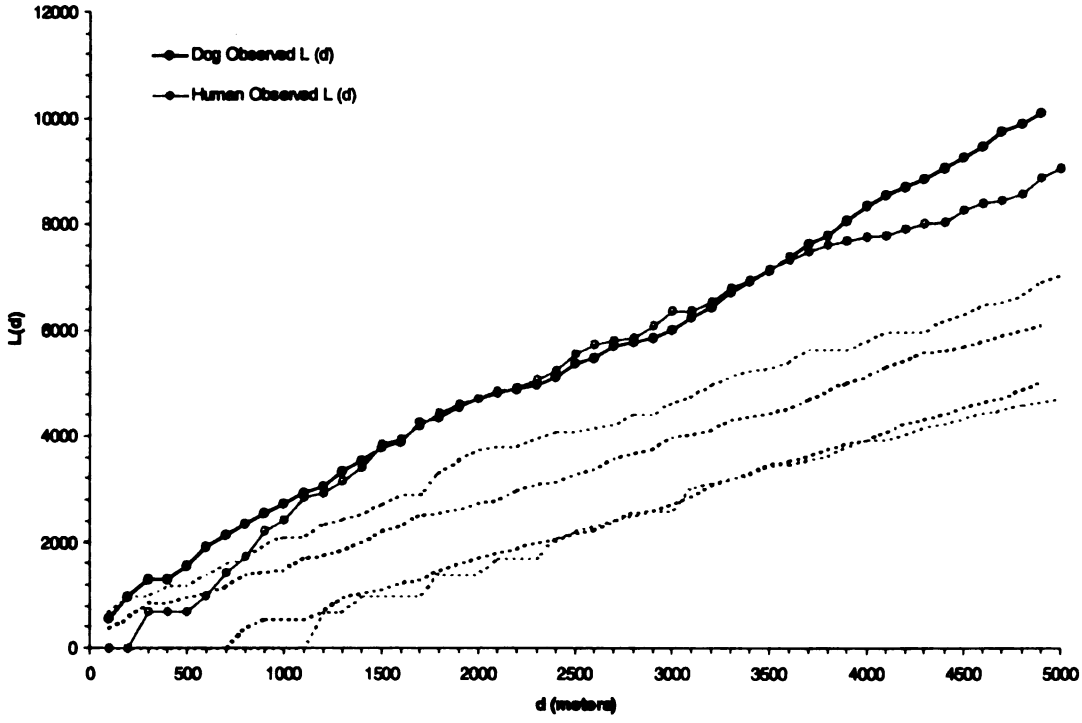

Dog Observed L (d)
Human Observed L (d)
L(d)
d (meters)
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000

Figure 4.2.

Figure 4.3. Graphical results of *K* function analysis, dog (black circles) and human (red circles) lymphomas, Ingham (top figure) and Oakland (bottom figure) Counties, Michigan, 1964-1994. Explanations of the plotted lines are as described in Figure 4.2. Note the plots of observed $L(d)_{human}$ and $L(d)_{dog}$ closely coincide in Ingham County, suggesting similar strength and significance of spatial clustering in each species, and so potentially a single underlying process jointly determining the geographic distributions in both species. In contrast, the results in Oakland County show clustering among the human cases that is much stronger than that among the dogs, arguing against such a jointly acting process in that county (Analogous plots for mammary neoplasms, melanomas in Oakland County, and spindle cell sarcomas, respectively, appear as Figures 6.4-6.8 in Appendix B. Plots of $\Delta L(d)$ vs. d [analogous to Figure 3.8] for human cases of all cancers under study appear as Figures 6.9-6.16 in Appendix B).


Dog Observed L (d)
Human Observed L (d)
L(d)
d (meters)
0
2000
4000
6000
8000
10000
12000
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000

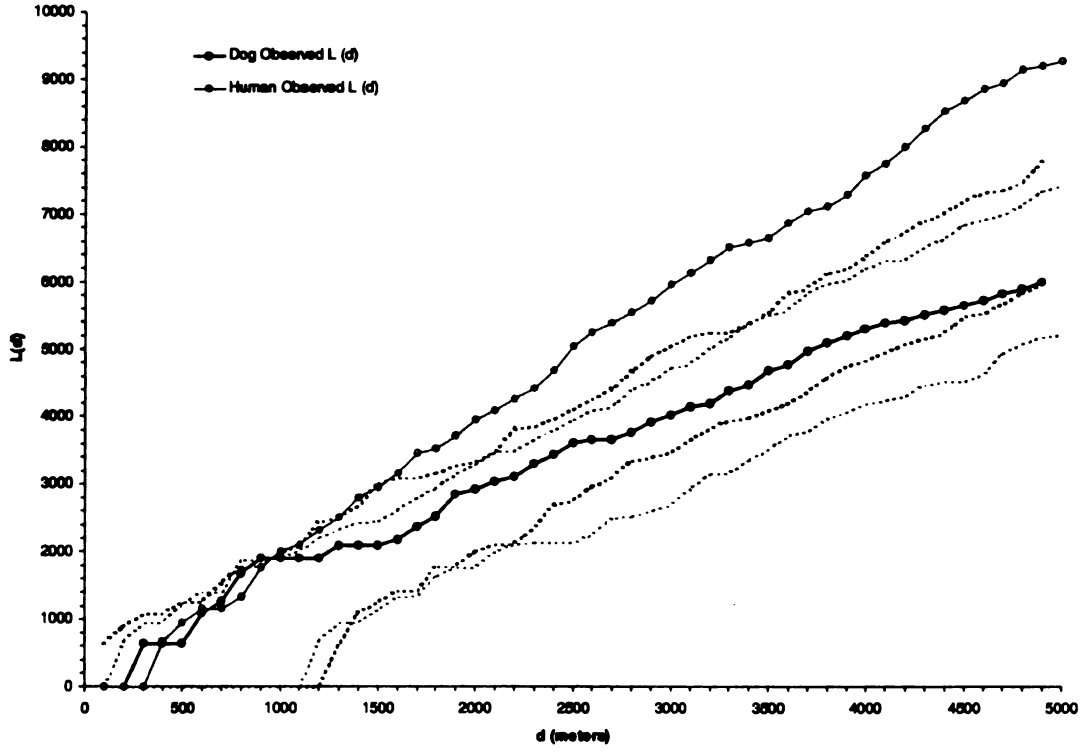

Dog Observed L (d)
Human Observed L (d)
L(d)
d (meters)
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000

Figure 4.3.

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

Summary and Conclusions

In the introduction to this dissertation, it was noted that studies of disease aggregation in both veterinary and human medicine are fertile ground for new research, albeit for different reasons. Studies of clustering of human diseases, while quantitatively advanced, have often concentrated on identifying causal factors of disease and linking them to environmental point sources. The draining theoretical and methodological rigor of that pursuit has meant that the use of clustering studies for descriptive purposes aimed at hypothesis generation has often been overlooked. Research into disease aggregation in veterinary medicine, and especially aggregation of companion animal diseases, by comparison, has received so little intellectual effort as to still be effectively in its period of genesis. Any research performed in this sphere where so little previous work exists has the opportunity to be of fundamental importance. Also, despite research activity initiated in the late 1960s and early 1970s to scrutinize the simultaneous occurrence of cancer in humans and companion animals, there has been little new investigation. The original efforts of Dorn and colleagues have been largely neglected over the intervening thirty years. Though many have speculated on the links between human and animal cancers of analogous biology, few have taken that fascination to the next level, that of interspecies comparisons based on simultaneous epidemiological study.

Recall that the objectives of the current study were to use methods novel

to veterinary medicine, methods which did not rely upon rate calculations, to describe and compare the spatial and temporal distributions of some biologically similar neoplasms in dogs and humans, assess whether those spatial distributions were independent between the species, and formulate some hypotheses to explain the findings. As has been seen over the course of this work, these objectives have been met. First, it has been shown that two widely applicable point pattern analysis methods used primarily in geography and ecology (second-order [*K* function] analysis and one-dimensional nearest-neighbor analysis), can be successfully applied to summary data accumulated in veterinary hospital case records and computerized registries of animal and human cancer cases. These methods found statistically significant geographic clustering of cases that varied by area, by species, and by type of cancer. The ability of these techniques to describe disease aggregation, now clearly demonstrated by this study, circumvents some of the long standing difficulties involved with enumerating populations-at-risk for companion animals, problems associated with calculation and extrapolation of incidence rates for small geographic areas, and the expense and difficulty of establishing and maintaining local pet censuses and companion animal disease registries over time. But to understand the full importance of these results is to realize that their use also shows potential application for similar investigations in human populations where it is not possible to enumerate populations-at-risk. As epidemiologists, our enviable position in societies where vast amounts of human population data are available may cause us to assume that such information is ubiquitous. But

on deeper consideration, it becomes obvious that vast regions of the world lack such data. Populations-at-risk may be impossible to enumerate in much of the third world. Moreover, history has shown that nefarious political motives have led some governments to suppress data for some at-risk population subgroups as a means of depriving them of essential health services. Indeed, even within our own supposedly advanced society, human population subgroups exist wherein the enumeration of an at-risk population may well be just as difficult as it it is for companion animal populations. Populations of urban homeless people, homosexuals and substance abusers come to mind as examples of such at-risk groups. Thus, techniques such as those demonstrated here on elusive animal populations for which little good data are available may have immediate relevance for human public health investigations as well. Commonalities such as these point to the inextricable linkages between human and animal health, and the importance of exploring them.

Second, it has been demonstrated that *K* function analysis can be used as a tool for the simultaneous study of cancer (and potentially many other diseases) in companion animals and humans, to reveal the dependency of the underlying processes which determine the geographic distribution of cases of disease. Even though the identification of the actual processes causing the diseases was, by design, not attempted or accomplished here, the demonstration of a broadly applicable method to determine significant interspecies dependencies allows delineation of those diseases for which the underlying processes are not acting independently from those where they are.

Such knowledge will allow limited research resources to be applied to the study of cancers and other diseases which are most likely to ultimately yield evidence of dependencies between the pathogenesis of disease in dogs and humans. It is those diseases, in turn, which may be most likely to mark the dog as an animal model, or as a sentinel species, of the human disease. Here, dogs had been previously identified as animal models for all the human cancers chosen for study, so it is not surprising that the spatial patterns in the two species were found not to be independent. But more importantly, now that the utility of these methods has been conclusively demonstrated, they may be applied in the future to diseases other than those studied here to explore whether the dog (or other species, for that matter) may be useful as an animal model or sentinel for those diseases as well. Thus, the demonstration that these methods are efficacious is of fundamental importance not only for its original contribution to the veterinary epidemiological literature, but also because it establishes a widely-applicable approach for screening population data to identify potential sentinel animals. If upon future application in other situations these methods suggest that the underlying processes driving the disease patterns in some animal species and humans are *in*dependent (in contrast to the findings here), then, clearly, there would be little reason to pursue that species as sentinels.

Third, these methods have been successfully used to generate (or cast doubt upon) hypotheses regarding the processes which underlie the spatial patterns formed by cases of canine and human neoplasia. These results now allow us to hypothesize that the spatial distributions of these biologically similar

cancers (and the processes that determine them) in dogs and humans are not independent of each other. The results of this study now also allow us to hypothesize that, at least for the cancers, areas and time periods studied, variations in human population density cannot always be assumed to explain the geographic distributions of cancer in dogs as has been previously suggested. Similarly, the spatial distributions of cancer in both species suggest the hypothesis that no single etiological process invariant to local environmental factors is determining the spatial occurrence of cancer. Based on the results of the current work, it can now also be hypothesized that the processes which determine aggregation of canine cancer do not appear to be uniformly related to such demographic factors as median household income, percent female population, median age or percent home ownership. Since all of these factors had been associated by previous studies with pet ownership, this study provides little support for the hypothesis that factors associated with pet ownership necessarily determine the geographic distribution of canine cancer case diagnoses. In addition, these results, though admittedly exploratory, suggest the hypothesis that the processes which result in the significant clustering of both human and canine neoplasia, where they occur, operate at a distance scale of less than 2000 meters. There is now evidence allowing us to hypothesize that those processes operate on the canine cases more strongly at shorter distances than they do on human cases living in the same areas during the same time periods. With respect to clustering in time, the study results now allow us to hypothesize that there is broad agreement between the processes causing

similar types of cancer in dogs and in humans to either cluster or disperse into a regular pattern over time. However, these results provide little support for the hypothesis that cases of similar cancers in dogs and humans cluster in time in the same areas. Finally, from the purely descriptive characteristics of the canine cases, the present study showed general agreement with previous descriptive cancer epidemiology of dogs, although it does suggest, not surprisingly, that hypotheses regarding breed predispositions for certain neoplasms that are based on observations of total case frequencies are apt to be misleading. Taking all of these findings into account, in the author's view, it can be stated quite defensibly that the declared goals of the study have been accomplished.

Recommendations for Future Research

However, no study stands long solely on its own merits, nor is any study without its limitations. This is perhaps nowhere more true than in epidemiology. Much of the joy of scientific inquiry lies, paradoxically, not in reflection on the results of a successful study, but in putting the new questions that those results inevitably raise into context, and framing them as subjects for new research. In addition to the future research avenues discussed at the end of Chapters 2 and 3, some others warrant attention.

Turning first to the joint spatial distributions of the studied cancers in dogs and humans, it has been observed here that across all eight cancer by county combinations studied, the spatial patterns of the canine and human cases' residences, and so the underlying processes which determine them, are not

independent of each other. Although this study, by design, does not allow us to say what those processes are, it does apparently show, perhaps for the first time recorded in the literature, what has long been speculated, namely, that the forces determining the geographic distribution of cancer in man and animals do not operate in isolation from one another. The clustering of human cancer cases in the direction of the canine cases implies a non-random spatial relationship between dogs and humans with the same cancer. However, what was *not* studied here, and so cannot be assessed, is whether the patterns of residences of humans without these cancers also have a non-random spatial relationship with the canine cases. If they do not, it would suggest the clustering of human cases towards the canine cases observed here is likely due to the disease process itself rather than some unrelated factor. That would be a remarkable finding indeed, with important implications for the dog as both an animal model and as a sentinel. Thus, one research approach now indicated is a replicatory study using these same methods on other data sets, a study which performs these analyses not only on residence addresses of cases, but also on addresses of non-diseased controls. In terms of methodology, such a study would employ methods identical to those discussed in Chapter 4, but compare additional pairwise combinations not studed here (say, dog cases vs. human controls, dog cases vs. dog controls and human cases vs. human controls). The results of that research would go a long way towards validation (or refutation) of the exploratory findings noted here. In addition, repetition of the methods used in this study is warranted to address questions concerning the variations in results

that might be seen in studies using other animal species as subjects, or studying other diseases. Would similar or unique relationships emerge if one studied other companion animals such as cats, or horses, humans having, in general, decidedly less exposure to the latter? How would the patterns found here change if zoonotic diseases were studied? Could these methods prove useful for studying diseases for which a zoonotic relationship has been postulated but not proven, *e.g.*, for bovine paratuberculosis and human Crohn's Disease? Using these methods to examine many different kinds of disease relationships may provide insight into whether some of the relationships hypothesized as a result of this study are an artifact of the method, or valid manifestations of disease attributable to particular etiologies. The wide applicability and relative inexpensiveness of the methods should facilitate the accumulation of such a body of studies as well. Clearly, this is an area ripe for new study.

It also plain from comparisons of the strength and statistical significance of the clustering patterns observed within each of the species, that the underlying processes determining the geographic distribution of cancer in dogs and humans, while not independent, do not effect the two species equally either, and vary by area and type of neoplasia. A multifaceted question raises itself here. What sorts of processes cause clustering of cancer case residences, do not vary independently between species, yet do vary by kind of neoplasm and area of study? Although speculative, it seems plausible that it is a combination of contributing factors at work, some of which may be absent from some areas, or which may effect one species more than the other in particular areas. It also

seems plausible that both biological and non-biological processes contribute to determination of the geographic patterns formed by the cancer cases. This study did not have the ability to distinguish between those factors which may have affected the likelihood of diagnosis or referral of the canine cases, as opposed to those which were actual etiologic factors for cancer. This leaves a knowledge gap begging to be bridged. Clearly, efforts to distinguish and define the nature of these underlying processes are another area where new research might be concentrated. As mentioned previously, replication of this study using a study population that is gathered from a larger number of animal clinics representing a wider geographic base, where the potential for referral biases are apt to be much smaller, would seem to be a good initial approach.

One fascinating potential subject for future research is investigation of the hypothesis that variation in the significance and intensity of interspecies clustering by area is related to differential sensitivity to a common etiologic exposure. Such a hypothesis would be a coherent explanation of how the spatial processes determining case clustering could at once be both non-independent, yet vary so much. It would also be consistent with the present study's finding that clustering tended to be stronger at shorter distances in dogs than in humans living in the same area, despite operating below an upper distance limit of 2000 meters in both species. If such a differential sensitivity were found, and dogs were more sensitive than humans (which would be consistent with many of the results of this study), it would be evidence of a characteristic essential to a sentinel animal, namely, greater sensitivity than

humans to an exposure held in common.[1] Methodologically, one might approach such an investigation with a study that postulates (based on the literature, say) a particular exposure as a risk factor for one of the cancers studied here. By gathering data on that factor, and stratifying the study population into some ordinal strata based on it, one could repeat the analyses conducted in this study within each strata, examining the results for evidence of variation in spatial patterns across levels of the exposure factor. Such would be an initial, though perhaps crude, approach to assessing a dose response relationship. As a variation on this approach, it might also be interesting to use these methods to study a rare human cancer with known and relatively specific environmental risk factors (*e.g.*, mesothelioma and asbestos exposure, hepatic angiosarcoma and vinyl chloride exposure, *etc.*) and compare the spatial distribution of the human cases with those of dogs (assuming adequate numbers of cases could be found among the latter).

A related issue of relevance to sentinel development also warrants comment. It has been previously argued[1] by this author that sentinel animals are likely to be most relevant and applicable to the narrow, well-defined exposure circumstances and diseases for which they have been rigorously developed. A more prevalent, though perhaps not explicitly stated, view of sentinel animals holds that animals can always be expected to be more sensitive than humans to a common exposure, and that this characteristic will hold regardless of the specific nature of the exposure or the pathologic outcome observed. Such a view has frequently been voiced by advocacy groups in the

environmental community. *If* the variations in spatial clustering of cases found in this study are attributable to differential sensitivity, then the finding that human cases cluster in some areas for some neoplasms while dog cases do not suggests that this latter view may be naïve. So, further research in the vein begun by this study may potentially shed light on some of the fundamental assumptions underpinning the concept of sentinel animals.

Another area shadowed by uncertainty and lending itself to future research efforts concerns the differences that might occur on repetition of this study with a less highly selected canine case population. As has been noted several times in this dissertation, cases derived from the Veterinary Medical Data Base are a population referred to a teaching hospital. It is quite reasonable to suggest that such a case population may capture a relatively small proportion of the total incident cases of disease, and that those cases may have characteristics which are not representative of the canine population as a whole. Thus, conclusions drawn from such a referral population may not be extrapolated with impunity to other dogs. Such a scenario, if it holds true, essentially constitutes a kind of selection bias. It has been argued in Chapter 2 that the nature of such a bias is likely to be less pronounced for neoplasia than for other diseases more easily diagnosed and treated by general veterinary practitioners. Indeed, the finding in the PCOP study that some 92% of owners reported willingness to pursue diagnosis of cancer in their dogs[2,3] suggests such an argument is not unreasonable. Yet it must also be admitted that the absolute existence and magnitude of such a selection bias in this study, and the effect it

might have had on the results, if any, is unknown. As such, new research might well be directed towards replicating this study using a more broadly-based canine population, perhaps a population where the cases were drawn from accessions to a diagnostic laboratory receiving histopathological submissions from private practitioners across the state. Alternatively, now that veterinary referral hospitals are more numerous than ever before, multicenter studies incorporating data from many such hospitals might help define the existence and characteristics of any referral biases that might be affecting the spatial results noted in this study. While still incapable of accounting for the characteristics of cancer cases that never receive veterinary care, it would still be expected to be more representative of the general dog population than was the study population examined in this study. It would also likely increase sample sizes, and so also help to overcome any unrecognized erroneous conclusions reached in this study due to a dearth of statistical power.

Speaking of power, another recommendation for future research which comes to mind involves the temporal analysis conducted here, of which comparatively little has been said. In this study, there was relatively little evidence of temporal clustering of the cases in either dogs or humans. It was pointed out in Chapter 3 that this was not entirely unexpected, due to the long and variable nature of cancer latency periods.[4] However, it has also been noted[5-7] that various clustering methods suffer not infrequently from a lack of power to detect clustering even when it really exists. Thus, it is also conceivable that the one-dimensional nearest-neighbor analysis employed in this study may

simply have lacked the necessary power to detect temporal clustering. It would be useful and relatively straightforward to execute a simulation study to test the power of the technique under various conditions. To the author's knowledge, no such study has yet been carried out. A final piece of statistical research that seems needed based on the experience gained in this study is the methodological development of techniques to test for statistical significance of concurrent temporal clustering in two or more species. In this study, significant temporal clustering was reported among both canine and human cases of mammary neoplasms in Wayne County, but no method was available to test whether the temporal patterns in the two species were independent or not. Such a technique could prove valuable in future investigations.

These, then, are but a few of the new research considerations issuing from the present work. In truth, the author would be genuinely pleased if the primary role of this work became the generation of questions. Questions beget interest, and interest begets new research and new hypotheses. Irrespective of whether the hypotheses posed in this study are validated, if they lead someone to questions which eventually lead to new insights, by however circuitous a route, they will have served their intended purpose fully and well. And so this dissertation, fittingly, ends where it began, with questions.

REFERENCES

REFERENCES

1. O'Brien DJ, Kaneene JB and Poppenga RH. The use of mammals as sentinels for human exposure to toxic contaminants in the environment. Environ Health Perspect 1993;99:351-68.

2. Lengerich EJ, Teclaw RF, Mendlein JM, Mariolis P, Garbe PL. Pet populations in the catchment area of the Purdue Comparative Oncology Program. J Am Vet Med Assoc 1992;200(1):51-6.

3. Teclaw R, Mendlein J, Garbe P, Mariolis P. Characteristics of pet populations and households in the Purdue Comparative Oncology Program catchment area. J Am Vet Med Assoc 1992;201:1725-9.

4. Whittemore AS, Friend N, Brown BW Jr, Holly EA. A test to detect clusters of disease. Biometrika 1987;74(3):631-5.

5. Besag J, Newell J. The detection of clusters in rare diseases. J R Stat Soc A 1991;154:143-55.

6. Waller LA. Statistical power and design of focused clustering studies. Stat Med 1996;15:765-82.

7. Wartenberg D, Greenberg M. Methodological problems in investigating disease clusters. Sci Tot Environ 1992;127(1,2):173-85.

APPENDIX A

Supplementary Figures, Chapter 3

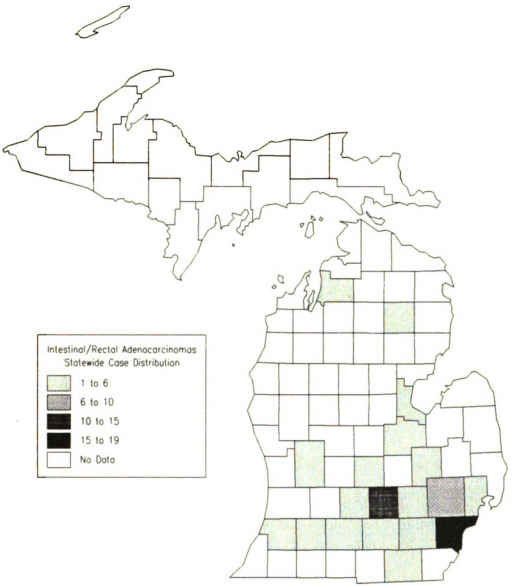


Figure 5.1: Case frequencies by county, intestinal/rectal adenocarcinoma, Michigan, 3/1/64 through 3/31/94

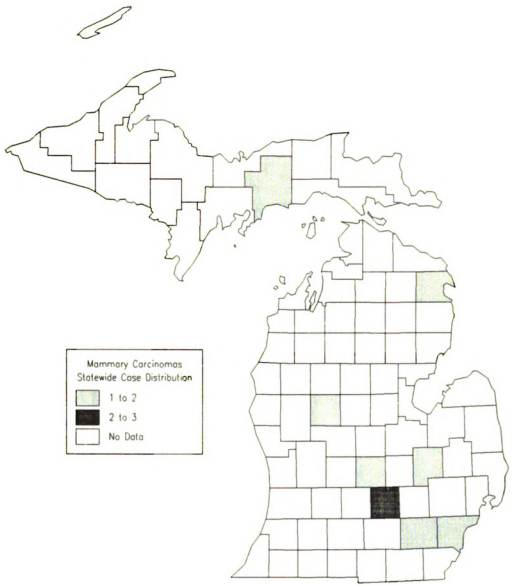


Figure 5.2. Case frequencies by county, mammary carcinoma, Michigan, 3/1/64 through 3/31/94

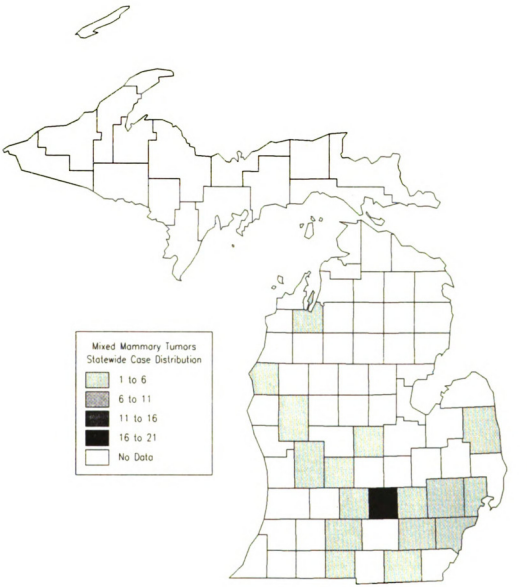


Figure 5.3. Case frequencies by county, mixed mammary tumor, Michigan, 3/1/64 through 3/31/94

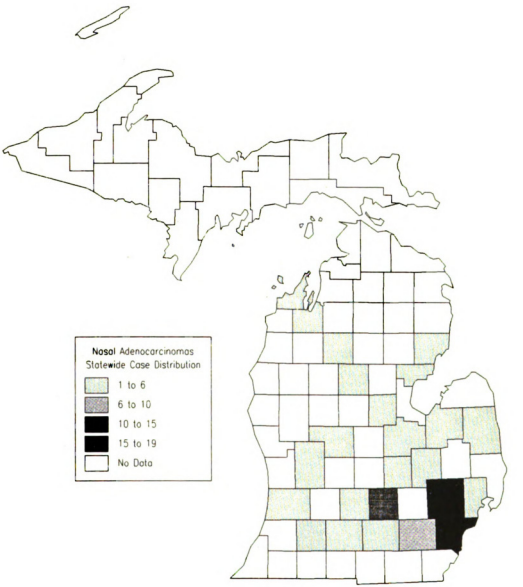


Figure 5.4. Case frequencies by county, nasal adenocarcinoma, Michigan, 3/1/64 through 3/31/94

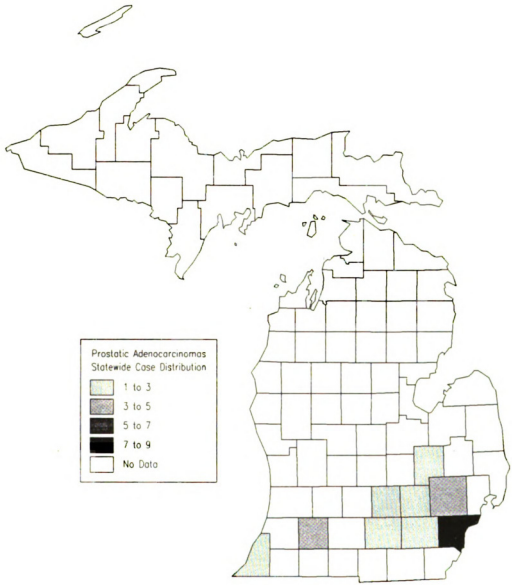


Figure 5.5. Case frequencies by county, prostatic adenocarcinoma, Michigan, 3/1/64 through 3/31/94

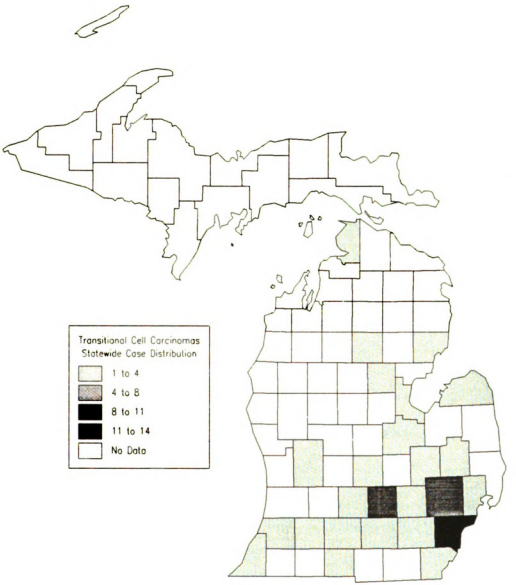


Figure 5.6. Case frequencies by county, transitional cell carcinoma of the urinary bladder and urethra, Michigan, 3/1/64 through 3/31/94

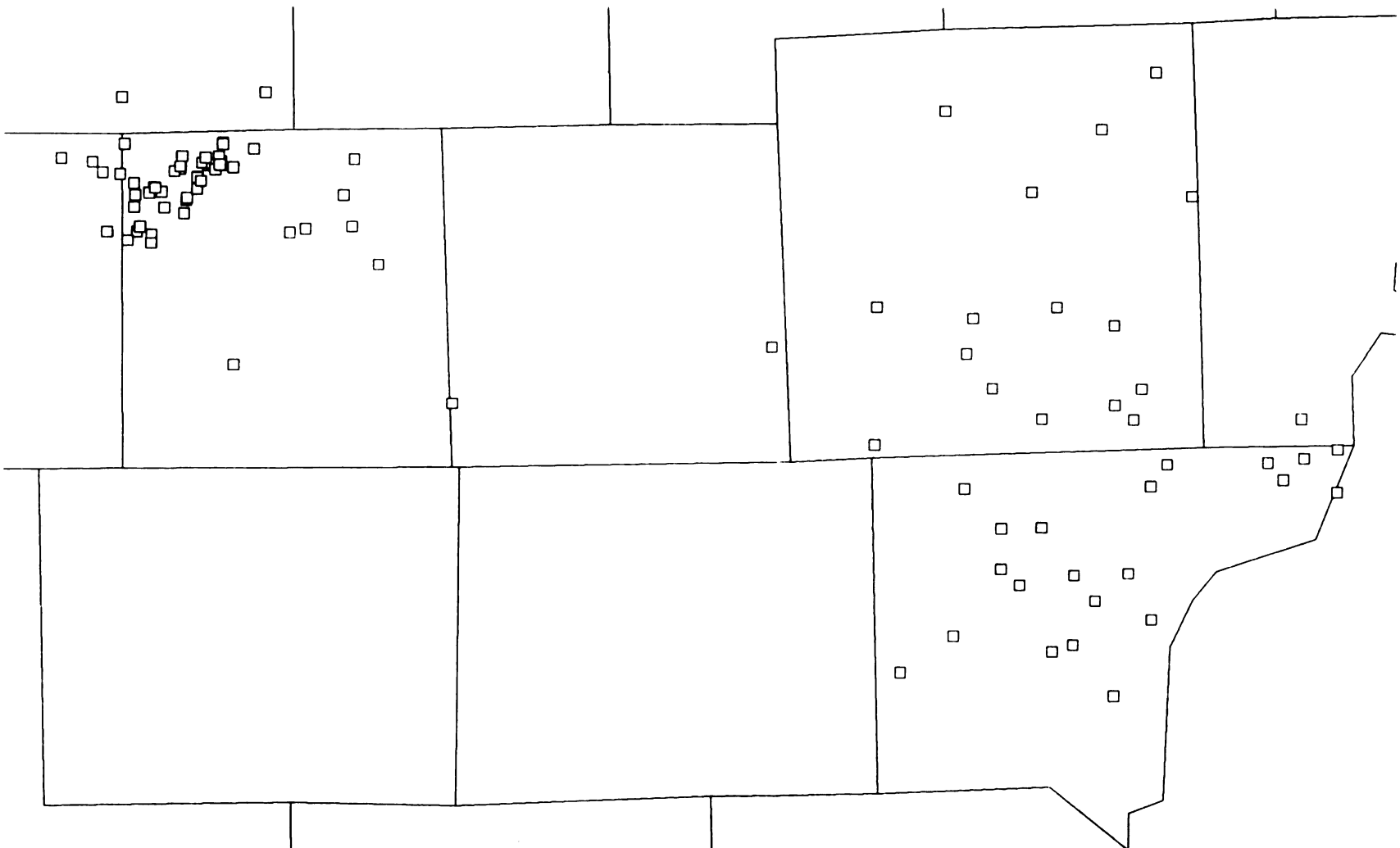


Figure 5.7. Distribution of canine mammary adenocarcinomas in Ingham, Oakland and Wayne Counties, Michigan, 1964 - 1994. Open squares mark locations of cases' residences at the time of diagnosis

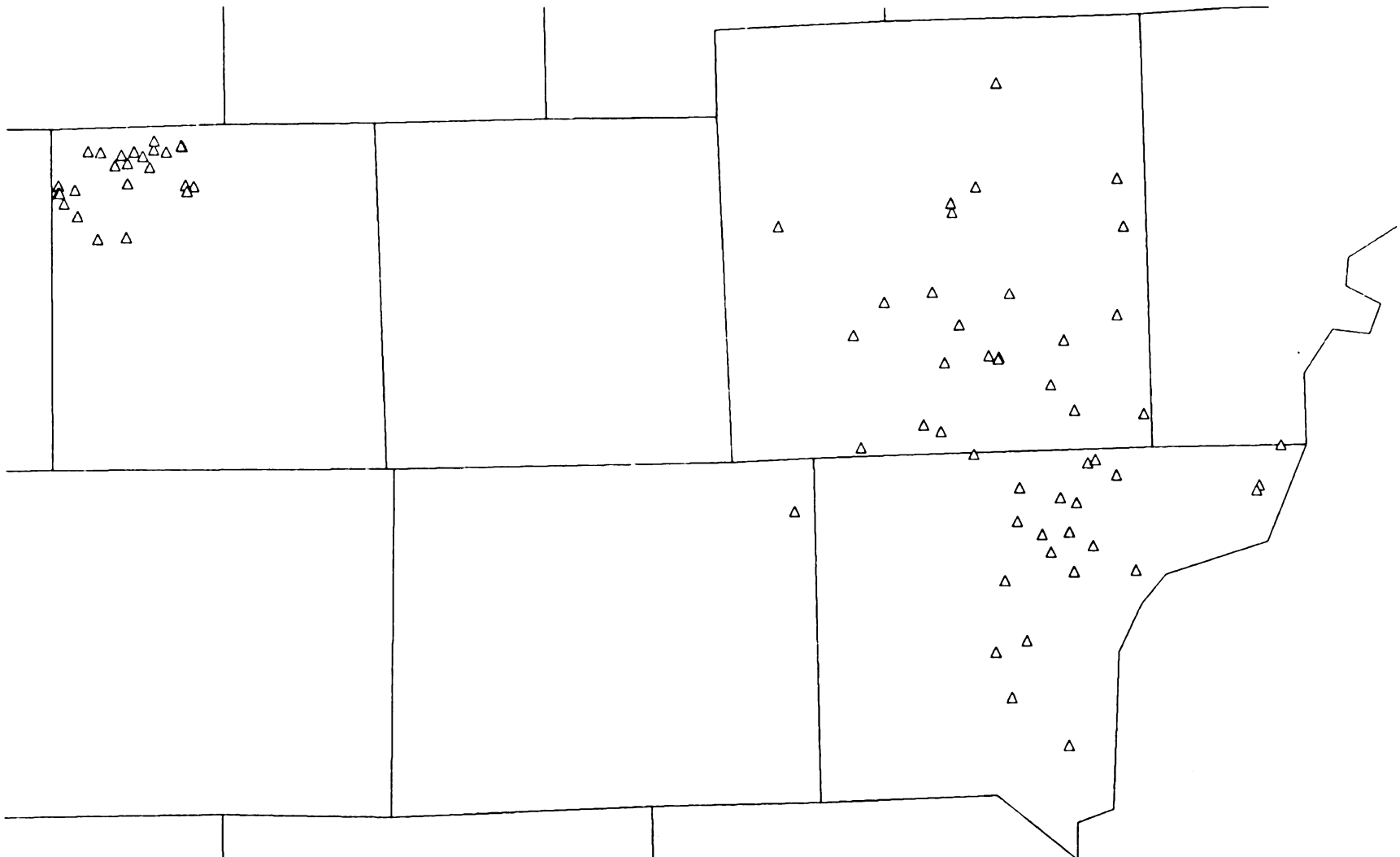


Figure 5.8. Distribution of canine melanomas in Ingham, Oakland and Wayne Counties, Michigan, 1964 - 1994. Open triangles mark locations of cases' residences at the time of diagnosis

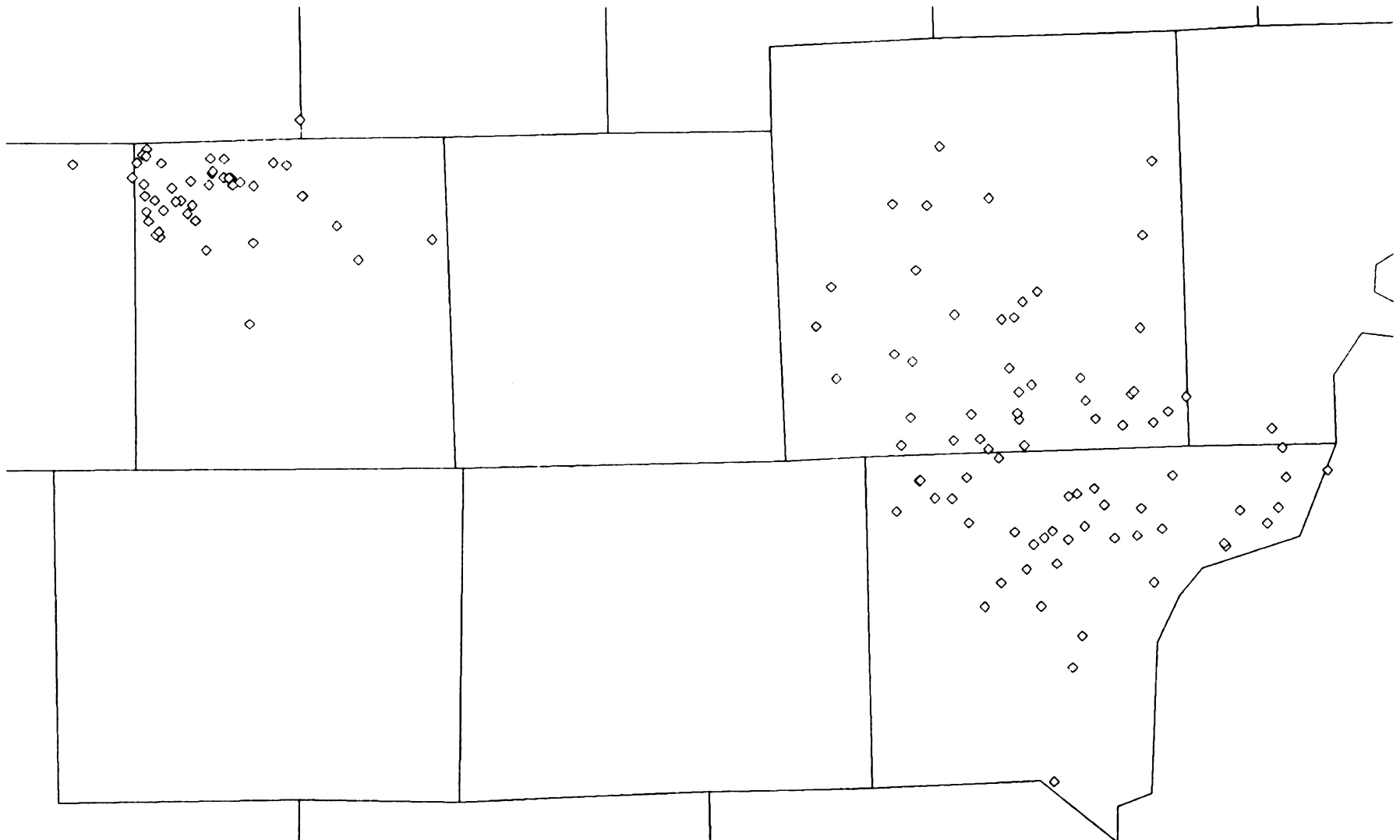


Figure 5.9. Distribution of canine spindle cell sarcomas in Ingham, Oakland and Wayne Counties, Michigan, 1964 - 1994. Open diamonds mark locations of cases' residences at the time of diagnosis

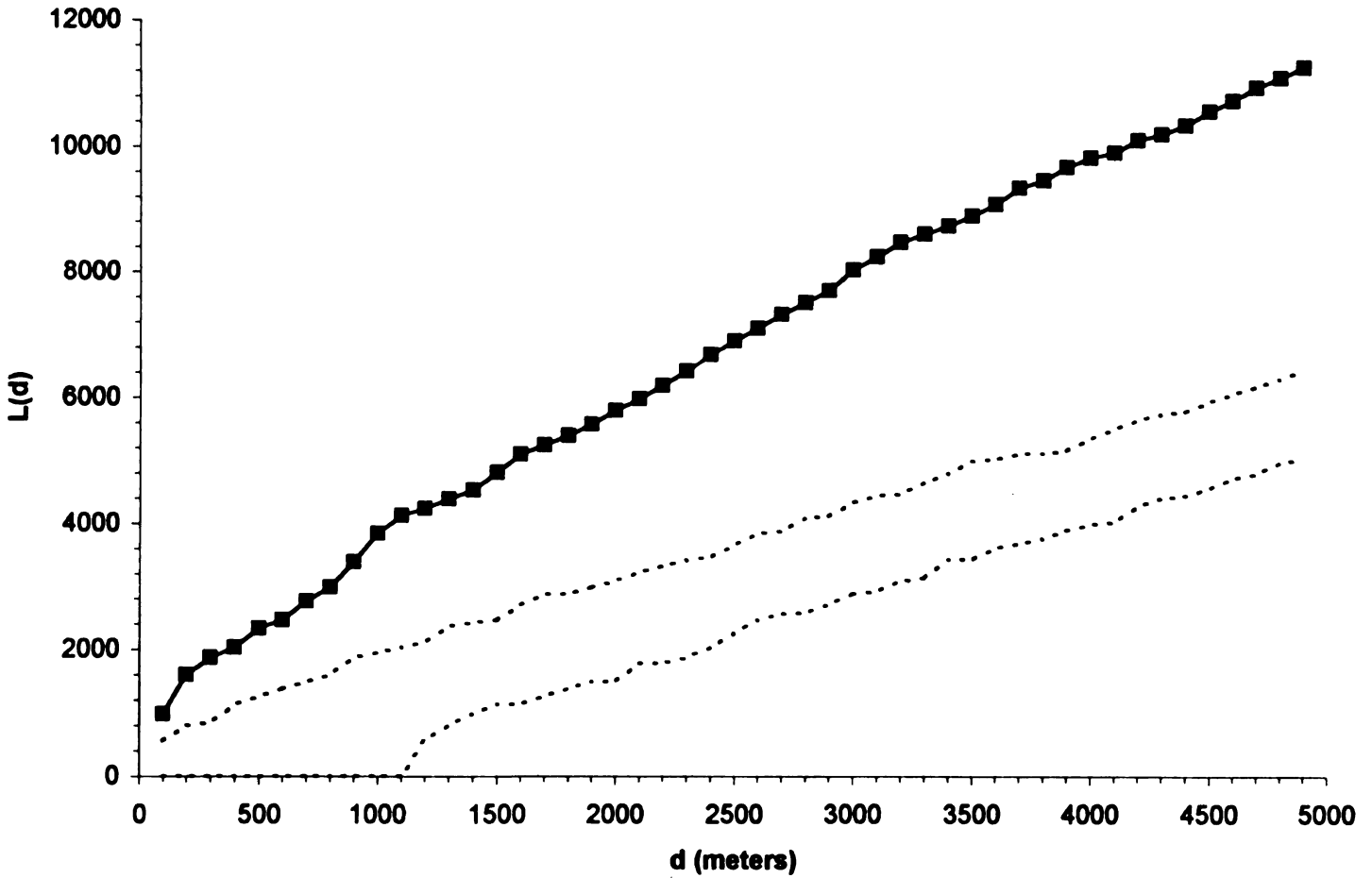


Figure 5.10. *K* function analysis graphical output, mammary adenocarcinomas, Ingham County

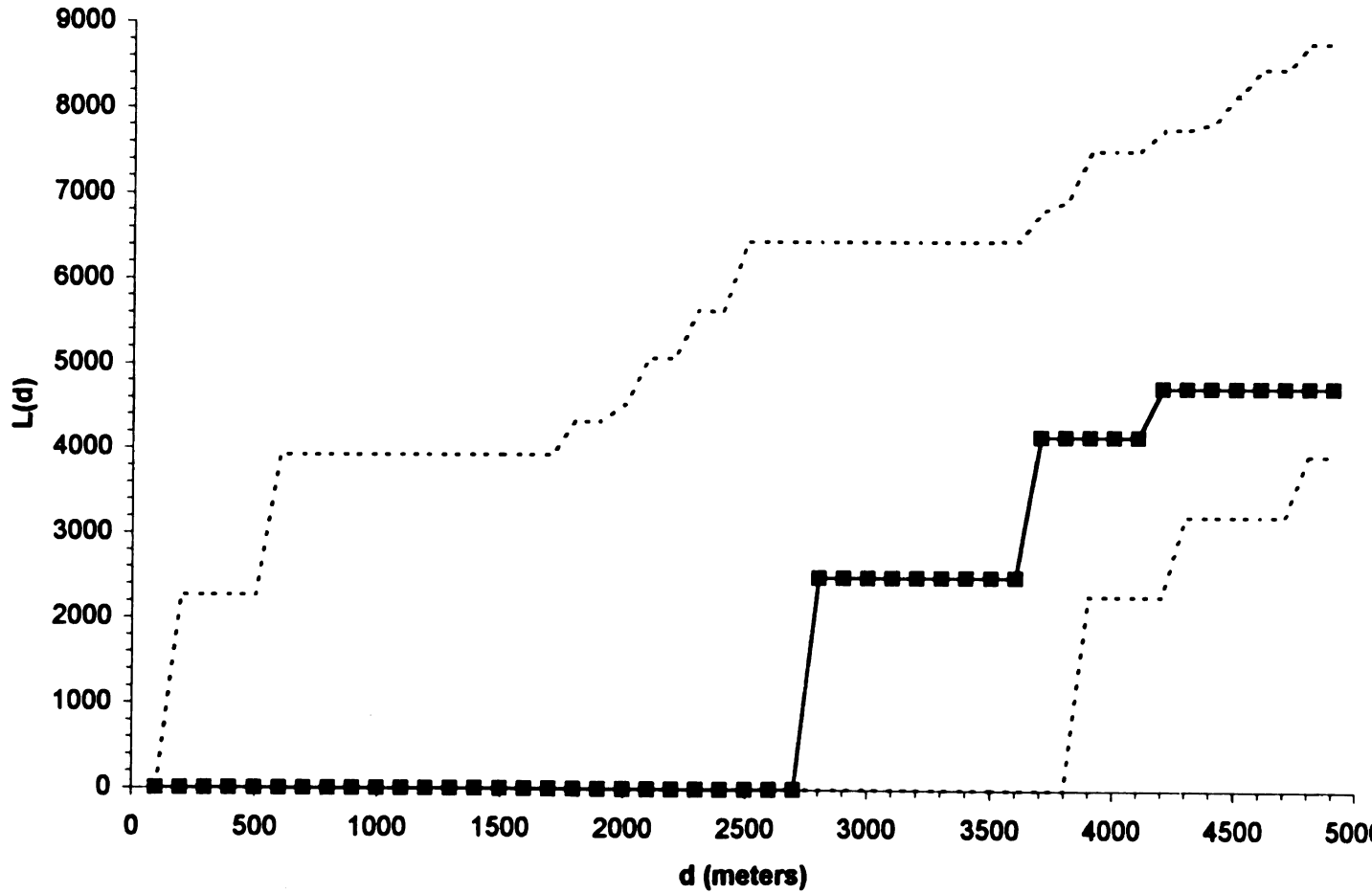


Figure 5.11. *K* function analysis graphical output, mammary adenocarcinomas, Oakland County

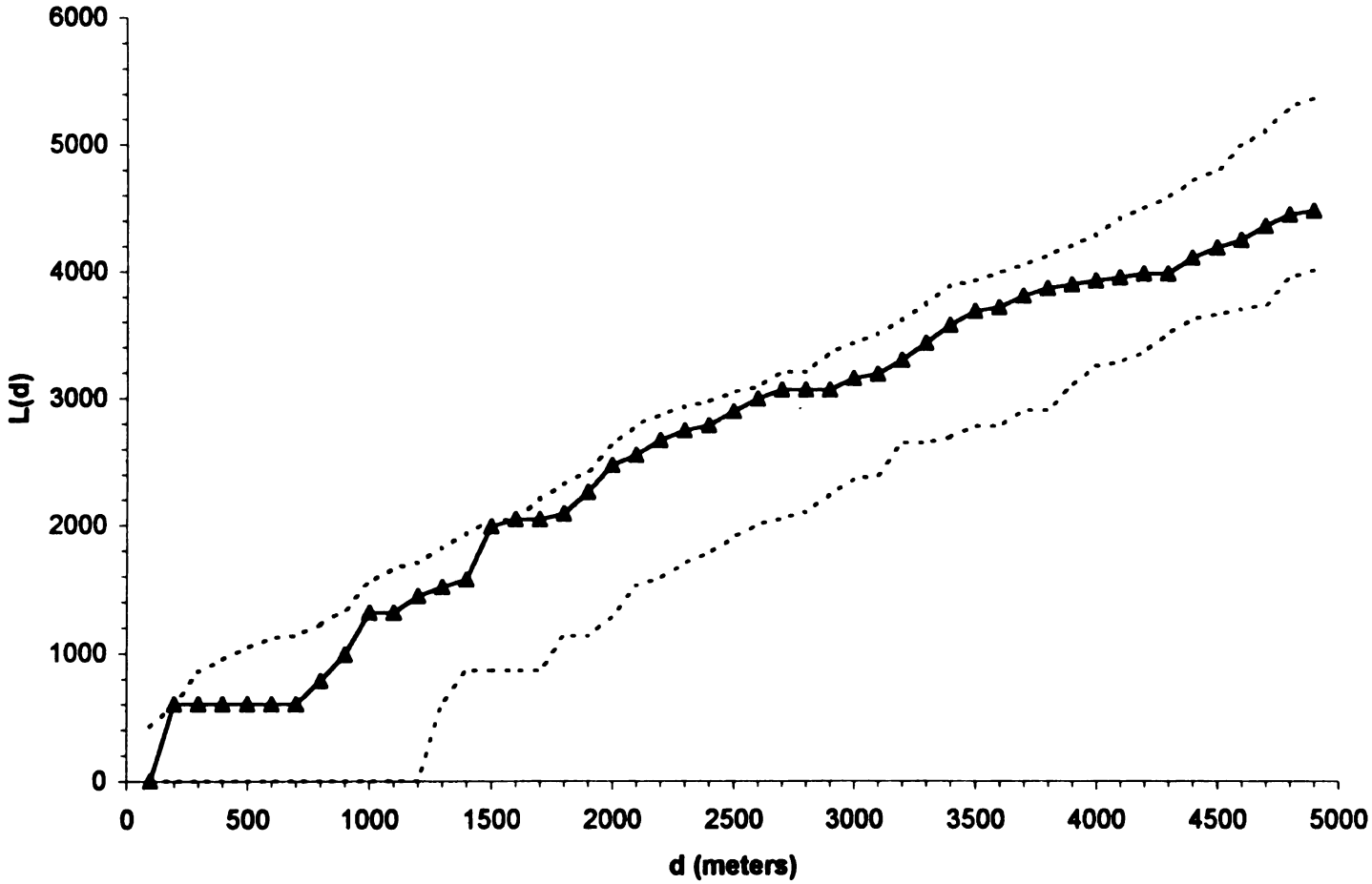


Figure 5.12. *K* function analysis graphical output, melanomas, Ingham County

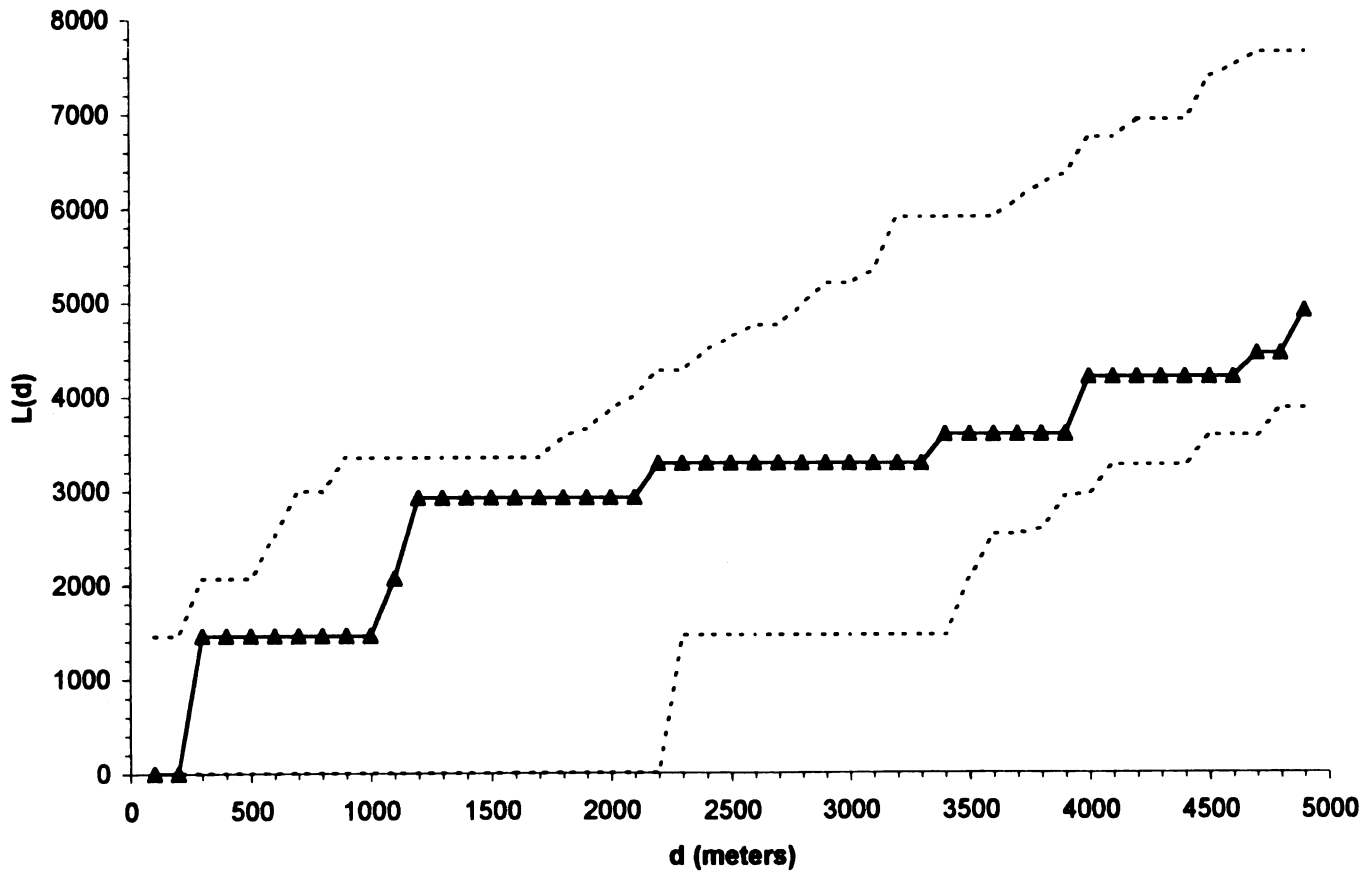


Figure 5.13. *K* function analysis graphical output, melanomas, Oakland County

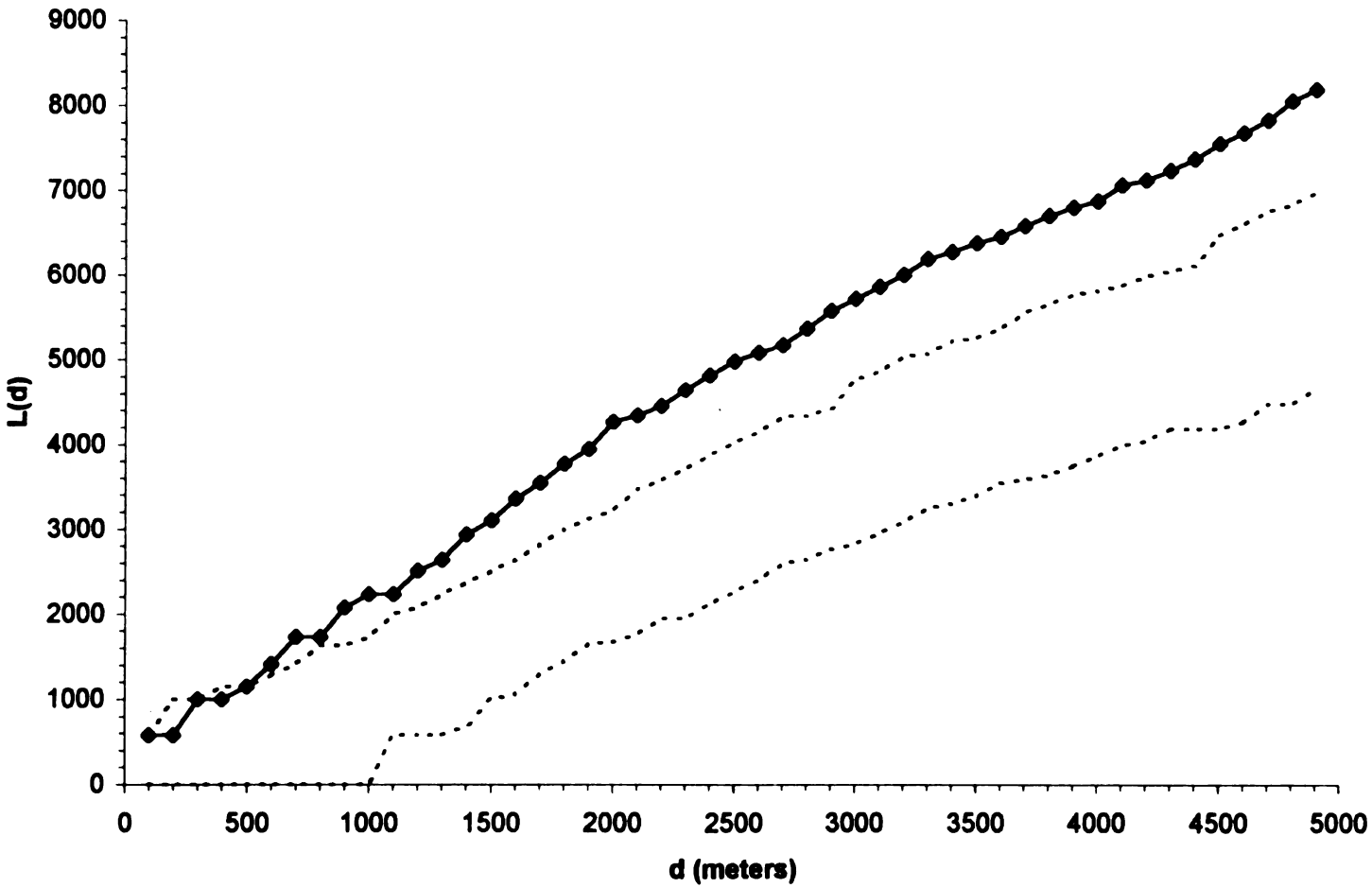


Figure 5.14. *K* function analysis graphical output, spindle cell sarcomas, Ingham County

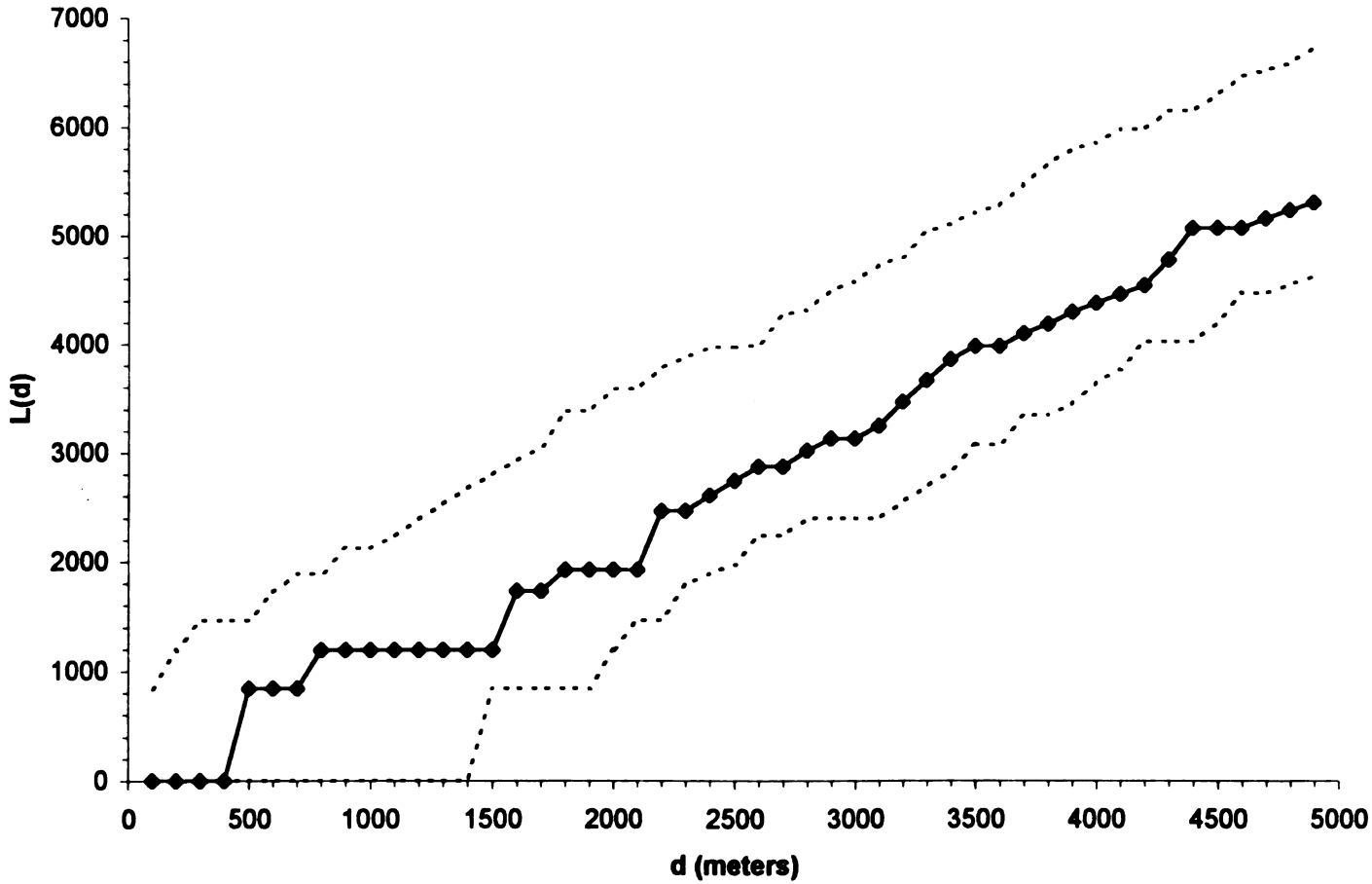


Figure 5.15. *K* function analysis graphical output, spindle cell sarcomas, Oakland County

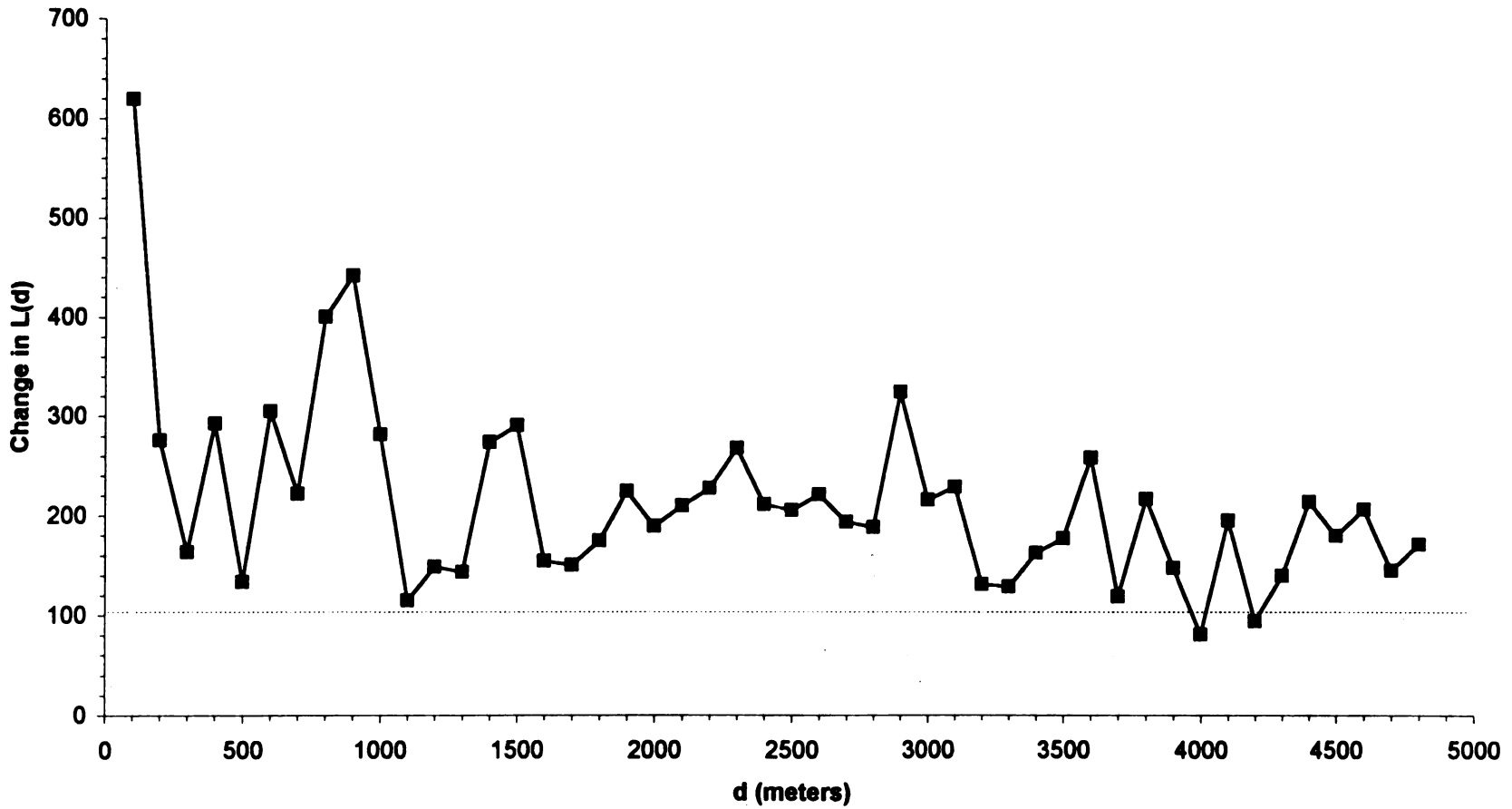


Figure 5.16. Plot of $\Delta L(d)$ vs. d, mammary adenocarcinomas, Ingham County

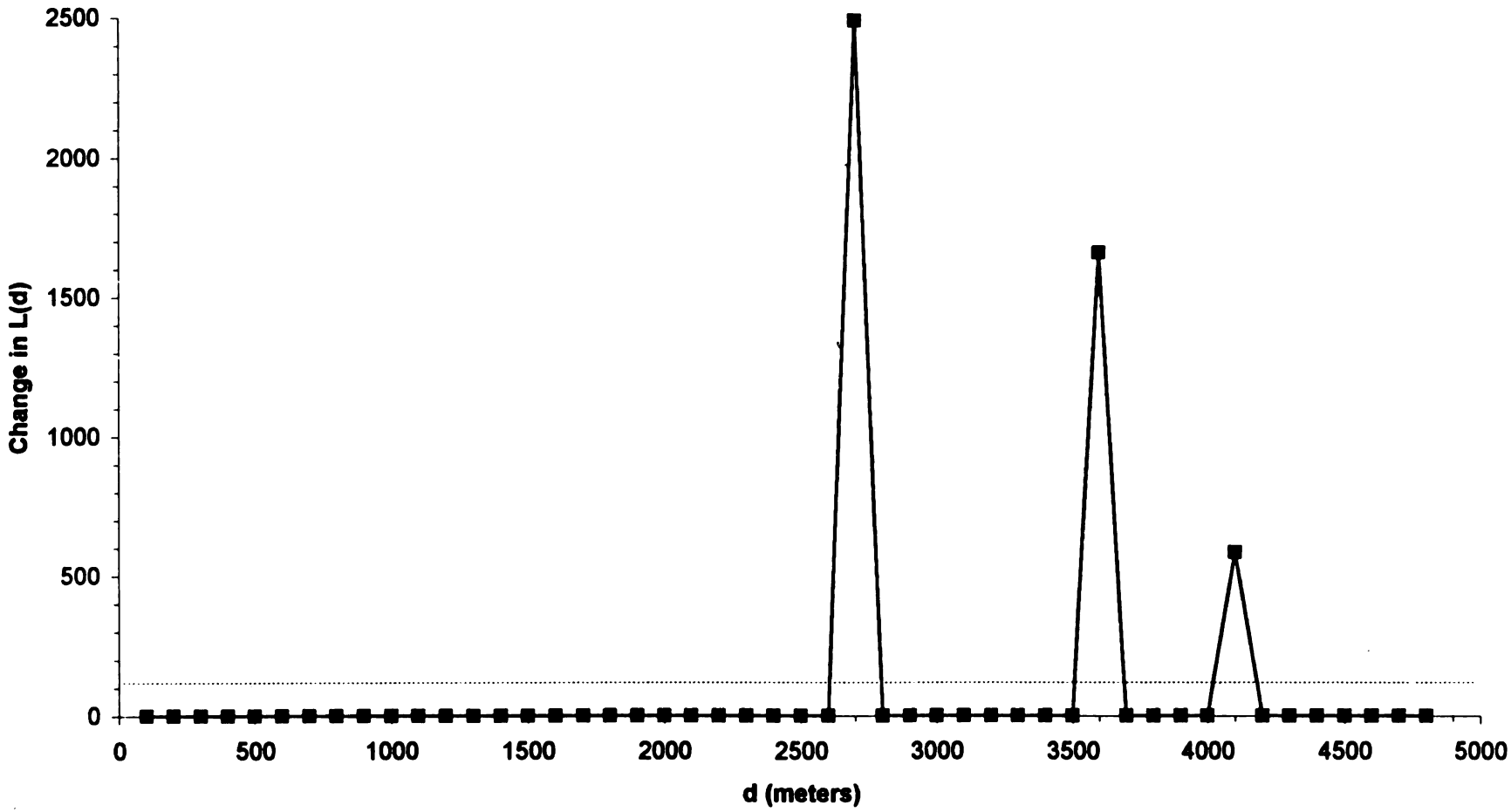


Figure 5.17. Plot of $\Delta L(d)$ vs. d, mammary adenocarcinomas, Oakland County

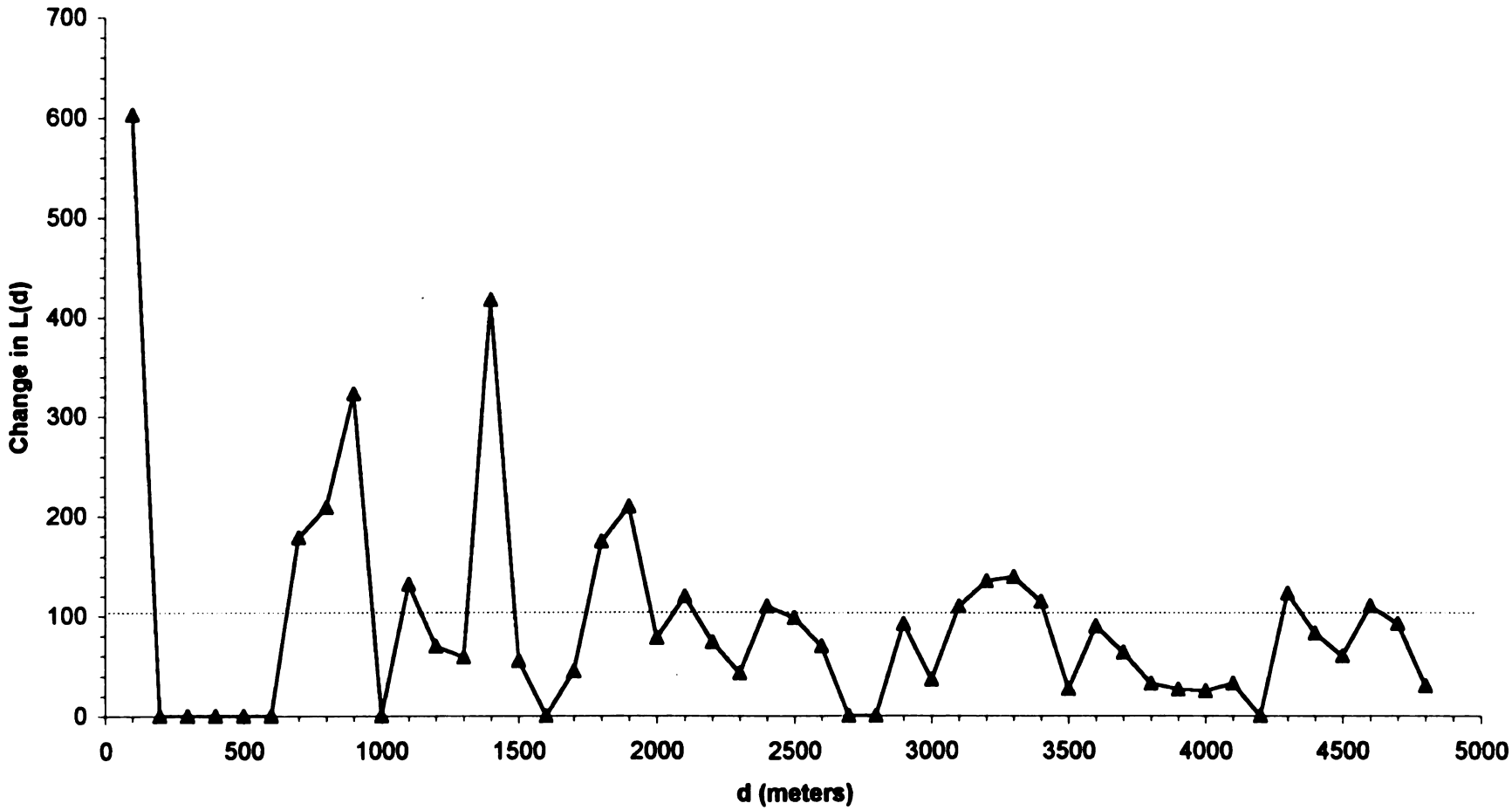


Figure 5.18. Plot of $\Delta L(d)$ vs. d, melanomas, Ingham County

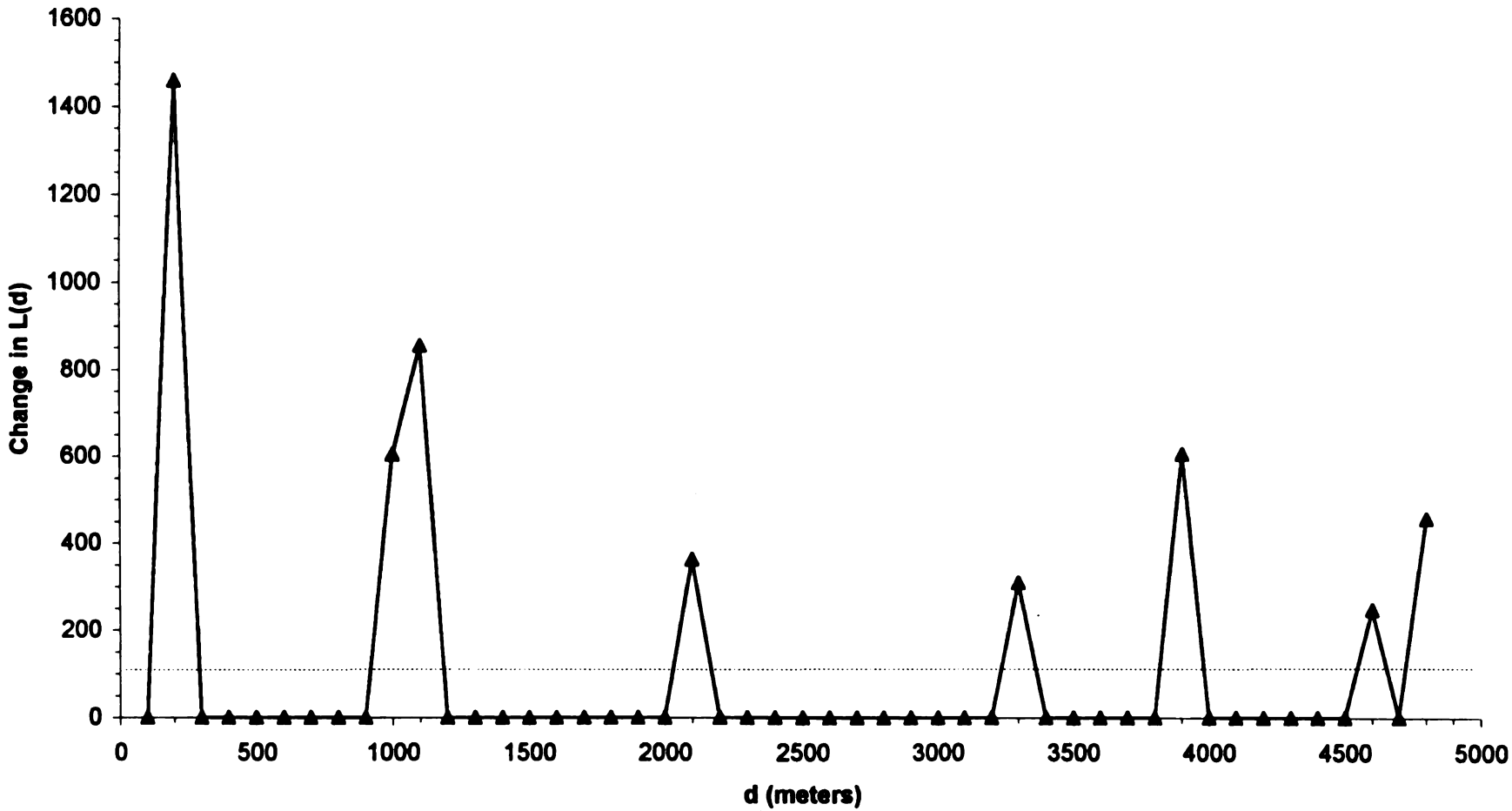


Figure 5.19. Plot of $\Delta L(d)$ vs. d, melanomas, Oakland County

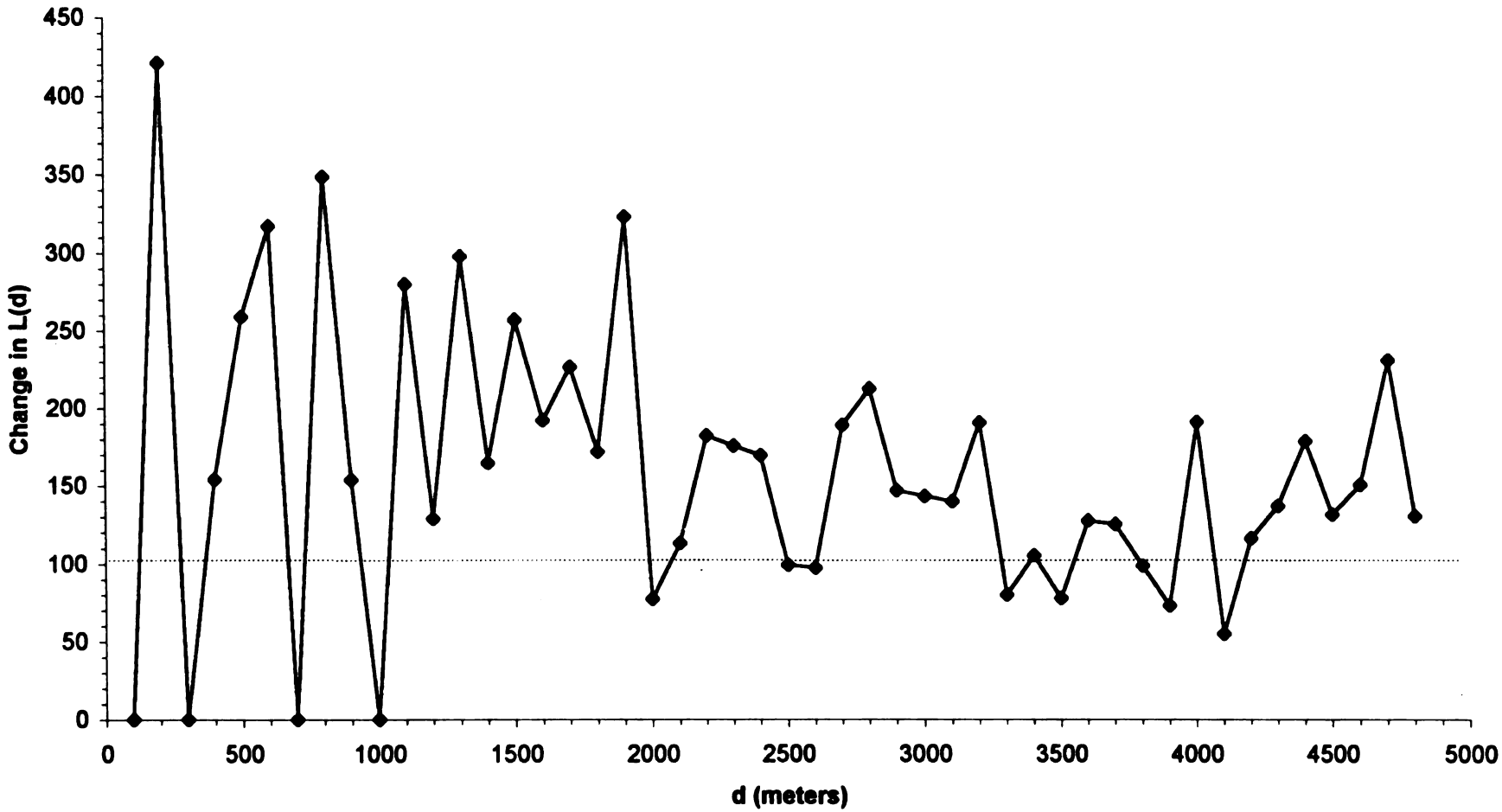


Figure 5.20. Plot of $\Delta L(d)$ vs. d, spindle cell sarcomas, Ingham County

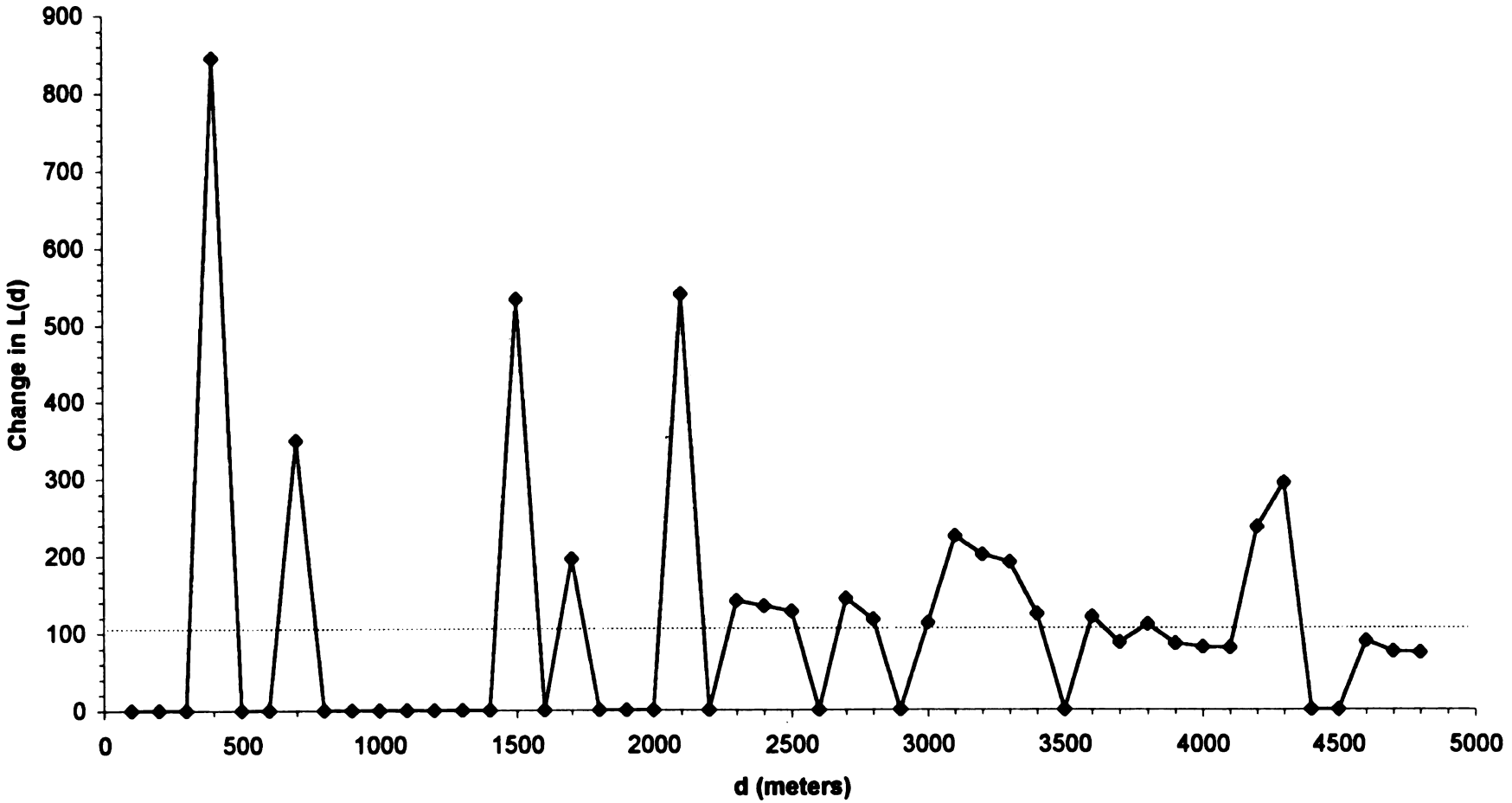


Figure 5.21. Plot of $\Delta L(d)$ vs. d, spindle cell sarcomas, Oakland County

APPENDIX B

Supplementary Figures, Chapter 4

Figure 6.1. Distribution of canine and human lymphomas in Ingham, Oakland and Wayne Counties, Michigan, 1964 - 1994. Open circles mark locations of case dogs' residences at the time of diagnosis, while shaded circles mark locations of human cases' residences at the time of diagnosis.

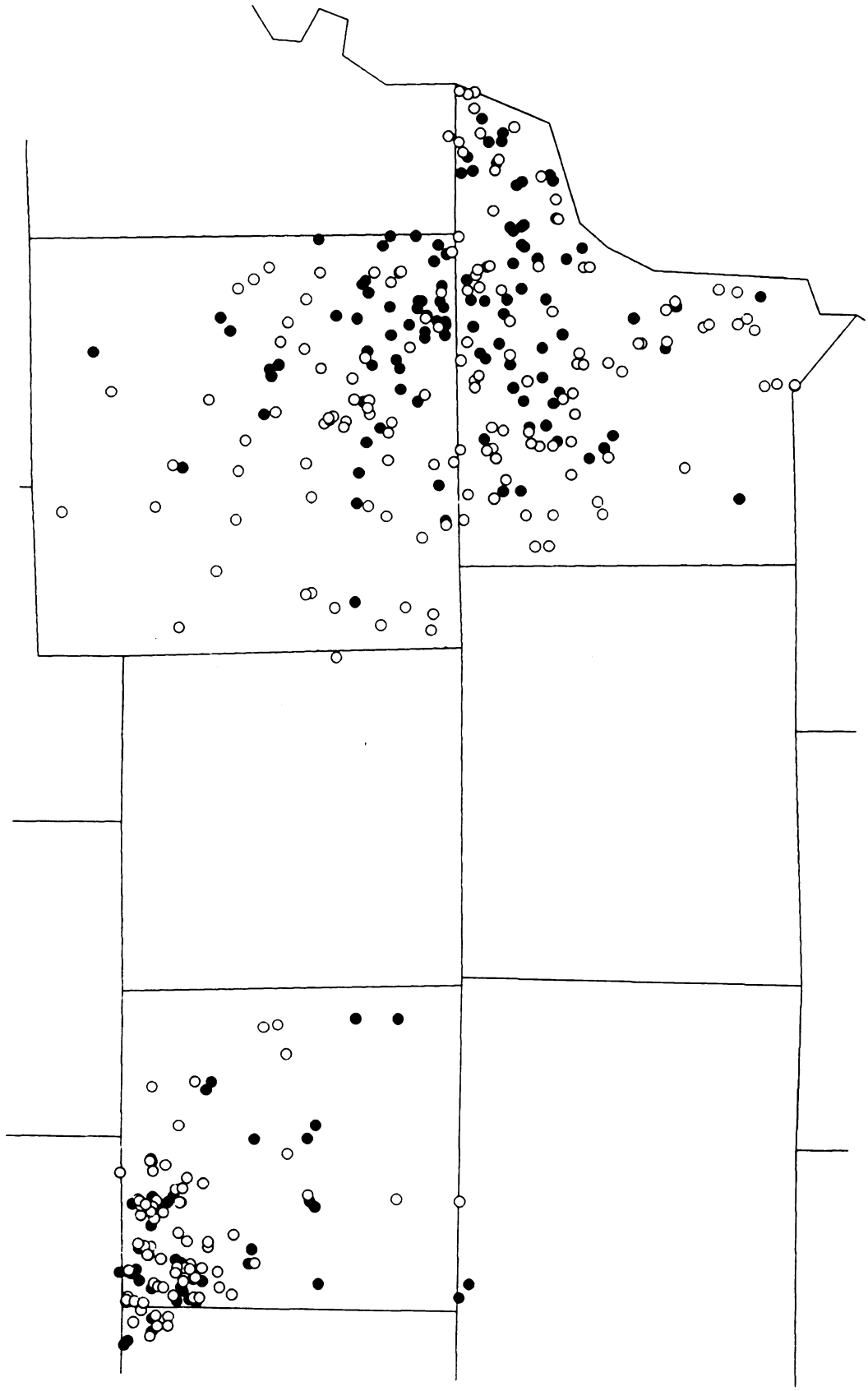


Figure 6.1.

Figure 6.2. Distribution of canine mammary neoplasms in Ingham, Oakland and Wayne Counties, Michigan, 1964 - 1994. Open squares mark locations of case dogs' residences at the time of diagnosis, while shaded squares mark locations of human cases' residences at the time of diagnosis.

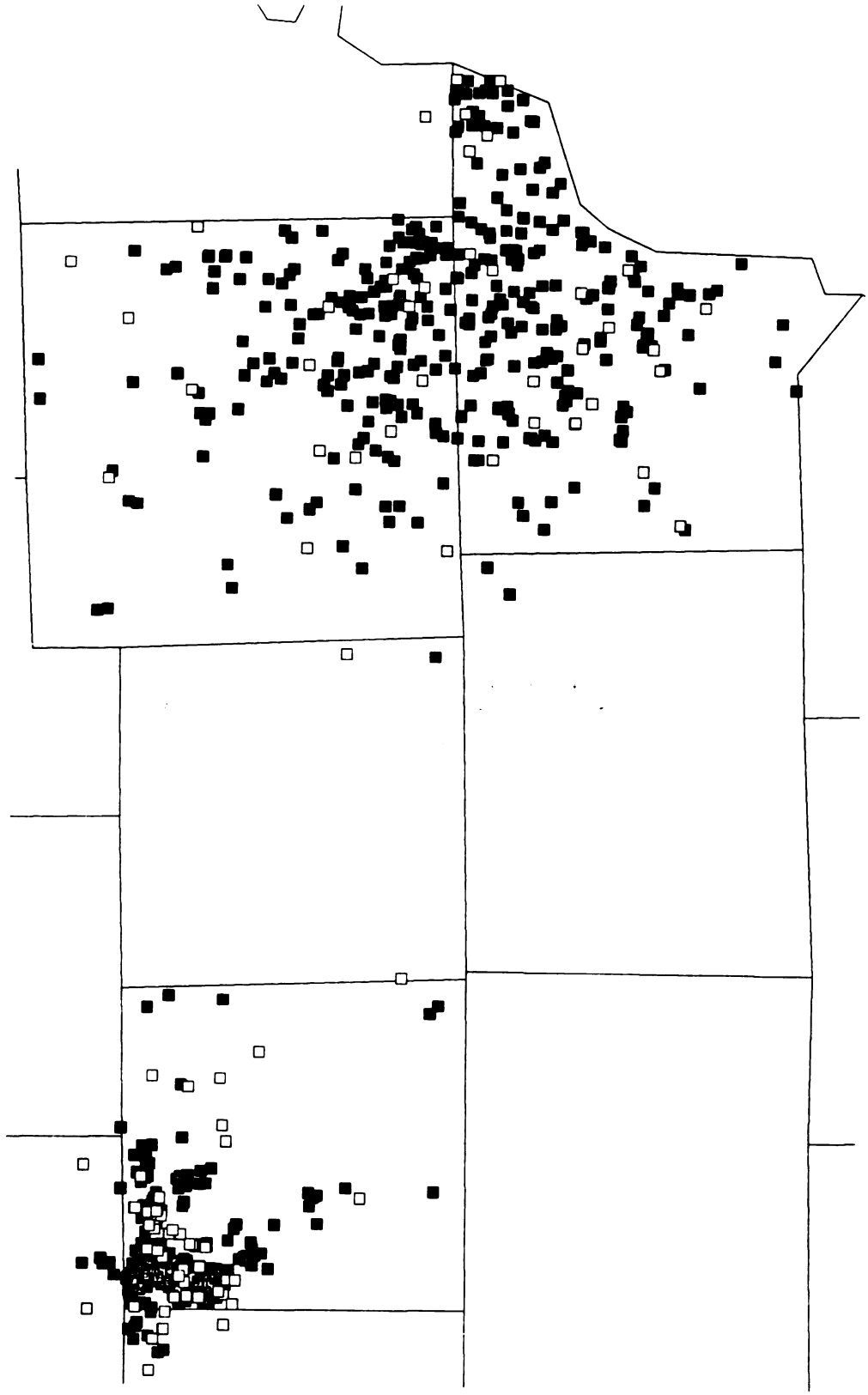


Figure 6.2.

Figure 6.3. Distribution of canine melanomas in Ingham, Oakland and Wayne Counties, Michigan, 1964 - 1994.. Open triangles mark locations of case dogs' residences at the time of diagnosis, while shaded triangles mark locations of human cases' residences at the time of diagnosis.

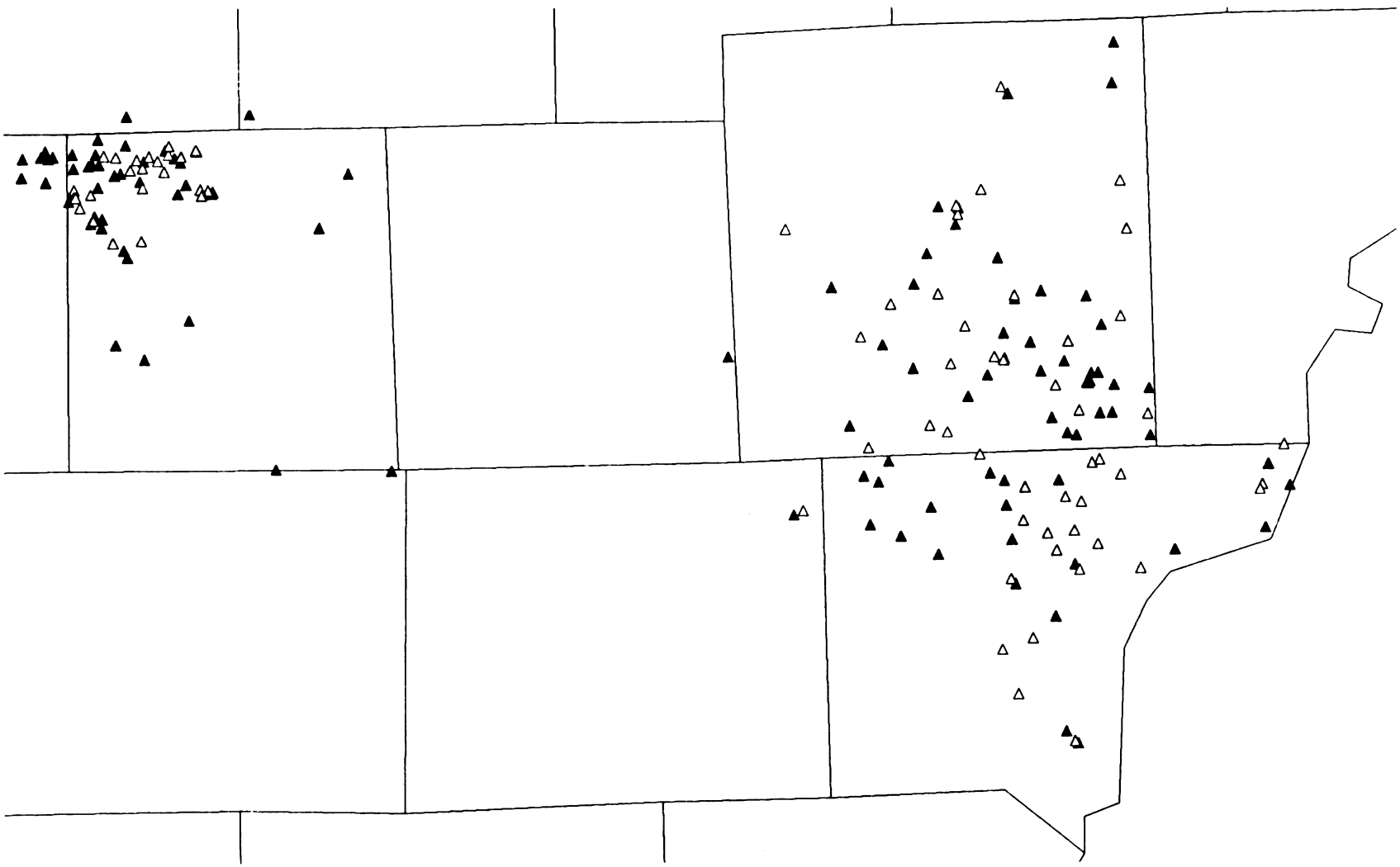


Figure 6.3.

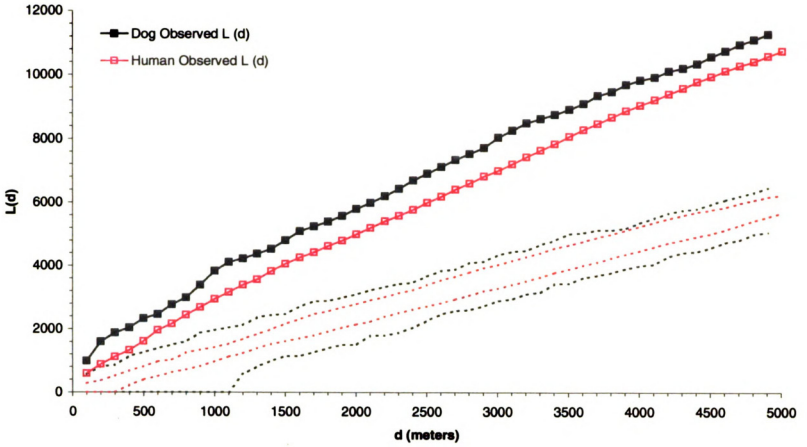


Figure 6.4. *K* function analysis graphical output, mammary adenocarcinomas, Ingham County

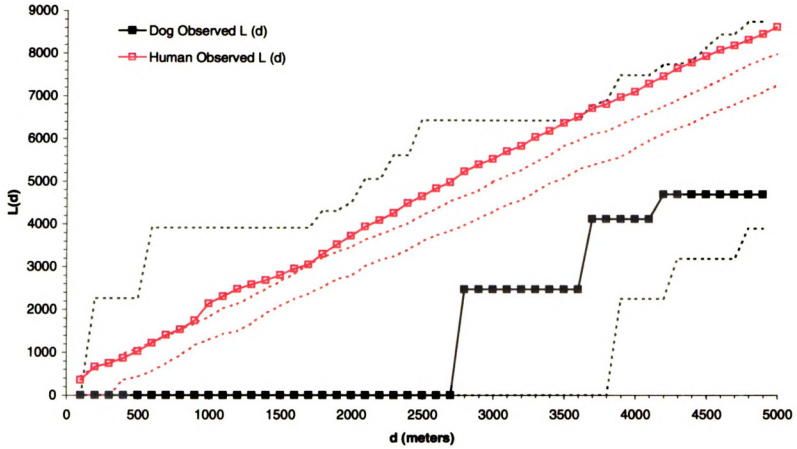


Figure 6.5. *K* function analysis graphical output, mammary adenocarcinomas, Oakland County

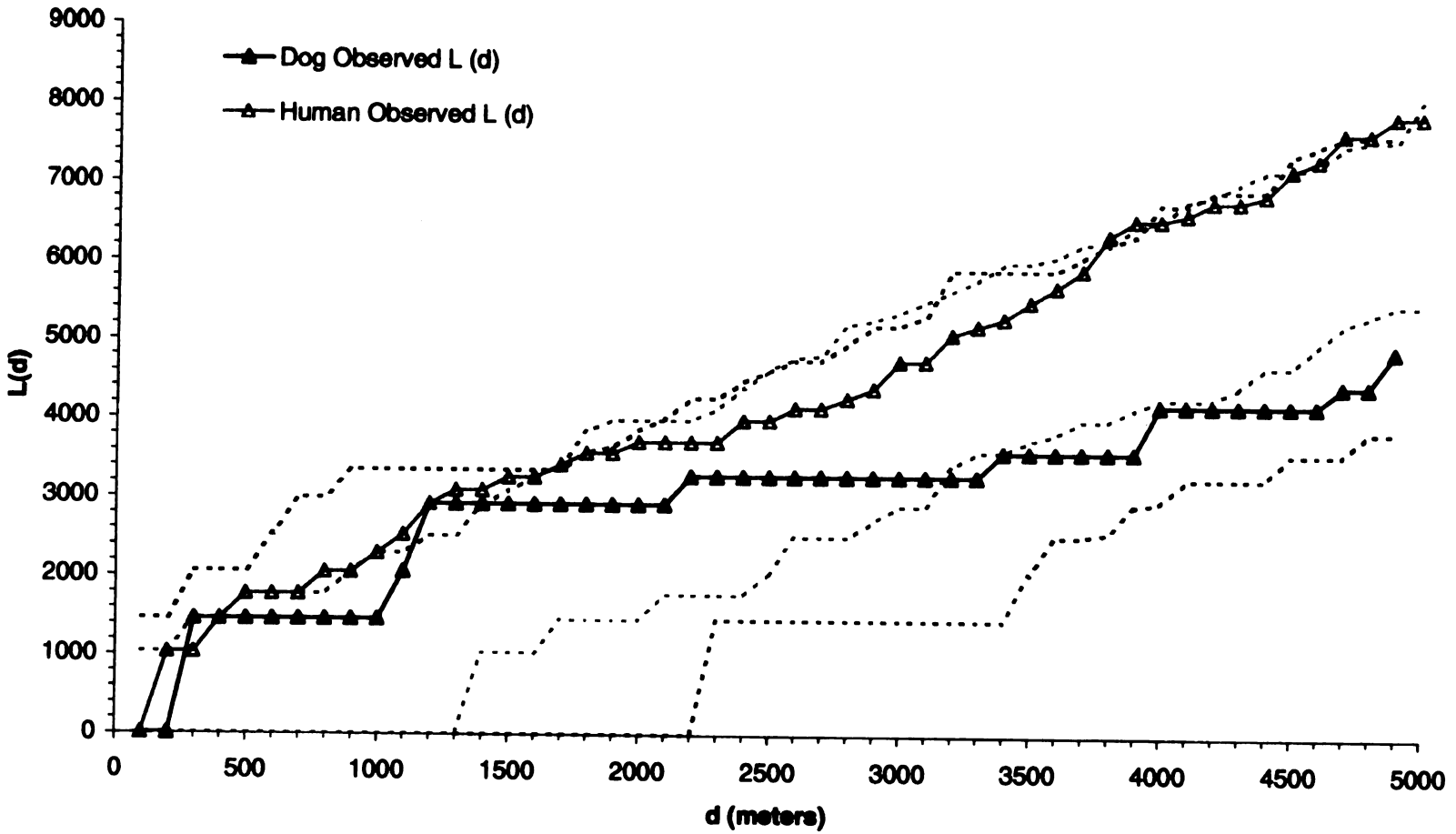


Figure 6.6. *K* function analysis graphical output, melanomas, Oakland County

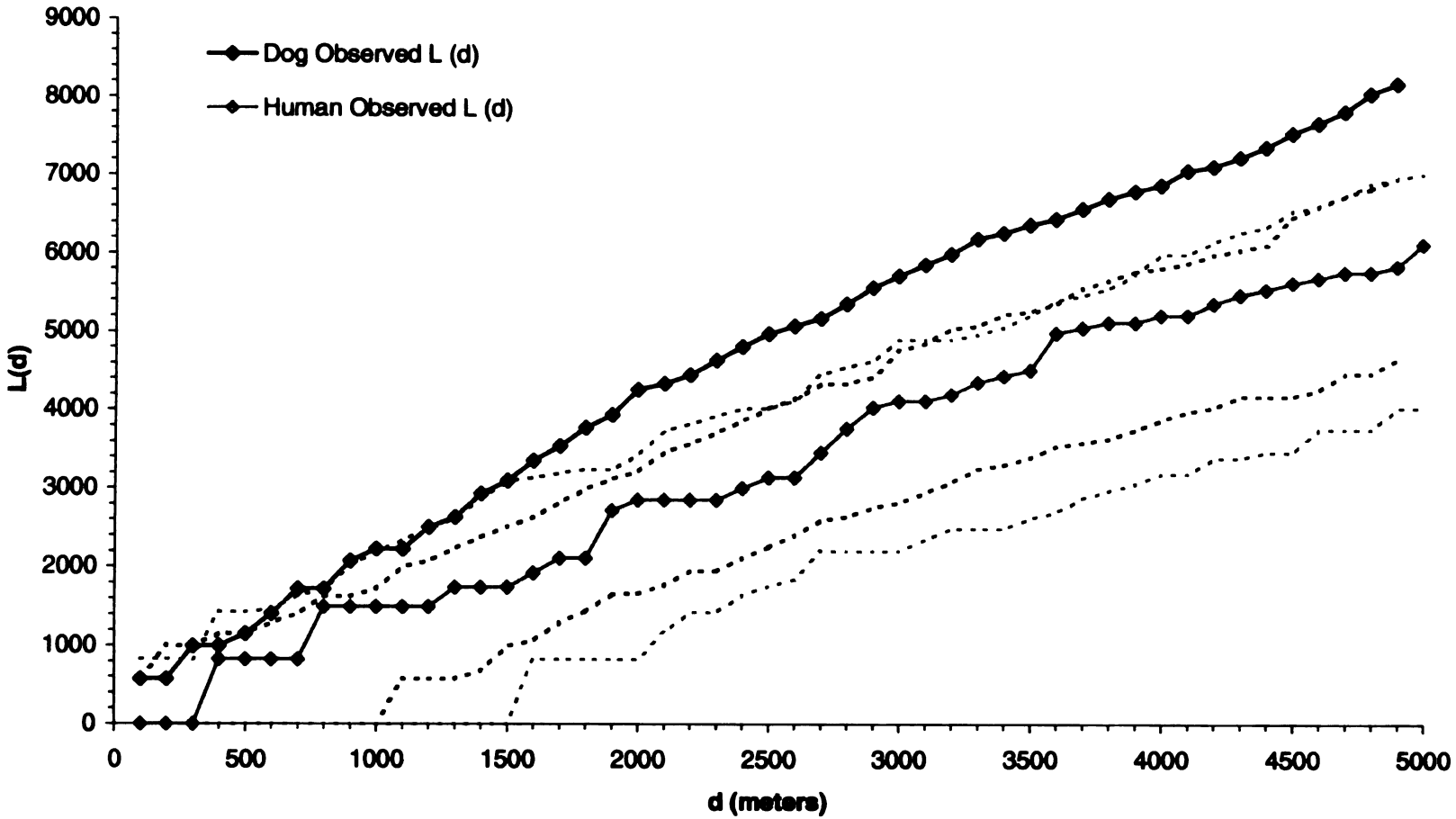


Figure 6.7. *K* function analysis graphical output, spindle cell sarcomas, Ingham County

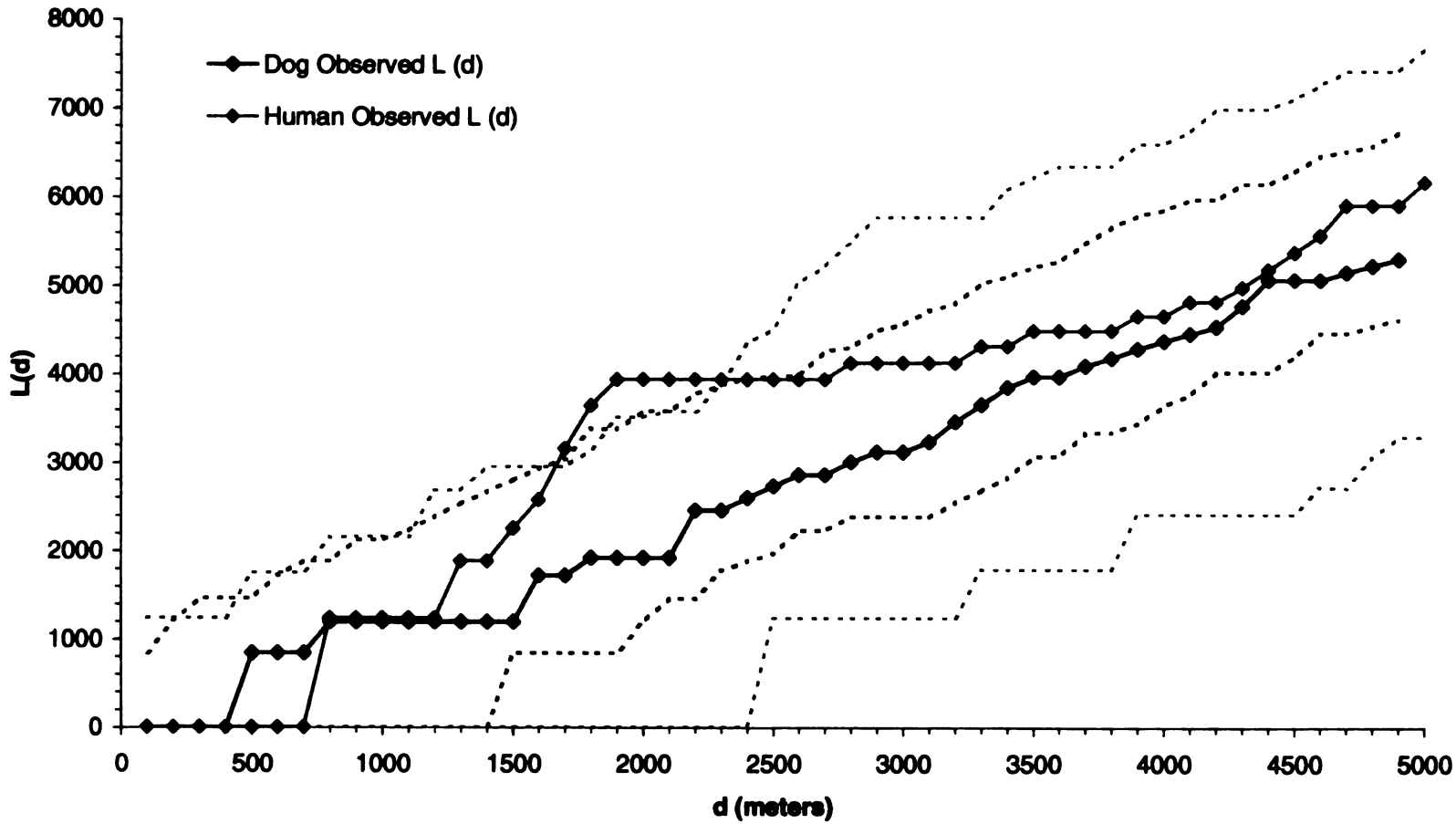


Figure 6.8. *K* function analysis graphical output, spindle cell sarcomas, Oakland County

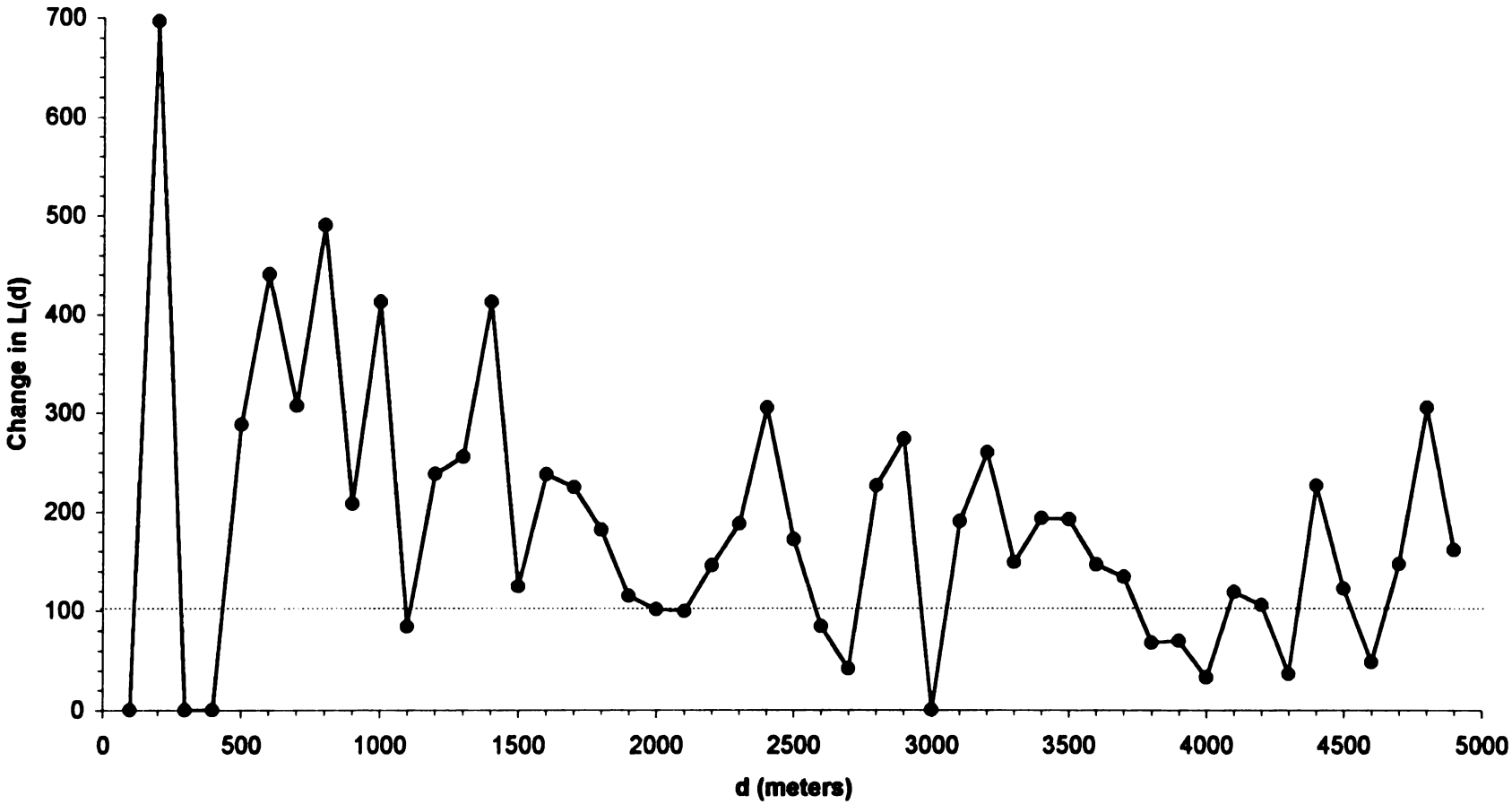


Figure 6.9. Plot of $\Delta L(d)$ vs. d, human lymphomas, Ingham County

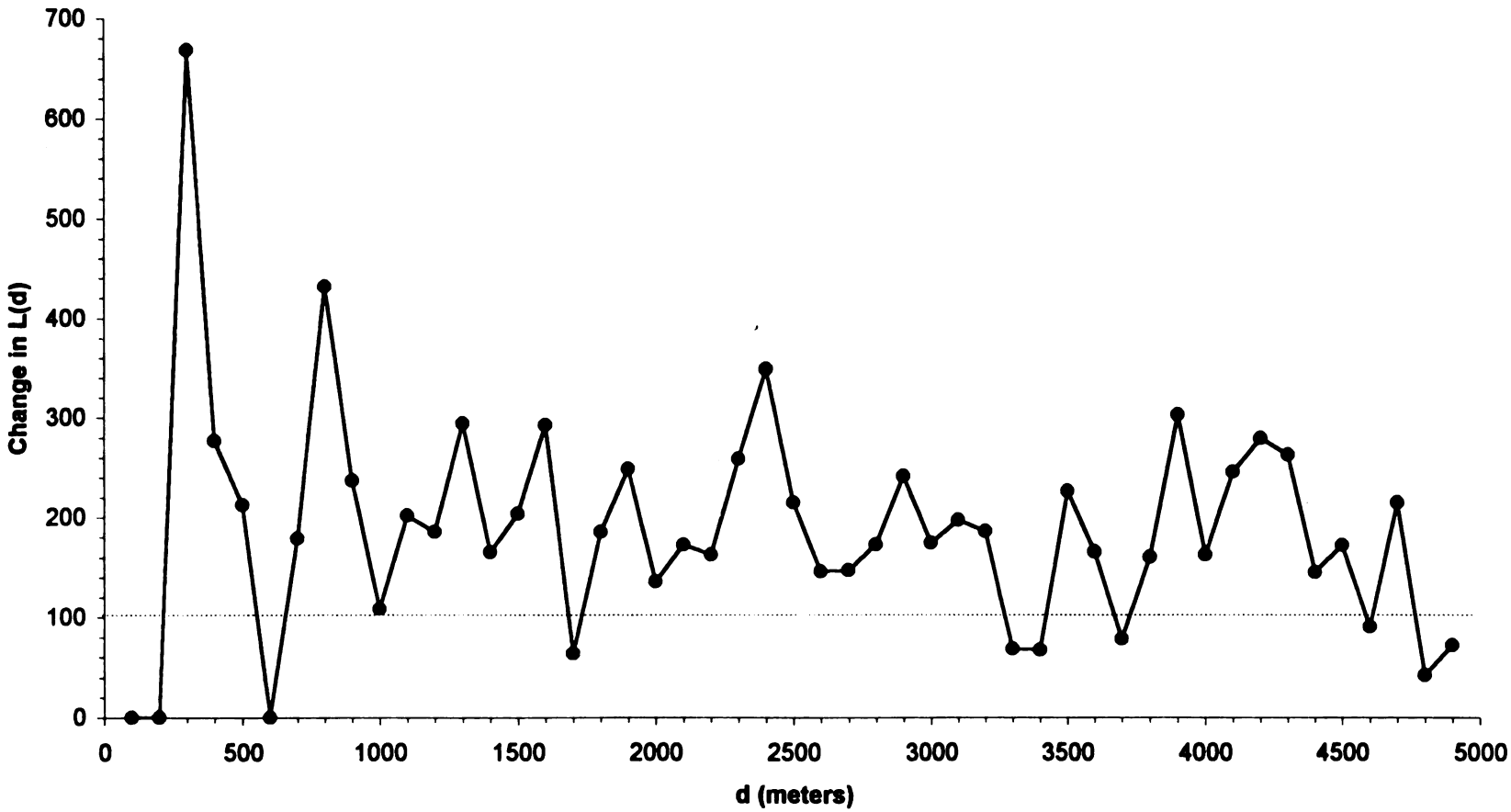


Figure 6.10. Plot of $\Delta L(d)$ vs. d, human lymphomas, Oakland County

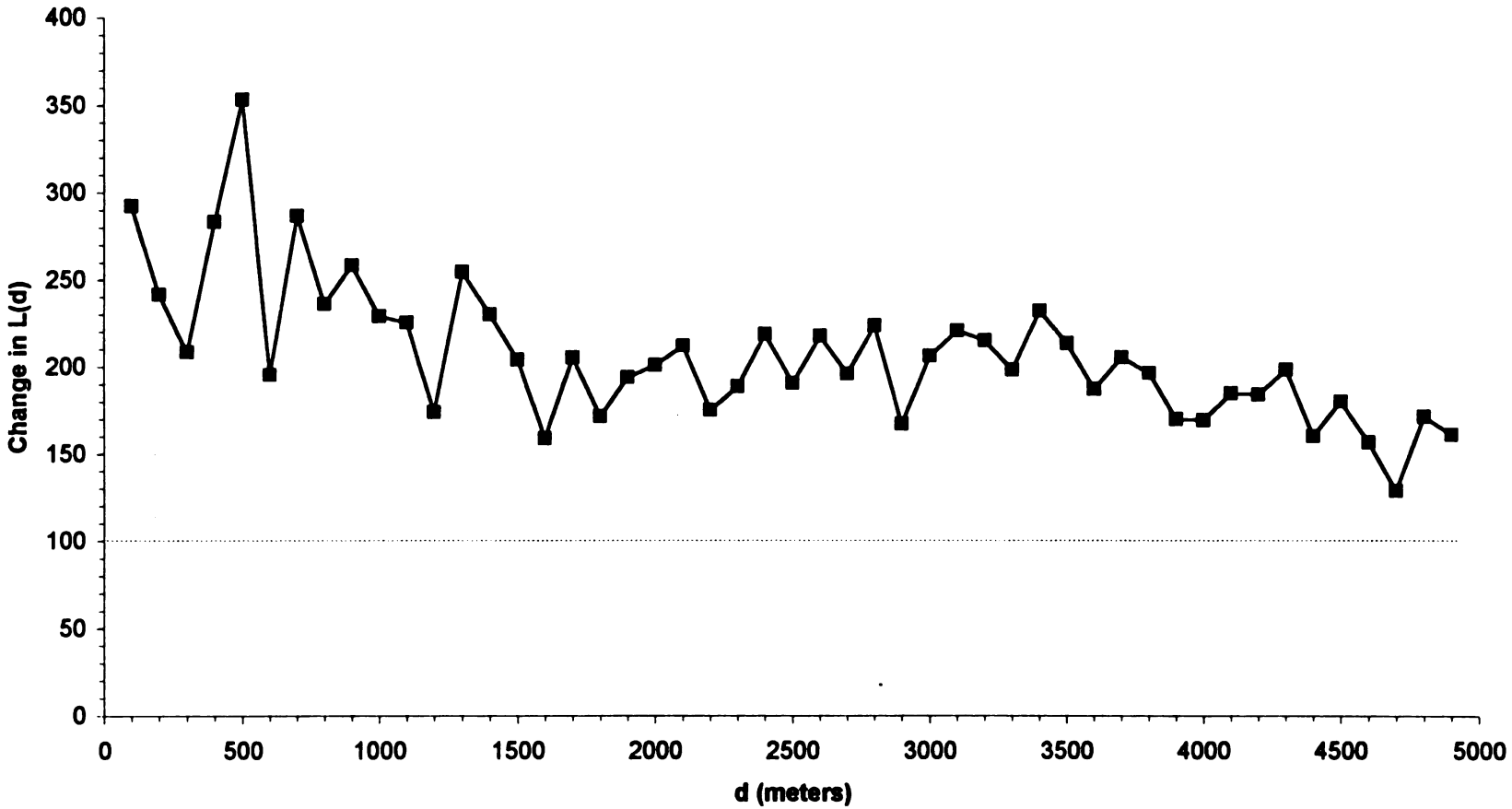


Figure 6.11. Plot of $\Delta L(d)$ vs. d, human mammary neoplasms, Ingham County

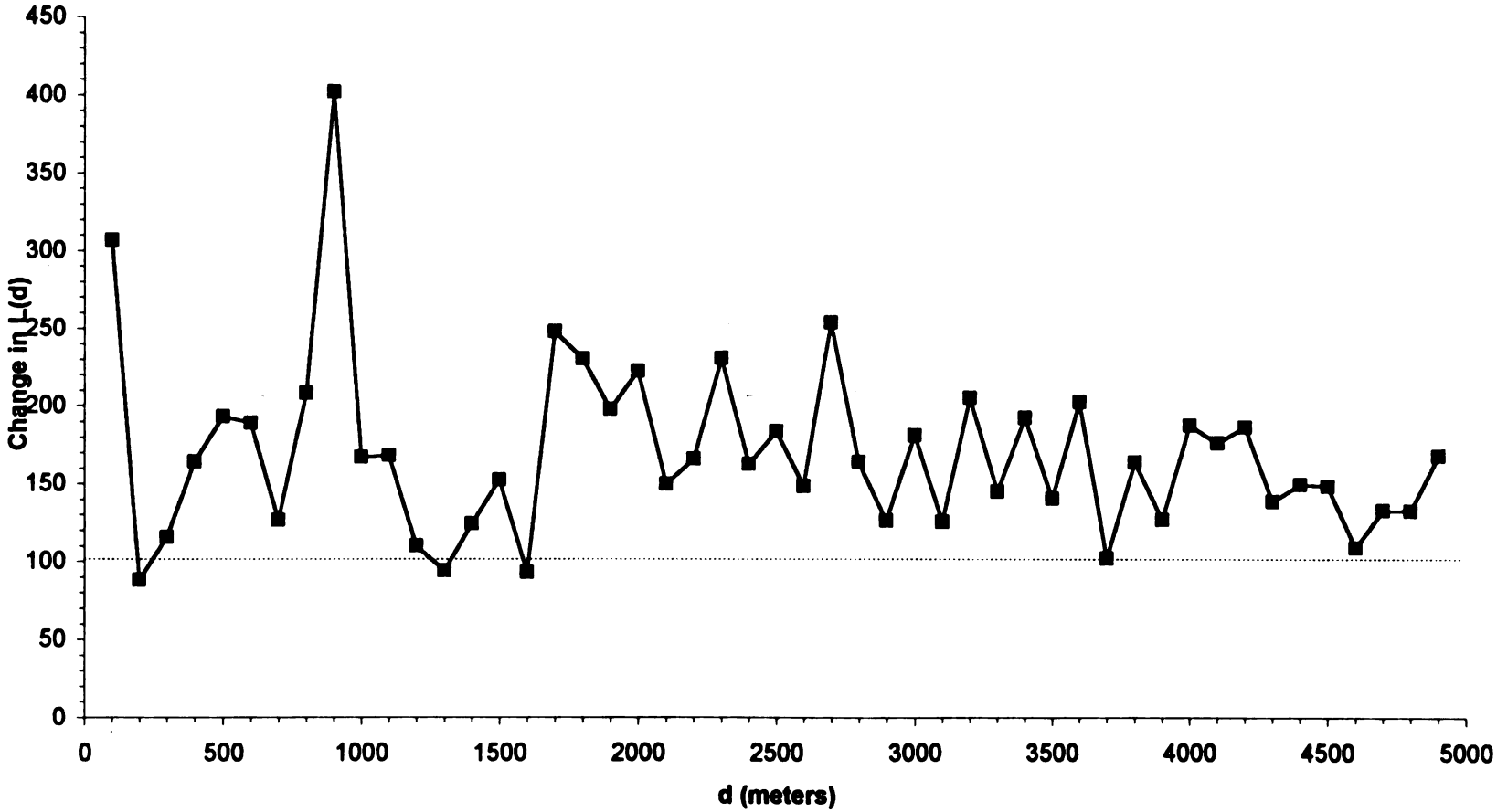


Figure 6.12. Plot of $\Delta L(d)$ vs. d, human mammary neoplasms, Oakland County

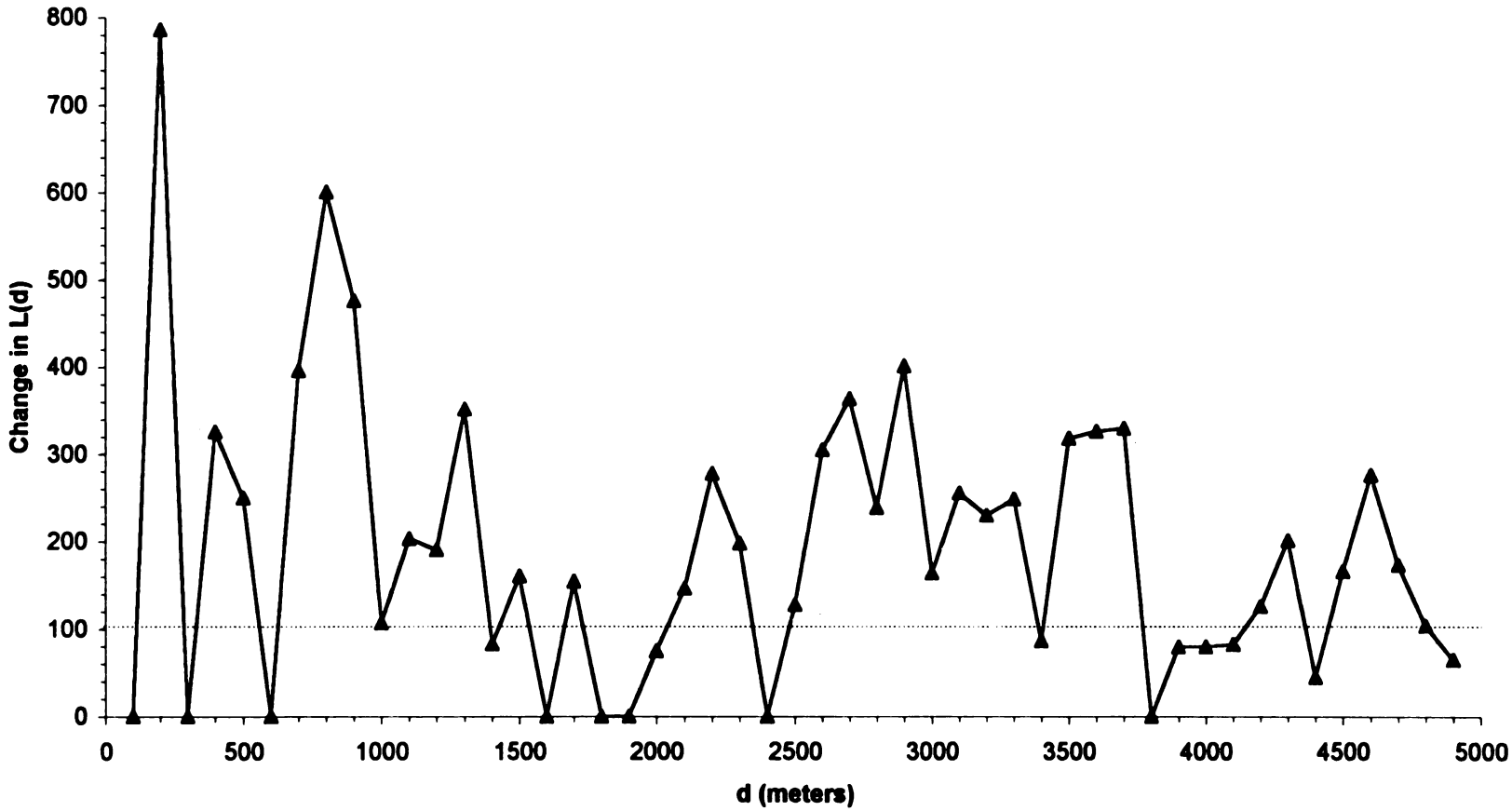


Figure 6.13. Plot of $\Delta L(d)$ vs. d, human melanomas, Ingham County

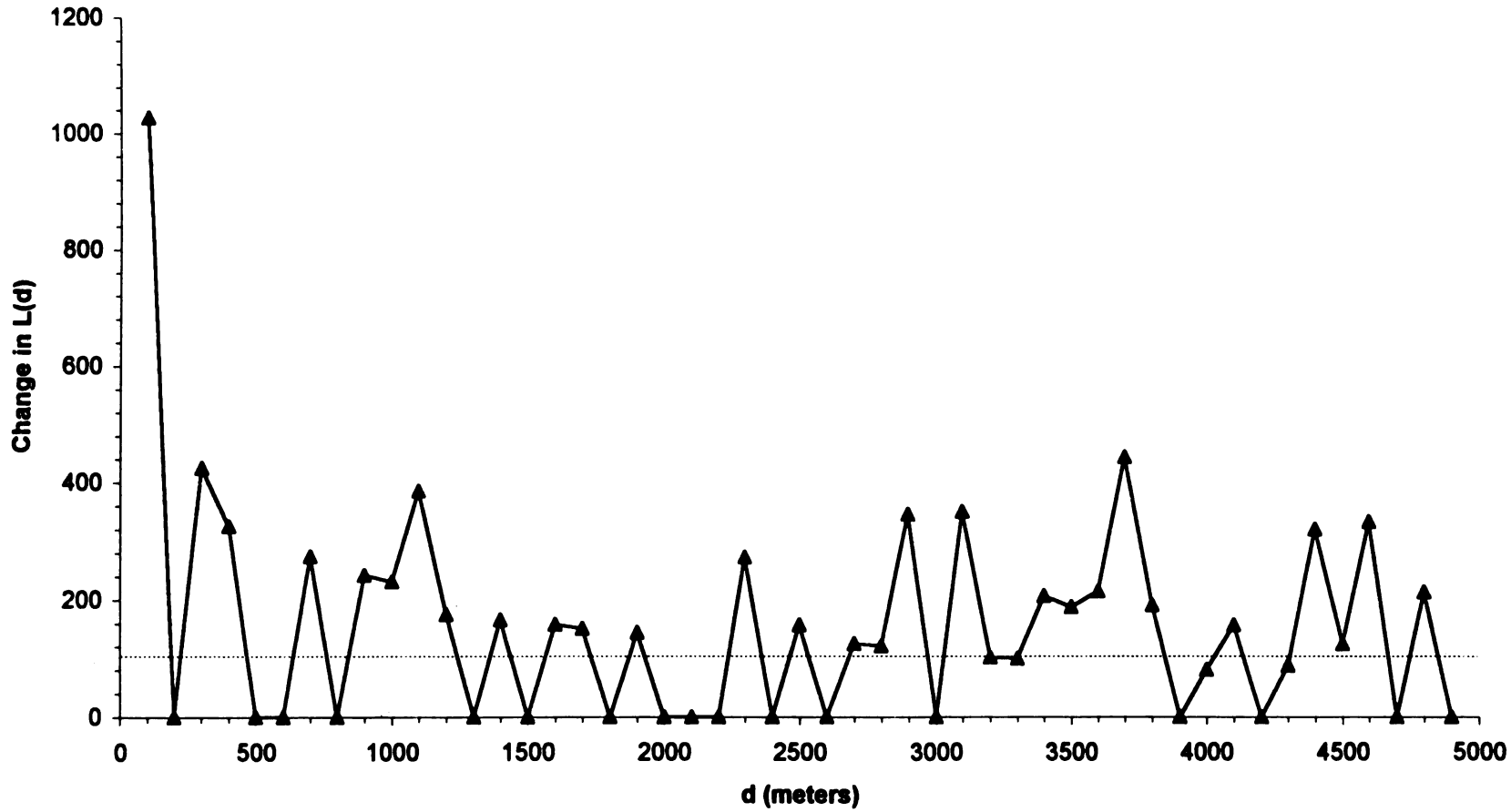


Figure 6.14. Plot of $\Delta L(d)$ vs. d, human melanomas, Oakland County

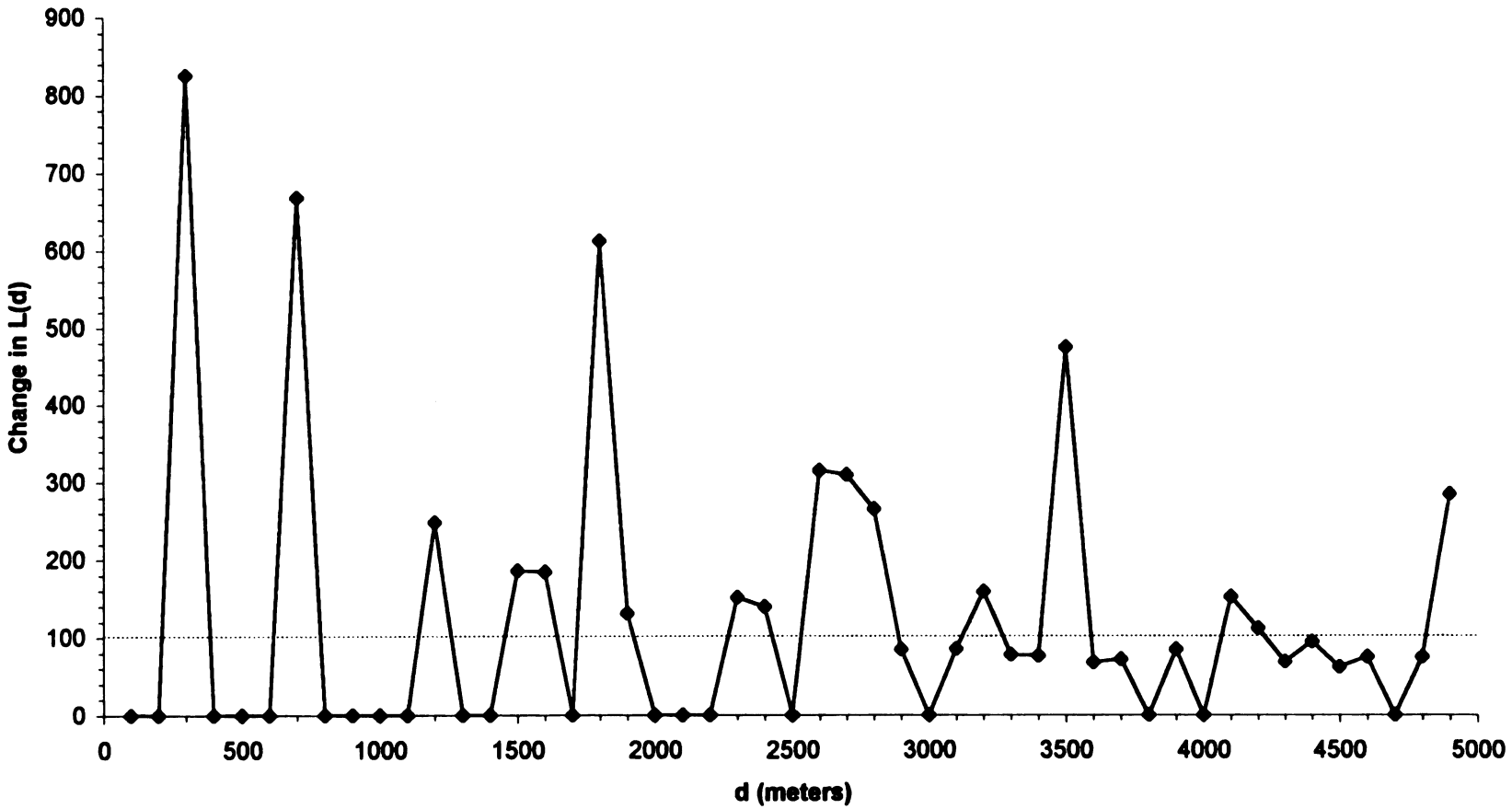


Figure 6.15. Plot of $\Delta L(d)$ vs. d, human spindle cell sarcomas, Ingham County

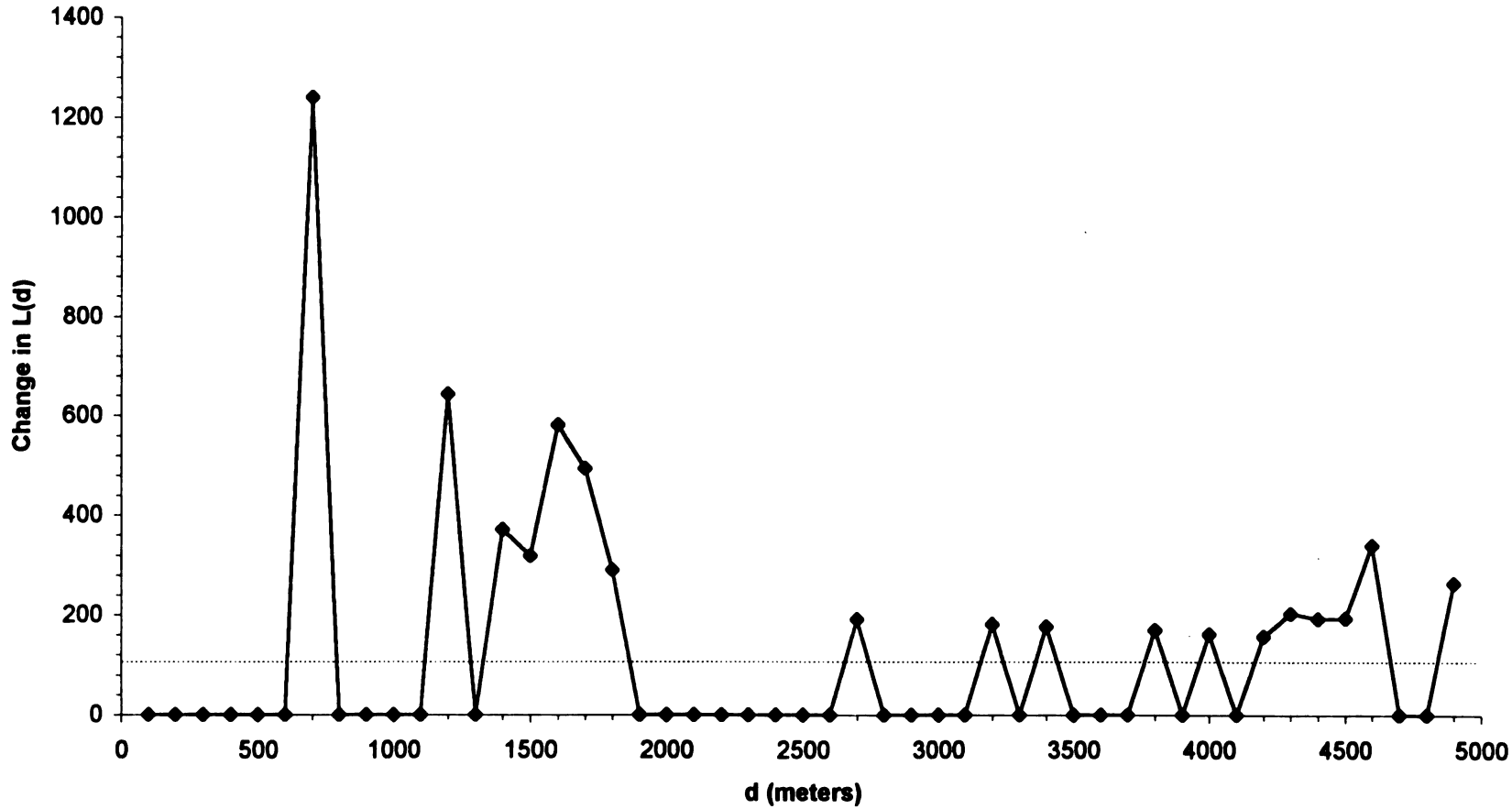


Figure 6.16. Plot of $\Delta L(d)$ vs. d, human spindle cell sarcomas, Oakland County

APPENDIX C

Basic Code to Generate Observed Nearest-Neighbor Distance for the One-Dimensional Nearest-Neighbor Temporal Analysis

```
C*************************************************************************************
C****** THIS PROGRAM COMPUTES OBSERVED NEAREST-NEIGHBOR
C****** DISTANCE FOR ONE-DIMENSIONAL NEAREST-NEIGHBOR
C****** ANALYSIS
C****** THIS ROUTINE WAS WRITTEN BY JAMES W. LLOYD, DVM, PHD AND
C****** DANIEL J. O'BRIEN, DVM ON OCTOBER 20, 1997
C*************************************************************************************


DEFINT I-N

INPUT "Filename = ", filename$
PRINT
INPUT "Number of observations = ", n

REDIM u1(n) AS SINGLE
REDIM u2(n) AS SINGLE

OPEN filename$ FOR INPUT AS #1

i = 1

sumu2 = 0

DO WHILE i <= n

  INPUT #1, u1(i)

  i = i + 1

LOOP

CLOSE #1

i = 1

DO WHILE i <= n

  IF i = 2 THEN
     u2(1) = u1(2) - u1(1)
     sumu2 = u2(1)
  END IF
```

```
 IF i >= 2 AND i < n THEN
      temp1 = u1(i) - u1(i - 1)
      temp2 = u1(i + 1) - u1(i)
      u2(i) = temp1
      IF temp2 < temp1 THEN u2(i) = temp2
   END IF

   IF i = n THEN
      u2(n) = u1(n) - u1(n - 1)
   END IF

   IF i > 1 THEN
      sumu2 = sumu2 + u2(i)
   END IF

   i = i + 1

LOOP



sumu2 = sumu2 / n

PRINT "Mean nearest neighbor distance = ", sumu2

END
```

APPENDIX D

Fortran Code for *K* Function Spatial Analysis

```
      PROGRAM SECSIM
C*****************************************************************
C**** THIS PROGRAM COMPUTES L(D) FOR WEIGHTED POINT DISTANCES
C**** AND CAN BE USED FOR SQUARES OR RECTANGLES OF ANY SIZE
C**** IT CAN ALSO GENERATE VALUES FOR A CONFIDENCE ENVELOPE
C**** WITH A SPECIFIED ALPHA LEVEL AROUND THE L(D) CURVE
C**** THIS VERSION WAS REVISED BY STEVEN M. LEFFLER, FEBRUARY
C**** 22, 1988 AND MODIFIED TO INCLUDED CONFIDENCE LEVELS BY L.
C**** HUNGERFORD
C*****************************************************************
      REAL X(1000), Y(1000), Z(1000), WTSUM(1000), ZVAL(1000)
      REAL RAD(500), LD(500), LDMAX(500), LDMIN(500)
      REAL SUMSQZ, AREA, MINX, MAXX, MINY, MAXY, MAXRAD, SUMZ,
STPSIZ
      REAL XRIGHT, XLEFT, YUP, YDOWN, RADIUS1,
RADIUS2,WEIGHT,ZSUM,DIST
      REAL NOBS,NTIMES,XDIST,YDIST,RADIUS,XDIVIDE,YDIVIDE,RANDZ
      INTEGER A,B,C,D,E,F,G,H,I,J,K,L,M,N
      INTEGER ALPHA,TEMPZ,SEED,CONFID,COUNT,CHOICE
      CHARACTER * 40 INNAME,OUTNAM
      DATA
SUMSQZ,YDOWN,MINX,MAXX,MINY,MAXY,MAXRAD,SUMZ,STEPSIZ /9*0.0/
      DATA XRIGHT,XLEFT,YUP,AREA,RADIUS1,RADIUS2,WEIGHT,ZSUM
/8*0.0/
      DATA DIST,NOBS,NTIMES,XDIST,YDIST,RADIUS /6*0.0/
      DATA A,B,C,D,E,F,G,H,I,J,K,L,M,N /14*0/
      DATA ALPHA,TEMPZ,SEED,CONFID,COUNT /5*0/
C*****************************************************************
C****************************** INTRODUCTION *********************
C*****************************************************************
C
      WRITE(*,*)'*****************************************************'
      WRITE(*,*)'*************** SECOND ORDER ANALYSIS ***************'
      WRITE(*,*)'*****************************************************'
      WRITE(*,*)'*************** BY STEVEN M. LEFFLER ****************'
      WRITE(*,*)'*****************************************************'
      WRITE(*,*)'******* WITH MONTE CARLO CONFIDENCE ENVELOPE
********'
      WRITE(*,*)'*****************************************************'
      WRITE(*,*)'*************** BY LAURA L. HUNGERFORD **************'
      WRITE(*,*)'*****************************************************'
      WRITE(*,*)'*                                                   *'
      WRITE(*,*)'*   THIS PROGRAM COMPUTES L(D) FOR WEIGHTED OR
*'
      WRITE(*,*)'*  NON-WEIGHTED POINT DISTANCES AND CAN BE USED
```

```
*
      WRITE(*,*)'*  FOR RECTANGLES OF ANY SIZE. IT CAN ALSO COMPUTE
*
      WRITE(*,*)'*  A CONFIDENCE ENVELOPE OF SPECIFIED SIGNIFICANCE
*
      WRITE(*,*)'*        AROUND THE L(D) VERSUS D CURVE          *'
      WRITE(*,*)'*                                                *'
      WRITE(*,*)'*   YOUR INPUT FILE SHOULD HAVE THREE COLUMNS OF
*
      WRITE(*,*)'*   DATA -- X, Y, AND Z. THE VALUE OF THE WEIGHT  *'
      WRITE(*,*)'*   SHOULD FALL IN THE Z COLUMN. FOR NONWEIGHTED
*
      WRITE(*,*)'*   POINT DATA, FILL THE Z COLUMN WITH ONES.      *'
      WRITE(*,*)'*                                                *'
      WRITE(*,*)'*   HIT RETURN TO CONTINUE.                      *'
      WRITE(*,*)'**************************************************'
      PAUSE
C*********************************************************************
C************************ READ IN FILENAMES ************************
C*********************************************************************
      WRITE(*,*)'WHAT IS THE NAME OF YOUR INPUT FILE? '
      WRITE(*,*)
      READ(*,1) INNAME
      WRITE(*,*)'WHAT IS THE NAME OF YOUR OUTPUT FILE? '
      WRITE(*,*)
      READ(*,1) OUTNAM
    1 FORMAT (A)
      OPEN (1,FILE=INNAME,STATUS='OLD')
      OPEN (2,FILE=OUTNAM,STATUS='NEW')
C*********************************************************************
C******************* DEFINING THE SCOPE OF ANALYSES ******************
C*********************************************************************
      WRITE (*,*) 'PLEASE GIVE THE MINIMUM AND MAXIMUM
COORDINATES'
      WRITE (*,*) 'FOR YOUR RECTANGLE'
      WRITE (*,*) 'MINIMUM X '
      WRITE (*,*)
      READ (*,*) MINX
      WRITE (*,*) 'MINIMUM Y '
      WRITE (*,*)
      READ (*,*) MINY
      WRITE (*,*) 'MAXIMUM X '
      WRITE (*,*)
      READ (*,*) MAXX
      WRITE (*,*) 'MAXIMUM Y '
```

```
      WRITE (*,*)
      READ (*,*) MAXY
      WRITE(*,*) 'WHAT IS YOUR MAXIMUM SEARCH RADIUS?'
      WRITE (*,*) '*** CAN BE NO LONGER THAN HALF THE SHORTEST SIDE
***'
      WRITE (*,*)
      READ (*,*) MAXRAD
      WRITE (*,*) 'WHAT IS THE DESIRED STEP SIZE?'
      WRITE (*,*)
      READ (*,*) STPSIZ
      WRITE (*,*) 'WHAT PERCENT CONFIDENCE LEVEL WOULD YOU LIKE?'
      WRITE (*,*) '***** 1, 2 OR 5 ARE COMMON ALPHA LEVELS *****'
      WRITE (*,*) '*** CHOOSE 100 IF YOU DO NOT WISH TO HAVE ***'
      WRITE (*,*) '***** THE CONFIDENCE ENVELOPE DETERMINED *****'
      WRITE (*,*)
      READ (*,*) ALPHA
      WRITE (*,*) '  GIVE A NONZER0 TWO DIGIT INTEGER'
      WRITE (*,*) 'WHICH DOES NOT OCCUR IN YOUR Z VALUES'
      WRITE (*,*)
 2    READ (*,*) TEMPZ
      IF (TEMPZ.EQ.0.OR.TEMPZ.GE.100) THEN
       WRITE (*,*) 'PICK A NONZERO TWO DIGIT INTEGER'
       WRITE (*,*)
       GOTO 2
       ELSE
        GOTO 5
       ENDIF
 5    CONTINUE
      WRITE (*,*) 'CHOOSE AN INTEGER BETWEEN 1 AND 100 TO SEED THE'
      WRITE (*,*) 'RANDOM NUMBER GENERATOR FOR CONFIDENCE
CALCULATIONS'
      WRITE (*,*)
 7    READ (*,*) SEED
      IF (SEED.EQ.0.OR.SEED.GE.100) THEN
       WRITE (*,*) 'PICK AN INTEGER BETWEEN 1 AND 100'
       WRITE (*,*)
       GOTO 7
       ELSE
        GOTO 9
       ENDIF
 9    CONTINUE
C**************************************************************************
C************************* READING IN THE DATA ************************
C**************************************************************************
      DO 10 A=1,1000
```

```
      READ (1,*,END=20) X(A), Y(A), Z(A)
      ZVAL(A)=Z(A)
      SUMZ=SUMZ+Z(A)
      SUMSQZ= SUMSQZ+(Z(A)*Z(A))
   10 CONTINUE
   20 NOBS=A-1
     CONFID=100-ALPHA
C*************************************************************************
C******** CALCULATE THE TOTAL AREA AND THE NUMBER OF SEARCHES
********
C*************************************************************************
     NTIMES = MAXRAD / STPSIZ
     AREA = ( MAXX- MINX) * ( MAXY- MINY)
C*************************************************************************
C******************************** WRITE ********************************
C*************************************************************************
     WRITE (*,*)'THE SUM OF ALL Z VALUES = ',SUMZ
     WRITE (*,*)'THE SUM OF THE SQUARE OF ALL Z VALUES = ',SUMSQZ
     WRITE (*,*)'THE TOTAL NUMBER OF OBSERVATIONS = ',NOBS
     WRITE (*,*) 'THE NUMBER OF CIRCLES IN EACH SEARCH = ',NTIMES
     WRITE (*,*) 'THE TOTAL AREA = ', AREA
     WRITE (*,*) 'THE PERCENT CONFIDENCE LEVEL = ', CONFID
     WRITE (*,*)
     WRITE (*,*)
     WRITE (2,*)'THE NAME OF THE INPUT FILE IS ',INNAME
     WRITE (2,*)'THE NAME OF THE OUPUT FILE IS ',OUTNAM
     WRITE (2,*)'THE SUM OF ALL Z VALUES = ',SUMZ
     WRITE (2,*)'THE SUM OF THE SQUARE OF ALL Z VALUES = ',SUMSQZ
     WRITE (2,*)'THE TOTAL NUMBER OF OBSERVATIONS = ',NOBS
     WRITE (2,*) 'THE NUMBER OF CIRCLES IN EACH SEARCH = ',NTIMES
     WRITE (2,*) 'THE TOTAL AREA = ', AREA
     WRITE (2,*) 'THE PERCENT CONFIDENCE LEVEL = ', CONFID
     WRITE (2,*)
     WRITE (2,*)
     IF (NOBS.EQ.SUMZ) THEN
       WRITE (2,*)'    DISTANCE        L(D)'
       WRITE (*,*)'    DISTANCE        L(D)'
      ELSE
       WRITE (2,*)'    DISTANCE        L1(D)'
       WRITE (*,*)'    DISTANCE        L1(D)'
     ENDIF
     WRITE (2,*)
     WRITE (*,*)
C*************************************************************************
C***** DETERMINING WHETHER TO USE THE ORIGINAL VALUES OR IF
```

```
THIS ******
C***** HAS ALREADY BEEN DONE, USING A RANDOM NUMBER
GENERATOR TO ******
C***** REASSIGN THE Z VALUES TO NEW POINTS TO BEGIN TO CREATE
THE *****
C********** PROBABILITY ENVELOPE AROUND THE CURVE **********
C*************************************************************
    DO 98 COUNT=0,CONFID
    IF (COUNT.EQ.0) THEN
     GOTO 15
     ELSE IF (COUNT.LE.CONFID) THEN
      ZSUM=0.0
      DO 11 N=1,NOBS
       Z(N)=TEMPZ
       WTSUM(N)=0.0
11     CONTINUE
C*************************************************************
C********** FINDING A RANDOM NUMBER **********
C*************************************************************
      DO 14 M=1,NOBS
12      CALL RANDOM(SEED,RANDZ)
        CHOICE=500*RANDZ
        IF (Z(CHOICE).EQ.TEMPZ) THEN
          Z(CHOICE)=ZVAL(M)
C*************************************************************
C********* IF THE Z VALUE IN THAT POSITION HAS ALREADY BEEN *********
C********** SELECTED, TRY A NEW RANDOM POSITION **********
C*************************************************************
          ELSE
           GOTO 12
          ENDIF
14      CONTINUE
      WRITE (*,*)'VALUES HAVE BEEN RANDOMLY REASSIGNED'
C*************************************************************
C********** IF THE CONFIDENCE LIMIT HAS BEEN REACHED **********
C********** GO TO THE END OF THE ANALYSIS **********
C*************************************************************
      ELSE
       GOTO 100
      ENDIF
15    CONTINUE
C*************************************************************
C********** BEGINING THE ANALYSIS **********
C*************************************************************
    DO 50 I=1,NOBS
```

```
        XLEFT=X(I)-MINX
        XRIGHT=MAXX-X(I)
        YUP=Y(I)-MINY
        YDOWN=MAXY-Y(I)
      DO 40 J=1,NOBS
C*************************************************************************
C**** WE BEGIN BY CALCULATING THE DISTANCE BETWEEN POINTS I
AND J *****
C*************************************************************************
          IF (I.EQ.J) GOTO 40
          XDIST=X(I)-X(J)
          YDIST=Y(I)-Y(J)
          DIST=SQRT(ABS((XDIST*XDIST)+(YDIST*YDIST)))
C*************************************************************************
C***** THEN WE FIND OUT IF POINT J LIES WITHIN THE SEARCH RADIUS
******
C*************************************************************************
          RADIUS2=0
          DO 30 B=1,NTIMES
            RADIUS1=RADIUS2
            RADIUS2=RADIUS1+STPSIZ
    IF (DIST.LT.RADIUS1.OR.DIST.GT.RADIUS2) THEN
      GOTO 30
C*************************************************************************
C********* IF POINT J DOES LIES WITHIN THE SEARCH RADIUS THEN
*********
C***** WE NEED TO CHECK FOR AND DEAL WITH ANY BORDER
CORRECTIONS. *****
C*************************************************************************
C***************** FIRST THE TWO BORDER CORRECTIONS. ******************
C*************************************************************************
    ELSE IF ((XLEFT.LT.DIST).AND.(YUP.LT.DIST)) THEN
      XDIVIDE=XLEFT/DIST
      YDIVIDE=YUP/DIST
      WEIGHT=1./(1.-((ACOS(XDIVIDE))+(ACOS(YDIVIDE))+1.57)/6.28)
    ELSE IF ((XLEFT.LT.DIST).AND.(YDOWN.LT.DIST)) THEN
      XDIVIDE=XLEFT/DIST
      YDIVIDE=YDOWN/DIST
      WEIGHT=1./(1.-((ACOS(XDIVIDE))+(ACOS(YDIVIDE))+1.57)/6.28)
    ELSE IF ((XRIGHT.LT.DIST).AND.(YUP.LT.DIST)) THEN
      XDIVIDE=XRIGHT/DIST
      YDIVIDE=YUP/DIST
      WEIGHT=1./(1.-((ACOS(XDIVIDE))+(ACOS(YDIVIDE))+1.57)/6.28)
    ELSE IF ((XRIGHT.LT.DIST).AND.(YDOWN.LT.DIST)) THEN
      XDIVIDE=XRIGHT/DIST
```

```
        YDIVIDE=YDOWN/DIST
        WEIGHT=1./(1.-((ACOS(XDIVIDE))+(ACOS(YDIVIDE))+1.57)/6.28)
C*************************************************************************
C*************** THEN THE ONE BORDER CORRECTIONS. ***********************
C*************************************************************************
    ELSE IF (XLEFT.LT.DIST) THEN
        XDIVIDE=XLEFT/DIST
        WEIGHT=1./(1.-((ACOS(XDIVIDE))/3.14159))
    ELSE IF (XRIGHT.LT.DIST) THEN
        XDIVIDE=XRIGHT/DIST
        WEIGHT=1./(1.-((ACOS(XDIVIDE))/3.14159))
    ELSE IF (YUP.LT.DIST) THEN
        YDIVIDE=YUP/DIST
        WEIGHT=1./(1.-((ACOS(YDIVIDE))/3.14159))
    ELSE IF (YDOWN.LT.DIST) THEN
        YDIVIDE=YDOWN/DIST
        WEIGHT=1./(1.-((ACOS(YDIVIDE))/3.14159))
C*************************************************************************
C************* OR ELSE WE HAVE NO BORDER CORRECTIONS.
**************
C*************************************************************************
    ELSE
        WEIGHT = 1
    ENDIF
C*************************************************************************
C******** THEN WE CALCULATE THE TOTAL WEIGHT AT EACH DISTANCE
********
C*************************************************************************
            WTSUM(B)=WTSUM(B)+(Z(I)*Z(J)*WEIGHT)
    GOTO 40
  30 CONTINUE
  40 CONTINUE
  50 CONTINUE
C*************************************************************************
C********************** NOW WE CAN CALCULATE L(D) ***********************
C*************************************************************************
    DO 60 C=1,NTIMES
      ZSUM=ZSUM+WTSUM(C)
      LD(C)=SQRT(ZSUM*AREA/(3.14159*((SUMZ*SUMZ)-SUMSQZ)))
      RAD(C)=REAL (C) * STPSIZ
  60 CONTINUE
C*************************************************************************
C***************** IF THIS IS THE ORIGINAL DATA ORDER ********************
C********************** THEN PRINT THE RESULTS ***************************
C*************************************************************************
```

```
    IF (COUNT.EQ.0) THEN
      DO 70 D=1,NTIMES
      WRITE (2,*)RAD(D),LD(D)
      WRITE (*,*)RAD(D),LD(D)
 70   CONTINUE
C*************************************************************************
C***** IF WE ARE DOING AN ENVELOPE THEN PRINT THE HEADINGS FOR
THE ****
C********** MAX AND MIN CURVES, OTHERWISE GO TO THE END
**********
C*************************************************************************
      IF (CONFID.EQ.0) THEN
       GOTO 100
       ELSE
        WRITE (*,*)
        WRITE (*,*)'CALCULATING CONFIDENCE ENVELOPE,'
        WRITE (*,*)'PLEASE WAIT!!!'
        WRITE (*,*)
        ENDIF
C*************************************************************************
C***** FOR THE FIRST ITERATION, SET THE MAX AND MIN VALUES EQUAL
******
C********** EACH OTHER AND THE FIRST RANDOM-BASED L(D)S
**********
C*************************************************************************
     ELSE IF (COUNT.EQ.1.AND.COUNT.LE.CONFID) THEN
       WRITE (*,*) COUNT
       WRITE (2,*)
       WRITE (2,72) (RAD(H), H=1,NTIMES)
       WRITE (2,*)
       WRITE (2,72) (LD(K), K=1,NTIMES)
 72     FORMAT (500F8.3)
       DO 80 E=1,NTIMES
         LDMAX(E)=LD(E)
         LDMIN(E)=LD(E)
 80     CONTINUE
C*************************************************************************
C***** REPLACE THE MAX AND MIN VALUES WITH THE NEW VALUES IF
THE ******
C******** NEW VALUE IS HIGHER OF LOWER THAN THE CURRENT VALUE
*********
C********** UNTIL THE CONFIDENCE LIMIT IS REACHED **********
C*************************************************************************
     ELSE IF (COUNT.LE.CONFID) THEN
       WRITE (*,*) COUNT
```

```
      WRITE (2,*)
      WRITE (2,88) (LD(L), L=1,NTIMES)
 88    FORMAT (500F8.3)
      DO 90 F=1,NTIMES
      IF (LD(F).LT.LDMIN(F)) LDMIN(F)=LD(F)
      IF (LD(F).GT.LDMAX(F)) LDMAX(F)=LD(F)
 90    CONTINUE
C*************************************************************************
C******* WHEN THE CONFIDENCE LIMIT IS REACHED PRINT THE RESULTS
*******
C****** TO GIVE THE MAX AND MIN CURVES OF THE CONFIDENCE
ENVELOPE ******
C*************************************************************************
      IF (COUNT.LT.CONFID) THEN
        GOTO 94
        ELSE
       WRITE (2,*)
       WRITE (2,91) COUNT
       WRITE (2,*)
       WRITE (2,*)'    DISTANCE        L(D)MAX       L(D)M
   XIN'
       WRITE (*,91) COUNT
       WRITE (*,*)
       WRITE (*,*)'    DISTANCE        L(D)MAX       L(D)M
   XIN'
       WRITE (*,*)
 91     FORMAT (1X,'THE CONFIDENCE LEVEL =',I4,'%')
        DO 92 G=1,NTIMES
         WRITE (2,*) RAD(G),LDMAX(G),LDMIN(G)
         WRITE (*,*) RAD(G),LDMAX(G),LDMIN(G)
 92      CONTINUE
        ENDIF
 94      CONTINUE
    ENDIF
 98 CONTINUE
100 CONTINUE
   STOP
   END
C*                                                              *
C*************************************************************************
C************** SUBROUTINE TO GENERATE RANDOM NUMBER
******************
C*************************************************************************
   SUBROUTINE RANDOM (SEED,RANDZ)
C*************************************************************************
```

```
C**** THIS SUBROUTINE GENERATES RANDOM VALUES BETWEEN 0.0
AND 1.0 *****
C******* USING AN INTEGER SEED.  IT IS TAKEN FROM D. M. ETTER, ********
C****** STRUCTURED FORTRAN FOR ENGINEERS AND SCIENTISTS, 2ND
ED *******
C*************** BENJAMIN/CUMMINGS PUBLISHING CO., 1987 ***************
C******** BASED ON AN ALGORITHM FROM S. D. STERNS, A PORTABLE
*********
C******* RANDOM NUMBER GENERATOR FOR USE IN SIGNAL
PROCESSING, ********
C********** SANDIA NATIONAL LABORATORIES TECHNICAL REPORT
***********
C*************************************************************************
      REAL RANDZ
      INTEGER SEED
C
      SEED=2045*SEED + 1
      SEED=SEED - (SEED/1048576)*1048576
      RANDZ = REAL(SEED + 1)/1048577.0
C
      RETURN
      END
C*************************************************************************
```