THE QUANTITATIVE EPIDEMIOLOGY OF CANINE NEOPLASTIC DISEASE: RISK FACTOR IDENTIFICATION USING DIAGNOSTIC HISTOPATHOLOGY DATA

by

Heather G. Richards BSc BVMS CertSAM MACVSc MRCVS

Thesis submitted for the Degree of Doctor of Philosophy in the Faculty of Veterinary Medicine, University of Glasgow.

Comparative Epidemiology and Informatics
Department of Veterinary Clinical Studies
University of Glasgow.

January 2003

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

This thesis is dedicated to my father, for his far-reaching support, and to my mother, for her legacy of courage and determination.

ABSTRACT

Research was undertaken to investigate the risk factors for neoplasia in two canine biopsy populations, the first originating from the Canine Infectious Diseases Research Unit (CIDRU) diagnostic histopathology service, based at the University of Glasgow Veterinary School (GUVS), and the second from the diagnostic histopathology service operated by a commercial organisation. Both provide histopathology reports to veterinary practitioners located throughout the United Kingdom and occasionally overseas. The studies were undertaken to determine the feasibility of using these data sources for meaningful epidemiological analyses of host-related risk factors for canine neoplasia. The analytical scope was expanded to explore the effect of submitting practice as a risk factor for canine neoplasia. Finally, spatial and spatio-temporal epidemiological techniques were applied to the CIDRU data to explore the significance of its geographical origin upon an outcome of neoplasia in a canine biopsy.

Records pertaining to canine biopsy submissions were extracted from both histopathology databases. The records were subjected to a hierarchical and iterative data cleaning process, which focused upon the main host-related variables of age, gender and breed of dog, when available, and biopsy site of origin. This procedure highlighted a number of important quality assurance issues in both datasets. The coding system used in the CIDRU database was found to be adequate for assisting data preparation, although there were issues relating to the lack of integral data checks and the use of free text input. The extensive use of free text input for the commercial dataset limited the amount of data content that could be prepared from this database for subsequent analysis.

Following establishment of data integrity, case-control studies of the cleaned datasets were performed. Multivariable logistic regression was used to assess the effect of the host-related risk factors of age, gender and breed of dog, when available, and site of biopsy, on the outcome of neoplasia in a biopsy submitted to the histopathology service. Similar results were produced in analyses of both datasets. The grouping of data by submitting veterinary practice was considered to cause violation of the assumption of independence for individual biopsies because of unknown practice-related factors associated with biopsy submission. Following the application of inclusion criteria to the data, a practice variable was entered into the host-related multivariable models first as a fixed-effect term, then as a random-effect term. The introduction of a variable for practice verified that group effects due to practice were significant in the data from both histopathology services.

Spatial and space-time analyses were conducted on the CIDRU dataset using spatial and space-time scan statistics. Graphical display of the results with a Geographical Information System (GIS) illustrated a trend for clusters with low risk of neoplasia diagnosis in biopsies submitted from the north of the UK compared to high risk clusters located in the south of the country. A number of individual practices caused significant subclusters within the main clusters, leading to the proposal of conducting practice-based research to investigate practice-related factors that influence tissue biopsy submission.

It was concluded that the histopathology databases provided data suitable for the epidemiological analysis of host-related risk factors for canine neoplasia. The findings of the studies suggest that future research should focus upon identification of factors within individual practices which directly influence the procurement of tissue for histopathological analysis, to accurately ascertain their effect upon the diagnosis of neoplasia in canine biopsies. Practice-based research may also provide insight into differences in geographical occurrence of canine neoplasia, which may lead to the generation of hypotheses regarding possible environmental factors contributing to the aetiology of canine neoplastic disease.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

The work in this thesis has only been possible due to the support, encouragement and guidance of many people. My journey into research has been a truly enlightening experience, the benefits of which reach far beyond the work itself and I am eternally grateful to all who have been involved.

The origin of this project can be traced back to an idea arising from my personal experiences of working in first-opinion veterinary practices throughout the UK. My wish to explore the epidemiology of canine neoplasia was given structure and direction by Professor Stuart Reid, without whom this project would not have been possible. As my supervisor, I thank him for his belief, support and guidance during the evolution of this thesis.

Fundamental to the production of this research is the dedication of the pathologists at the University of Glasgow Veterinary School to the provision of a diagnostic histopathology service to first-opinion veterinary practitioners. I am particularly grateful to Dr. Pauline McNeil of the Department of Veterinary Pathology for allowing access to the histopathology data held by the Canine Infectious Diseases Research Unit, and for her assistance with histopathological terminology during the early phases of this project. I also wish to thank Dr. Hal Thompson for providing historical information about the CIDRU recording systems and development of the CIDRU database. Sandy Young of Specialist Services Limited provided useful comments regarding the CIDRU database structure. I wish also to acknowledge Dr. Katie Knox for her research involving the hospital database at GUVS. My understanding of the structure and function of the CIDRU database was significantly enhanced by her work.

The staff of Idexx Laboratories have provided much support and encouragement for this research from its outset. I am indebted to Ivan Jennings, who during his role as Director of Idexx Laboratories UK, allowed access to the histopathology data utilised in Chapter 5. Dr. Irene McCandlish was instrumental in the initiation of this project and my sincere thanks go to her for offering advice and critically appraising sections of this thesis. Nick Davison performed the data extraction from the Idexx database and provided the screen displays shown in Chapter 5. His patience with explaining the database structure, answering my catalogue of questions and his review of the database description in Chapter 5 are very much appreciated. My thanks also go to Dr. Rob Rahaley, current Director of Idexx Laboratories, for his interest and support of this work, and Nick Carmichael, for his understanding. The contribution made by Duncan Little with his assistance in the preparation of the Idexx data is also gratefully acknowledged.

I wish to extend my warmest thanks to the members of the Comparative Epidemiology and Informatics group, and the Information Services Unit with whom I have had the pleasure of sharing an office over the years. In particular, I would like to thank David Irvine for his support in the use of the CIDRU database system, and his assistance with data extraction during the early phases of the study. More recently, Graeme McCoombe and Drew McConnell have provided valued assistance with all things computer-orientated. I am particularly grateful to Dr. Giles Innocent for his interest in, and advice on the statistical

elements of this work. Our discussions were inevitably productive and I thank him for providing direction when the way forward seemed obscure. My thanks also go to Dr. Dominic Mellor for his assistance in the use of Geographical Information Systems and the production of the maps presented in Chapter 6. I also extend my heartfelt appreciation to Vicki Dale, Tom Irwin and Helen Ternent who were always ready with words of encouragement and ways to bring laughter into any situation. Their friendships during the triumphs and tragedy have been invaluable.

This project began in an office at the University of Strathclyde, where Colin Naismith and I embarked on our PhD journeys. Our utilisation of a common data source led to the sharing of core elements of our projects, and the development of a friendship that I will always treasure. His spirit lives on.

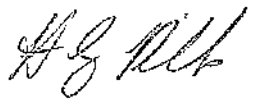
While working full-time during the last stages of this thesis, I have been extremely appreciative of the support and patience of my colleagues at Idexx Laboratories, as they inevitably shared the experience. I also wish to thank the many friends in Glasgow who have provided me with a place to stay during my numerous visits in the concluding stages of this work.

I would like to thank my family for their enduring support, regardless of where in the world I have been during the creation of this thesis. Their presence on the end of a telephone provided reassurance and solace when needed most. The few months I spent in Australia during my PhD journey were memorable for their wonderful encouragement and I look forward to the time when I may share the completion of this experience with them in a physically shared space. Finally, I would like to thank James, for being there.

The Royal College of Veterinary Surgeons Trust Fund and the University of Glasgow provided funding for this work.

AUTHOR'S DECLARATION

The work presented in this thesis was performed solely by the author except where the assistance of others has been acknowledged. It has not been submitted in any form for another degree or professional qualification.



Heather G. Richards

Some of the work presented in this thesis has been the subject of the following publications and presentations:

Richards, H.G., McNeil, P.E., Thompson, H. and Reid, S.W.J. (2001) An epidemiological analysis of a canine biopsies database compiled by a diagnostic histopathology service. *Preventive Veterinary Medicine*, **51**, 125-136.

Richards, H.G., McNeil, P.E., Thompson, H. and Reid, S.W.J. (1999) An epidemiological analysis of canine neoplasia as submitted to a veterinary histopathology service. Paper given at the 53rd Scientific Meeting of the Association of Veterinary Teachers and Research Workers, held in Scarborough, 29 March-1 April, 1999

Richards, H.G., McNeil, P.E., Thompson, H. and Reid, S.W.J. (1999) An epidemiological analysis of canine neoplasia as submitted to a veterinary histopathology service. Paper given at the 80th Annual Meeting of the Conference of Research Workers in Animal Diseases, held in Chicago, USA, 7-9 November, 1999.

Richards, H.G., McNeil, P.E., Thompson, H. and Reid, S.W.J. (2000) An epidemiological analysis of a canine biopsies database compiled by a diagnostic histopathology service. Paper given at the Australian College of Veterinary Scientists Science Week, held in Surfers Paradise, Australia, 6-8 July, 2000.

In addition, the following papers have been accepted for presentation at the 46th Annual Congress for the British Small Animal Veterinary Association, to be held in Birmingham, in April 2003:

Richards, H.G., McNeil, P.E., Thompson, H. and Reid, S.W.J. Do geographical clusters of canine neoplasia occur in the UK?

Richards, H.G., McCandlish, I.A.P., Hamilton, J.M., McDougall, D.F., Holden, K.E., Johnson, H.M., Jennings, I. and Reid, S.W.J. Epidemiological analysis of a canine biopsies database compiled by a commercial diagnostic histopathology service.

LIST OF ACRONYMS

2,4-D	2,4-dichlorophenoxyacetic acid
5-FU	5-fluorouracil
AD	Anno domini
ACTH	Adrenocorticotropin hormone
BC	Before Christ
CANR	California Animal Neoplasm Registry
CDDP	Cisplatin
CI	Confidence interval
CIDRU	Canine Infectious Diseases Research Unit
DNA	Deoxynucleic acid
EBV	Extra-binomial variation
FE	Female entire
FN	Female neutered
GIS	Geographical Information System
GUVS	University of Glasgow Veterinary School
ICD	International Classification of Diseases
ICD-O	International Classification of Diseases - Oncology
mG	milligauss
ME	Male entire
MN	Male neutered
NCVM	Norwegian College of Veterinary Medicine
NCCR	Norwegian Canine Cancer Registry
OR	Odds Ratio
PCOP	Purdue Comparative Oncology Program
PFMA	Pet Food Manfacturer's Association
PIN	Prostatic intraepithelial neoplasia
RNA	Ribonucleic acid
SCC	Squamous cell carcinoma
SD	Standard deviation
SNDO	Standard Nomenclature of Diseases and Operations
SNOMED	Systematised Nomenclature of Medicine
SNOMED CT	Systematised Nomenclature of Medicine - Clinical Terms
SNOMED RT	Systematised Nomenclature of Medicine - Reference Terminology
SNOP	Standard Nomenclature of Pathology
SNOVET	Systematised Nomenclature of Veterinary Medicine
SNVDO	Standard Nomenclature of Veterinary Diseases and Operations
TCC	Transitional cell carcinoma
VMDB	Veterinary Medical Data Base
WHO	World Health Organisation

CHAPTER 1

GENERAL INTRODUCTION AND REVIEW OF THE LITERATURE

1.1 INTRODUCTION

The growth and development of veterinary practice in the companion animal sector in the last thirty years has reflected the increased importance of pets in our society. Neoplasia in the dog is of major significance in terms of animal health and welfare, and is an emotive subject when encountered by pet owners. Research advances in the field of canine neoplasia have both increased knowledge of the disease process and provided better therapy options, for which there is an increasing demand (Withrow, 1996). The emergence of better diagnostic facilities in veterinary practices and the willingness of owners in more affluent societies to pay for the necessary costs of diagnosis have resulted in substantial accumulation of data pertaining to canine neoplasia.

The review begins with exploring the role of the dog in human society from earliest records, leading to the important profile as a companion and worker that the dog has today. This is followed by a description of the pathogenesis of canine neoplasia and the laboratory procedures required to confirm its diagnosis in an individual. A review of the major epidemiological studies of canine neoplasia is then presented, with emphasis upon the presence or absence of a defined background canine population and the data sources utilised for these studies. Specific risk factor studies for various types of canine neoplasia are detailed to demonstrate the scope of research in this field, while considering their strengths and weaknesses. The importance of the dog as a sentinel and animal model for human neoplastic disease is discussed and finally, the study of the geographical distribution of neoplastic disease is briefly described, highlighting the requirements for such inquiry. The limited epidemiological investigations of this element of canine neoplasia are also reviewed.

1.2 EVOLUTION AND DOMESTICATION OF THE DOG

For many years there was speculation regarding the origin of the domestic dog, including whether some types of dogs were derived from the wolf and others from the jackal (Lorenz, 1954), or whether an Asiatic wild dog was the progenitor of the domestic dog (Zeuner, 1963). These authors based their theories on evidence from studies of behaviour and vocalization of the wild species. In the past two decades evidence based on numerical taxonomy, behaviour, morphology and genetic distance analysis has shown convincingly that the principal ancestor of the dog must be the wolf, *Canine lupus* (Wayne, Nash and O'Brien, 1987; Wayne *et al.*, 1989). However, controversy has remained, given that the wolf, jackal and coyote can interbreed and produce fertile offspring (Clutton-Brock and Jewell, 1993).

The exact dating for the domestication of the dog is still unclear. The earliest archaeological find of a domesticated dog, as defined by anatomical criteria, is a small mandible from a late Paleolithic grave at Oberkassel in Germany (Benecke, 1987). It is dated to 14,000 years ago, which is 2,000 years earlier than the sites in western Asia, where canid remains have been identified as what would now be referred to as being a dog (Tchernov and Valla, 1997).

Analyses of mitochondrial DNA (mtDNA) control region sequences from both wolves and domestic dogs showed the sequences to have considerable diversity and support the hypothesis that wolves are ancestors of dogs. However, most dog sequences belonged to a divergent monophyletic clade sharing no sequences with wolves. The sequence divergence within this clade led the authors to suggest that the actual origin of the dog dates to more than 100,000 years before the present, rather than 14,000 years before present as suggested by the archaeological record (Vila *et al.*, 1997). However, a second research group, looking at mtDNA from dogs in Asia, Europe, Africa and arctic America, found that while most dogs shared a common gene pool, genetic diversity was highest in East Asia, suggesting domestication of this species began in this region. This finding, and the pattern of phylogenetic variation, suggested an origin for the domestic dog of approximately 15,000 years ago (Savolainen *et al.*, 2002).

The association between man and wolves appears to have been present since the Middle Pleistocene period. Wolf bones have been found in early hominid sites dated to more than 200,000 years ago at Zhoukoudian in China (Olsen, 1985), and at the 400,000-year-old site of Boxgrove in Kent, England (Clutton-Brock and Jewell, 1993; Pitts and Roberts, 1997). These finds showed that the sites of occupation and hunting activities of the two social predators, humans and wolves, must have overlapped. Bone remains show that wolves were killed by hunters, probably for their skins for clothing, and the wolf meat may then have been consumed by the hunters. However, not all interaction was necessarily hostile and it is possible that a passing interdependence may have developed. Clutton-Brock (1994) suggests that wolves may have been socialised into human company because humans may transgress barriers which otherwise keep species apart. Wolf pups may have been part of a human community as playmates with children, and as scavengers on the periphery of settlements, feeding on human debris. There are societies today, on islands in the Pacific Ocean and in Southeast Asia, where pups are fed and suckled by women and become their playthings (Tuan, 1984; Serpell and Paul, 1994). As humans followed their instinct to dominate other beasts, once a wolf was in a subordinate role, this may have led to certain behavioural and anatomical changes towards domestication of this carnivorous species (Clutton-Brock and Jewell, 1993).

The earliest archaeological evidence of the transition of the dog from an associate of the human settlement to an animal for which obvious affection was involved in the human-dog relationship, is depicted in a Nartufian burial site at Ein Mallaha in Israel, dating from 12,000 years ago. At this site, the skeleton of an aged human was found buried with a puppy, approximately 5 months old. The human skeleton lay on its right side, with its hand resting on the thorax of the puppy, signifying a companion-type relationship. The animal was too large to have been a jackal and could only have been a tamed wolf, or a "dog" (Davis and Valla, 1978; Clutton-Brock and Jewell, 1993).

1.3 HISTORY OF PET OWNERSHIP

1.3.1 The Ancient World

We think of the dog today as a companion animal, however this animal has had many other associations with human culture. In some early civilizations, the dog was associated with worship, such as the Egyptian canine funerary god, Anubis (Schwabe, 1994). The dog was associated with death by the ancient Greeks and in Mesopotamia, but it appears that beliefs gradually evolved to those where the dog could ward off death, prevent death and promote healing (Leach, 1961). The Mesopotamian healing goddess Gula had a dog as a healing companion, and dogs were important in Greek healing temples. The Zoroastrian Persians had a more positive relationship linking the dog with death. This society believed that a deceased person needed to be gazed upon by a dog for the soul to enjoy the afterlife. Thus, among these Persians all dogs received unusual respect, protection and care (Schwabe, 1994).

The dog became one of man's earliest companions. There is evidence in Africa and Asia that dogs were members of early societies, and the ancient Greeks and Romans depict dogs in their paintings of daily life. Many epitaphs from the third/second centuries BC to fourth/fifth centuries AD express mourning at the loss of loved canine companions (Bodson, 2000). Tombstones or sarcophagi were used, sometimes engraved with a portrait and epitaph. Sometimes funerary offerings similar to those found in human graves were also made. The practice of animals being represented on their owners' own tombstones first appeared in archaic Greece, implicitly recognising animal companionship as an important feature of the deceased person's life (Bodson, 2000). Epitaphs of this period suggest that strong and selfless affection for animals should not be considered a uniquely modern phenomenon (Serpell and Paul, 1994).

1.3.2 The Middle Ages

During the Middle Ages, there was a class division of attitude towards dogs, which appears to be due to the influence of religious and secular authorities. The number of dog breeds also increased dramatically during the thirteenth to the fifteenth centuries AD. A driving

force behind the development of particularly the hunting breeds was the emergence and establishment of the aristocracy, to whom hunting was a symbol of power and status (Menache, 2000). The use of certain breeds for certain aspects of the various hunts was part of the ritual. Many breeds were named after the beast they were bred to hunt, such as the Deerhound and Wolfhound. Law restricted hunting and the ownership of hunting dogs to the upper class, beginning with laws in the eleventh century that made it a crime for poor people to keep greyhounds (Ash, 1927). Medieval kings indirectly caused the detachment of hunting from its former role as a principal source of subsistence, to one that was linked with the aristocracy.

Lap dogs, too small to be of any value to the hunt, became popular with the female nobility, while the men lavished attention on the more masculine hunting hounds. The upper classes therefore, seem to have kept pets for a variety of reasons. Although companionship was important to some, others viewed their dogs as emblems of social status and a way of advertising their power (Savishinsky, 1983).

Dogs were not accepted as loyal companions or essential workers by all levels of society. Religious and secular authorities encouraged ordinary working people to view animals as useful objects, only present for their ability to assist man. For human beings to develop attachments to companion animals and to anthropomorphise or humanise the animals was a threat to the foundations of religious belief (Serpell, 1986).

1.3.3 1700 - 1900 A.D

The popularity of animals as pets increased in England during the seventeenth to eighteenth centuries, as the attitudes of the religious authorities relaxed and a change in thinking by the prominent philosophers of the time exerted influence on the common populace. The most significant changes in attitudes to animals occurred during the eighteenth and nineteenth centuries, particularly with regard to animal welfare. By 1800, animal welfare had become a concern to the educated middle-classes and a series of bills was proposed in the House of Commons to rule out various forms of cruelty to animals (Ritvo, 1994). Though initially defeated, within 50 years certain blood sports, such as dog-fighting and cock-fighting were prohibited. The changing views towards animal welfare

were accompanied by a tremendous increase in the popularity of pets, which at this time spread downward from the aristocracy and into the middle-classes. Dog breeding became a pastime, with the first official recorded dog show taking place in 1854 in Newcastle, for setter and pointer breeds only. By the end of the century, dog shows were established throughout the United Kingdom, and the role of the dog as a companion animal had become the primary reason for dog ownership (Kennel Club, 2002a).

1.3.4 Present day

Present-day attitudes to pet-keeping demonstrate the strong attachment and empathy that many people feel towards their animal companions. It is evident that the keeping of companion animals in the UK is now a widespread and entrenched feature of life. It satisfies human interests and is not ordinarily contrary to the social or public interest (Council for Science and Society, 1988). According to a report by the Pet Food Manufacturer's Association, nearly 50% of households in the United Kingdom owned at least one pet in 2001. Dogs are now owned by 4.8 million households (Pet Food Manufacturers Association, 2002).

It is generally accepted that dog ownership is associated with clinically significant health effects in people, including improved survival after a coronary event, as shown by the studies of Friedmann *et al.* (1980) and Friedmann and Thomas (1995) . Short-term psychological and physiological benefits of pet ownership include reduction of blood pressure and mitigation of psychological indicators of anxiety. However not all studies of pet ownership have supported the hypothesis that pet ownership will be beneficial to the physiological and psychological well-being of the owner. Rajack (1997) found that pet ownership was not related to six-month survival incidence for people with previously diagnosed cardiovascular disease.

Dogs have been regarded as helpers in activities such as herding, guarding and hunting for centuries. Assistance dogs are now trained to help people with a range of physical and mental disabilities, to enable affected individuals to have more independent living. Such dogs include guide dogs for the blind, 'hearing' dogs for the deaf, and seizure alert dogs for sufferers of epilepsy (Pfaffenberger *et al.*, 1976; Weiss and Greenberg, 1997). As well

as having important working roles with the disabled, dogs fulfil essential roles in the police force as 'sniffers' for the detection of illegal narcotics (Dumonceaux and Beasley, 1990), as trackers for missing persons and may assist in the location of victims of disasters, both natural such as avalanches and earthquakes, and man-made (Duhaime *et al.*, 1998).

1.4 AETIOLOGY OF NEOPLASIA

1.4.1 Introduction

A neoplasm can be defined as an abnormal tissue that grows by cellular proliferation more rapidly than normal, continues to grow after the stimuli that initiated the new growth cease, shows partial or complete lack of structural organization and functional coordination with the normal tissue, and usually forms a distinct mass of tissue which may be either benign or malignant (Stedman, 2000). Benign neoplasms are encapsulated, non-invasive masses caused by overproliferation of specific cell types that resemble the cell of origin. The characteristics of a malignant neoplasm are considered to be the ability to grow in an uncontrolled manner; ability to invade surrounding tissues and ability to spread to distant sites (metastasize) in the body. The term "cancer" typically refers to malignant neoplasia. The ability for pathologists to distinguish between a benign and malignant growth is imperative as the prognosis and treatment recommendations are markedly different. Benign growths are sometimes, but not always precursors of malignant growths (London, 2000; King, 2000).

Carcinogenesis is a multistep mechanism resulting from the accumulation of errors in vital regulatory pathways which involve the individual cell and its micro-environment. A loss of the normal regulatory events that control cellular growth and proliferation is thought to arise from mutations in genes encoding the regulatory process (King, 2000; Morris and Dobson, 2001b). Although two genetic alterations may be sufficient, as many as five may be necessary for malignant transformation (London, 2000). Tissue-specific and cellular-specific factors as well as other gene products mediate the processes of differentiation, growth and apoptosis, or programmed cell death. Alterations in these gene products can lead to premalignant changes, benign tumours or malignancy. The most common genes to

be affected by such mutations are either proto-oncogenes or tumour suppressor genes (London, 2000; Morris and Dobson, 2001b; Cullen, Page and Misdorp, 2002).

1.4.2 The main genes

Proto-oncogenes are normal cellular sequences of DNA which function to regulate cell growth, proliferation, differentiation and transcriptional activation in normal cells. Proto-oncogenes become oncogenes when their level of expression or gene production is altered, activating them to behave in an aberrant way. Several different kinds of mutations can affect the function of a proto-oncogene, including point mutation, deletion, translocation or amplification, all if which may result in constitutive expression or activation of the resultant protein (London, 2000).

Tumour suppressor genes are responsible for products that act to restrict or inhibit cell proliferation. When tumour suppressor genes are lost or malfunction, cell proliferation can occur in an uncontrolled fashion, leading to the development of a tumour. Unlike oncogenes, which are usually activated through somatic mutations, suppressor gene mutation may be found in germ cells, allowing the defect to be passed from one generation to the next. Offspring acquiring the mutated suppressor gene are more likely to develop cancer because of the pre-existing defect in cell growth regulation. The classic example of a tumour suppressor gene is the p53 gene, which is involved in multiple central cellular processes, including transcription, DNA repair, genomic stability, senescence, cell cycle control and apoptosis. The p53 gene can be inactivated by structural mutations, interactions with viral products, and endogenous cellular mechanisms in many cancers (Hahn, 1998). Mutations in the p53 gene are estimated to occur in as many as 60% of human tumours (Prokocimer and Rotter, 1994; Alevizopoulos *et al.*, 1997).

Mutations of the p53 gene are found in a number of spontaneous canine tumours and may contribute to increased cytogenetic alterations and tumour formation (Veldhoen *et al.*, 1999). Alterations in expression of p53 have been reported in canine osteosarcoma (Sargartz *et al.*, 1996; van Leeuwen *et al.*, 1997; Levine and Fleischli, 2000), thyroid carcinoma (Devilee *et al.*, 1994), circumanal gland adenoma (Mayr *et al.*, 1997), lymphoma (Veldhoen *et al.*, 1998), mammary tumours (van Leeuwen *et al.*, 1996; Chu *et*

al., 1998; Mayr *et al.*, 1998; Veldhoen *et al.*, 1999) and mast cell tumours (Jaffe *et al.*, 2000; Ginn *et al.*, 2000).

1.4.3 Principles of Carcinogenesis

A multi-step process of cancer development is a well-accepted concept, in which mutations accumulate over time, eventually allowing outgrowth of neoplastic cells (Farber, 1984; Barrett, 1993; Beckmann *et al.*, 1997). The first step, known as initiation, induces a permanent and irreversible change in the DNA of the affected cell. DNA synthesis is required for fixation of the initiated state. This stage alone is not sufficient to induce neoplastic transformation, and initiated cells cannot be distinguished from other cells in the surrounding environment. Such initiated cells may then be acted upon by promoting agents that cause reversible tissue and cellular changes. These agents are not capable of inducing neoplastic transformation unless they act on previously initiated cells. Promoting agents may induce changes in cellular morphology or mitotic rate, however they do not alter the genome itself. Promotion normally takes place over a prolonged period of time during which there is near continuous exposure of the promoting agent. Either an initiated cell, or a cell affected by promoting agents may be acted upon by progressing agents which induce significant alterations in the genome that may affect cell growth or invasiveness, leading to the development of cells exhibiting a malignant phenotype and metastatic potential (London, 2000; MacEwen, Khanna and Radinsky, 2001). The process induced by progressing agents causes irreversible genetic alteration.

Multi-step carcinogenesis can occur through several main pathways and more than one pathway may be involved in the generation of a particular tumour. Spontaneous occurrence of point mutations, chromosomal translocations and gene amplifications may occur. DNA repair mechanisms, if inadequate, may lead to permanent heritable DNA changes (Cotran, Kumar and Collins, 1999). It is likely that spontaneous mutations accumulate over a lifetime which may partially explain why many tumours arise in the mature or aged individual (Ershler and Longo, 1997).

There are a number of neoplastic diseases in humans which have been shown to be the result of an inherited genetic defect. Most exhibit recessive inheritance and involve tumour

suppressor genes. The most well-described of these is retinoblastoma, a tumour of the eye in young children (Greger *et al.*, 1994). New genes have also recently been identified that predispose individuals to certain types of cancers, such as BRCA mutations and the development of breast cancer (Kent, O'Donoghue and O'Hanlon, 1995; Cotran, Kumar and Collins, 1999). No specific hereditary genes have been identified in domestic animals but there appear to be a number of tumour types that seem to show breed predispositions, notably sarcoma (Wood *et al.*, 2000b; Morris and Dobson, 2001b). The Boxer is mentioned in many publications as a breed with higher incidence (Priester, 1967; Cohen *et al.*, 1974; Priester, Goodman and Theilen, 1977; Arnesen *et al.*, 2001) and has been suggested by one author as suitable candidate for genetic study (Misdorp, 1996). A familial incidence for certain types of neoplasia has been shown within certain breeds, such as the Bull Mastiff and lymphoma (Onions, 1984) and the Bernese Mountain dog and malignant histiocytosis (Padgett *et al.*, 1995). Selective breeding practices may have produced populations of purebred dogs with defects in tumour suppressor genes, DNA repair mechanisms or deficient immunosurveillance (London, 2000).

There are biological agents capable of inducing neoplasia, such as the RNA retroviruses and DNA tumour viruses, some parasites (e.g., *Spirocerca lupi*) (Johnson, 1992) and certain hormones (e.g. oestrogen) (Schneider, Dorn and Taylor, 1969). In companion animal medicine, feline leukaemia virus is of greatest clinical importance (Rosenberg *et al.*, 1981; Miyazawa and Jarrett, 1997; Flynn, Hanlon and Jarrett, 2000). This slow-acting retrovirus induces malignancy by insertion of provirus near or into a proto-oncogene, which may disrupt normal gene regulation. To date, no equivalent retrovirus has been identified in the dog, despite extensive research. The canine papovavirus is an example of a DNA tumour virus (Shimada, Shinya and Awakura, 1993).

Many chemical compounds, both naturally occurring and synthetic, have been identified as being capable of inducing neoplasia. In most cases, repeated exposure over a long period of time is required to cause DNA damage, thus there is a long latency between exposure and tumour development. In humans, probably the most well known are polycyclic aromatic hydrocarbons which are present in tobacco smoke, combusted fossil fuels and cooked meats (Cullen, Page and Misdorp, 2002). Experimental studies have proved that pulmonary neoplasms can be produced in dogs by bypassing the nasal filtration mechanism

to allow direct inhalation of cigarette smoke (Auerbach *et al.*, 1970). Other examples of chemical carcinogens for dogs include nitrosamines and tumour formation in the lung (Benfield *et al.*, 1981; Benfield *et al.*, 1986) and liver (Hirao *et al.*, 1974). Carcinomas in the canine urinary bladder have been induced by 3,3'-dichlorobenzidine (Stula *et al.*, 1978).

There are a range of physical factors which are capable of inducing neoplasia, including ultraviolet and ionising radiation, as well as foreign bodies and fibres, such as asbestos. Ultraviolet radiation can induce specific DNA changes such as pyrimidine dimers and point mutations, and is known to be a cause of several different cancers in animals and humans, including squamous cell carcinoma (Dorn *et al.*, 1971), melanoma in humans (Fears *et al.*, 2002; Landi *et al.*, 2002) and cutaneous haemangiosarcoma in dogs (Nikula *et al.*, 1992). Electro-magnetic radiation (X-rays, gamma rays), particulate radiation (electrons, protons, neutrons, alpha particles) and heavy ions can induce strand breaks, point mutations, deletions and chromosomal fragmentation. In human medicine, chronic leukaemias, thyroid tumours and breast cancer, and others, have been linked to radiation exposure (Inskip *et al.*, 1997). Radiation-induced tumours have also been recognised in dogs after treatment of primary tumours with radiation therapy (Gillette *et al.*, 1990; London, 2000; Dickinson *et al.*, 2001).

1.5 DIAGNOSTIC TECHNIQUES

1.5.1 Introduction

Neoplasia may produce many signs in an affected animal, ranging from visible lesions and superficial masses, to internal masses detectable by a qualified veterinarian, or a broad spectrum of pathophysiological signs with no obvious palpable evidence of abnormal cell growth. There are numerous pathological techniques used to confirm the presence of neoplastic disease, once suspected by an owner or veterinarian. The diagnosis of neoplasia is most commonly based upon the microscopic examination of haematoxylin and eosin stained tissue sections, histopathology (Morris and Dobson, 2001a), though the cytological examination of exfoliated or aspirated cells can be diagnostic for some tumour types (Morrison and De Nicola, 1993). Once cells have been identified as neoplastic, the aims of the pathologist are to determine the cellular origin of the neoplasm and to define its cellular

characteristics to determine if the tumour is benign or malignant. In some cases, special stains, immunohistochemistry (Griffey *et al.*, 1993) or electron microscopy (Madewell, Griffey and Munn, 1992) may be needed to identify the cell of origin (Morrison and De Nicola, 1993). The clinician also relies upon the pathologist to provide information regarding mitotic activity, completeness of tumour removal and other information important for prognosis, so that appropriate treatment strategies may be selected. Immunophenotyping (Fisher *et al.*, 1995), flow cytometry (Madewell *et al.*, 1991) and image analysis (Griffey *et al.*, 1998) are specialised diagnostic procedures which have been found to provide useful prognostic information in some human and canine cancers. The identification and measurement of "tumour markers", being antigens associated with certain malignancies, is a rapidly expanding area of cancer diagnosis research. Finally, the diagnosis of some neoplasms is made by indirect methods such as endocrinological tests. An important example in the dog is the diagnosis of functional pituitary neoplasia, as procurement of suspected neoplastic tissue is a difficult and dangerous procedure. These procedures will be reviewed in the following section.

1.5.2 Histopathology

Biopsy is frequently the definitive step in cancer diagnosis. The examination of suspected tissues by a trained histopathologist can provide a definitive diagnosis in the majority of cases, as well as evaluate tumour grade, depth of invasion of lymphatics, blood vessels or adjacent tissue structures, detection of *in situ* lesions and premalignant changes and assessment of surgical margins for remnant disease. Routine histological specimens are a 5 µm-thick slice of formalin-fixed and paraffin-embedded tissue. Approximately 90% of oncological cases in humans can be diagnosed by light microscopy using haematoxylin and eosin stains (Pfeiffer and Wick, 1991). This percentage is probably a close estimation of the situation in veterinary medicine as well (Powers, 1996). In the remaining 10% of cases, special stains or special procedures may be helpful (Section 1.5.4).

Where suitable cryostat facilities are available, frozen tissue sections can be made during surgical procedures where rapid diagnosis is wanted. Samples are quick-frozen using liquid nitrogen, sectioned on a cryostat, fixed, stained, and examined within 10 to 15 minutes. These samples are inferior to those processed into paraffin, thus diagnostic accuracy may

suffer. This service is available at few veterinary institutions, and few veterinary pathologists are experienced with the interpretation of frozen tissues (Powers, 1996).

There are three main tissue biopsy techniques, with interpretation of the histopathological findings being dependent on the technique used and the clinical information provided with the sample (Withrow, 2001). Needle core biopsy utilizes a coring instrument which cuts a core of tissue from a mass. Only local anaesthesia with or without sedation is required in most cases (Martin, 1993). Although the samples are small, tissue architecture is preserved. However, this biopsy technique is suboptimal for cystic, inflamed, poorly defined, deep or non-palpable masses.

Incisional biopsy involves the sampling of tissue at the junction of the normal tissue and mass to evaluate tumour behaviour. This technique is useful for large masses and those in sites where complete removal is impossible or would require complex surgery. Incisional biopsy is more invasive and generally needs general anaesthesia, but with appropriate tissue sampling, should yield definitive diagnosis where a pre-excision diagnosis is needed (Salisbury, 1998; Prymak, 1999).

Excisional biopsy involves the removal of the entire mass with a surrounding margin of normal tissue. This technique may result in diagnosis and treatment with a single surgery however this ideal outcome is usually only possible for smaller skin masses. Complete removal of the mass is determined by microscopic examination of the surgical margins. This is essential given that some tumours may present clinically as discrete masses, however invasion of the adjacent tissues is evident at the cellular level. Only excisional biopsies are both diagnostic and therapeutic (Martin, 1993; Withrow, 2001).

Accurate and informative biopsy reports are dependent upon numerous factors associated with the submitted tissue. Samples must be of adequate size, be representative of the lesion and be adequately fixed. Careful handling of the biopsy throughout the sampling procedure is required to avoid crush artifact, or fragmentation, which may result in distortion of the cell architecture and individual cell morphology (Goldschmidt, 1993; Powers, 1996; Morrison, 1998).

The formalin-fixed biopsy sample is currently the most common sample type in veterinary pathology, and is likely to remain so for clinical samples, because of the need to rapidly establish a diagnosis based on criteria obtained from histopathologic examination. Formalin-fixed samples are also suitable for special staining, however for other adjunctive procedures, other fixatives and tissue preservation techniques may be preferred (Morrison and De Nicola, 1993). For those neoplasms that require further diagnostic procedures in order for the pathologist to provide the most accurate diagnostic and prognostic information, frozen tissue is optimal for immunohistochemistry and cytogenetic analysis, and glutaraldehyde is the preferred tissue fixative for ultrastructural analysis. Communication between the clinician and the diagnostic pathologist is important to ensure that samples are collected and handled in the most appropriate manner, particularly for neoplasms that pose diagnostic challenges (Goldschmidt, 1993; Henderson and D'Andrea, 1993; Madewell and Griffey, 2000).

1.5.3 Cytology

Cytological examination has proved to be extremely valuable in certain organs in which the clinical signs suggest the presence of cancer (Madewell, 1987). The medium of interpretation of cytology specimens obtained from tissue impression, aspiration, scraping or fluid sedimentation is a smear of cells. The primary advantages of cytology are speed, low cost and, when interpreted by a competent cytologist, reliability. In addition, cell samples may be obtained by aspiration biopsy from organs and structures that are not easily accessible to a surgical biopsy (Koss *et al.*, 1984). Cytological specimens are routinely examined by clinical pathologists, rather than histopathologists (Morrison and De Nicola, 1993).

The major disadvantage of this diagnostic technique is that it cannot provide information about invasiveness, due to the disruption of the normal tissue relationships which inevitably occur due to the sampling and preparation methods for cytological examination of cells. Cytological interpretation is commonly used to characterise cells as normal, benign, malignant or inflammatory. Due to the diversity of microscopic patterns that may be seen in samples obtained by cytological techniques, in most instances, histopathological

confirmation is needed to obtain a definitive diagnosis and provide important prognostic information (Morrison and De Nicola, 1993).

1.5.3 Adjunctive procedures

1.5.3.1 Special Stains

Special stains are commonly used to assist in the diagnosis of certain poorly differentiated tumours (Misdorp, 1987). Stains may allow documentation of the presence of cell products, such as the use of toluidine blue or giemsa to identify granules in poorly differentiated mast cell tumours (Cullen, Page and Misdorp, 2002), or the use of a melanin bleach or iron stain to help distinguish between haemosiderin and melanin in suspected cases of melanoma (Powers, 1996).

1.5.3.2 Immunohistochemistry

Immunohistochemistry is also used for the evaluation of poorly differentiated neoplasms (Sandusky, Carlton and Wrightmann, 1987). It is a staining procedure using antibodies to identify different cellular components, such as specific intermediate filaments, secretory substances, or proteins and may be performed on specimens fixed in formalin and processed into paraffin blocks, or on frozen tissue (Powers, 1996). Immunohistochemistry often allows the precise determination of the histogenesis of a neoplasm that would otherwise be characterised as undifferentiated on the basis of morphologic examination (Madewell and Griffey, 2000).

1.5.3.3 Immunophenotyping

Immunophenotyping allows the classification of cases of lymphoma into subtypes by cytofluorographic analysis using a panel of antibodies (Fisher *et al.*, 1995;Cangul, Teske and Ingh, 1998). This is now routinely available at certain veterinary diagnostic laboratories and is considered to be of prognostic significance in cases of canine lymphoma.

1.5.3.4 Flow cytometry

Flow cytometry is an analytic procedure that can be used to evaluate cell suspensions obtained from suspected neoplastic masses. In human medicine this procedure is frequently used to diagnose and occasionally to monitor for recurrence of various tumours such as bladder carcinoma (Powers, 1996). The process is used to determine the DNA content or ploidy of cells and is based on the principle that malignant cells tend to be aneuploid, i.e. abnormal DNA content, while normal tissue, benign tumours and reactive tissues are usually diploid, i.e. normal DNA content. Occasionally, benign tumours and reactive tissue can be aneuploid and malignant cells can be diploid. In some instances, aneuploidy can be predictive of survival time. Flow cytometry has been used to evaluate tumours in dogs, including osteosarcomas (LaRue *et al.*, 1994), mast cell tumours (London *et al.*, 1996) and lymphomas (Sanz *et al.*, 1998).

1.5.3.5 Electron microscopy

Electron microscopy may help to identify specific cellular features that may be useful in distinguishing poorly differentiated tumours; however as with other special procedures, the initial diagnosis of neoplasia is made with routine histopathology. It is not useful for distinguishing benign from malignant cells in many cases because the magnification is too high and the tumour pattern in the tissue is not evident (Pfeiffer and Wick, 1991). It must be performed on extremely small representative tumour samples, preferable preserved in glutaraldehyde. Due to its requirement for specialised equipment and technical support, this procedure is only available at a limited number of diagnostic laboratories (Powers, 2001).

1.5.3.6 Image analysis

Microvessel density within tumours is considered to be a measure of angiogenesis, and has been found to be a useful prognostic indicator of tumour behaviour for some tumour types in humans, for example, cutaneous melanoma. Computer-assisted image analysis is a technique by which tumour vessel density may be assessed. This method allowed objective assessment of tumour vessel density in canine mammary tumours, and was found to be

greater in malignant tumours when compared with histologically benign tumours (Harari *et al.*, 1995; Restucci, Vico and Maiolino, 2000). Tumour vessel density might be a useful measurement to assist in prediction of clinical outcome, however routine histopathologic examination of tissue biopsies must be performed to establish diagnosis and to localise exact areas of tissue invasion if present. If the experience in human medicine is a guide for veterinary medicine, then it is probable that image analysis may find application in the veterinary diagnostic laboratory (Griffey *et al.*, 1998).

1.5.4 Tumour Markers

A tumour marker may be considered a laboratory measurement of a substance or a process that provides clinically useful information with regard to tumour diagnosis and patient management (Madewell, 1997). Few putative tumour markers have been described in veterinary medicine, however the rapid advances in immunology and cell and molecular biology in the past decade have provided methods for the critical examination of biological processes associated with neoplasia, and the contrast of abnormalities of cell function with those of homeostasis. A few of these methods have undergone clinical appraisal, and are currently available for routine use.

1.5.4.1 Hormones

Alterations in plasma hormone concentrations and their biological effects may alert the veterinarian to suspect certain types of neoplasms. The alteration of plasma cortisol concentration using dexamethasone inhibition and/or adrenocorticotropin (ACTH) stimulation is used in a variety of procedures for the diagnosis of functional pituitary and adrenocortical tumours (Madewell, 1997). Due to the difficult location of the pituitary gland, histopathological confirmation by tissue biopsy is dangerous and impractical. Thus, functional pituitary tumours are only confirmed by tissue biopsy at necropsy, if performed.

1.5.4.2 Tumour-associated antigens

In both veterinary and human medicine, the discovery of quantitative and qualitative methods to identify and measure antigens associated with malignancy is a growing field

(Carpinito *et al.*, 1996; Pectasides *et al.*, 1996). A noninvasive diagnostic test, Bard BTA (bladder tumour-associated antigen; Bion Diagnostic Sciences, Inc.) was first tested in human patients with transitional cell carcinoma (TCC). The test is a qualitative, rapid, latex agglutination test utilising antibodies to a bladder tumour-associated glycoprotein complex that is detectable in the urine of patients with TCC. The Bard BTA test has recently been shown to be sensitive for the detection of canine TCC with its suggested use being in routine screening of geriatric, at-risk patients to aid in early detection of disease and guide further diagnostics (Borjesson, Christopher and Ling, 1999). Histopathological confirmation remains essential for definitive diagnosis of TCC.

1.5.5 Molecular oncology

Molecular oncology studies involve the isolation of DNA, RNA or protein from tumour cells and the analysis of their components for abnormalities related to the presence, absence or amplification of a specific gene or its products or other alterations such as chromosome translocation (Griffey, Kraegel and Madewell, 1998; Veldhoen *et al.*, 1999; Madewell and Griffey, 2000). Cytogenetics remains limited in veterinary medicine because of the specialised expertise and equipment required, and is currently not used as a routine tool in the diagnosis of canine neoplasia.

1.6 EPIDEMIOLOGICAL STUDIES OF CANINE NEOPLASIA

1.6.1 Introduction

Published observations on the occurrence and types of canine neoplasia first appeared in the veterinary literature in the twentieth century. The initial stimulus for such dissemination of information appeared to have been the comparative aspects of neoplasia in domestic animals and humans (Withers, 1939; Cotchin, 1951). The first publications were surveys conducted by veterinary pathologists, who described the details of neoplasms submitted to their diagnostic histopathology services. The sources of the neoplasms included dogs seen at the clinical department of a veterinary college (Withers, 1939; Knight and Douglas, 1943), samples submitted from dogs seen by private practitioners, or a combination of the two (Cotchin, 1951; Cotchin, 1954). Although canine practice was a

minor component of the veterinary curricula and professional life during this time period, these pioneers in the field commented that publication of surveys of canine neoplasms may provide a basis for prognosis as well as potentially providing information regarding their aetiology (Withers, 1939; Cotchin, 1951; Cotchin, 1954). It was also noted that recording systems used in the veterinary field were inadequate for the collection of data suitable for analytical epidemiological studies, and that it was impossible to determine what percentage or segment of the animal population was represented by the tumour submissions (Cotchin, 1951).

Unlike the human census, a comprehensive, nationwide demographic survey of the canine population has never been undertaken in any country. However there have been studies designed to provide estimates of the canine population for defined geographical regions and time periods, mainly in the United States (Dorn and Chaulk, 1964; MacVean *et al.*, 1978; Lengerich *et al.*, 1992; Teclaw *et al.*, 1992) and the United Kingdom (Thrusfield, 1989; Gregory, 2000). Many of these surveys have provided the baseline population, or denominator data, required to determine incidence, prevalence and risk of neoplasia and associated characteristics in that population. However, in general, animal population surveys are expensive and time-consuming to conduct. They are also unlikely to account for the entire canine population, unless all subpopulations are accounted for outside the household pet dog population, such as roaming strays, and dogs in animal welfare shelters. For both of the major population-based studies of the incidence of canine neoplasia, veterinary practitioners were recruited to submit all samples of suspected neoplasms to a diagnostic pathology service. Thus, the results for these studies were either adjusted for the likelihood of dogs not receiving veterinary attention (Dorn *et al.*, 1968) or were only applicable to the dogs in veterinarian-using households (MacVean *et al.*, 1978).

Other data sources for epidemiological studies of canine neoplasia have included medical records from first-opinion and referral veterinary hospitals (Priester and Mantel, 1971; Cohen *et al.*, 1974; Reid-Smith, 1999), the Veterinary Medical Database (see below), laboratory canine colonies (Auerbach *et al.*, 1970; Nikula *et al.*, 1992), veterinary insurance companies (Wood *et al.*, 2000b; Dobson *et al.*, 2002), veterinary pathology departments (Goldschmidt and Shofer, 1992; Yamagami, 1996), animal shelters (Dorn *et al.*, 1966) and canine licensing records (Cohen, Booth and Sussman, 1959; Robinson,

1968). In the absence of accurate general population data, the case-control study has become the most popular analytical method because the study population may be a subset of the general population, so long as strict criteria are followed for the selection of cases and more importantly, controls (Kelsey, Moore and Glickman, 1998). Where efficient systems for the recording, storage and retrieval of data exist, hospital, colony or pathology records can allow these types of studies to be undertaken efficiently and inexpensively. However, the findings from these studies, strictly speaking, may only be applied to the study population, because the source population (e.g., dogs attending a veterinary clinic/hospital) comprises an undefined fraction of the total population. The case-control study design is also reliant on the accuracy and to a lesser extent, precision of existing data records, which may have been created for a primary purpose other than epidemiological study (Breslow and Day, 1980). This salient point may render a data source as unsuitable for analysis unless significant time is spent tracing back to the data origins and making the necessary corrections, thus increasing the running costs of the intended study. Another disadvantage of this study design is its susceptibility to the introduction of biases (Breslow, 1982; Lilienfield and Stolley, 1994). For example, the types of cases seen at a hospital/clinic will be influenced by factors such as the resident veterinarians' expertise, availability of specialist diagnostic equipment and whether the hospital/clinic is a first-opinion or referral veterinary practice.

1.6.2 Nomenclature and Coding

For efficient analysis of epidemiological data to be possible, adequate data collection, storage and retrieval systems are needed. Advances in computer technology have resulted in vast amounts of data being present in computerised databases in many facets of the veterinary profession. With regard to the type of data needed for epidemiological studies of disease, standardized methods to allow the conversion of medical terminology into a form suitable for computer analysis are needed. The typical approach is to encode the information using some standard terms.

In human medicine, the World Health Organization (WHO) publishes the standardized nomenclature and coding system used in cancer registries worldwide, the International Classification of Diseases – Oncology (ICD-O), currently in its 10th revision. The ICD-O

was created in 1976 using the long-established International Classification of Diseases(ICD), and the more comprehensive Systematized Nomenclature of Human Medicine (SNOMED), first published by the American College of Pathologists in 1975. The goal of the ICD coding and classification system is to allow morbidity and mortality data from different countries to be systematically collected and statistically analyzed. Most other major classification systems endeavour to make their systems compatible with ICD, so that data coded in these systems can be mapped directly to ICD codes. SNOMED has undergone further refinements with the launch of SNOMED Reference Terminology (SNOMED RT®) in 2000, designed to facilitate the health care field's transition from paper records to electronic records. In 2002, SNOMED Clinical Terms (SNOMED CT®) First Release became available, which provides integrated medical and veterinary reference terminology. With its multilingual framework, it is designed to facilitate the capture, sharing and analysis of health data worldwide (Anonymous, 2002a).

At present there is no nomenclature and coding system universally accepted by the veterinary profession. Perhaps the most widely used system today is The Standard Nomenclature of Veterinary Diseases and Operations (SNVDO) (Priester, 1964; Priester, Wade and McKay, 1977), created in 1964 by researchers in the Epizoology Section at the National Cancer Institute. The SNVDO was partly derived from the American Medical Association's Standard Nomenclature of Diseases and Operations (SDVO) (Thompson and Hayden, 1961). The Systematized Nomenclature of Human and Veterinary Medicine - SNOMED International (Cote *et al.*, 1993) was initially known as the Standard Nomenclature of Pathology (SNOP) (College of American Pathologists, 1971), which was itself drawn from the SNDO. Thus SNVDO codes can be assigned to headings from the ICD-O, because the two systems share the SNDO as a common ancestor. Prior to SNOMED International, the Systematized Nomenclature of Veterinary Medicine (SNOVET) (Palotay, 1983) had been created, based on the second edition of SNOMED. A survey of North American veterinary teaching hospitals performed in 1995 showed that the SNVDO was the most commonly used nomenclature/coding system, with SNOMED, SNOVET, ICD-9 and in-house systems also being used. Some schools were using only free text for items stored in their electronic medical data systems (Pollari, Bonnett and Bamsey, 1996b). In recent years, there has been attempts to map the SNVDO codes to SNOMED terminology. In 1998, this was approximately 90% done. However, it was

reported that problems with term definition and ambiguity were slowing the process (Case, 1998). With the release of SNOMED CT®, guidelines for its use in veterinary medical systems have been recently suggested (Wilcke, Livesay and Zimmerman, 2002).

1.6.3 Data Sources for Epidemiological Studies of Canine Neoplasia

Epidemiological studies of canine neoplasia vary from small-scale studies concentrating on a particular tumour type, to large-scale, population-based surveys undertaken over several years. Many studies have utilized data collected for the specific purpose of epidemiological study; others have involved interrogation of databases in the veterinary domain which contain the required information, however the data was collected for another primary purpose. Examples of the latter include pathology databases and private practice electronic medical record databases. Several important data sources of canine neoplasia will be briefly discussed, illustrating some of the approaches taken in its investigation during the past half century.

1.6.1.1 Population-based data sources

1.6.1.1.1 California Animal Neoplasm Registry

The California Animal Neoplasm Registry (CANR) in Alameda and Contra Costa counties, California was established in July 1963, to collect morbidity information and to provide a source of animals with neoplasia for analytical studies undertaken by the California Cancer Field Research Program (Taylor, 1965; Schneider, 1975). The registry is located at the Department of Epidemiology, School of Veterinary Medicine, University of California, Davis, and holds information on more than 30,000 tumour cases, collected during the years of the registry's activity. The registry was based upon data submitted from recruited veterinary practices, supplied with standardised neoplasm case report forms and specimen mailing materials. The pathology laboratory processed, at no charge, the biopsy specimens sent to the registry and relayed histopathological results to the participating practitioners. Because the registry was based upon submissions from veterinarians, the population-at-risk estimates were adjusted for the likelihood of animals not receiving veterinary attention (Dorn *et al.*, 1968a; Dorn *et al.*, 1968b). Follow-up surveys of the

numbers and demographic characteristics of the animal population in the geographical region covered by the registry were performed to ensure accurate baseline population data was available (Schneider and Vaida, 1975).
The histological types and primary sites of the neoplasms were coded according to the SNVDO and the Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death (ICD), Seventh Revision (Priester, 1964). Both coding systems were used to permit a broad use of registry data in comparative studies with other animal and human data (Dorn *et al.*, 1968a).

Case reports, morbidity information and analytical studies were published using this registry (Dorn *et al.*, 1967; Dorn *et al.*, 1968a; Dorn *et al.*, 1968b; Dorn and Schneider, 1976b; Bender *et al.*, 1982; Bender, Dorn and Schneider, 1983; Bender, Dorn and Schneider, 1984). Data from the registry were used to demonstrate that spaying a dog under 2.5 years of age had a protective effect against mammary cancer (Dorn *et al.*, 1968b; Schneider, Dorn and Taylor, 1969). Estimates of cancer incidence on in dogs and cats according to age, sex and breed derived from the registry have been widely quoted since their first publication. The cancer incidence in humans in the same demographic area was compared (Dorn, 1968; Schneider, Dorn and Klauber, 1968). Other published reports compared specific cancer types common to dogs and humans (Schneider, 1970a; Schneider, 1970b). The use of animal models for the study of breast cancer in women, genetic effects on the development of cancer, and animal cases of leukaemia and lymphoma as possible models for the human disease was also discussed (Frye, Dorn and Taylor, 1967).

1.6.1.1.2 Tulsa Registry of Canine and Feline Neoplasms

The Tulsa Registry of Canine and Feline Neoplasms was established at the College of Veterinary Medicine, Oklahoma State University in 1972 and data were collected for approximately 5 years. The registry shared important characteristics as the CANR, being a population-based tumour registry that represented a defined geographical area. Biopsy or necropsy samples of suspected neoplasms were submitted by all veterinarians in the study area and only tumours histologically confirmed by the registry's pathologists were included. The specimens were submitted with information on the owner (name and

address); animal case history; the veterinarian's observations and treatments; the status of the case at 2, 6, 12, 18, and 24 months after surgery; the size and type of tumour; and other specifics necessary to properly identify the tumour (MacVean *et al.*, 1978). A census of the dog and cat populations of all animal hospitals in the study region was made to provide a "veterinarian-using" baseline population, from which incidence rates were determined. Thus, certain animal sub-populations were excluded, similar to the CANR.

The Tulsa Registry used the SNVDO to code all cases except mammary tumours. A new classification scheme was developed for canine mammary tumours using data from the registry, and the value of cytology in the early diagnosis of mammary neoplasms was demonstrated (Monlux *et al.*, 1977). The registry also provided data for statistical analyses demonstrating the frequencies of certain cancers in pets (MacVean *et al.*, 1978).

1.6.1.1.3 Purdue Comparative Oncology Program

The Purdue Comparative Oncology Program (PCOP) began in 1979 at the School of Veterinary Medicine, Purdue University. Initial and bi-annual follow-up (until an animal's death) data on naturally occurring tumours in animals is collected by the program, which provides histopathologic diagnosis free of charge for any suspected neoplasm submitted by one of more than 200 participating veterinarians. The veterinarians are asked to submit samples from 100% of their suspected cancer patients. Approximately 75% of the entries in the PCOP involve animals in two counties of Indiana. An estimate of the pet population (baseline population-at-risk) in the catchment area of the PCOP was obtained by a telephone survey conducted in 1988 (Lengerich *et al.*, 1992). However, to date, no cancer incidence rates have been published using data derived from the registry. The data collected by the PCOP are also incorporated into the Veterinary Medical Database (Section 1.6.3.2.1).

The work of the PCOP concentrates on providing care for pet animals with specific types of malignancies that are similar to their human counterpart in histopathological appearance, biological behaviour and response to therapy, e.g. bladder, bone and prostate cancer. Tumour-bearing pet dogs are referred to the program by primary-care veterinarians. With the consent of their owners, many dogs enter clinical trials designed not only to assist

their disease, but also potentially to contribute to the development of better therapies for humans with cancer (Knapp and Waters, 1997).

1.6.1.1.4 Norwegian Canine Cancer Registry

The Canine Cancer Registry at the Norwegian College of Veterinary Medicine/Central Veterinary Laboratory was established to show the relationship between hereditary and environmental factors in the development of cancer. The first recommendation for establishment of a national cancer registry for animals in Norway was made by the Collegium of the Norwegian College of Veterinary Medicine (NCVM) in 1984. Following a pilot study in 1989 and an invitation to veterinarians in 3 of 19 Norwegian counties, the definitive project to collect data for the cancer registry was begun in 1990. A fourth county was included in 1991, and data collection continued until March 1998. Participating veterinarians submitted formalin-fixed specimens of all suspected neoplasms for histopathological examination (provided free of charge) at the Department of Pathology, NCVM, and if a diagnosis of neoplasia was made, all data were recorded in a specially designed database (Nordstoga *et al.*, 1997).

Results of epidemiological analyses utilising this data have recently been published and comparative studies with the human Cancer Registry of Norway are planned (Arnesen *et al.*, 2001).

1.6.3.1.5. Cancer Registry (Missouri)

The cancer registry at the University of Missouri was a 2-year pilot program, begun in June 2000. The researchers were only accepting samples from certain breeds during the pilot study (Flat-coated Retriever, Standard Schnauzer, Mastiff, Doberman Pinscher, American Cocker Spaniel, English Cocker Spaniel, Dachshund, Boston Terrier, Beagle, Labrador Retriever, German Shorthaired Pointer). The cancer registry was initiated for the purpose of evaluating DNA collected from frozen cancer tissue and from formalin-fixed cancer tissue to investigate the influence of genes on the occurrence of cancer in dogs (Thornburg, 2001).

1.6.1.2 Referral hospital data sources

1.6.1.2.1 Veterinary Medical Database

The most important referral hospital data source currently collecting data is the Veterinary Medical Database (VMDB), which was established in 1964 by the National Cancer Institute, beginning at the College of Veterinary Medicine, Michigan State University. The VMDB is a multi-institutional veterinary clinical data storage and retrieval facility. Since its inception, 26 veterinary schools in the United States and Canada have participated in the program, with six actively contributing information in 2002. The participants in the program agree to provide complete abstracts using a standardised format for presenting animal, regardless of their complaint. The VMDB is now housed at the School of Veterinary Medicine, Purdue University and is the largest compilation of medical records for domestic animals, holding over 6 million records (Anonymous, 2002b).

All data are abstracted by the participating institutions and diseases are coded according to the SNVDO (Priester, 1975). The following information is available in the VMDB for each record: participating institution, patient identifying number, species, breed, sex and neuter status, age, weight, up to five diagnoses, diagnostic procedures completed, treatments, length of stay, date of discharge, discharge status, and attending clinician. Age at surgical neutering is not recorded on the standardised abstract form, unless this was the procedure relevant to the animal's visit.

Many publications pertaining to canine neoplasia have resulted from data collected by the VMDB, and have been comprehensively reviewed by Kelsey *et al.* (1998). The VMDB has significant advantages as a data source for epidemiological studies, having a standardised coding scheme, easily accessed and analysed data, and a relatively large sample size, allowing subgroup analyses (Ru, Terracini and Glickman, 1998). Most studies are of the case-control type, using animals with diagnoses unrelated to the diagnosis of interest as controls. However, there are also distinct limitations to studies using VMDB data. The structure of the VMDB is responsible for immediate selection bias as the source population is derived from the segment of the general animal population attending participating veterinary teaching hospitals, so more complex conditions requiring treatment unattainable

from most first opinion private practices are likely to be over-represented. Other influences may be socio-economic, with less affluent owners being unable to afford referral hospital fees. Data from these cases thus remain in the private practice sector, regardless of the presenting complaint. To the author's knowledge, to date, there have been no estimates regarding the representativeness of the VMDB population to pet dogs. As is characteristic of the major population-based studies (CANR, Tulsa Registry), the non-veterinarian-attending pet dogs, roaming strays and animal shelter subpopulations are also not represented. Other potential limitations of using the VMDB for epidemiological studies include undetermined diagnostic accuracy; lack of uniform diagnostic criteria; coding of continuous variables such as age and weight as categorical and lack of information of extrinsic variables (Ru, Terracini and Glickman, 1998).

1.6.1.3 Other data sources

There has been a trend in human and veterinary medicine to utilise existing computerised sources of data (Egenvall *et al.*, 1998), such as hospital records and medical insurance databases. However, as with all attempts to use secondary data, quality and completeness of the available information is of utmost concern. Computerisation of primary practices has allowed ease of collection and storage of data from this source. However, many commercial computer packages are designed for accountancy purposes rather than epidemiological research, and may have inadequate data retrieval capabilities to allow extraction of essential patient and diagnostic data. There has been one major study utilising private veterinary practice records for epidemiological studies of canine neoplasia which has demonstrated selected practice management software systems to be capable of supporting data collection on patient population and neoplasia occurrence. The authors are optimistic for this data source to become increasingly available as practice management systems become more sophisticated and user-friendly (Reid-Smith *et al.*, 2000).

A veterinary insurance database provided the raw data for recent studies of canine neoplasia incidence in the UK (Wood *et al.*, 2000b; Dobson *et al.*, 2002). Relative risk figures were calculated using the number of dogs insured by the company as the baseline population. However, although this data source provided a large number of records, disadvantages included the over-representation of younger age groups, and the lack of

histopathological confirmation of neoplasia for all cases, because the diagnosis was accepted as that written on the insurance claim forms submitted by the veterinarian in charge of the case.

1.7 RISK FACTORS FOR NEOPLASIA IN THE DOG

1.7.1 Introduction

Risk factors for neoplasia in the dog may be considered to be either host-related (e.g. age, breed), or extrinsic to the host, such as contagion, diet and environmental exposures. The earliest studies of risk factors were performed using data from the CANR and Tulsa Registry of Canine and Feline Neoplasms (Section 1.6.3.1) and concentrated on host-related risk factors for certain tumour types. Stratification and the Mantel-Haenszel procedure (Mantel and Haenszel, 1959) were utilised to control for confounding in the statistical analyses. Advances in computing technology and statistical software development over the past three decades now allow multivariable analytical techniques to be used routinely to examine the effects of confounders and effect modifiers in epidemiological studies.

The case-control study design has become the key epidemiological approach to studies of risk factors for canine neoplasia, with the VMDB (1.6.3.2.1) being used as the major data source. During the following review of risk factor studies, the data source has been the VMDB or other referral veterinary hospitals, unless otherwise stated. Thus, results and conclusions from the studies remain referable to the referral, rather than the general, canine population.

1.7.2 Host related risk factors

1.7.2.1 Age

Similar to findings in the human population, increasing age has been shown by many researchers to be a significant risk factor for the occurrence of neoplasia in dogs. This was shown using data from the CANR (Dorn *et al.*, 1968b; Priester and McKay, 1980) and

referral hospital data sources (Moulton, 1990; Ru, Terracini and Glickman, 1998). Although the majority of tumours occur in older dogs, certain tumour types are more common in the younger age groups, such as histiocytoma, papilloma, lymphoma, osteosarcoma and fibrosarcoma (Mulvihill and Priester, 1978). One study reported that the three most common sites for neoplasia in dogs less than 6 months old were the haemopoietic system, brain and skin (Keller and Madewell, 1992). Some rare tumours, such as osteochondroma of the trachea, routinely arises in dogs less than 1 year of age with active osteochondral ossification (Hahn and Anderson, 2000).

1.7.2.2 Breed

For hundreds of years, dog breeding has been a pastime for many societies. Although selective breeding has resulted in the creation of breeds conforming to official standards, with selective breeding has also occurred the emergence of breed-specific predispositions to disease, including neoplasia. Selective breeding practices may have produced populations of purebred dogs with defects in tumour suppressor genes, DNA repair mechanisms or deficient immunosurveillance (London, 2000). As previously mentioned (Section 1.4.3), certain breeds are recognised as having an increased risk for the development of certain tumour types. One breed in particular, the Boxer, consistently appears to have an overall increased risk for neoplasia (Priester, 1967; Cohen *et al.*, 1974; Priester, Goodman and Theilen, 1977; Misdorp, 1996; Arnesen *et al.*, 2001). The flat-coated Retriever has been the subject of recent research because this breed seems to be at increased risk for the development of soft-tissue sarcomas (Morris *et al.*, 2000; Morris *et al.*, 2002).

1.7.2.3 Gender

Sex-specific neoplasms obviously occur, and the incidence of these neoplasms is further influenced by the presence or absence of sex specific hormones, an important consideration given the common practice of neutering dogs in the pet population. Entire male dogs are reported to have a higher incidence rate of perianal gland neoplasms than either castrated males or females (Goldschmidt and Shofer, 1992). It is considered that endocrine factors play an important role in the development of particularly canine

mammary neoplasia, the most common form of neoplasia in the bitch (Rutteman and Misdorp, 1993). Entire females are reported to be at sevenfold greater risk of mammary tumour development than bitches spayed at less than two years old. One of the earliest significant findings arising from epidemiological studies of neoplasia in the dog was the protective effect of ovariohysterectomy (spaying) against mammary cancer in the bitch shown using data from the CANR (Dorn *et al.*, 1968b; Schneider, Dorn and Taylor, 1969).

1.7.2.4 Body characteristics

Height, weight and skull shape have been reported as being associated with altered risk for the development of certain tumour types. Large and giant breeds are reported as having high risk for osteosarcoma (Tjalma, 1966; Ru, Terracini and Glickman, 1998). Crosses of these breeds also show increased risk, indicating that size is the significant factor, rather than pure breeding. It has been proposed that the increased weight, height and growth rates of large and giant breeds can cause increased compression and stress to growing points of bones, and subsequent multiple minor trauma to sensitive cells may play a role in malignant development (Straw, 1996).

It has been speculated that skull shape may influence the risk of intranasal neoplasia in the dog. One large series showed that long-nosed breeds had the highest risk, breeds with medium length noses and dogs of mixed breed an intermediate risk, and short-nosed breeds the lowest risk (Hayes, Wilson and Fraumeni, 1982). It has been suggested that there may be increased deposition of inhaled carcinogens on the nasal mucosa due to the mechanism for filtration of airborne particulates in long-nosed dogs (Reif and Cohen, 1971).
Canine and human breast cancer share some histological and biological characteristics, and epidemiological studies in humans have shown that a high-fat diet and obesity increase the risk of breast cancer. Sonnenschein *et al.* (1991) performed a hospital-based case-control study of body conformation, nutritional factors and canine breast cancer, which showed that neither a high fat diet nor obesity 1 year before diagnosis increased the risk of breast cancer, and that thinness as a puppy in spayed dogs was significantly associated with a reduced risk of breast cancer.

Several studies have reported that dogs with cryptorchidism have a markedly increased risk for testicular seminomas and Sertoli cell tumours (Reif and Brodey, 1969; Hayes and Pendergrass, 1976; Reif *et al.*, 1979; Hayes *et al.*, 1985).

1.7.3 Exogenous risk factors

1.7.3.1 Contagion

To date, only one tumour-associated virus has been verified in the dog, a papovavirus which causes papillomatosis, primarily in young dogs (Bregman *et al.*, 1987; Calvert, 1990). Although retroviruses have been shown to be causal agents for lymphoma in other species such as cats, poultry and cows, no infectious cause for lymphoma has been proven to exist in the dog, despite extensive research (Onions, 1980; Tomley *et al.*, 1983; Sykes, King and Cooper, 1985; Theilen, 1997).

Transmissible venereal tumour is a contagious venereal tumour, usually transmitted at coitus. The tumours apparently have a cellular mode of transmission, and are seen most commonly in young, roaming, sexually active dogs (Batamuzi, Kassuku and Agger, 1992) . A viral aetiology has been investigated but not verified. Virus particle have been observed, but the tumour has not been transmitted by cell-free filtrates (Sapp and Adams, 1970; Cohen, 1985).

1.7.3.2 Metallic implants

Metal-related malignant tumours have been reported in both the human and the veterinary literature (Stevenson *et al.*, 1982; Hughes *et al.*, 1987; Stevenson, 1991). A case-control study has been conducted that assessed the possible role of metallic implants in the development of local, systemic and remote site cancer in the dog (Li *et al.*, 1993). Multivariable analyses showed that a significant protective effect on tumours other than bone and soft tissue tumour development was associated with internal fixation, and no significant positive association was found between metal materials and the development of bone and soft tissue tumours. However, incomplete and ambiguous information on type of fixation used, and the location and site of tumour development was recognised as a

criticism of the study. The authors concluded that further study was needed to investigate the local effect of metallic implants on the development of bone and soft tissue tumours.

1.7.3.3 Diet

Certain dietary factors have been assessed in studies of canine mammary cancer. No association was found between fat content of the diet and increased risk for the development of mammary cancer in the observational study by Sonnenschein *et al.* (1991). High dietary fat intake and table food was actually protective when cases and the non-cancer controls in the study were compared. A study of dietary protein content in conjunction with a low fat intake was associated with increased survival with mammary cancer (Shofer *et al.*, 1989). In a more recent case-control study conducted in Madrid, Spain, the multivariable analysis showed that older age, obesity at 1 year of age, and a high red meat intake were independently and significantly associated with the risk of canine mammary tumour and dysplasia development (Perez *et al.*, 1998).

1.7.3.4 Environmental causes

1.7.3.4.1 Tobacco smoke

Exposure to tobacco smoke in the home was found to be associated with an increased risk of neoplasia of the nasal cavity in long-nosed dogs, and a decreased risk in short and medium nosed breeds. In the long-nosed breeds, the greater the number of total cigarettes smoked in the household, the greater was the risk for nasal cancer (Reif, Bruns and Lower, 1998). The hypothesis suggesting that long-nosed dogs have a nasal filtration mechanism that may lead to increased deposition of particulate carcinogens at this site (Reif and Cohen, 1971) (Section 1.7.2.4) was reiterated by the authors of this more recent study.

An experimental study showed that direct smoking by dogs (via a tracheotomy) was significantly associated with the occurrence of bronchoalveolar tumours (Auerbach *et al.*, 1970), an unsurprising finding given the known carcinogenic potential of smoking in humans. A later case-control study of lung cancer in dogs, using records derived from two veterinary teaching hospitals, found only a weak association between exposure to tobacco

smoke in the home (indirect smoking) and canine lung cancer. However, skull shape was found to be a confounder in the study, with the increased risk being restricted to breeds with short and medium length noses. This finding suggested that the filtration mechanism in long-nosed breeds may exert a protective influence on the lung (Reif *et al.*, 1992).

1.7.3.4.2 Herbicides

The use of 2,4-dichlorophenoxyacetic acid (2,4-D) herbicides on lawns, or the use of a lawn care company to treat lawns, was found to be weakly associated with the occurrence of lymphoma in dogs living at those residences (Hayes *et al.*, 1991). This study was criticised because of poor measurement of exposure, methods of statistical analysis, inadequate information gathering to address potential confounders and other issues (Carlo *et al.*, 1992). However, it has been proven that dogs living in areas recently treated with 2,4-D absorb measurable amounts for several days following its application (Reynolds *et al.*, 1994). Further study of canine lymphoma and herbicide application at the dog's residence is needed to confirm or refute the presence of an association.

1.7.3.4.3 Electric and magnetic fields

There has been one study examining the association between residential exposure to magnetic fields and the risk of canine lymphoma, which used data derived from a veterinary teaching hospital. Dogs living in homes with very high current codes had the highest risk of developing lymphoma. Moderate, imprecise increases in risk were found for residence in a home with a sidewalk, backyard, or front yard magnetic field of 2.0 mG or greater, but not for indoor measurements (Reif, Lower and Ogilvie, 1995). Further studies are warranted to increase evidence for or against these findings.

1.7.3.4.4 Asbestos

It is well known that mesothelioma in humans is strongly associated with exposure to asbestos. This neoplasm occurs rarely in the dog. In the only case-control study examining its occurrence in dogs it was found that an asbestos-related occupation or hobby of a household member was significantly associated with mesothelioma (Glickman *et al.*,

1983). Lung tissue from three dogs with mesothelioma and one dog with squamous cell carcinoma of the lung had higher levels of chrysolite asbestos fibres than lung tissue from control dogs. The authors proposed that epidemiological studies of spontaneous tumours in dogs might provide insight into the role of environmental factors in human cancers.

1.7.3.4.5 Ultraviolet and ionizing radiation

Long-term exposure to the ionizing effects of sunlight was shown to result in solar dermatosis of abdominal, non-pigmented skin leading to documented increases in cutaneous haemangioma, haemangiosarcoma and squamous cell carcinoma in members of a Beagle dog colony (Nikula *et al.*, 1992). The development of squamous cell carcinoma of the nasal planum has also been correlated with ultraviolet exposure and lack of protective pigment (Hargis, 1981).

Experimental studies have shown that in dogs receiving irradiation for soft tissue sarcomas and in dogs exposed to intra-operative irradiation of the lumbar spine, the risk for developing osteosarcoma is increased (Taylor *et al.*, 1981; Gillette *et al.*, 1990). Many types of radiation-induced neoplasia in the dog have been reported (Rebar *et al.*, 1980; Moulton, Rosenblatt and Goldman, 1986; Taylor *et al.*, 2000).

1.7.3.4.6 Industrial activity

Some of the earliest reported observations of canine neoplasia described tonsillar squamous cell carcinoma as a tumour type which appeared to be associated with dogs living in urban areas (Withers, 1939; Cotchin, 1954; Cohen, Brodey and Chen, 1964; Ragland III and Gorham, 1967). It was hypothesised that the existence of air pollutants in an industrialized metropolitan area could be significant in the occurrence of this neoplasm. A study of proportional morbidity ratios of various types of cancer performed by Hayes *et al.* (1981) showed a significant positive correlation between spontaneous canine bladder cancer frequency and the degree of industrial activity in various parts of the United States.

1.7.3.4.7 Miscellaneous exposures

Dogs exposed to topical insecticides (flea and tick dips) were shown in one study to have an increased risk for bladder cancer (Glickman *et al.*, 1989) with the risk being enhanced by obesity. Obesity was likely to increase risk because many of the insecticides under study were highly lipophilic, however no one chemical type of flea and tick dip accounted for the increased risk. The need to measure the concentration of 'inert' substances found in the insecticidal preparations in fatty tissue of study subjects was commented upon, particularly as many of those substances are known carcinogens.

Toxicity and epidemiological studies have shown that ovarian steroid hormones and their synthetic derivatives can enhance mammary cancer development in dogs (Kwapien *et al.*, 1980; Rutteman and Misdorp, 1993).

1.8 THE DOG AS A SENTINEL FOR HUMAN NEOPLASIA

The role of the dog as a sentinel for human exposure to environmental contaminants has been proposed by many researchers, because it is a species that shares its environment closely with man, is likely to live a natural lifespan (unlike traditional laboratory animals) and can be of a suitable size for assessing levels and effects of toxicants. The dog's potential as a sentinel to assess risks to humans of neoplasia secondary to toxic environmental exposure has been of particular interest because they develop certain tumours with biological behaviour and histological characteristics which are similar to their counterparts in man, and have shorter latency periods for tumour development than longer-lived humans. Observations on dogs are also less confounded by variations in behaviour, particularly the major confounders of human cancer study, being smoking, alcohol consumption and occupation-related factors. Studies which have been performed with specific reference to the dog as a sentinel include those by Hayes *et al.* (1981) (Section 1.7.3.4.6), Glickman *et al.* (1983) (Section 1.7.3.4.4), Glickman *et al.* (1989) (Section 1.7.3.4.7), Reif *et al.* (1992) (Section 1.7.3.4.1) and Reif *et al.* (1995) (Section 1.7.3.4.3). Few observational studies have been published in recent years, though this topic and related studies have been reviewed many times in the literature (Buck, 1979; Garbe, 1988; O'Brien, Kaneene and Poppenga, 1993; Bukowski and Wartenberg, 1997). However

there appears to be little funding available for studies to evaluate environmental exposures and cancer incidence in dogs, in comparison to work on molecular and genetic aspects of cancer research (Vastag, 1999).

1.9 THE DOG AS AN ANIMAL MODEL FOR HUMAN NEOPLASIA

The characteristics of the dog that have led many to recommend it as a sentinel species for human disease are also those that have promoted its use as an animal model for human neoplasia. Specifically, dogs develop spontaneous neoplasms some of which have histological characteristics and biological behaviour similar to their human counterparts, typically live a near-natural lifespan, share the environment with their owners, and have a shorter latency period for tumour development so may be studied over a shorter period of time. The dog is also of a suitable size for studies involving drugs and/or irradiation, or surgical research (Gardner, 1996).

The dog has been proposed as a model for investigation of the aetiology of appendicular osteosarcoma because the neoplasm's biological behaviour, epidemiology and histopathology is similar to that in humans. The role of p53 in tumour formation is the subject of extensive research (Section 1.4.2). One investigation of p53 mutations in canine osteosarcomas showed that the locations and types of mutations found in the canine tumours were nearly identical to those reported in human cancer (Johnson, Couto and Weghorst, 1998). Other studies have emphasized the role played by alterations of p53 in canine osteosarcoma, as is well described in the human equivalent (Sagartz *et al.*, 1996; van Leeuwen *et al.*, 1997; Mendoza *et al.*, 1998).

The role of p53 in canine mammary neoplasia has been explored (van Leeuwen *et al.*, 1996; Veldhoen *et al.*, 1999; Schafer *et al.*, 1998), with researchers making comparisons between the role of p53 in the development of canine and human mammary neoplasia. Canine mammary tumours also contain hormone receptors, like human mammary tumours. This feature has been led to the suggestion that canine mammary tumours may be suitable models for hormone-dependent human breast carcinoma (Martin *et al.*, 1984; Mol *et al.*, 1996).

Lymphoma in dogs shows biological, pathological, and clinical similarities to non-Hodgkin's lymphoma in humans (Teske, 1994). Structural similarities have been shown between certain canine genes and their human equivalent, which in humans, have been shown to be highly significant in the development of non-Hodgkin's lymphoma (Chaganti, Mitra and LoBue, 1992).

The dog is the only species apart from man to develop spontaneous prostatic carcinoma with significant frequency. High grade prostatic intraepithelial neoplasia (PIN) is recognised as the most likely precursor to human prostate cancer. High grade PIN is also recognised in dogs (Waters and Bostwick, 1997a; Waters *et al.*, 1997b; Aquilina *et al.*, 1998; Waters, 1999). The authors of these studies suggest that canine prostatic cancer may be a useful animal model for studying prostatic carcinogenesis and progression.

Perhaps of greater importance than its role as a model for investigation of tumour aetiology, is the role of comparable canine neoplasms as models for the evaluation of new forms of cancer therapy. To many dog owners their pet is an important family member, worthy of receiving treatment for disease regardless of cost. However, the costs of developing cancer treatments for sole use of the veterinary profession is prohibitive. The costs may be offset by developing clinical trials in the dog which may be used directly for the benefit of the animal under treatment, although their primary purpose is to evaluate therapeutic regimes of interest to medical oncologists. With this approach, the requirements of patients and researchers in both veterinary and medical disciplines are fulfilled. The PCOP (Section 1.6.3.1.3) is one program that was specifically created to determine which forms of pet animal cancer could serve as suitable models of human cancer. As part of this program, owners of dogs with cancers under study in the program are asked to consent to their dogs participating in clinical trials designed to evaluate novel therapeutic approaches. Potential benefits of participation include: helping their pet; adding to the understanding of the disease that will benefit other dogs; and potentially contributing to the development of better therapies for humans with cancer (Knapp and Waters, 1997).

Canine lymphoma is commonly treated with combination chemotherapy, and as in human patients, dogs with high grade tumours tend to respond better to chemotherapy (MacEwen,

1990). Numerous therapeutic trials in dogs have been reported (Teske, 1994; Zemann *et al.*, 1998; Ogilvie *et al.*, 2000).

Researchers in the PCOP recently evaluated the use of a cisplatin/piroxicam treatment regime in dogs with transitional cell carcinoma of the bladder. Invasive canine bladder cancer is similar to its human counterpart in histopathological characteristics, biological behaviour and response to cisplatin and carboplatin therapy. Results of the trial were that the combination therapy induced remission more frequently than cisplatin alone in a canine model of human invasive TCC, however strategies to reduce renal toxicity would need to be developed prior to evaluation of cisplatin/piroxicam in humans or general use of the treatment in pet dogs (Knapp *et al.*, 2000).

Canine sun-induced squamous cell carcinoma (SCC) has been cited as representing a useful animal model to evaluate new therapeutic modalities for possible human applications. An investigation based upon this tumour type concluded that sustained-release chemotherapy using intralesional 5-FU/epi gel and CDDP/epi gel therapeutic implants was effective in treating canine sun-induced SCC of the skin (Kitchell *et al.*, 1995).

Not all studies of canine models for potential cancer therapies have targeted a specific tumour type. Cesano *et al.* (1996) evaluated the possible toxicity and efficacy of a type of cell therapy, being the systemic administration of lethally irradiated TALL-104 cells in the absence of exogenous interleukin 2, in 19 dogs, all with advanced, refractory malignancies of various histological types. The conclusion of this study was that the treatment may be regarded as a safe and promising adjuvant type of treatment for advanced cancer patients.

There are numerous reviews discussing the role of the dog as an animal model (Pierrepoint, 1985; MacEwen, 1990; Hahn *et al.*, 1994; Gardner, 1996). As well as the tumour types discussed in this section, canine tumours highlighted as being potential models for human neoplasia by reviewers have included oral melanoma, nasal tumours, soft tissue sarcomas and lung tumours.

1.10 GEOGRAPHICAL DISTRIBUTION OF CANINE NEOPLASIA

The major population-based studies of canine neoplasia, the CANR and Tulsa Registry (Section 1.6.3.1) were based upon the animal populations within defined geographical regions, so that a population-at-risk for disease occurrence could be determined. The primary aim of the researchers was to determine the incidence of pet animal neoplasia (canine and feline), types of neoplasia and to investigate host-related risk factors for particular cancer types. Another priority of the team was to conduct comparative studies of human and pet animal neoplasia incidence in the region, such as those performed by Schneider *et al.* (1968) and Schneider (1970). This was possible because a human cancer registry, the Alameda County Cancer Registry, was operational in the same area. Although these studies were carried out within defined geographical regions, detailed spatial and temporal analyses of canine cancer were not performed. A prime reason for this would have been the lack of powerful computer hardware and modern statistical software, which are required for such investigations.

Knowledge of the spatial distribution of diseases provides useful information in aetiological research and in the implementation of preventive activities in community health. The search for "clusters", where disease occurrence is aggregated in space and/or time, has provoked widespread discussion and debate among medical epidemiologists, particularly with reference to cancer (Rothman, 1990; Heederik, 1994; Elliott, Martuzzi and Shaddick, 1995; Aldrich and Sinks, 2002). Mainstream media reportage of so-called "cancer cluster alarms", in geographical areas where there is a perceived environmental threat to human health, have been paramount to the debate (Olsen, Martuzzi and Elliott, 1996). Public perception of increased cancer occurrence associated with living in the vicinity of nuclear power establishments has led to numerous epidemiological studies in locations around the world. Studies of childhood leukaemia occurrence near nuclear sites have recently been reviewed by Laurier and Bard (1999); other examples include the investigation of brain cancer in Los Alamos, Mexico (Kulldorff *et al.*, 1998). There have also been many cancer investigations in areas of Europe known to be contaminated by the 1986 Chernobyl nuclear reactor accident (Michaelis *et al.*, 1996; Tondel *et al.*, 1996; Ivanov *et al.*, 1997).

Methods for the analysis of the spatial aggregation of health events have received growing attention under the pressure of public opinion concern and as tools for the identification of potential risk sources, for monitoring relevant geographical areas and for public health decisions. There are numerous reviews of spatial data analytical techniques available (Waller and Jacquez, 1995; Gatrell and Bailey, 1996; Jacquez, 1996b; McKenzie, 1999; Moore and Carpenter, 1999; Ward and Carpenter, 2000b; Carpenter, 2001).

The study of spatial distribution of disease requires spatially referenced data. It has become standard practice for postcode (zipcode) data to be collected by health authorities, because the postcode may provide the spatial reference data necessary for spatial analyses and infectious disease tracing and surveillance. In the UK, the Central Postcode Directory links postcodes to an Ordnance Survey grid reference and an electoral ward (in England and Wales) or a postcode sector (in Scotland). Geocoding is the process by which postcode data are converted to grid reference or x, y-coordinate data. The development of geographic information systems (GIS) over the last 20 years has provided a powerful and rapid ability to examine spatial patterns and processes. A GIS is an integrated set of computer hardware and software tools to capture, store, edit, organize, analyse and display spatially referenced data (Bailey and Gatrell, 1995). Spatial analyses of human neoplasia have become commonplace in the last decade due to the computerisation of spatial data through the use of GISs, and the development of spatial data analytical techniques.

Although there has been continued interest expressed in the role of the dog as a potential sentinel for human disease, including neoplasia (Section 1.8), spatial and temporal analyses of canine neoplasia occurrence are still extremely rare in the literature. Reasons for this are likely to include the difficulty of determining an accurate population at risk for a particular region without time-consuming and expensive animal census studies, and the lack of suitable spatial datasets. To date, there have been two publications from an exploratory study designed to compare the distributions of some selected, biologically similar cancers in dogs and humans living in the same geographical region in Michigan, USA during the same time period (O'Brien *et al.*, 1999; O'Brien *et al.*, 2000). Canine cancer data was sourced from the VMDB (Section 1.6.3.2.1). GIS and point-pattern analysis were used in the initial study, which found that significant spatial clustering occurred by county and type for four selected canine cancers. No definitive temporal patterns could be

demonstrated for the cancer cases under study (O'Brien *et al.*, 1999). The authors suggested that processes determining the aggregation of canine cancer cases do not act in a spatially uniform manner.

The second study performed by O'Brien *et al.* (2000) used the same canine data, plus human cancer case records obtained from separate sources depending on the Michigan county of residence. Human and canine cancer cases were mapped using a GIS and *k*-function spatial analysis and nearest-neighbour temporal analyses were performed on the residence addresses and dates of diagnosis/discharge of the subjects. A stratified random sample of human incident cases of comparable cancers diagnosed during the same time period from the same counties as the canine cancer cases was used because of insufficient monetary sources to permit geocoding of all the available human records. The results of this unique study suggested that processes determining spatial aggregation of cases in dogs and humans were not independent of each other, did not act uniformly over different geographical areas, operated at spatial scales <2000 m regardless of species and tended to act upon dogs more strongly at shorter distances than on humans. No definitive interspecies concurrence of temporal clustering was found (O'Brien *et al.*, 2000).

1.11 THESIS OUTLINE

The aim of this thesis was to interrogate canine histopathology data compiled by two diagnostic histopathology services for quantitative risk factor analyses of the occurrence of neoplasia in their biopsy populations. Data compiled by both histopathology services originated from first-opinion veterinary practices. The second chapter of this thesis gives a detailed description of the database operated by the Canine Infectious Diseases Research Unit (CIDRU) in the Department of Pathology at the University of Glasgow Veterinary School. This histopathology service provided the data used for studies presented in Chapters 3, 4 and 6 of this thesis. The preparation of this data source is described in Chapter 3, highlighting the importance of data quality assurance to subsequent epidemiological analyses. Chapter 4 presents the application of logistic regression methodology to the dataset defined in Chapter 3. The data recording system of a commercial diagnostic histopathology service is presented in Chapter 5, followed by interrogation and statistical analysis of data pertaining to canine biopsies submitted to the

service. Chapter 6 outlines the application of a spatial and space-time scan statistic designed to detect disease clustering to the CIDRU dataset. The results are graphically displayed using a Geographical Information System, with discussion focusing on potential causes of different geographical patterns of canine neoplastic biopsy occurrence. Finally, general conclusions pertaining to the use of histopathology record databases for epidemiological study of canine neoplasia are detailed in Chapter 7. These studies represent the first large-scale epidemiological analyses of canine biopsy data from diagnostic histopathology databases in the UK.

CHAPTER 2

THE UNIVERSITY OF GLASGOW EXTERNAL VETERINARY HISTOPATHOLOGY SERVICE

2.1 INTRODUCTION

The use of large computerised medical databases for epidemiological research on human neoplasia has been recognised for many years. Cancer registries were developed specifically for the purpose of cancer research and represent one of the most well-known types of medical databases in existence today (Muir, Demaret and Boyle, 1985; Joslin, 1990; Bell *et al.*, 1995). In recent years there have been significant technological advances which now allow linkage of other sources of human cancer data, such as pathology and hospital medical record databases, directly to cancer registries via electronic networking of information (Moss, Smith and Nicholas, 1997).

The number of computerised veterinary databases has been increasing steadily in the past few decades, particularly due to the computerisation of veterinary practice records. Other sources of computerised data exist in veterinary diagnostic laboratories and insurance companies. Although the software for many of these sources has not been designed with epidemiological studies in mind, their suitability for this purpose has been investigated (Pollari *et al.*, 1996a; Pollari, Bonnett and Bamsey, 1996b; Bonnett *et al.*, 1997; Egenvall *et al.*, 1998). Epidemiological studies of canine neoplasia have been performed in some cases (Reid-Smith *et al.*, 2000; Wood *et al.*, 2000a; Dobson *et al.*, 2002).

This thesis is based upon data from two computerised sources of veterinary histopathology data. One source is maintained by an academic institution, the University of Glasgow; the other is run by a commercial veterinary diagnostic laboratory. All data originates from first opinion or charity veterinary practices. This chapter describes the database utilised by the Department of Veterinary Pathology at the University of Glasgow Veterinary School, as well as its data collection, storage and retrieval systems. The daily operation of the histopathology service is also detailed.

2.2 THE UNIVERSITY OF GLASGOW HISTOPATHOLOGY SERVICE

2.2.1 Introduction

Glasgow Veterinary College was founded in 1862, and incorporated into the University of Glasgow in 1949. Previously part of the Faculty of Medicine, Veterinary Medicine became an independent Faculty in 1968. The University of Glasgow Veterinary School (GUVS) operates within the Faculty of Veterinary Medicine as a teaching and research institute, and incorporates a state-of-the-art veterinary referral hospital. Animals are referred from first-opinion veterinarians located throughout Scotland and northern England, as well as occasionally further afield.

The Department of Veterinary Pathology within the Faculty of Veterinary Medicine has an international reputation for excellence in research and is highly regarded in the veterinary profession for its teaching curriculum. As well as providing teaching and research programs, the department is essential to the framework of the referral service, offering a diagnostic histopathology and necropsy service, for animals seen by clinicians at GUVS. Pathology services are also available from the department for first-opinion veterinary practices, animal welfare agencies and as an alternative opinion source for other diagnostic veterinary laboratories.

2.2.2 Early recording systems

Since its inception, the department has run a diagnostic histopathology and necropsy service for veterinarians in first-opinion practice. From the 1950s until 1978, the service operated on an unofficial basis, becoming known among practitioners only by word of mouth. There were no formal submission guidelines, so information pertaining to tissue samples, such as species and site of origin, age, gender and breed of animal, and biological behaviour of the submitted tissue, was supplied at the discretion of the practitioner.

A record number system was in place within the referral hospital at GUVS where each animal referred to the hospital was allocated a unique identification number upon its arrival, and retained this number for all investigative procedures and for subsequent visits,

if required. The Department of Veterinary Pathology identified all samples originating from outside the referral hospital by generating one hospital number, and giving this number a unique extension for each sample. All records were paper-based, with each pathologist maintaining their own records system. The only centralised recording system prior to 1978 was in the form of a "day-book", in which abstracts of each biopsy report were recorded. The diagnosis and unique identification number of each report were recorded under headings corresponding to the site of tissue origin, using a coding system devised for this purpose. The day-book allowed rapid identification of all diagnoses and record numbers for tissues originating from a particular site.

2.2.3 Computerisation

The discovery of parvovirus by pathologists in the Department of Veterinary Pathology in 1978 brought the department to the international attention of the veterinary profession. Due to the virulence of this disease and the recognised expertise at GUVS, the number of samples submitted by practitioners increased dramatically. This placed great pressure on the department's paper-based recording systems and provided the incentive to develop a revolutionary, computerised recording system. Due to the foresight of the pathologists working at GUVS, a relational database structure was chosen at a time when this technology was in its infancy. The pathologists recognised the advantages and flexibility of such a system which would support efficient data storage, report production and allow rapid search and retrieval of information. The decision was also made to create a referral unit to handle samples received from external practitioners, and the Canine Infectious Diseases Research Unit (CIDRU) was established in 1978.

The CIDRU database system was designed to run in conjunction with the existing paper record system, rather than replace it. With the outbreak of parvovirus, the most common test requested by practitioners was parvovirus serology, and the numbers of submitted blood samples far exceeded the numbers of tissue biopsies received. Consequently, the first records to be entered into the CIDRU database were those generated by serology requests. By 1986, all histopathology and necropsy requests were also entered into the database, and the computerised system continues today. The submission form remains an essential component of the reporting system, and all paper records from 1986 to the present day are

stored in numerical order in filing cabinets in the department, once the reporting process has been completed.

With computerisation, quick and efficient search and retrieval of records from the relational database, based on sample criteria such as site of origin, was possible. Thus, the "day-book" was made redundant in 1986. Paper-based logbooks are still in use to catalogue daily submissions at sample reception (Section 2.4.1), and in the histopathology laboratory (Section 2.4.2).

2.2.4 The CIDRU pathology record system

Designers of database systems are concerned with issues such as the efficient use of computer memory, the design of filing systems that permit rapid retrieval of data from massive stores, and methods for ensuring the safety and security of the stored data (Coiera, 1997). The initial file structure of the CIDRU database was developed with regard to the Unit's immediate requirement of handling many samples (and therefore data) associated with parvovirus diagnosis. Of major importance were files for veterinary practice and animal identification, and for sample identification for serology, biopsy, necropsy, and isolation procedures.

2.2.5 File Structure

The file structure was designed by a software development team in direct consultation with the pathologists in charge of the CIDRU diagnostic service. Initially, the pathologists identified general categories to be used in the pathology database, with each relating to both an "Animal" and "Vet" file. The general categories, or "lower" files, were identified as "serology", "isolation", "biopsy" and "necropsy", containing information on the results of samples received for examination in each of the four areas (Figure 2.1). The core file structure was unchanged during the time period of sample collection for this thesis, although some alterations to the software program were made to simplify certain areas of data entry. Major restructuring of the database, including its integration into the internal GUVS referral hospital database, is currently underway.

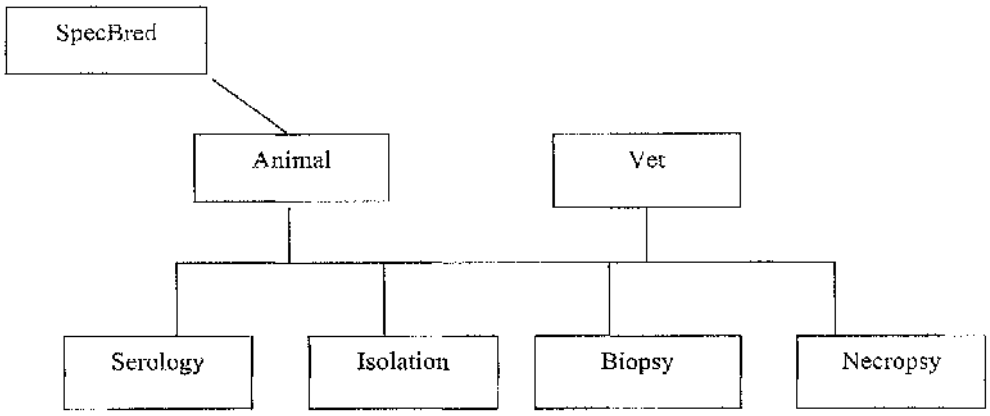


Figure 2-1 Representation of the layout of files which comprise the CIDRU pathology database system. The Animal and Vet files link to the lower files by a unique reference number generated in the lower files e.g. "*histref*" in the Biopsy file.

The "Vet" file holds details of all veterinary practices using the CIDRU diagnostic service, being the practice name, full postal address (including postcode), telephone and facsimile numbers, and account status. A recent addition to this file is a field to record practice e-mail addresses. Each practice is allocated a unique identifying number (*vetref*) at the time of its first sample submission. Only this number is recorded in the lower files as a pointer to the "Vet" file, since, obviously, one practice may submit many samples to the service over a period of time. This prevents the recording of the same information many times over and is an efficient use of disk space.

The "Animal" file holds a number of fields containing information on animal identification – name, species, breed, date of birth, gender, owner identification, and vaccination history. A record and unique identification number (*animalrec*) for each animal is generated by data entry into the lower files. To ensure accurate coding of species and breed data, a separate file "SpecBred", was created and linked to the "Animal" file. In the "SpecBred" file, each species is given a numeric code and then subdivided into general and more detailed breed categories which are also numerically coded: thus, a Labrador Retriever dog is classed as 04 (DOG), 74 (RETRIEVER), (05) LABRADOR. The numeric codes are easily found as the file is set up to allow searching in designated key fields on the data entry screen by the names of the species and breeds, as well as by the numeric codes themselves. Examples of the coding system used in this file are given in Table 2-1.

Each of the lower files holds a number of fields containing information on sample identification. The lower files are linked to the "Vet" file by the *vetref* field, and the "Animal" file by the *animalrec* field. The "biopsy" file was the lower file interrogated for the purpose of this thesis; the following description refers to this file.

Breed1 description	**Breed2 description**	**Breed1 code**	**Breed2 code**
Doberman		26	0
Boxer		12	0
German Shepherd		32	0
German Shepherd	X	32	99
Retriever	Labrador	74	5
Retriever	Golden	74	4
Spaniel	English Springer	87	7
Spaniel	Cocker		6
Terrier	Jack Russell	92	16
Terrier	Border	92	5
Terrier	X	92	99
X*		99	0

Table 2-1 Examples of common canine breed codes and the crossbreed code which are used for data storage in the "SpecBred" file of the Canine Infectious Diseases Research Unit pathology database. There are 207 different breeds which have been individually coded.

When a biopsy sample arrives at CIDRU, it is given a unique reference number, *histref*. The *histref* number forms the main identifying factor in the "biopsy" file, and under each *histref* number the following details are recorded: date sample sent; date sample received; date sample results reported; duty pathologist; *vetref*; *animalrec*; animal name, date of birth and gender; whether the animal is alive or not; requesting veterinarian and whether a previous biopsy from the animal has been submitted. Also recorded are the clinical details as supplied on the submission form accompanying the sample, being: clinical diagnosis; lesion appearance, site, size; how long the lesion has been present, and whether other tests (e.g. bacteriology, mycology) have been requested on samples accompanying the tissue biopsy(ies). The "biopsy" demographic data entry screen (Figure 2-2) contains some fields

with drop-down menus to allow searching for information contained in linked files, and to ensure standard recording of data. Specifically, the correct *vetref* may be found by searching on veterinary practice name or town, and an animal's file may be found by searching on animal or owner name. If there have been no previous submissions originating from the animal, the entry of the animal's name and owner surname automatically creates a new record with a unique identification number (*animalrec*) in the "Animal" file. Drop-down menus are available on the species and breed data entry fields, to ensure accurate data entry of this information from the "SpecBred" file. Other fields have drop-down menus that are hard-coded into the data entry screen, again to ensure consistency of data entry. Fields with such menus are duty pathologist identification and sex. The drop-down menus were created because the need for easier and more efficient data entry in certain fields was recognised as a way to improve data accuracy during an early data quality assessment.

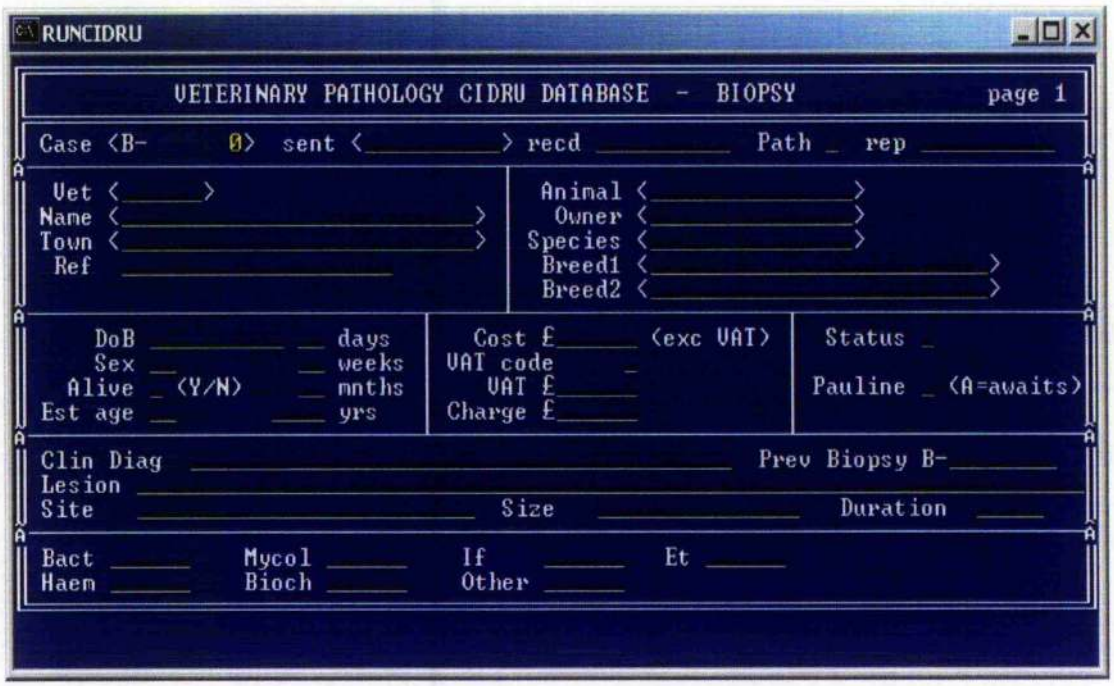


Figure 2-2 Demographic data entry screen for the "Biopsy" file in the Canine Infectious Diseases Research Unit pathology database.

Other details recorded under each *histref* are: slide number; special stains and corresponding results; histopathological diagnosis (two fields available); site and class codes for each diagnosis; report, and prognosis. These data are manually entered by the

duty pathologist on a second data entry screen (Figure 2-3). The coding systems for site and class of diagnosis are used to prevent unnecessary repetition within the database, and to facilitate record searches. The location of the lesion is recorded using a numerical code in the site field. This number refers to one of 50 possible sites, as specified in Table 2-2. There are four possible classes of diagnosis, each also designated by a numerical code (Table 2-3). The coding system allows searching of the database for biopsy cases from particular sites and of a particular type, for example, "site=29" and "class=1" identifies all cases originating from bone that are neoplastic.

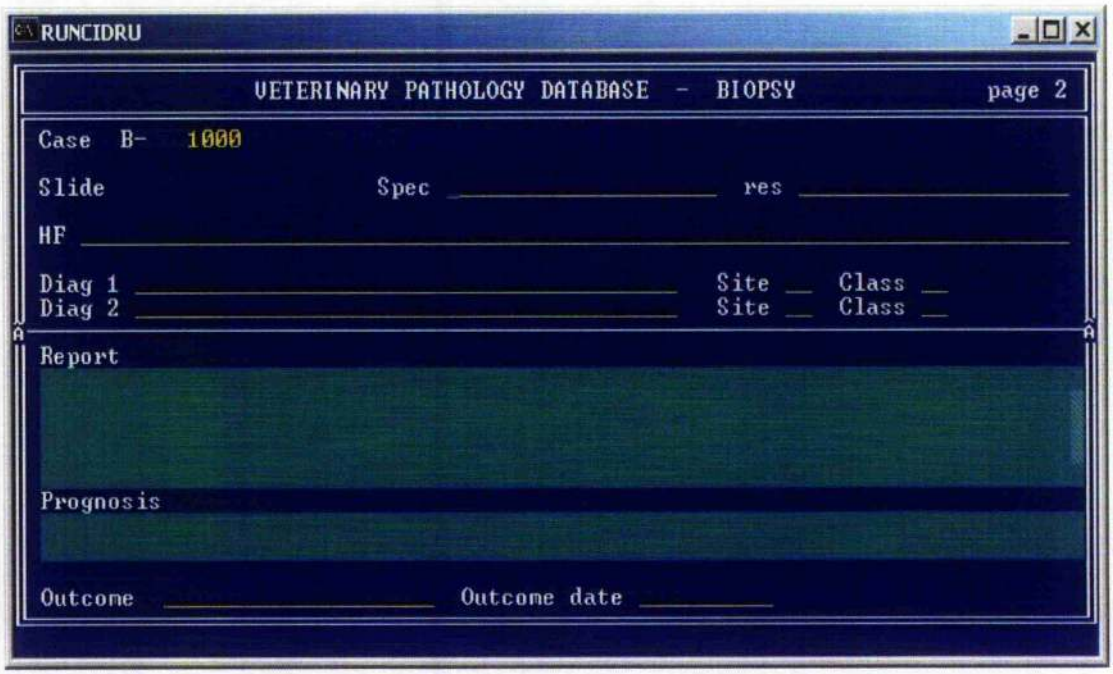


Figure 2-3 Data entry screen for details accompanying biopsy submission and histopathology report generation for the CIDRU histopathology service.

2.2.6 Software

The CIDRU database system has been developed using DataFlex, a commercially available, multi-user, relational database system. At the time of its introduction, DataFlex could support up to 250 datafiles each with up to 9 indexes and each containing up to 255 different data elements per record. The system was installed and customised for the Canine Infectious Diseases Research Unit's purposes by an independent company[1]. The multi-user

[1] Specialist Services (Scotland) Limited

facility of the database could support simultaneous access of the database by up to 15 people at a time. Advances in software, hardware and networking technologies now allow simultaneous access by up to 250 people. Permission to access the CIDRU DataFlex database is granted by Faculty information technology staff, following approval of the applicant by the Head of the Department of Veterinary Pathology.

Site code	Site description	Site codes	Site description
0	Multiple System	25	Adrenal
1	Mouth	26	End. Pancreas
2	Pharynx	27	Placenta
3	Oesophagus	28	Muscle
4	Forestomachs	29	Bones
5	Stomach	30	Joints
6	Small Intestine	31	Tendons
7	Large Intestine	32	Reproductive Male
8	Anus	33	Reproductive Female
9	Liver	34	Mammary Glands
10	Exocrine Pancreas	35	Prostate
11	Salivary Glands	36	Kidney
12	Tonsil	37	Lower Urinary
13	Upper Respiratory	38	Blood & Bone Marrow
14	Lower Respiratory	39	Thymus
15	Larynx	40	Spleen
16	Brain	41	Lymphatic
17	Spinal Cord	42	Cardiovascular
18	Peripheral Nerves	43	Skin
19	Autonomic Ns	44	Connective Tissues
20	Eye	45	Serosae
21	Ear	46	Septicaemia-Toxaemia
22	Pituitary	47	Feet
23	Thyroid	48	Foetus
24	Parathyroid	49	Other

Table 2-2 Codes used to identify site of diagnosis used by pathologists in the Canine Infectious Diseases Research Unit database.

CIDRU currently uses a version of DataFlex that runs under DOS. The program can be viewed either in a DOS window (shell) or full screen, under the operating systems in use on GUVS personal computers (Microsoft Windows 95, 98, NT and 2000 (Microsoft Corporation)). Given the rapid increase in the number of Windows-based software

programs and their widespread end-user acceptance, Visual DataFlex software has been used to create a Windows-based database environment for end-users. The implementation of this version of the database by the personnel of CIDRU is imminent.

Class code	Class description
1	Neoplasms
2	Specific Disease
3	Congenital
4	Others

Table 2-3 Codes used to identify class of diagnosis used by pathologists in the Canine Infectious Diseases Unit database.

Within DataFlex resides a Query language which at the time the CIDRU database system was created, was considered to be one of the most advanced available. Any query on a lower file automatically makes available all data held in related files. For example, a pathologist looking at data held in the "biopsy" file is able to select biopsy cases of a particular species and/or breed ("SpecBred"/"Animal"), referred from a particular veterinarian ("Vet") and submitted within a given period of time ("Biopsy"). The query could be further extended to extract biopsy records with a diagnosis of a particular type ("Biopsy"), or diagnoses of a particular class ("Biopsy"). The output from any Query may be directed to a screen, printer, or a file which can be subsequently accessed by a personal computer.

The Query language described above was used to extract the data used in this thesis. The above Query language has recently been replaced by the Crystal 7 query language, which is part of an internationally used database reporting system in a high quality Windows environment.

2.2.7 Hardware

Since the introduction of the relational database system, there have been significant developments in information technology and components of the system, particularly the hardware.

Initially, the CIDRU computer database was held on one of the networked "clusters" at the University of Glasgow Veterinary School. There were six clusters in total – CIDRU, feline virus unit, routine pathology, experimental pathology, biochemistry and medicine. Each cluster comprised a microcomputer supporting several visual display units (VDUs) and printers. The clusters were linked by cable to form a local area network (LAN), allowing information held on different clusters to be accessed from the user's own VDU.

Rapid advances in computer technology have been reflected in the upgrades made to the hardware of the computer system since its establishment. The clusters as previously described have been superceded by the use of a dedicated file server (Novell NetWare 4.11), maintained on GUVS premises. This server holds all data files that were previously held on the clusters' microcomputers. Networking of staff desk-top computers has resulted in all users of the CIDRU database having immediate access to the database, and the hard drive size of the server currently comfortably exceeds levels essential for smooth database operation and data storage.

2.2.8 Back-up

Back-up facilities were initially provided by a tape deck system, with six tapes. This allowed new information stored on the database to be saved at the end of each day.

Back-up of the server is now performed with an Enterprise Back-up Unit, which consists of a tape autochanger, tape library system and Enterprise back-up software. An incremental data back-up is performed 4 nights per week, with a full back-up performed one night per week. The incremental tapes are retained for one week and the weekly tapes are kept for one month on the premises. The last full back-up tape of each month is retained for 3 months and is stored off-site. In the event of server failure, only data from that day would

need to be re-entered. If a catastrophic event did occur, the maximum data loss would be five days, as long as the weekly tapes on the premises remained intact.

2.3 OPERATION OF THE HISTOPATHOLOGY SERVICE

From 1978 to 1990, the diagnostic histopathology service in CIDRU was directed by two pathologists, in addition to their teaching and research commitments. In 1990, another pathologist became involved in the external histopathology service, with a particular interest in dermatology cases. Currently, there are five pathologists responsible for the service, operating on a rota system, which allows them to fulfil their other commitments to teaching and research.

There are important factors affecting the accuracy of each biopsy report generated from samples received by the CIDRU histopathology service, involving many personnel and procedures. The decision by the veterinarian to biopsy a lesion can only be reached after thorough consultation with the animal's owner, who will be relying upon the veterinarian to correctly evaluate the patient. The veterinarian must decide upon the correct biopsy technique (Chapter 1, Section 1.5.2), and ensure that the biopsy is representative of the lesion, is handled carefully at the time of procurement and is correctly fixed, if rapid delivery to the laboratory is impossible. The sample container must be correctly labelled and packaged, and the sample accompanied by an accurate, concise and complete description of the clinical aspects of the case, including full signalment (species and breed, gender, neuter status, age) of the patient.

With the creation of CIDRU in 1978, a specific request form was developed to encourage practitioners to provide pertinent information about their submissions (Appendix 1). A comprehensive form was developed in 1998 (Appendix 2), which is in use today.

At the laboratory, marking each sample and form with the same identifying number is paramount to efficient data flow, as is accurate data entry of supplied information by the administration staff. Other important factors at the laboratory include expert processing of the tissue by the histopathology staff, and timely delivery of prepared slides to the veterinary pathologist. The pathologist must then carefully assess the tissue sections and

provide a precisely worded report which considers the clinical picture. Finally, the report needs to be quickly dispatched to the veterinary practice.

The passage of samples through the CIDRU histopathology laboratory, and generation of data within the electronic pathology record culminating in report generation and delivery to the veterinary practice, are outlined in the following sections.

2.3.1 Sample reception

The majority of tissue samples are posted to CIDRU; some are hand-delivered by local veterinary practitioners. All samples are handled by trained personnel on arrival at the Unit, where contents of each delivery are checked to ensure that each specimen container has an accompanying submission form, and that the details on the container match those on the form. Both items are given a unique reference number (*histref*), and this reference number, the name of the submitting veterinary practice and the date of the sample's arrival are recorded in a logbook. The current submission form (Appendix 2) is photocopied, and the photocopy delivered to administration. When the form was one of those in triplicate format (Appendix 1), one copy was removed and delivered to the CIDRU administration staff.

2.3.1.1 Data entry

The CIDRU administration staff initiate the computerised pathology record for each submission on receipt of the copies of the original forms. As described in 2.2.5, details are transcribed from the submission form to the first data entry screen (Figure 2.2).

2.3.2 Sample preparation procedure

The samples and the original copy of the submission form are taken to the histopathology laboratory's preparation room, where the samples are inspected by histopathology technicians to ensure that there is adequate tissue for processing and that the submitted tissue is properly fixed (Chapter 1, Section 1.5.2). The *histref* and the number of tissues received with each request form are recorded in paper-based laboratory record logbook.

The most appropriate portion of tissue for processing is selected. Some tissue types require additional treatments prior to routine processing - an example is bone, which must first be softened by decalcification. Each portion of tissue is trimmed to fit into processing cassettes and processed into paraffin blocks overnight. A 2-part laboratory identification number is allocated to each tissue, the first part being the year, and the second corresponding to the sequential number of the tissue of those processed by the laboratory that year. For example, the 2310th tissue processed by the laboratory in 2001 would be given the laboratory number 01/2310. There may be multiple laboratory numbers associated with one *histref*, and these laboratory numbers are also recorded in the logbook.

The blocks are sectioned on a microtome, the sections mounted on slides and stained. The prepared slides are labelled with the *histref* and laboratory number and delivered, with the original submission form, to the pathologist within 36 hours of receipt of the sample, unless prior procedures have been required. Slides are routinely stained with haematoxylin and eosin. However, they may be returned to the laboratory technicians for special staining if required by the pathologist following his/her preliminary examination of slides.

2.3.3 Microscopic examination

Following examination of the prepared slides, the duty pathologist reviews the clinical details given on the submission form and if possible, writes a full report describing the microscopic findings (Chapter 1, Section 1.5.2), making a diagnosis and providing prognostic information.

If further procedures are deemed necessary, such as special staining or immunohistochemistry (Chapter 1, Section 1.5.4), a provisional report may be written and sent to the veterinary practice with a final report pending the results of further procedures.

2.3.4 Report generation

The report is entered directly into the CIDRU database. The accuracy of the data entry on the first screen (Figure 2-2), containing the demographic information entered by the administration staff, is checked against the information on the original submission form by

the pathologist. The second screen is used for report generation (Figure 2-3). Fields are available for the pathologist to enter two diagnoses with corresponding codes for diagnosis site and class (Tables 2-2 and 2-3). Free text fields are available to enter the report and information helpful for prognosis.

2.3.5 Report delivery to practitioner

Once the pathologist has completed the assessment of the slides and entered the report, the original submission forms are given to the administration staff. The original forms are matched with the administration copy, and a hard copy of the report is printed for delivery to the veterinary practice. The report is typically faxed to the practice on the day the samples are reported, and a copy placed in the post. In most cases, a report is with the practice within 48 to 72 hours of the sample's arrival at the Unit.

2.4 DATA EXTRACTION FROM THE CIDRU DATABASE

The most simple and efficient data extraction from the CIDRU computer system is achieved through the built-in query language, described in Section 2.2.6. Through this system, all the files containing details of samples processed by the Unit can be interrogated. Queries can be run on any of the files, with the level of available information differing depending on which file is selected. Most commonly undertaken are the interrogations of the lower files, from which information within the "Vet", "Animal" and "SpecBred" files may also be assessed. However, two lower files may not be interrogated simultaneously, i.e. serology results, maintained in the "serology" file, may not be extracted at the same time as biopsy results, maintained in the "biopsy" file. Means of conducting such specialised interrogations are possible, but demand the use of customised programs which can be compiled within the DataFlex system or export of data to a different format.

When employing the query facility within the CIDRU system, the user is led through a series of screens at which various selections must be made in order to identify which file is to be interrogated, what field information is to be output, which fields held within that file are required for the output requirements, and how the data are to be output.

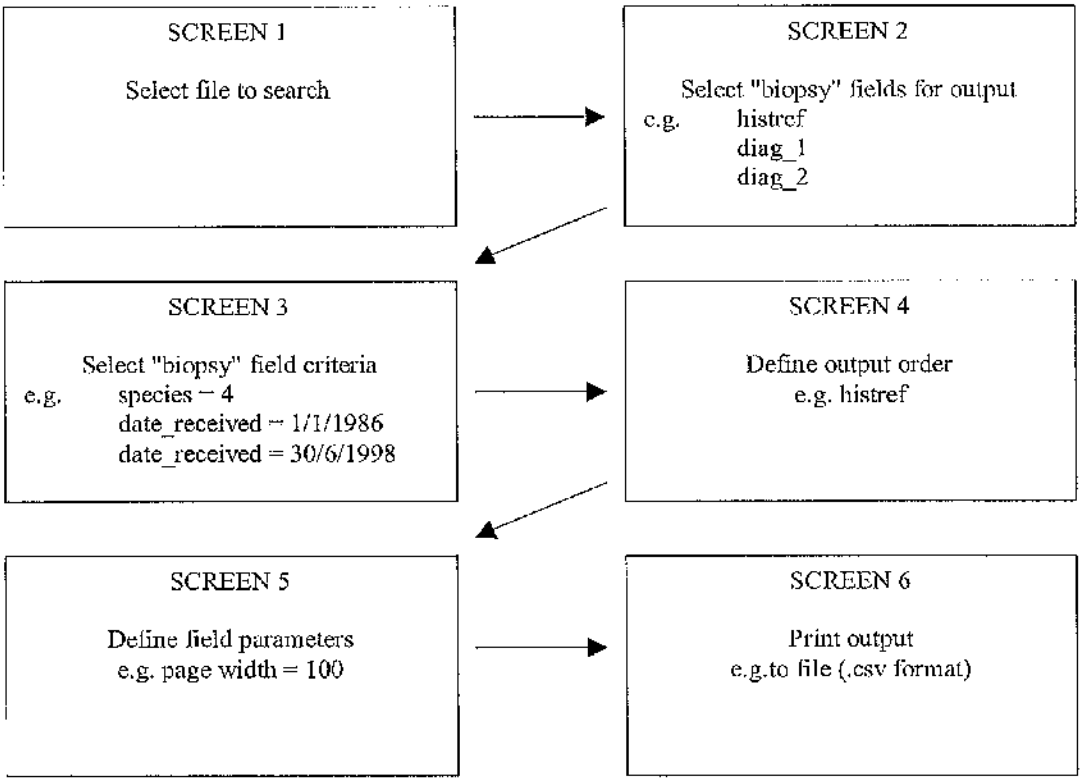


Figure 2-4 Representation of query screens within the CIDRU pathology database system where user must make selections of file to search and fields for output, specifying which fields are to be queried, providing definitions for the output order and field parameter requirements, and stating where the output of the query is to be printed.

Figure 2-4 outlines six key screen selections which are required, using an example of searching within the "biopsy" file, for the histology references (*histrefs*) and diagnoses of samples received between 1/1/1986 and 30/6/1998, from dogs only.

First, the file to be interrogated must be selected. In this example, "biopsy" must be selected because it contains the biopsy results. Second, the fields containing the data to be output must be selected. For the above example, "*histref*", "diag_1" and "diag_2" are chosen. Third, the field selection criteria may be stipulated, if necessary. To interrogate the database with respect to dogs only, it is at this stage that a "selection" screen where "species = 4" would be selected. Choosing "date received" prompts where the specific dates of interest may be defined. Selecting "date received 1/1/1986" and "date received

30/6/1998" fulfil the requirements of the original query. Fourth, the order in which the data are to be output must be instructed, and "*histref*" is a logical choice. Fifth, parameters, such as "page width" may be altered to accommodate the fields containing the data. In the above example, increasing the page width from 75 (default) to 100 is advisable because "diag_1" and "diag_2", being free text fields, may contain more than 75 characters. Finally, instructions as to where the data are to be printed are required. Data may be printed to the screen, but more usefully may be printed to an output file on the hard drive of the user. If "file" is selected, then further screen prompts with respect to desired filename and file format, such as "comma delimited" or "printable report" ensure that the data will be readily accessible to the user. Once the output file is stored on the user's hard disk, it can be manipulated as needed by the user.

2.5 DISCUSSION

Major categories of mistakes in histopathology laboratory are technical mistakes and mistakes in interpretation of the tissue (Bonfiglio and Terry, 1983). Mistakes may occur if the technician mislabels sample containers, paraffin blocks or prepared slides, or does not process all critical tissue. If the tissue is misprocessed due to equipment malfunction or poor technique, artifacts may occur making the processed tissue impossible to interpret.

Mistakes in interpretation can occur. It is to be expected that a second opinion on difficult cases is required. If the diagnosis does not fit the clinical picture, the clinician should consider asking the pathologist for a second opinion (Henderson and D'Andrea, 1993).

A submission form for practitioners to fill in and to accompany the tissue sample, was developed early on, and as the wish for more complete information about the animal and lesion under examination has been wanted, the form has undergone changes to encourage optimal information transfer from the practitioner responsible for the animal, and the information provided to the pathologist. As mentioned in Chapter 1, Section 1.5.2, information regarding the gross appearance and biological behaviour of the lesion can be crucial to the pathologist being able to provide the most accurate prognosis.

The forms for the submission of biopsies to the diagnostic histopathology service were unaltered for many years. However with the reconstruction of the diagnostic pathology service in 1998, forms were altered to encourage referring veterinarians to provide more comprehensive information about the samples needed to be studied. The latest form contains many fixed space areas for data recording and boxes for free text (Appendix 2).

The data examined for the purpose of this thesis was collected by the Department of Veterinary Pathology from January 1986 to June 1998. The number of canine biopsy submissions received by the external histopathology service during January 1986 to December 2001 are displayed graphically in Figure 2-5. The histogram clearly illustrates the substantial variation in submission numbers over time. This is likely to reflect the changing face of veterinary practice during the past five decades, and the birth of many privately owned and run veterinary histopathology diagnostic services within the past 15 years. During 2002, the GUVS external histopathology service received an average of 100 tissue submissions per month, of which 70 were canine in origin.

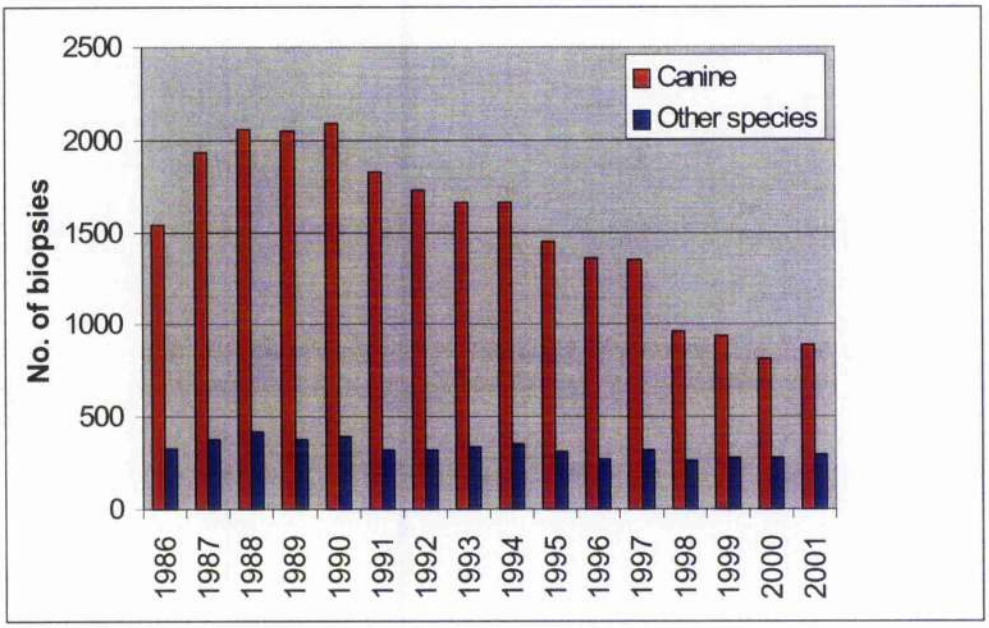


Figure 2-5 Biopsy submissions from dogs and all other species combined to the CIDRU histopathology service from January 1986 to December 2001.

In conclusion, the CIDRU database allows for detailed demographic and pathological data storage within a relational database system that provides efficient data storage and

retrieval. Targeted data retrieval is possible due to coding systems. The simplicity of coding systems removes the need for expensive training programs to assist users in their implementation. A criticism of the current systems is that their unique nature reduces the depth to which direct comparisons of these data may be made with histopathological data from sources using other coding and nomenclature systems, such as the SNVDO and SNOMED (Chapter 1, Section 1.6.2). However, the CIDRU database provides readily accessible data on pathological samples. Its user-friendly structure enhances its suitability as a data source for veterinary epidemiological research in the fields of neoplasia and infectious disease.

CHAPTER 3

THE CANINE INFECTIOUS DISEASES RESEARCH UNIT DATABASE - PREPARATION OF THE STUDY DATASET

3.1 INTRODUCTION

Neoplasia has been identified as one of the most common causes of disease and death in dogs today. The increased life expectancy of the dog due to widespread vaccination and improved veterinary care, greater owner awareness and expectation has paralleled the disease changes seen in the human populations of the developed world, with cancer becoming one of the most common disease conditions (Bonnett *et al.*, 1997; Michell, 1999).

Most epidemiological studies of canine neoplasia in recent years have been performed using populations of dogs derived from veterinary referral teaching hospitals. The most common data source is the Veterinary Medical Data Program (VMDP) (Chapter1, Section 1.6.3.2.1). The California and Tulsa Animal Neoplasm Registries (Dorn *et al.*, 1968a; Dorn *et al.*, 1968b; MacVean *et al.*, 1978) (Chapter 1, Sections 1.6.3.1.1, 1.6.3.1.2) were created with data derived from first opinion practices, a rare data source for researchers in this field. These data were used to calculate age, sex, breed and site specific incidence rates for many types of neoplasia and are still quoted widely, although it is now over 30 years since some of the data were collected. It is likely that incidence rates have changed during this time due to the advancements in methods of detection and diagnosis in the field. Canine cancer registries are still rare, although the investigation of cancer in the individual is now commonplace in veterinary practice. Thus, there is considerable data on the clinical aspects of canine neoplasia held by veterinary practices, and many sources of histopathological data can be found in veterinary pathology institutions, both in the academic and private sectors of the profession (Goldschmidt, 1993).

Advances in information technology over the past three decades now mean that veterinary pathology departments and diagnostic laboratories have computerised their pathology

records, many using databases specifically created for the purpose. Computerisation can greatly improve the efficiency with which information can be gained, and leads to the collection of vast datasets. The large increase in computerisation of veterinary practices has created real opportunities for practice-based research (Pollari *et al.*, 1996a; Reid-Smith *et al.*, 2000). However, as with all attempts to use secondary data, quality and completeness of available information is of utmost concern (Egenvall *et al.*, 1998).

Data entry precedes data utilisation, so the reliability of the information stored depends not only on the quality of the supplied data, but also the quality of data input. Most database software is largely based on the premise that data entered into the system are valid. Automated control routines on data entry tend to be limited and most often are confined to simple data type checks. For some databases, such as those serving research or diagnostics on pathology, quality control procedures at data entry are more important than for other database types, to ensure that the integrity of pathology data codes, ranges and relationships between data fields are maintained (Fleege, van Diest and Baak, 1992).

Types of data errors include incorrect spelling, omission of characters or numbers, incorrect format, ambiguous data type definition and incorrect alignment within a data field. These type of errors can be detected relatively easily. Errors may occur in numeric data if digits are omitted or their correct order is altered. These numeric data are more troublesome because a typing error does not lead to an obviously wrong notation but merely to a different number. It becomes even more complicated when mistakes produce eligible data that nonetheless may be false (Fleege, van Diest and Baak, 1992).

Data errors may arise due to transcription, translation, or omission. Every time information is recorded or transcribed (either onto paper or to computerised format) there is a possibility for errors to occur (Pollari *et al.*, 1996a). A certain level of generated coding/manipulation errors is expected in all datasets and are considered to be inconsequential in large population studies. However, data cleaning must be undertaken to consider a dataset suitable for further analyses. For such a potentially effort-intensive yet informative activity, data cleaning is rarely discussed and is not studied empirically in the literature, in which the data user's perspective is generally underrepresented (Maudsley and Williams, 1999).

To remove the selection bias associated with the use of a teaching hospital canine population such as the VMDP data, data collected by a histopathology service which serves first opinion veterinary practices were selected for the current study. This chapter introduces the proposed study design and details the data cleaning and coding procedure used to prepare data originating from the CIDRU histopathology service (Chapter 2) for subsequent epidemiological analyses. The quality of data received by the service and the common data errors encountered by the data user are described. As well as the procedure being integral to data validation, information gained from the procedure was also considered to be potentially useful to the database developers for improving in-built data quality checking, and to assist all users of the database.

3.2 MATERIALS AND METHODS

3.2.1 Data Source

The study dataset was derived from the CIDRU database, operated by the Department of Veterinary Pathology at the University of Glasgow Veterinary School (Chapter 2). The CIDRU external histopathology service receives samples from locations throughout the UK and occasionally from overseas. Most samples originate from privately owned or charity veterinary practices, with some being submitted from veterinary schools, diagnostic laboratories, or other non-practice sources. Within the database, each biopsy has a separate record. For multiple biopsies from an individual animal, several unique biopsy identification numbers (*histref*) link to a unique animal identification number. Each record is initiated using information supplied by veterinary personnel on a purpose-designed submission form (Appendices 1 and 2), one of which should accompany every sample.

3.2.2 Data Extraction

Using the query language incorporated into the CIDRU relational database system (Chapter 2, Section 2.2.6) all information pertaining to canine biopsy samples submitted to the external veterinary histopathology service, from 1/1/1986 to 30/6/1998, was extracted from the "Biopsy", "Animal", "SpecBred" and "Vet" files. The variables of particular interest for subsequent analyses were the host-related variables of age, gender ("sex"),

breed of dog, and site of biopsy. The selections were based on results from previous studies (Dorn *et al.*, 1968b; Cohen *et al.*, 1974; MacVean *et al.*, 1978). In addition, submitting practice was considered to be a potentially important variable in this dataset, because practice-related factors could influence biopsy submission to the service. Table 3-1 shows which fields were marked for data export. The extracted data was subsequently saved in comma-delimited file format on the author's hard drive.

Biopsy	**Animal**	**SpecBred**	**Vet**
Histref	Animal_rec	Species	Vetref
Date_received	Animal_ref	Breed_1	Vet_name
Pathologist	Sex	Breed_2	Vet_address_1
Vet_ref	Owner		Vet_address_2
Animal_rec	Dob		Vet_address_3
Animal_ref			Vet_town
Prev_biopsy			Vet_county
Clinical_dx			Vet_postcode
Site			
Diagnosis_1			
Site_1			
Class_1			

Table 3-1 Fields selected for data extraction from the CIDRU database.

3.2.3 Data cleaning

Microsoft Access 97 and Microsoft Excel 97 (Microsoft Corporation) were used to prepare the dataset for epidemiological analysis. Initially, the output file was imported into Microsoft Excel 97 and using spreadsheet functions, age (in years) of the dog was calculated as the difference between the "Dob" (date of birth) field and the "Date_received" field. The file was then imported into Microsoft Access 97, which immediately identified three records with data type errors. On closer inspection of these records, it was noted that all three records were affected by the same error, being that the correct information had been placed sequentially in the incorrect field. Once these records were corrected, it was possible to interrogate the entire extracted dataset using multiple

queries in Microsoft Access 97, to clean the data in preparation for epidemiological analyses.

Because a known diagnosis would be crucial to further analyses, the "Diagnosis_1" field was initially inspected for completeness of data entry, and records where this information was missing were immediately removed. A hierarchical and iterative approach to the data cleaning procedure consisting of sequential queries was developed, focusing on the main variables of age, sex and breed of dog, and site of biopsy. Coding procedures for the categorical variables were implemented to assist this procedure and incremental corrections were made to the study dataset to facilitate progress. Type and location of the submitting veterinary source were also assessed.

There were two stages to the data cleaning procedure. First, the data underwent preliminary inspection to locate obvious inconsistencies, e.g., missing or nonsense values. For example, those biopsy records with missing or nonsensical values in the "age" field were located and removed before performing the next sequential query based upon data in the "Sex" field. Data were then stratified to conduct logic checks, e.g., the use of valid gender, breed and site codes, correct gender recorded for biopsies originating from gender-specific sites and correct gender and site recorded for gender-specific diagnoses. For example, for all biopsy records given the "Site_1" code 32, identifying those biopsies originating from tissues of the male reproductive system (Chapter 2, Table 2-2), their free text "Site" field was checked for biological agreement, and it was verified that the "Sex" field contained the appropriate male gender code "ME" (male entire) or "MN" (male neutered) (Section 3.2.4.2). To complete the checks described above, inspection of the free text field, "Diagnosis_1", was required because a detailed morphological coding system for diagnosis is not used in the CIDRU database. Due to the size of the dataset and time constraints, the full computerised histopathology reports held in the CIDRU database, rather than the original paper records, were reviewed during this process to further correct discrepancies.

In order to prepare a practice variable for consideration in further analyses, it was necessary to review the coding of veterinary practice (*vetref*) in the "Vet" file. Checking for duplicates and making corrections where necessary were performed using a combination of Microsoft Excel and Microsoft Access. Practice address details were

checked using the Royal Mail Postcode Directory (1998), and where practice was not listed, the telephone number recorded in the CIDRU database was utilised for this purpose.

3.2.4 Research design

Preparation of the dataset was undertaken to enable a retrospective case-control study to be conducted (Chapter 4). The preliminary aim of the proposed study was to assess the likelihood of a submitted biopsy being neoplastic, while controlling for host factors of age, gender and breed of dog, and site of biopsy.

3.2.4.1 Definition of cases and controls

Criteria for a biopsy to be selected as a case were known age, gender and breed of dog, known biopsy site, and a histopathological confirmation of neoplasia. The same criteria as for the cases were used for the selection of a control population from the study population, though with a histopathological confirmation of a non-neoplastic diagnosis.

Selection of cases and controls was assisted by the "Class_1" coding system (Chapter 2, Table 2-3), and data filtering techniques in Microsoft Excel were used to improve the efficiency of visual inspection of the "Diagnosis_1" field where the "Class_1" field was inaccurate. For example, all biopsy records containing one of the three non-neoplastic "Class_1" codes (Chapter 2, Table 2-3) were filtered from those records containing "Class_1" code 1 (neoplasms). A second filter was then applied to the "Diagnosis_1" free text field, allowing efficient inspection of data in this field to confirm the entry of a non-neoplastic diagnosis for each record. Only biopsies received from UK first opinion or charity veterinary practices, with a confirmed histopathological diagnosis were considered for inclusion in the study. Where more than one biopsy had been received from a dog during the study period, the biological independence of these biopsies was assessed, thus making the unit of investigation the biopsy, rather than the dog.

3.2.4.2 Coding of independent variables

The categorical independent variables, sex, breed and site, were subdivided for further analysis. Gender was grouped using the coding system adopted by database users, into four categories: male entire (ME), male neutered (MN), female entire (FE) and female neutered (FN).

Specific breed level analysis was restricted to the top seven represented purebreeds, being: Labrador Retriever, German Shepherd, English Springer Spaniel, Boxer, Cocker Spaniel, Jack Russell Terrier and Doberman Pinscher. All other purebreeds were designated as "other", with crossbreeds being placed in a separate category. The non-specific breed classifications, "Retriever", "Terrier", "Spaniel" and "Collie" were commonly used in records entered during the early years of the study period, presumably because only this limited information was provided on the submission forms. Chi-squared analyses for association were performed to determine whether samples from dogs with a non-specific breed classification could be pooled with the "other" or "crossbreed" reference groups. These analyses were performed using Minitab 12.21 (Minitab Inc. State College, PA).

Site of biopsy was coded using the numerical system designed by the pathologists responsible for creating the CIDRU relational database. The number in the site field refers to one of 50 possible sites (Chapter 2, Table 2-2). The number of site categories was reduced by combining categories based on biological reasoning, to allow maximal inclusion of records in more detailed site analyses. Combining certain sites also produced groupings similar to those used by researchers when conducting previous studies of the epidemiology of canine neoplasia. Specifically, sites 43 (skin), 44 (connective tissue) and 47 (feet) were combined to form a skin/connective tissues category; sites 1 (mouth) and 2 (pharynx) were combined to form an oropharynx category; sites 32 (male reproductive), 33 (female reproductive) and 35 (prostate) were combined to form a reproductive category; sites 6 (small intestine) and 7 (large intestine) were combined to form an intestine category, and sites 12 (tonsil), 39 (thymus) and 41 (lymphatic) were combined to form a lymphatic category. Biopsies coded as site 8 (anus) were recoded as either intestine (if the lesion was rectal in origin), or skin (if the lesion was epithelial in origin). Following the

reduction of the site codes, site was grouped according to the top eight represented biopsy sites, with all biopsies from other sites being placed into a separate category.

3.3 RESULTS

3.3.1 General

During the study period, a total of 21 371 canine biopsies were submitted to the external histopathology service at GUVS from 542 sources located throughout the United Kingdom, and three sources located overseas (Malta, Abu Dhabi, Dubai).

3.3.2 Data cleaning

The percentage of records found to be ineligible for inclusion in the statistical analysis of records from all sources was 14.1% (3005/21 371). Missing or nonsense values were the major causes of record exclusion (1892/3005, 63%), followed by an "inconclusive" diagnosis being recorded for the biopsy (882/3005, 29.3%). Because the study was to be based upon biopsies received from first-opinion veterinary practices in the United Kingdom, those received from the three overseas sources (n =10) and from non-first-opinion practice sources (n = 221) were also excluded (231/3005, 7.7%).

For each host-related variable of interest, i.e., age, gender, breed of dog and site of biopsy, data checks were performed during the data cleaning process and corrections made by inspecting the full histopathology report where this was necessary. Coding errors were detected in 0.91% (167/18366) of records, 66 (39.5%) of these being major miscoding of site, i.e., not closely related to the correct code, and 54 (38.6%) being miscoding of gender status, i.e., MN rather than ME; FN rather than FE. Certain site categories were more likely to contain incorrect entries, such as pituitary (2/2, 100%), adrenal (5/11, 45.5%), lower respiratory (3/12, 25%) and kidney (6/52, 11.5%). There were also site categories for which there were no entries following corrections (forestomachs, autonomic nervous system, pituitary, placenta, septicaemia/toxaemia, foetus).

There were no records containing illegal breed codes, however there were a significant number (2396/18366,13%) where only the first breed code ("Breed_1") was recorded. Chi-squared comparison of cases and controls in each of the non-specific breed groups with purebreeds and crossbreeds resulted in records from Retrievers of unknown breed being grouped with purebreed Retrievers (^{2}=2.08, df=1, P=0.15), records from Collies of unknown breed being grouped with Collie crossbreeds (^{2}=0.01, df=1, P=0.93) and records from Terriers of unknown breed being grouped with Terrier crossbreeds (^{2}=0.28, df=1, P=0.6). Records from Spaniels of unknown breed were grouped with "other" breeds because there were significant differences between this group and purebreed Spaniels (^{2}=6.25, df=1, P=0.01) as well as with crossbreed Spaniels (^{2}=7.35,df=1, P=0.007).

According to the "Class_1" code criteria (Chapter 2, Table 2-3), 2.3% (430/18366) of records contained illegal code entries. Coding errors were detected in 3.8% (438/11665) of the neoplasm category, and neoplasms were incorrectly coded in 1.8% (113/6271) of entries in the other categories (specific disease, congenital, other). Where coding discrepancies were found, records were selected as cases or controls by inspection of the free text "Diagnosis_1" field.

Following data cleaning, 11 422 neoplastic biopsies were eligible as cases and 6944 controls were selected for further analysis. The total of 18366 biopsies originated from a total of 16877 dogs, with 6.9% (1158/16877) of dogs contributing multiple biopsies. Where more than one biopsy was submitted from the same dog, the number of biopsies that were not of a biologically distinct nature was 0.01% (213/18366). Due to the negligible effect this was expected to have upon further analyses, it was considered acceptable to retain these records. Figure 3-1 summarises the handling of the CIDRU dataset.

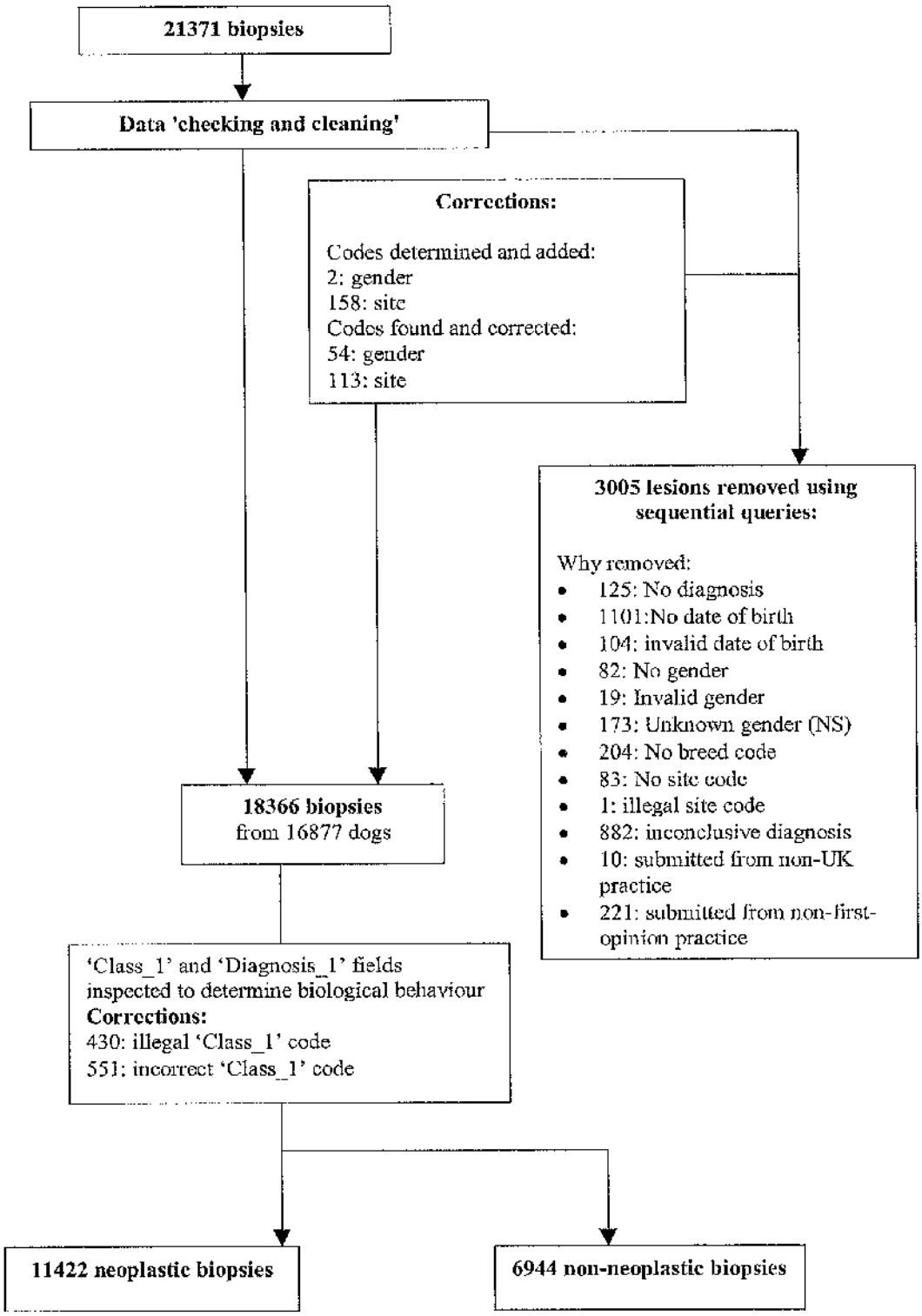


Figure 3-1 Summary of handling of the Canine Infectious Diseases Research Unit study dataset.

Initial inspection of the veterinary practice data revealed that 464 practices, or *vetrefs* had contributed biopsies to the study dataset. Closer inspection showed that 29 practices had been allocated more than one *vetref* code; 26 practices had two *vetref* codes, and 3 practices had 3 *vetref* codes. Thus 32/464 (6.9%) of the *vetref* codes were representing a practice incorrectly. Of these cases, reference to the original *vetref* code for the practice was made in the "vet_town" field in 31/32 (96.9%) cases. Reasons for more than one *vetref* code being allocated to a practice included the allocation of codes to individual veterinarians within the same practice (12/32, 37.5%), duplicated veterinary practice details (11/32, 34.4%) and change in veterinary practice name (3/32, 9.2%). Following correction, 436 veterinary sources were verified as contributing biopsies to the study dataset.

The "vet_postcode" field was of particular interest because of this field's ability to be used to obtain geographical co-ordinates for practice location. Of the original 464 *vetref* codes, 39 postcodes (39/464, 8.4%) were missing, 49/464 (10.6%) postcode fields contained illegal entries and 9/464 (1.9%) contained incomplete entries. Following identification of the 436 correct veterinary sources, their postcodes were entered, corrected or completed with assistance of the Royal Mail Postcode Directory (1998). Particular problems were encountered with postcodes of practices in Aberdeenshire, the majority of which had altered during the study time period. Further investigation of this finding revealed that this region had been subjected to a major review of its postcode sectors in 1996/97, resulting in the creation of many new postcodes. Where needed, the practices were contacted by telephone to confirm their postcodes if the information could not be gained by other means. This procedure also revealed some retired practitioners, which explained the absence of some practice addresses in the postcode directory.

3.3.3 Qualitative observations on data quality

The date of birth ("Dob") field was problematic, causing a large number of records to be excluded due to missing or nonsense values (1205/3005, 40.1%). This demonstrated an insufficiency of range and safety checks at data entry in the database software. The lack of data type checks was noted for other fields, e.g., where numbers were placed in text fields

and text placed in numerical fields. Examples of fields where this occurred were: "Sex", "Site", "Site_1", "Diagnosis_1" and "Class_1".

The major site categories involving in site miscoding were gender-related, i.e., male reproductive (code 32) and female reproductive (code 33). Many of these miscodings may have been unintentional, given the close numerical similarity of the codes, but they were viewed as major misclassifications due to the extreme biological dissimilarity of the site represented by the code.

While checking the site code validity of records by cross-referencing with gender and diagnosis data, it was noted that the coding of certain diagnoses was problematic. The site coding system could allow the same lesion from the same site to be given more than one code, with more than one code being correct. For instance, perianal hepatoid adenomas were classified as originating from skin (code 43), connective tissue (44), anus (8), or male reproductive (32). Of these four codes, the last was incorrect in all instances, however any of the other three could be viewed as correct, if the adenoma was from the anus. It was due to diagnoses of this nature that the decision was made to combine certain site categories (Section 3.2.5).

The errors uncovered in the "Vet" file revealed a lack of integral data quality checks in this area of the database, which allowed noteworthy duplication of veterinary practice data. The allocation of codes to individual veterinarians, while recording another *vetref* code in the "vet_town" field, showed that review of the field structure in the "Vet" file was needed in order to ascribe easily accessible demographic and geographical data for each biopsy directly from the database.

Consideration could be given to creating an additional field to allow recording of the submitting practitioner's name, rather than allocating a separate code for each practitioner within a practice. Duplication of veterinary practice address details may be avoided by introducing integral data quality checks into the software. These types of problems may be avoided by reviewing data entry procedures with database user personnel, to increase awareness of the type of errors that may be inadvertently introduced into the database.

3.4 DISCUSSION

The overall aim of the proposed study involved the host factors of age, gender and breed of dog, the site from which a biopsy was taken and the ability to use a practice-based variable in further analyses. The need for prioritisation of these data enabled a systematic approach to the data cleaning process to be developed, in order to retain the maximum number of records for statistical analyses. The hierarchical and iterative data cleaning method gave consideration to which missing data could be most easily retrieved from the content of the database, without requiring the time-consuming and effort-intensive tasks of searching of the paper records, or contact with the submitting practice. The removal of records with incomplete demographic and/or diagnostic data was inevitable with this approach. However, where summarised topographical data was missing or incorrect, inspection of the electronic histopathology report was, in many cases, an extremely useful and time-effective method of finding the required information.

Data errors may occur whenever there is entry or transfer of information. It is unrealistic to expect a perfect dataset wherever human processes are involved in the data collection but this should not dissuade researchers. Explicitly tackling technical concerns about routine data may actually help to combat conceptual and emotional reservations held by data users that data may be too 'dirty' and therefore unsuitable for further study (Yates and Davidge, 1984). Although firm estimates of allowable error rates have not been established, there should be some effort to quantify the quality of the data source in all studies involving secondary data (Ray *et al.*, 1992; Meehan *et al.*, 1995; Egenvall *et al.*, 1998).

The number of excluded records in this dataset (3005/21371, 14%) was considered acceptable to validate the remaining dataset for further analyses. The cleaning strategy must be customized to minimize effort, balancing feasibility of data recovery versus the missing or incorrect data's overall importance to the intended study. The major misclassifications were mostly detected by logic checks. Cleaning strategies should focus on such checks, as should methods to prevent the introduction of errors into the database. Prevention involves avoiding their introduction, primary prevention, and early detection by in-house quality assurance, secondary prevention (Maudsley and Williams, 1999). Primary prevention methods would be aimed at all users of the database. This could involve staff

training to avoid systematic coding errors, and a protocol where practitioners were contacted by reception staff if incomplete data were provided with a submission to the histopathology service. Secondary prevention could be aimed at increasing the level of integral checks built into the database software, to avoid, e.g., nonsense date sequences, or incorrect gender coding for biopsies from gender-specific sites.

In any study using veterinary diagnostic information, accuracy and consistency in the use of nomenclature are vital (Bonnett *et al.*, 1997). This study utilises a coding scheme for classification of primary site of neoplasms developed by members of the Department of Veterinary Pathology at GUVS. This coding system is simple in comparison to others such as the SNVDO and SNOMED (Chapter 1, Section 1.6.2). These published coding and nomenclature systems can involve a high initial costs due to training and software requirements (Feigl *et al.*, 1981; Moss, Smith and Nicholas, 1997). The simplicity of the CIDRU site coding system is an advantage in that no formal training is required. However, its simplicity also carries disadvantages, some of which were shown by the data cleaning process. These include the numerical closeness of codes of distinctly different biological entities, which may have contributed to the number of topographical code errors due to inadvertent keystrokes. There were also problems of overlap between codes, so that a single lesion could potentially, in some cases, be given any one of up to three codes. Review of the site coding system could be considered by pathologists to increase uniformity of coding procedure and completeness of data in this field. Discussion may also highlight the potential for simple coding errors to be entered into the database.

The previously mentioned coding systems incorporate coding for morphology, which significantly increases their complexity. The CIDRU database contains a general morphology coding system, recorded in the "Class_1" field (Chapter 2, Table 2-3), and all diagnoses are entered into the CIDRU histopathology database as free text. The accuracy of coding neoplasms with the CIDRU general morphology coding system (Class_1) was highly acceptable, with minimal corrections required. However, there were problems associated with illegal code entry and other class categories showed problems with overlap between codes. Strategies to reduce inadvertent keyboard entry errors, such as integral range limitation checks, could be employed to minimise some of these data error types. Ultimately, for the purpose of this study, the class coding system and free text entry of

"Diagnosis_1" led to extensive, labour-intensive querying of the diagnosis field. This was necessary during the data cleaning process to identify those biopsies without a definitive histopathological diagnosis, to check gender-specific codes as well as to select cases and controls for further study. Review of the class coding system and developing a standardised nomenclature for diagnosis would significantly improve data handling efficiency for data users. This would be particularly helpful if the researcher's interest was associated with extraction of records with a non-neoplastic diagnosis.

The "vet_postcode" field caused the most difficulties during the cleaning of veterinary practice data. Highlighting the importance of postcode as a source of geographical information suitable for spatial analyses (Chapter 1, Section 1.10) is necessary to ensure that postcode information is obtained by the data entry personnel if it has not been recorded, or is incomplete, on the submission form. Increasing the data users' awareness of the effects of keystroke error may also assist with reducing the number of illegal entries in the postcode field. Regular review of postcode entries may also assist with identifying independent alterations to these data, such as those caused by the re-structuring of postcodes in Aberdeenshire by the Royal Mail in 1996/97.

The data cleaning approach was effective in producing a study dataset suitable for further analyses, and identified a number of important quality assurance issues. There remain concerns associated with the cost of alterations to the database software to prevent coding errors, and with the development of a cleaning procedure with defined levels of allowable coding errors, so that further epidemiological analyses will not be invalidated. The need to examine the free text diagnosis field also requires the data user to have some prior knowledge of histopathological terminology, or direct communication with a pathologist to assist with terminology queries.

Data cleaning preceding epidemiological analysis and *ad hoc* studies are best at detecting registration problems in human cancer registries. Cancer registration data need "to be seen to be of high quality" to be valued and inspire confidence (Brewster, 1995).

Assessment of the quality of some veterinary data sources has been addressed for the purpose of selected studies (Pollari *et al.*, 1996a; Egenvall *et al.*, 1998; Dobson *et al.*,

2002). The findings of data users may provide insight into developing efficient, time-effective methods of improving data quality. The documentation of data quality issues is an essential component of increasing a database's potential for epidemiological research.

CHAPTER 4

APPLICATION OF LOGISTIC REGRESSION TO A CANINE BIOPSY POPULATION COMPILED BY A DIAGNOSTIC HISTOPATHOLOGY SERVICE

4.1 INTRODUCTION

Risk factors for canine neoplasia can be classed as intrinsic or host-related, and extrinsic or environmentally-related (Chapter 1, Section 1.7). The largest population-based studies of intrinsic risk factors for canine neoplasia are the CANR (Dorn *et al.*, 1968a; Dorn *et al.*, 1968b) (Chapter 1, Section 1.6.3.1.1) and the Tulsa Registry of Canine and Feline Neoplasms (MacVean *et al.*, 1978) (Chapter 1, Section 1.6.3.1.2). In the last decade, the Norwegian Canine Cancer Register (NCCR) been developed and continues today as a country-wide canine cancer register (Arnesen *et al.*, 2001) (Chapter 1, Section 1.6.3.1.4).

Many other studies of specific canine neoplasms and their risk factors have been performed using data from the Veterinary Medical Database (VMDB) (Chapter 1, Section 1.6.3.2.1) which unlike the aforementioned studies, contains information derived from a referral population. Studies of first-opinion data (such as the CANR, Tulsa Registry and NCCR) have been reliant upon recruitment and co-operation of veterinary practitioners in the designated study areas to ensure submission of all suspected neoplasms to the investigators. To encourage submission of all suspected lesions, histopathology was offered free of charge by the investigators in all of the above studies. Possible sources of error were recognised by the NCCR researchers, particularly bias in the collection methods. These included bias towards the submission of easily accessible lesions, and the possibility of bias in the submission of samples from particular breeds (Arnesen *et al.*, 2001). However, no comment was made regarding owner-related biases, such as their willingness for their pets to undergo biopsy procedures for suspect lesions.

Multivariable techniques such as multivariable logistic regression are widely recognised in veterinary epidemiology, and have been used in studies investigating possible risk factors for certain types of canine neoplasia in particular sites, such as bone cancer (Ru, Terracini

and Glickman, 1998), bladder cancer (Glickman *et al.*, 1996) and mammary cancer (Sonnenschein *et al.*, 1991). However, multivariable logistic regression has not been used as a method for assessing risk or odds of neoplasia being diagnosed in a biopsy sample controlling for the effects of age, gender and breed of dog, as well as site of biopsy.

The primary objective of this study was to assess the risk or odds of neoplasia being diagnosed in a biopsy sample while controlling for certain host-related risk factors. Because the dataset was derived from a diagnostic histopathology service, without prior communication with the veterinary practitioners responsible for biopsy submissions, it was hypothesised that the data might be aggregated by practice. To control for confounding by practice, fixed and random-effect regression analysis was used.

4.2 MATERIALS AND METHODS

4.2.1 Data Source

The study dataset, data preparation, study design and definitions for cases and controls have been described previously (Chapter 3).

4.2.2 Statistical analysis

4.2.2.1 Univariable analysis of host-related variables

An initial screening univariable logistic regression was performed to identify those variables that had little or no association with the outcome of neoplasia. All variables significant at P 0.25 were considered eligible for inclusion in a multivariable analysis (Hosmer and Lemeshow, 1989).

4.2.2.2 Multivariable analysis of host-related variables

The effect on the outcome of neoplasia of the combination of risk factors identified in the screening univariable analysis was evaluated using multivariable logistic regression. The conditional distribution upon which multivariable logistic regression is based, is the

binomial (chi-square) distribution as opposed to the normal distribution applicable to linear regression analysis. The specific form of the logistic regression model used in this study, was as follows:

$$P(X) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_k x)}} \qquad \text{equation 4-1.}$$

where P(X) represents the conditional probability of neoplasia given the collection of independent variables, x_I, and where β_0 and the β_i's are constant terms representing unknown parameters which are estimated based on the data set.

The logistic model has become popular in epidemiological studies where the outcome of interest is dichotomous, e.g. presence/absence of disease (Breslow and Day, 1980; Hosmer and Lemeshow, 1989). It has gained widespread acceptance in case-control studies, in which results are interpreted in terms of estimates of odds ratios. Odds ratios quantify the effect of exposure to risk factors in an individual with the disease of interest, which for rare diseases, provide an estimate of the relative risk. Odds ratios for this study were estimated by fitting a logistic regression model to the data set. Computation of these ratios used an alternative form of the logistic model, the logit form, which is derived from the logistic model by the logit transformation, g(x), defined as:

$$g(x) = \ln\left[\frac{P(X)}{1-P(X)}\right] \qquad \text{equation 4-2.}$$

which after substitution of the logistic model form of P(X), simplifies to:

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_k x_k \qquad \text{equation 4-3.}$$

This logit form gives an expression for the log odds of neoplasia for a biopsy with a specific set of independent explanatory variables. β_i, known as the regression coefficient, represents the change in log odds for one unit change in the variable x_i, when all other variables are fixed. Central to the consideration of the multivariable logistic regression model, is estimation of the regression coefficient for each independent variable and testing

for the significance of each (Hosmer and Lemeshow, 1989). Maximum likelihood estimation is used to calculate the logit coefficients. This method seeks to maximize the log likelihood which reflects how likely it is (the odds) that the observed values of the dependent or outcome variable may be predicted from the observed values of the independent explanatory variables.

Variables considered for inclusion in the regression analysis were those that were significant at $P \leq 0.25$ in the univariable analysis. Dummy variables were generated for categorical variables with more than two levels. Beginning with the full model of variables significant at $P \leq 0.25$, a backward elimination procedure was used to refine the multivariable model. Variables remained in the model if removal of that variable resulted in a significant change ($P < 0.05$) in the likelihood ratio test statistic, assessed as a chi-square distribution. Biologically meaningful two-way interaction terms among the independent variables were examined after identification of the main effects.

The assumption of linearity in the logit for the continuous variable, age, was checked by grouping "age" into intervals, calculating the mean of the outcome variable (neoplasia) for each interval and creating a scatter plot of this mean versus the midpoint of each interval. The Box-Tidwell transformation was also performed, by adding the term xLn(x) to the model and assessing the coefficient associated with this term (Hosmer and Lemeshow, 1989).

The goodness-of-fit of the final main (host-related) effects model was assessed by the Hosmer-Lemeshow statistic (Lemeshow and Hosmer, 1982). A series of regression diagnostics were also inspected to assess fit of the model over the entire covariate pattern. These included the square of the standardised residuals, the deviance, the leverage and the predicted probability. Influence of individual covariate patterns was assessed by examining the delta beta ($\Delta\hat{\beta}$) values, also known as the influence diagnostic:

$$\Delta\hat{\beta}_j = \frac{r_j^2 h_j}{1 - h_j}$$

equation 4-4.

(Hosmer and Lemeshow,1989)

where $\Delta\hat{\beta}$ represents the influence diagnostic, j represents individual covariate patterns, r^2 represents the square of the standardised residuals and h represents the leverage.

Inspection of the regression diagnostics allowed identification of covariate patterns that were poorly fit to the regression model and those with strong influence on the estimation of the coefficients. Biopsies with the most influential covariate patterns were removed from the dataset and the significance of the variables retained in the model re-evaluated. The criteria adopted for intervention was a change of greater than 20% in any of the variable coefficients and/or evidence of statistical instability. The fitting, assessment and diagnostics were repeated until a final main effects model was attained in which all variables were considered to be of biological and statistical significance. Adjusted odds ratios and 95% confidence intervals were obtained for each risk factor retained in the final multivariable model, by exponentiation of the regression coefficients.

Due to the heterogeneous distribution of covariate patterns within the multidimensional covariate space, the interaction term involving site of biopsy and breed of dog did not converge to satisfy the tolerance criterion. A stratified approach was then adopted. For each of the eight specific site categories individually, univariable analyses as already described were carried out to assess the significance of age, gender and breed on the outcome of neoplasia in a biopsy from that site. Where appropriate, multivariable models were then created and assessed in the same manner as for the main (host-related) effects model. All statistical analyses were performed using Minitab Release 12.21 (Minitab Inc., State College, PA).

4.2.2.3 Inclusion of practice variable

4.2.2.3.1 Univariable analysis

Univariable logistic regression was performed to assess the association of practice with the outcome of neoplasia. Due to the large number of zero cells in the covariate matrix, it was decided to alter the inclusion criterion of practice so that only those contributing more than five neoplastic and more than five non-neoplastic biopsies were eligible for inclusion in a

multivariable analysis which contained a variable representing practice. Data relevant to these practices and their associated biopsies, hence and hereafter referred to as the "reduced dataset", underwent a univariable screening procedure as previously described (Section 4.2.2.1). In addition, two-sample t-tests for the continuous variable, age, and chi-squared analyses for the categorical variables (gender and breed of dog, sites of biopsy) were performed to examine the data for patterns of biopsy submission between practices, according to the variables of interest.

4.2.2.3.2 Multivariable analysis

Use of multivariable logistic regression requires the assumption that the subjects with identical sets of risk factors are independently and identically distributed. However, where the disease data has arisen from populations which are organised into groups, such as herds, flocks or pens, it is likely that there are factors inextricably linked to the individual groups. Such factors include differences in exposures, management practices, genetics, feed or other, sometimes difficult to quantify, differences between the groups. Group effects may also result from intragroup correlation, for instance, if the disease under study is contagious. With these group effects, an additional source of extra-binomial variation (EBV), or overdispersion, is introduced into a logistic regression model. This occurs because of the groups are heterogeneous with respect to disease risk, so animals in different groups with identical risk factors do not have the same probability of disease. Extra-binomial variation causes violation of the independence and identical distribution assumptions, leading to biased variance estimators and spurious statistical significance (Curtis *et al.*, 1993).

One method of accounting for the EBV introduced by group effects is to model the group variable as a fixed-effect. Including group in this way assumes that risk is fixed for a given group and necessitates inclusion of k-1 dummy variables for k groups in the logistic regression model. For large numbers of groups, this can be cumbersome to deal with analytically. The epidemiological concern with this method, being that these group effects are not quantities that can be fixed, has also been recognised (McDermott, Schukken and Shoukri, 1994).

An alternative approach is treat group effects as random-effects, so that variation from unmeasured and unmeasurable (but clustered) sources can be accounted for (Curtis *et al.*, 1993). Random-effects models have an extra constant parameter which models the excess variation introduced by heterogeneity of risk between groups. There are several random-effects models that can be used to account for group data, and the choice of which to use is dependent upon whether both group- and individual-level data are available, and the level of *a priori* knowledge of the problem being studied. Where group- and individual-level data are to be examined, if data are considered to be indistinguishable, i.e., covariate or factor levels do not vary within a given cluster or group, logistic-normal or logistic-binomial random-effects logistic regression are available. For distinguishable data, where covariate values are known to vary with a given cluster or group, the logistic-binomial model is used.

In this dataset, biopsies were grouped by submitting practice. It was likely that at the practice level, there would be "group" effects, such as veterinarians' individual choices regarding when and which lesions to biopsy, which biopsies were to be submitted to a diagnostic histopathology service and which service to use, of the many available in the UK. Other factors, such as the percentage of practice work that was small animal-oriented and urban versus rural practice location may have also contributed to EBV in a multivariable logistic regression model of data derived from different practices.

To examine the level of EBV (overdispersion) in the reduced dataset, a multivariable logistic regression analysis was first performed, in an identical manner to that described above for the analysis of host-related variables in the full dataset (Sections 4.2.2.1, 4.2.2.2). The resultant main-effects model was then extended with the inclusion of practice, initially as a fixed effect. However, given that a large number of practices were represented in the reduced dataset, it was decided that a more appropriate method of accounting for EBV may be to treat practice effects as a random-effect, and a logistic-binomial random-effects logistic regression analysis was performed. The results of the fixed- and random-effects models were then compared with the main-effects model of host-related variables. The P value for the EBV parameter in the logistic-binomial model was obtained using a likelihood ratio test based upon model deviance. The analyses containing practice as a fixed-effect were performed using Minitab Release 12.21 (Minitab

Inc., State College, PA). The analysis containing practice as a random-effect was performed using EGRET for Windows, version 2.0.3 (Cytel Software Corporation, 1999).

4.3 RESULTS

4.3.1 Univariable analysis

4.3.1.1 Univariable analysis of host-related variables

The summary statistics for the cases and controls by age and by subcategory of the categorical independent variables identified for investigation are presented in Table 4-1. The results of the univariable screening of independent variables demonstrated that age, gender and breed of dog and site of biopsy were all significantly associated with a histopathological diagnosis of neoplasia (Tables 4-2 and 4-3).

4.3.1.2 Univariable analysis of practice

Visual inspection and summary statistics of the distribution of biopsy submission by practice showed marked variation in the numbers of biopsies contributed from individual practices (1 to 1938 biopsies, median 4, interquartile range 1 to 107). The number of neoplastic versus non-neoplastic lesions contributed from each practice also varied widely. For example, there was one practice that contributed 114 biopsies, of which only four were neoplastic. Univariable logistic regression confirmed that the practice variable was significantly associated with the outcome of neoplasia. There were 329 practices contributing 5 neoplastic and 5 non-neoplastic lesions (329/436, 75.5%); these practices contributed 10% (1830/18366) of biopsies in the study dataset. Because fitting these practices to a logistic regression model would cause unstable parameter estimates, it was decided to exclude these practices and their biopsies from any analyses which included a practice variable. Univariable logistic regression of practice in the reduced dataset (16536 biopsies from 107 practices) showed practice was still significantly associated with the outcome of neoplasia. Output of this analysis is presented in Appendix 3.

	Controls		Cases	
Number	6944	100%	11422	100%
Age (years)				
1st quartile	3.5		6.5	
Median	7		9	
3rd quartile	9.5		11	
Gender				
Male entire	3252	46.83%	4576	40.06%
Male neutered	257	3.70%	352	3.08%
Female entire	2558	36.84%	5090	44.56%
Female neutered	877	12.63%	1404	12.29%
Breed				
Labrador Retriever	706	10.17%	1491	13.05%
German Shepherd dog	601	8.65%	695	6.08%
English Springer Spaniel	329	4.74%	552	4.83%
Boxer	261	3.76%	516	4.52%
Cocker Spaniel	227	3.27%	489	4.28%
Jack Russell Terrier	230	3.31%	379	3.32%
Doberman	276	3.97%	251	2.20%
Crossbred	1124	16.19%	2346	20.54%
Other	3190	45.94%	4703	41.17%
Site of biopsy[a]				
Skin[a]	5138	73.99%	6496	56.87%
Mouth/pharynx	315	4.54%	876	7.67%
Mammary gland	284	4.09%	2217	19.41%
Lymphatic system	216	3.11%	493	4.32%
Liver	160	2.30%	49	0.43%
Spleen	106	1.53%	125	1.09%
Reproductive system	185	2.66%	549	4.81%
Intestine	132	1.90%	135	1.18%
Other	408	5.88%	482	4.22%

[a]includes biopsies of epithelial, mesenchymal and melanocytic cell origin

Table 4-1 Summary statistics for case-control data for canine biopsies compiled by the CIDRU histopathology service during 1/1/1986 – 30/6/1998.

	Controls	Cases	P-value	Odds Ratio	95% CI Lower	Upper
Number	6944	11422				
Constant						
Age (years)			<0.00	1.17	1.16	1.18

Table 4-2 Univariable analysis of the continuous variable, age, investigated for association with an outcome of neoplasia being diagnosed in a canine biopsy submitted to the CIDRU diagnostic histopathology service.

	Controls	Cases	P-value	Odds Ratio	95% CI Lower	Upper
Gender			<0.00			
Male entire[1]	3252	4576		1		
Male neutered	257	352		0.97	0.82	1.15
Female entire	2558	5090		1.41	1.32	1.51
Female neutered	877	1404		1.14	1.03	1.25
Breed			<0.00			
Labrador Retriever[1]	706	1491		1		
German Shepherd	601	695		0.55	0.48	0.63
English Springer Spaniel	329	552		0.79	0.67	0.94
Boxer	261	516		0.94	0.79	1.11
Cocker Spaniel	227	489		1.02	0.85	1.22
Jack Russell Terrier	230	379		0.78	0.65	0.94
Doberman	276	251		0.43	0.36	0.52
Crossbred	1124	2346		0.99	0.88	1.11
Site of biopsy			<0.00			
Skin[1a]	5138	6496		1		
Mouth/pharynx	315	876		1.36	1.18	1.57
Mammary gland	284	2217		3.86	3.35	4.46
Lymphatic system	216	493		1.4	1.17	1.67
Liver	160	49		0.16	0.11	0.22
Spleen	106	125		0.55	0.41	0.72
Reproductive system	185	549		1.43	1.19	1.73
Intestine	132	135		0.63	0.48	0.82

[1]referent category
[a]includes biopsies of epithelial, mesenchymal and melanocytic cell origin

Table 4-3 Univariable analysis of categorical variables investigated for association with an outcome of neoplasia being diagnosed in a canine biopsy submitted to the CIDRU diagnostic histopathology service.

4.3.2 Multivariable analysis

4.3.2.1 Main effects model of host-related variables

The final main-effects model of the host-related variables, containing age, gender, breed of dog and site of biopsy is shown in Table 4-4. Increasing age was associated with an increased risk (odds) of neoplasia in a biopsy (OR=1.13; 95%CI 1.12, 1.14). Compared to the Labrador Retriever (the referent category for breed) the Boxer was the only dog breed of those specified to be associated with an increased risk of neoplasia in a biopsy (OR=1.3; 95%CI 1.06, 1.59). Biopsies from the oropharynx (OR=1.36; 95%CI 1.18, 1.57), mammary gland (OR=3.86; 95%CI 3.35, 4.46), lymphatic system (OR=1.4; 95%CI 1.17, 1.67) and reproductive systems (OR=1.43; 95%CI 1.19, 1.73) had an increased risk of a neoplastic diagnosis compared to biopsies from the skin and connective tissues (the referent category for site), when controlled for age, gender and breed of dog. Biopsies from neutered females (OR=0.83; 95%CI 0.74, 0.93), German Shepherd dogs (OR=0.63; 95%CI 0.53, 0.74), English Springer Spaniels (OR=0.67; 95%CI 0.55, 0.81), Jack Russell Terriers (OR=0.65; 95%CI 0.52, 0.81), Dobermans (OR=0.51; 95%CI 0.41, 0.64), crossbreeds (OR=0.77; 95%CI 0.67, 0.88) and other non-specified breeds (OR=0.67; 95%CI 0.60, 0.75) showed a decreased risk of neoplasia (when controlled for other host-related factors). Decreased risk was also shown for biopsies from the liver (OR=0.16; 95%CI 0.11, 0.22), spleen (OR=0.55; 95%CI 0.41, 0.72) intestine (OR=0.63; 95%CI 0.48, 0.82) and other non-specified sites (OR=0.66; 95%CI 0.56, 0.76), when controlled for age, gender and breed of dog.

Sequential exclusion of the most influential covariate patterns, as determined by inspection of delta beta values, did not significantly improve goodness-of-fit of the model. The Hosmer-Lemeshow statistic of the final model was 2.2 with eight degrees of freedom (P = 0.97), indicating excellent model fit.

Variable	Coefficient	SD	P-value	Odds Ratio	95%CI Lower	95%CI Upper
Constant	-0.46					
Age	0.14	0.00	0.00	1.15	1.14	1.16
Gender						
Male entire[1]	0.00			1		
Male neutered	-0.11	0.09	0.23	0.90	0.75	1.07
Female entire	0.01	0.04	0.81	1.01	0.94	1.09
Female neutered	-0.14	0.05	0.01	0.87	0.78	0.96
Breed						
Labrador Retriever[1]	0.00			1		
German Shepherd	-0.49	0.08	0.00	0.61	0.53	0.71
English Springer Spaniel	-0.32	0.09	0.00	0.72	0.61	0.86
Boxer	0.29	0.09	0.00	1.33	1.11	1.60
Cocker Spaniel	-0.01	0.10	0.95	0.99	0.82	1.21
Jack Russell Terrier	-0.47	0.10	0.00	0.63	0.51	0.77
Doberman	-0.66	0.10	0.00	0.52	0.42	0.63
Crossbred	-0.15	0.06	0.01	0.86	0.76	0.97
Other	-0.42	0.05	0.00	0.66	0.59	0.73
Site of biopsy						
Skin[1a]	0.00			1		
Mouth/pharynx	0.55	0.07	0.00	1.74	1.52	2.00
Mammary gland	1.64	0.07	0.00	5.16	4.50	5.91
Lymphatic system	0.56	0.09	0.00	1.76	1.48	2.08
Liver	-1.63	0.17	0.00	0.20	0.14	0.27
Spleen	-0.33	0.14	0.02	0.72	0.55	0.94
Reproductive system	0.36	0.09	0.00	1.88	1.58	2.25
Intestine	-0.21	0.13	0.01	0.81	0.63	1.04
Other	-0.21	0.07	0.00	0.81	0.70	0.93

[1]referent category

[a]Includes biopsies of epithelial, mesenchymal and melanocytic cell origin

Table 4-4 Coefficients, odds ratios and 95% confidence intervals from the multivariable logistic regression model containing the host-related main effects for factors associated with the risk of neoplasia being diagnosed in a canine biopsy submitted to the CIDRU diagnostic histopathology service.

4.3.3 Site-specific models of host-related variables

4.3.3.1 General

It was possible to create multivariable models for data contained in all but two (liver and spleen) of the eight specific site categories. For the mammary gland category, only biopsy samples from the mammary gland tissue of females were used in the model-building

process, based on the biological knowledge that the risk of a mammary gland abnormality is highly correlated with female gender.

4.3.3.2 Univariable analysis

The results of the univariable screening of independent variables for each of the eight specific site categories showed that age was significantly associated with a histopathological diagnosis in all cases with the exception of biopsies originating from the spleen. Gender was significantly associated with the outcome in the mouth/pharynx and reproductive categories (Table 4-5). Breed was significantly associated with the outcome in four categories (skin, mouth/pharynx, mammary gland, reproductive) (Table 4-6). For two categories, liver and spleen, the significance of breed on the outcome could not be assessed due to sparsity of data within the breed groupings. Specifically, there were no neoplastic biopsies of the liver submitted from an English Springer Spaniel, Boxer or Doberman. No non-neoplastic biopsies of the spleen were submitted from a Boxer or Doberman, no neoplastic splenic biopsies were submitted from a Cocker Spaniel, and a single neoplastic splenic biopsy was submitted from a Doberman.

Tissue origin	Gender	Controls	Cases	P-value	OR	95% CI Lower	95% CI Upper
Mouth/pharynx				**0.01**			
	Male entire[1]	155	475		1		
	Male neutered	16	28		0.57	0.3	1.08
	Female entire	97	297		1	0.75	1.34
	Female neutered	47	76		0.53	0.35	0.79
Reproductive system				**0.01**			
	Male entire[1]	128	402		1		
	Male neutered	3	1		0.11	0.01	1.03
	Female entire	47	141		0.96	0.65	1.4
	Female neutered	7	5		0.23	0.07	0.73

[1]referent category

Table 4-5 Univariable analysis of gender: Significant associations with an outcome of neoplasia being diagnosed in canine biopsies of specific tissue origins submitted to the CIDRU diagnostic histopathology service.

Tissue origin	Breed	Controls	Cases	P-value	OR	95% CI Lower	95% CI Upper
Skin[a]				**<0.00**			
	Labrador Retriever[1]	544	978		1		
	German Shepherd	449	373		0.46	0.39	0.55
	English Springer Spaniel	238	300		0.70	0.57	0.86
	Boxer	223	395		0.99	0.81	1.20
	Cocker Spaniel	160	290		1.01	0.81	1.26
	Jack Russell Terrier	160	189		0.66	0.52	0.83
	Doberman	234	148		0.35	0.28	0.44
	Crossbred	811	1342		0.92	0.80	1.05
Mouth/pharynx				**<0.00**			
	Labrador Retriever[1]	30	125				
	German Shepherd	27	44				
	English Springer Spaniel	12	29				
	Boxer	7	42				
	Cocker Spaniel	7	34		1		
	Jack Russell Terrier	16	18		0.97	0.82	1.15
	Doberman	2	19		1.41	1.32	1.51
	Crossbred	73	227		1.14	1.03	1.25
Mammary gland				**<0.00**			
	Labrador Retriever[1]	27	188		1		
	German Shepherd	13	133		1.47	0.73	2.95
	English Springer Spaniel	15	135		1.29	0.66	2.52
	Boxer	4	38		1.36	0.45	4.13
	Cocker Spaniel	12	116		1.39	0.68	2.85
	Jack Russell Terrier	8	109		1.96	0.86	4.46
	Doberman	15	41		0.39	0.19	0.80
	Crossbred	52	389		1.07	0.65	1.77
Reproductive system				**<0.00**			
	Labrador Retriever[1]	14	37		1		
	German Shepherd	20	32		0.61	0.26	1.39
	English Springer Spaniel	7	26		1.41	0.50	3.96
	Boxer	9	9		0.38	0.12	1.15
	Cocker Spaniel	9	19		0.80	0.29	2.18
	Jack Russell Terrier	2	19		3.59	0.74	17.48
	Doberman	6	9		0.57	0.17	1.89
	Crossbred	23	128		2.11	0.99	4.50

[1]referent category
[a]includes biopsies of epithelial, mesenchymal and melanocytic cell origin

Table 4-6 Univariable analysis of breed: Significant associations with an outcome of neoplasia being diagnosed in canine biopsies of specific tissue origins submitted to the CIDRU diagnostic histopathology service.

4.3.3.3 Multivariable analysis

Results of the site-specific models of the host-related variables are presented in Table 4-7, showing the direction of the differences in odds ratios for a biopsy submitted from a certain breed from a particular site, relative to the Labrador Retriever breed. In particular, highest odds ratios (OR ≥ 2) for an outcome of neoplasia were seen for biopsies of oropharyngeal origin from Dobermans (OR=2.59, 95%CI 0.56, 12.04), biopsies of lymphatic origin from Boxers (OR=2.19, 95%CI 0.55, 8.78) and biopsies of reproductive origin from Jack Russell terriers (OR=2.12; 95%CI 0.41, 10.86) and crossbreed dogs (OR=2.34; 95%CI 1.02, 5.37). The lowest odds ratios (OR ≤ 0.3) for an outcome of neoplasia were seen for biopsies of oropharyngeal origin from Jack Russell terriers (OR=0.28; 95%CI 0.12, 0.62) and intestinal biopsies from German Shepherd dogs (OR=0.20; 95% CI 0.06, 0.74).

Breed [1]	German Shepherd	English Springer Spaniel	Boxer	Cocker Spaniel	Jack Russell Terrier	Doberman	Crossbred
Site of biopsy							
Skin[a]	↓	-	↑	-	↓	↓	-
Mouth/pharynx	↓	↓	↑	-	↓↓	↑↑	↓
Mammary gland	↑	-	↑	↑	↑	↓	-
Lymphatic	↑	↓	↑↑	↓	-	↑	-
Reproductive system	-	↑	↓	-	^^	-	↑↑
Intestine	↓↓	↓	↓	↓	↓	↓	↓

[1] Referent breed is Labrador Retriever
[a] Includes biopsies of epithelial, mesenchymal and melanocytic cell origin

↑↑	High (≥2.0)
↑	Increased (≥1.3 and <2.0)
-	No change
↓	Decreased (≤0.7 and >0.3)
↓↓	Low (≤0.3)

Table 4-7 Comparison of odds ratios for site-specific analyses of canine biopsies submitted to the CIDRU diagnostic histopathology service.

4.3.4 Statistical analysis with inclusion of practice variable

4.3.4.1 Multivariable analysis with practice as a fixed-effect

The main-effects model of the host-related variables for the reduced dataset described in Section 4.2.2.3 is presented in Table 4-8. As in the full main-effects model, increasing age was shown to be associated with an increased odds of neoplasia in a biopsy, when controlled for gender and breed of dog and site of biopsy. The directions of odds ratios for neoplasia in biopsies from specific genders and breeds of dog, and sites of biopsy, were identical to those shown in the full main-effects model (Table 4-4).

Variable	Coefficient	Std deviation	P-value	OR	95%CI Lower	95%CI Upper
Constant	-0.20					
Age	0.13	0.01	0.00	1.13	1.12	1.15
Gender						
Male entire[1]	0.00			1		
Male neutered	-0.10	0.09	0.30	0.91	0.75	1.09
Female entire	0.05	0.04	0.20	1.05	0.97	1.14
Female neutered	-0.16	0.06	0.01	0.86	0.77	0.95
Breed						
Labrador Retriever[1]	0.00			1		
German Shepherd	-0.43	0.08	0.00	0.65	0.55	0.76
English Springer Spaniel	-0.32	0.09	0.00	0.73	0.61	0.88
Boxer	0.34	0.10	0.00	1.4	1.15	1.71
Cocker Spaniel	-0.04	0.10	0.71	0.96	0.78	1.18
Jack Russell Terrier	-0.44	0.11	0.00	0.64	0.52	0.79
Doberman	-0.63	0.11	0.00	0.53	0.43	0.67
Crossbred	-0.18	0.07	0.01	0.84	0.74	0.95
Other	-0.38	0.06	0.00	0.69	0.61	0.77
Site of biopsy						
Skin[1a]	0.00			1		
Mouth/pharynx	0.36	0.07	0.00	1.43	1.24	1.65
Mammary gland	1.42	0.07	0.00	4.14	3.6	4.76
Lymphatic system	0.32	0.09	0.00	1.38	1.16	1.64
Liver	-1.79	0.17	0.00	0.17	0.12	0.23
Spleen	-0.54	0.14	0.00	0.58	0.44	0.77
Reproductive system	0.45	0.09	0.00	1.57	1.31	1.88
Intestine	-0.40	0.13	0.00	0.67	0.52	0.87
Other	-0.39	0.07	0.00	0.67	0.58	0.78

[1]referent category
[a]Includes biopsies of epithelial, mesenchymal and melanocytic cell origin

Table 4-8 Coefficients, odds ratios and 95% confidence intervals from the multivariable logistic regression model containing the host-related effects for factors associated with the risk of neoplasia being diagnosed from a canine biopsy submitted to the CIDRU diagnostic histopathology service (reduced dataset).

Practice was then added as a fixed-effect to the model. The statistics associated with the covariates of prime interest, being age, gender and breed of dog, and site of biopsy are presented in Table 4-9. The full model, including statistics associated with practice, is presented in Appendix 4. The addition of practice as a fixed-effect caused a significant reduction in model deviance (P = 0.001), indicating that a degree of EBV was accounted for by the practice term. Detailed examination of outliers and their effect upon goodness-of-fit of the model was not conducted because this was not considered necessary for the purpose of the analysis. Examination of the coefficients and 95% CIs in the model

including practice as a fixed-effect revealed similar results for the host-related covariates as were apparent in the reduced dataset main-effects model (Table 4-8).

Variable	Coefficient	Std deviation	P-value	OR	95%CI Lower	95%CI Upper
Constant	-0.19					
Age	0.12	0.01	0.00	1.13	1.12	1.14
Gender						
Male entire[1]	0.00			1		
Male neutered	-0.17	0.1	0.09	0.85	0.70	1.03
Female entire	0.06	0.04	0.15	1.06	0.98	1.15
Female neutered	-0.18	0.06	0.00	0.83	0.74	0.93
Breed						
Labrador Retriever[1]	0.00			1		
German Shepherd	-0.5	0.08	0.00	0.63	0.53	0.74
English Springer Spaniel	-0.40	0.1	0.00	0.67	0.55	0.81
Boxer	0.26	0.1	0.01	1.3	1.06	1.59
Cocker Spaniel	-0.09	0.11	0.39	0.91	0.74	1.12
Jack Russell Terrier	-0.43	0.11	0.00	0.65	0.52	0.81
Doberman	-0.67	0.11	0.00	0.51	0.41	0.64
Crossbred	-0.26	0.07	0.00	0.77	0.67	0.88
Other	-0.40	0.06	0.00	0.67	0.60	0.75
Site of biopsy						
Skin[1a]	0.00			1		
Mouth/pharynx	0.31	0.07	0.00	1.36	1.18	1.57
Mammary gland	1.35	0.07	0.00	3.86	3.35	4.46
Lymphatic system	0.36	0.09	0.00	1.4	1.17	1.67
Liver	-1.85	0.17	0.00	0.16	0.11	0.22
Spleen	-0.61	0.14	0.00	0.55	0.41	0.72
Reproductive system	0.36	0.09	0.00	1.43	1.19	1.73
Intestine	-0.46	0.13	0.00	0.63	0.48	0.82
Other	-0.42	0.08	0.00	0.66	0.56	0.76

[1]referent category
[a]Includes biopsies of epithelial, mesenchymal and melanocytic cell origin

Table 4-9 Coefficients, odds ratios and 95% confidence intervals from the multivariable logistic regression model containing the host-related effects and practice as a fixed-effect for factors associated with the risk of neoplasia being diagnosed from a canine biopsy submitted to the CIDRU diagnostic histopathology service (reduced dataset - individual practice results shown in Appendix 4).

4.3.4.2 Multivariable analysis with practice as a random-effect

Based upon the *a priori* knowledge that biopsies were grouped, or clustered by practice, a logistic-binomial regression model was used to fit a random-effects term for practice to the reduced dataset main-effects model (Table 4-10). Analysis of the data by covariate and

practice did not prove there were distinguishable patterns of biopsy submission between individual practices, therefore the model was built assuming data were indistinguishable. Inclusion of the random-effects term caused an unexpected large increase in model deviance, indicating that the term caused a poorer model fit. Because of this result, further comparisons of the effect of the random-effects term upon the main-effects covariates between the main-effects model and that including the random-effects term were not made.

Variable	Coefficient	Std deviation	P-value	OR	95%CI Lower	95%CI Upper
Constant	-6.28	0.09	<0.01	<0.01	0.00	0.00
Age	0.15	0.00	<0.01	1.16	1.15	1.18
Gender						
Male entire[1]						
Male neutered	-0.61	0.13	<0.01	0.54	0.42	0.71
Female	0.09	0.06	0.1	1.1	0.98	1.22
Female neutered	-0.64	0.08	<0.01	0.53	0.45	0.62
Breed						
Labrador Retriever[1]			<0.01			
German Shepherd	-0.60	0.11	0.02	0.55	0.44	0.68
English Springer Spaniel	-0.27	0.11	0.26	0.77	0.61	0.96
Boxer	-0.14	0.12	0.08	0.87	0.68	1.11
Cocker Spaniel	-0.23	0.13	<0.01	0.79	0.61	1.03
Jack Russell Terrier	-0.56	0.14	<0.01	0.57	0.43	0.75
Doberman	-0.55	0.17	<0.01	0.57	0.41	0.8
Crossbred	-0.34	0.08	<0.01	0.71	0.61	0.84
Other	-0.33	0.07	<0.01	0.72	0.62	0.83
Site of biopsy						
Skin[1a]						
Mouth/pharynx	-0.17	0.09	0.06	0.84	0.7	1.01
Mammary gland	0.52	0.07	<0.01	1.69	1.48	1.92
Lymphatic system	-0.38	0.12	<0.01	0.68	0.54	0.86
Liver	-0.94	0.28	<0.01	0.39	0.22	0.68
Spleen	-0.28	0.22	0.2	0.75	0.49	1.16
Reproductive system	0.68	0.11	<0.01	1.98	1.58	2.47
Intestine	-0.7	0.22	<0.01	0.5	0.32	0.76
Other	0.06	0.11	0.62	1.06	0.85	1.32
Random term						
%SCL	2.26	0.03				

[1]referent category

[a]Includes biopsies of epithelial, mesenchymal and melanocytic cell origin

Table 4-10 Coefficients, odds ratios and 95% confidence intervals from the multivariable logistic regression model containing the host-related main effects and practice as a random-effect for factors associated with the risk of neoplasia being diagnosed from a canine biopsy submitted to the CIDRU diagnostic histopathology service (reduced dataset).

4.4 DISCUSSION

Although the primary objective of this study was to assess host-related risk factors for canine neoplasia, it was important to recognise and take account of the influence of the data source for the study, being first-opinion practices that were not approached prior to data collection.

Results from the host-related effects model (Table 4-4) showed that age, the Boxer breed and four of the selected sites, *viz* mouth/pharynx, mammary gland, lymphatic system, and the reproductive system, were significantly associated with an increase in the odds of a biopsy sample having a neoplastic diagnosis. An increase in the odds of neoplasia with increasing age is in agreement with other studies (Priester and Mantel, 1971; Moulton, 1990; Dobson *et al.*, 2002). This finding was also observed in the site-specific analyses except for biopsies originating from the spleen. For this site, univariable analysis showed no significant difference between the mean ages of dogs from which non-neoplastic and neoplastic biopsies were submitted, 9.2 and 9.8 years, respectively. This is in agreement with a previous study of splenic disease, which showed that the most common non-neoplastic and neoplastic lesions of the spleen were found at virtually the same mean age in dogs, 10.5 and 10.4 years, respectively (Spangler and Culbertson, 1992).

In this study it was elected to fit gender using four categories in common with some previous authors (Bender, Dorn and Schneider, 1984; Cook *et al.*, 1993). It can be argued that a more valid approach is to fit gender and neutering status as two effects together with the interaction term gender x neutering status. However, from a biological point of view, the interpretation of interaction terms is non trivial and the four gender classes adopted does reflect the parlance and biological entity dealt with by clinicians and may be more readily interpreted.

Given that the referent category for gender in the host-related effects model was male entire, there was a marginal decrease in the odds of neoplasia in samples submitted from neutered females, when controlling for age, breed and site. The site-specific models showed a decrease in the odds of neoplasia in biopsies submitted from the mouth/pharynx and reproductive systems in neutered dogs of both sexes. The most common site of

neoplasia in the male reproductive system is the testis, and in the female, vaginal and vulvar neoplasms are the second most common reproductive tumours after the mammary gland. Subjective data indicate that the most common of these tumours in the female, the leiomyoma, may be hormone dependent (Klein, 1996). Thus, it is unsurprising that a decreased odds of neoplasia was found in biopsies from the reproductive system. The reason for the decreased odds of neoplasia in biopsies from the mouth/pharynx of neutered animals is unclear. A male predisposition for certain oral neoplasms has been previously reported (Dorn and Priester, 1976a) although no comment was made regarding neutering status of the dogs in the study.

Only biopsies from female dogs were considered in the mammary gland specific analyses. This study found no significant difference in the odds ratio for a diagnosis of neoplasia in mammary gland samples from entire or neutered females. Although spaying is known to have a protective effect against mammary cancer, this effect is present only if the dog is younger than 2.5 years at the time of spaying (Dorn *et al.*, 1968b; Schneider, Dorn and Taylor, 1969). Age at time of spay was not known for the neutered females in this study.

There were significant differences in the breed odds ratios, both in the host-related effects model (Table 4-4) and the site-specific models (Table 4-7). In the first model, a biopsy from the Boxer breed was shown to have an odds ratio 1.3 times that of the Labrador Retriever, the referent breed. All other breeds selected for more detailed analysis showed no or decreased odds of neoplasia in a biopsy, compared to the risk of neoplasia in a biopsy from a Labrador Retriever. Various authors have shown the Boxer breed to have an increased risk of neoplasia (Howard and Neilson, 1965; Priester and Mantel, 1971; Nordstoga *et al.*, 1997; Arnesen *et al.*, 2001) and the breed has been suggested as a good candidate for further genetic study (Misdorp, 1996). Interestingly, this study found the Boxer to have a decrease in odds of neoplasia in biopsies of the reproductive system and intestine (Table 4-7). There were also no neoplastic biopsies of the liver and only two neoplastic biopsies of the spleen submitted from Boxers in this dataset.

The study by Dorn *et al.* (1968b) identified the German Shepherd breed as having an increased risk of neoplasia of the mouth and pharynx, compared to other purebreeds examined in the study. The German Shepherd was found to have a decreased odds ratio,

with the Doberman breed having the highest odds of neoplasia in this site, when compared to the Labrador. An important factor contributing to this apparent disagreement in breed susceptibility to oral-pharyngeal neoplasia may be that only malignant neoplasms were included in the California counties survey, whereas both benign and malignant tumours were eligible as cases in this multivariable analysis. Oral-pharyngeal biopsies from the Jack Russell Terrier breed also had a significantly lowered odds of neoplasia in our study. This breed has not been singled out by previous authors for detailed study, so no comparisons of findings were available.

The skin was chosen as the referent site for the logistic regression analyses and biopsies of epithelial, mesenchymal and melanocytic cell origin were included in this category. In the final model (Table 4-4), biopsies from the mammary gland were nearly four times more likely to be neoplastic than those from the skin, and there were significant increases in odds ratios for neoplasia in biopsies from the oropharynx, lymphatic system and reproductive system, when controlled for age, gender and breed. Over 80% of biopsies included in the study population originated from the skin or mammary gland (Table 4-1), which is likely to reflect the relative ease of detection of abnormalities and subsequent sampling of tissues from these superficial sites. The high proportion of biopsies submitted from these sites was similar to datasets presented by Dorn *et al.* (1968b), MacVean *et al.* (1978) and Nordstoga *et al.* (1997).

As well as methodological differences between this study and others examining host risk factors for canine neoplasia, this study utilized a coding scheme for classification of primary site of neoplasms developed by members of the Department of Veterinary Pathology at GUVS (Chapter 2, Section 2.2.4.1). Most epidemiological studies in the field of canine neoplasia have used the SNVDO (Chapter 1, Section 1.6.2) as the standard coding system. The California counties study used both the SNVDO and the ICD (Chapter 1, Section 1.6.2). Without a global standard, it is likely that there will be differences in the classification of tumour sites which prevents accurate direct comparison between studies. Unlike the two coding systems mentioned, the coding scheme for the GUVS database also does not include a specific coding for whether a tumour is benign or malignant, thus the ratio of benign to malignant neoplasms has not been derived for the current study.

Statistical instability resulted when the interaction of site and breed was fitted to the model, due to heterogeneous distribution of covariate patterns within the covariate space. For this reason, in the current study, biologically meaningful interaction terms among the independent variables were not assessed in any of the models. This finding is an example of the "sparse data problem" (Rothman and Greenland, 1998). Despite the large size of the dataset, there were still categories, such as biopsies submitted from a certain breed and site, which contained too few samples to allow meaningful analysis. It was this same problem that prevented multivariable logistic regression analyses for biopsies stratified by the site categories liver and spleen. Strategies for handling empty cells in any contingency table include collapsing the categories of the independent variable in some sensible fashion to eliminate the zero cell or eliminating the category completely (Hosmer and Lemeshow, 1989). Since biopsies from the liver and spleen of dogs are biologically unique for only one site category, being the organ of origin, collapsing the categories was not feasible on biological grounds. Elimination of these categories was also considered to be an inappropriate strategy because biopsies from these organs were present in sufficient numbers to warrant their inclusion as specific sites for investigation.

Although there were instances where one or more of the independent variables were shown to be insignificantly associated with a histopathological diagnosis of neoplasia in a particular site category, all independent variables were included in the model-building process for the six sites where multivariable analyses were possible. This decision was based on biological, rather than statistical, significance of the independent variables.

Logistic regression model development including only the host-related variables assumed that each individual occurrence or biopsy from a particular practice, was independent and identically distributed. However, this assumption was unlikely to be valid because of the many factors related to the procurement of a biopsy. The decision to include practice as a fixed-effect was based upon the *a priori* knowledge that biopsies were aggregated by practice, so extra-binomial variation and possibly intra-practice correlation were likely be present in the data. The decision to exclude practices contributing low numbers of biopsies was to preempt preliminary concerns of practice-related analyses being invalidated due to the "sparse data problem", if data from all practices were considered. Fitting a group parameter (for aggregated data) as a fixed-effect has been performed in studies of

production animal epidemiology where group effects due to herd, flock or pen factors needed to be included in the risk factor analysis (Dohoo and Martin, 1984; Curtis *et al.*, 1988; Mousing *et al.*, 1990). The limitations associated with fixed-effect models include instability when large numbers of groups are present (Mousing *et al.*, 1990). However, this was not apparent when practice was included as a fixed-effect, which may be a reflection of the large size of this dataset. Interestingly, there was also little variation in the odds ratios for neoplasia with the inclusion of the practice variable. Another limitation of fixed-effect models is their inability to account for all levels of clustering that may be present within a dataset. The choice of methods to account for clustering in studies of animal populations is dependent on considerations such as the objectives of the study and the assumptions that can be made about the nature of the correlation structure (McDermott, Schukken and Shoukri, 1994). Including practice as a random-effects term was considered to be a more appropriate method of accounting for EBV in the data than considering practice as a fixed effect, because many practices were involved. The deterioration in model fit with the introduction of a random-effects term in the logistic-binomial model for indistinguishable data, was an unexpected result given that the biopsy population originated from discrete practices. Following strict definition guidelines, the data would be considered distinguishable, because the covariates of age, gender and breed were quantifiable. However, it was not possible to prove there were distinct patterns of biopsy submission from individual practices based upon the age, gender or breed of dog, or site of biopsy, so a valid argument for considering the data indistinguishable was made using the assumption that individual practices will see similar canine populations, thus covariates would not vary between practices.

Rapid advances in computer software development in the past decade now allow for detailed multi-level analyses of data which may have many levels of aggregation. This is a consideration for this dataset, and may be an approach which could assess risk factors for neoplasia not only at the biopsy level, but also at the animal and practice level.

For this study, the selection bias associated with data derived from dogs seen at referral establishments, such as the data from the VMDB, was removed by using records derived from dogs attending first opinion veterinary practices as the source of the biopsy population. Studies based on VMDB data have the advantage of being referable to a dog

population because they may use the population of dogs attending the referral establishments as the population at risk. These studies also tend to be biased towards cancers that are to some extent treatable (Kelsey, Moore and Glickman, 1998). Although referral bias may have been removed, by selecting histopathology data source, the results may only be reported with reference to a biopsy population, rather than the general dog population, given that detailed information of the general dog population from which the biopsies were obtained is unavailable.

The use of this first opinion source has many other associated biases, such as the influence of an individual veterinary surgeon's approach to abnormalities detected by the owner or on clinical examination. The veterinary advice received is of paramount importance in the owner's decision regarding management of their pet. Owners may be more ready to investigate abnormalities detected in the younger animal, and highly suspicious lesions may not be investigated in the older animal due to the owner's unwillingness to subject their animal to a surgical procedure, particularly if general anaesthesia is required. It is the interaction between the veterinary surgeon and owner that will lead to biopsy in these cases. A survey of veterinarian's attitudes and choice of treatment for selected neoplasms has been carried out in Australia (Peaston and Watson, 1995). However, the veterinarians involved in the survey were not asked specifically about the decision-making process that led to the diagnosis of the cancers on which comments regarding treatment were assessed. A survey of this aspect of the collection of data pertaining to neoplasia is warranted. At the owner level, permission for histopathology examination of a biopsy may not be granted due to financial constraints. One major distinction between this study and other large-scale studies of canine neoplasia, such as the CANR, Tulsa Registry of canine and feline neoplasia and the NCCR, is that all histopathology costs were borne by the owner or practice, and not by the researchers. The level of impact this situation had on the number of biopsies submitted for histopathological examination to the CIDRU database is likely to have been significant but remains extremely difficult to quantify.

There may also be bias associated with the pathologist and histopathology laboratory. Veterinary pathologists play a critical role in the treatment of neoplasia in the dog by providing accurate diagnostic information to veterinarians so that a prognosis can be determined and adequate treatment provided (Powers, 1996). Therefore, the veterinarian

requiring a histological determination must collect and correctly submit a representative sample of the lesion to the pathologist, who in turn, has the responsibilities of providing a rapid response, a histological diagnosis and any further information important for determining clinical treatment and prognosis (Madewell, 1987; Goldschmidt, 1993). Additionally, the interaction between the veterinarian and pathologist may influence the choice of biopsies for submission. Although veterinary histopathology laboratories provide vast amounts of data pertaining to canine neoplasia, biopsy populations obtained from this source may not accurately reflect the distribution of cancers present in the general population of dogs because of biases associated with the owner, the consultant veterinarian, and the pathology laboratory itself.

In conclusion, a multivariable approach has been applied to a dataset derived from a population of canine biopsies submitted by first-opinion and charity veterinary practices. Although mass screening methods for early detection of neoplasia are unlikely ever to be applied to animal populations, further knowledge of host risk factors for neoplasia in certain sites may improve the probability for early diagnosis (Madewell, 1987). The findings of the host-related risk factor analysis confirm and quantify many of the perceived risks for neoplasia based on previous clinical descriptive and univariable investigations. The effect of grouping by practice was explored by entering practice as a fixed-effect and random-effect in separate models of the host-related variables, to account for extra-binomial variation introduced by the clustered nature of the data. Although considered to be the most appropriate method for dealing with EBV and intra-group correlation, logistic-binomial regression analysis was unrewarding for this dataset, which may reflect its covariate distribution. The study raises questions relating to the process by which a canine biopsy population is accumulated, and exposes the need for detailed investigation of practice-related factors pertaining to the submission of biopsy samples to a diagnostic histopathology service. The results may also ultimately provide a means for prioritising biopsy examination protocols in busy histopathology laboratories.

CHAPTER 5

INTERROGATION AND STATISTICAL ANALYSIS OF A COMMERCIAL VETERINARY HISTOPATHOLOGY DATABASE

5.1 INTRODUCTION

There are many computerised sources of canine neoplasia data held in the UK. The histopathology databases operated by veterinary academic institutions and private veterinary diagnostic laboratories can be considered as secondary sources, because their primary role is to provide a diagnostic service for veterinary practitioners requiring histopathological diagnoses and prognostic information on tissue samples submitted from animals under their care. Other secondary sources of canine neoplasia data include pet insurance company databases. Epidemiological analyses of such a database have been performed (Wood *et al.*, 2000a; Dobson *et al.*, 2002). However, a criticism of these analyses is that diagnostic information within the database under study was accepted as that provided on insurance claim forms, so histopathological confirmation of neoplastic diagnoses was not verified.

The structure and function of the CIDRU histopathology database and epidemiological analysis of canine biopsy data held within that database have been presented in Chapters 2, 3 and 4. The current chapter explores the histopathology database held by a commercial veterinary diagnostic laboratory service. The laboratory provides a comprehensive ranges of services required by the general veterinary practitioner, including gross pathology, histopathology, cytology, microbiology, haematology, biochemistry, and endocrinology.

Prior to this project, data within the laboratory's database have not been utilised for an epidemiological study of canine neoplasia, because the priorities of the laboratory lie with its commitment to high-quality diagnostic work and providing a comprehensive and accessible service for the veterinarian. Great emphasis is placed upon rapid turnaround time for sample processing and a high level of communication with their clients. Although the database is a vast resource of information, data interrogation is limited primarily to that

involved with administration of the laboratory's services, to ensure that client expectations and needs are fulfilled. The importance of smooth running of the database to the business is paramount, and a dedicated database administrator is employed to manage all computing systems on the premises.

This chapter describes the hardware and software of the laboratory database and the functioning of the histopathology service, from sample submission to data entry, report generation and delivery of results to the veterinarian. A statistical analysis of a subset of the data is presented and compared to the results obtained from the CIDRU database. The unique aspects of this database system, and its strengths and weaknesses for use in large-scale epidemiological studies are also discussed.

5.2 MATERIALS AND METHODS

5.2.1 The laboratory pathology record system

The designers of this laboratory's database system were concerned with the general issues of efficient use of computer memory and rapid retrieval of data from storage. However of prime importance were the reporting and accounting capabilities of the system. The file structure has been developed to handle vast numbers of samples involving many disciplines, and to allow rapid dissemination of results to referring veterinarians. The goals of optimum speed and accuracy in the recording of practice data and sample details have led to the implementation of more sophisticated data capture systems in some data handling areas during the lifetime of the laboratory.

5.2.1.1 File structure

The file structure is based upon a proprietary database format, created by William Woodard Associates and is known as the WinPath LIMS (Laboratory Information Management System). The primary goal of the designers of the original version of the database, created in the 1970s, was efficient data storage, because at that time computer hard drives were of extremely small size by today's standards. All data for this analysis

were collected by the laboratory when running WinPath database version 5 and it is this file structure that is described.

Data are stored in a series of individual databases and program files are used to link data contained in one database to another for validity purposes. Each database has a fixed structure as produced by the original designers. Within the structure, there are specific databases where the contents may be modified, but not the data fields - fields cannot be added, deleted or modified in any way. Within other databases, field use may be modified, and field combination is possible to make longer fields. Fields may be activated or deactivated by the database manager as required. Individual records are created by data entry into the fields of each relevant database. The histopathology service uses the system databases storing veterinary practice data, patient data and laboratory test data. Facsimile data and accounts data are not held in a separate database, but are stored as individual files on a central directory.

Each record in the veterinary practice database is 512 bytes in length. Within that record, information is stored in a linear format, in fixed positions along the 512 bytes. All fields are used in this database and no modifications to the fields are possible. As well as practice name and address details, the stored information in each record includes a unique identifying practice code, which is used by other databases in the system.

The database of particular importance to this study is the patient database, or "patient data directory". Records are stored in blocks of 1000 within laboratory number-specific files, which reside within the directory. Each record can be a maximum of 8000 bytes, with information stored in fixed positions along the record's length. Within this database's structure are many data fields, created and named by the database designers to be suitable for use by medical laboratories working in the National Health Service (NHS) in the UK. Fields have been selected from the original framework by the veterinary laboratory staff to allow fulfillment of its specific requirements for a "patient" record. Each record actually refers to a sample, rather than the individual from which a sample was taken. Of the approximately 120 fields available, 90% are currently in use. Some fields have been used for an alternative purpose from that specified (by name) by the database designers to accommodate laboratory-specific information.

The first 32 bytes of each record contain key data, which is used as an indexing system by system-specific search programs to link the record's information to other databases in the system. The use of key data allows the search programs to rapidly locate data, without the need to fully decompress records. Key data includes: date the sample was received; record status; pathologist code and practice code. The key data in each record is copied to a KYS file, which is used by the search programs. The next 33 to 63 bytes of each record contain overflow "pointers" with these data being copied to an OVF (overflow) file (see below). These sections of the record remain uncompressed so the information can be accessed by the search program files needed to verify information within the record that is common to other databases, e.g., practice code data in the practice database.

The remainder of each record length is subdivided into sections for the storage of demographic data, test data and results. Data is stored in fixed positions in 1024 byte blocks. If a record is greater than 1024 bytes in length, the first 1024 bytes is placed into a DAT (data) file for that record and the remainder is placed into OVF (overflow) files in 512 byte blocks. Pointers to the positions of the OVF files are placed at position 33 to 63 bytes of the record to ensure all information pertaining to that record can be retrieved. Finally, 16 bytes of each record contains information reflecting the status of the record in each of 16 possible laboratory sections involved in the record (e.g. histopathology, cytology, haematology), i.e., whether the record is incomplete, saved, authorised or printed. These data are similarly copied into a separate file, the IPL file. All but the first 80 bytes of each record can be compressed. This linear file structure with fixed positions for specific data along the length of each record and using data compression programming ensures that hard disk space is used extremely efficiently. However, this type of database structure is far more complex than the relational database structures, such as the CIDRU database (Chapter 2), which are commonly in use today.

The compressed section of each record contains information for sample identification: labref, date sample received, date sample reported, owner surname, animal name, gender, age, sample type and veterinary practice code. Other information includes: personnel responsible for data entry, fee +/- fee adjustment, test format code, test status and personnel responsible for sample handling and reporting. There are fields available for entry of the clinical history accompanying the sample. However, the primary use of these

fields by the administration staff is to notify the veterinary pathologists of particular enquiries, or requirements for the sample, made by the submitting veterinarian. There are fixed position fields in each record for the storage of results information. For histology, these fields are: diagnosis, histology report, comment and prognosis.

The laboratory test database contains details on all tests and procedures offered by the laboratory, such as sample requirements, turnaround time and costs. Each test has a unique test format code and it is this code which is used by programs in the system to validate the test information specified for each sample in its record held in the patient data directory.

When each record is completed and authorised, the reporting program refers to the client database to see which output mechanism is to be used (hard copy, facsimile or email) to deliver the report to the veterinary practice. If facsimile or email is required, it creates a flat ASCII file of the report. This version is then submitted to the FaXServe5 system of automated facsimile delivery, via the FaxServe5 interface program, for final delivery of the report.

5.2.1.2 Software

The William Woodard Associates LIMS used by the laboratory is DOS-based, running under various DOS versions and within DOS sessions. It is viewed via the operating systems in use on the laboratory staff's personal computers (PCs), being Microsoft Windows 95, 98, 2000 or Me (Microsoft Corporation). All PCs in the laboratory are networked to allow simultaneous access by licensed users to the system. Permission to access the database system is granted by the database administrator following approval by the Director of the laboratory.

The software suite includes a program that enables staff to perform simple searches of the contents of the system. Searching is focused on data stored in the KYS files (which act as indexes), which then allows the remainder of the record to be retrieved, if matches are found. This search function is set up to enable staff to answer telephone queries from clients on specific samples, and is unsuitable for extensive data retrieval. Specific

programs must be written by the database administrator to query and extract large amounts of data from the database system, such as was needed for this study.

5.2.1.3 Hardware

A client server network operating system is in use, with staff PCs being networked to the core component of the laboratory's computer network, a dedicated file server (Novell Netware 3.12, upgraded to 3.2 for year 2000 compliance) maintained on the laboratory premises. Other Novell Netware servers in the system include the server dedicated to providing the automatic facsimile transmission of laboratory reports and a server located off-site, linked by fibre-optic cable to the rest of the system. This off-site server is an integral component of the data back-up system.

5.2.1.4 Back-up

There are four hardware components to the data back-up system managed by the database administrator: a file server off-site which is a copy of the core file server, a PC on-site, and a notebook computer which is taken to another off-site location at the end of each working day. Prior to the commencement of each working day, a program creates a zipped version of all data on the core file server and copies it to the PC and to the copy file server. Every 30 minutes throughout the day, all changed data from that day are copied from the core file server to the PC and copy server; every 60 minutes, an additional zipped file of all changed data is made and copied to the PC and copy server. Once daily, all changed data from the core file server is copied to the notebook computer, which is taken to a remote location overnight. Once weekly, a program creates a zipped version of all data on the core file server and copies it to the PC, to act as a baseline for all subsequent hourly zipped files of changed data. If catastrophic core server failure was to occur, maximum data loss would be the changed data from the 30 minute time period prior to the server's failure.

5.2.2 Operation of the histopathology service

The laboratory's histopathology service is the responsibility of a team of experienced veterinary pathologists. From an initial single pathologist, there are now six full-time or

part-time veterinary histopathologists, and seven full-time or part-time technical staff. Technical staff are also responsible for the preparation of cytology specimens. The client services staff and administration staff provide support to all laboratory departments, and play a key role in the functioning of the histopathology service, by regulating the interface between veterinary practitioners and the pathologists.

The procurement of samples for submission to the laboratory's histopathology service is influenced by the same factors as described in Chapter 2, Section 2.3. Additionally, the laboratory price structure for histopathology allows the examination of up to four separate tissue types for a fixed price. This may encourage veterinarians to provide all relevant tissues for the thorough investigation of a disease process. For example, investigation of clinical signs emanating from intestinal dysfunction often requires multiple samples of the intestinal tract to be examined to reach a diagnosis. Having a price structure that encourages multiple submissions also allows assessment of multiple lesions from the same animal. This is advantageous in many ways: the animal may have many lesions sampled under one general anaesthetic; it is cost-effective to the owner regarding anaesthesia, surgery and histology expenses; and veterinarians are able to encourage owners to have, for example, multiple skin or mammary lesions from their pet assessed histologically, rather than having to select a lesion based only on gross appearance and taking the risk of not diagnosing a malignancy.

The laboratory has a specific submission form for tissue samples to encourage the provision of clinical information relevant to the sample(s) by the veterinarian (Appendix 5). The laboratory also supplies practice-specific barcode adhesive labels free of charge to the practice, to be placed on all submission forms. Barcode technology is utilized widely in the laboratory for sample identification, from data entry through to sample processing in the laboratory.

When a practice contacts the laboratory with the intention of submitting samples for the first time, a starter kit is posted to the practice with information regarding sample submission requirements, price lists, and appropriate containers for the transport of pathology samples. For histopathology, a variety of sizes of screw-top cylindrical plastic containers with a pre-measured quantity of 10% formalin, pre-paid postage envelopes and

submission forms are supplied free of charge. The supply of these items is continued (free of charge) following confirmation that the practice is using the service and supplies. The sizes of the formalin pots are designed to encourage submission of appropriately fixed specimens for handling by the histopathology laboratory staff. Within the starter pack, guidelines are given regarding the sample size suitable for the quantity of formalin supplied in each sample pot, in order to ensure adequate tissue fixation (Chapter 1, Section 1.5.2).

At the laboratory, smooth progression of samples from reception to the processing laboratory, to the pathologist and ultimately, production and distribution of a report to the submitting practitioner is achieved by a highly regulated sample flow system. An outline of these procedures is given below.

5.2.2.1 Sample reception

The vast majority of submissions to the laboratory (greater than 98%) are received by post or courier. The post is opened and inspected by a trained team of personnel who are responsible for the identification of received samples, their correlation with the accompanying submission form and the labelling of all contents in each delivery. For histopathology samples, each submission form is given a unique identification number (*labref*), and all tissues corresponding to the case described on the submission form are given the same identification number.

5.2.2.2 Sample preparation procedure

Samples for histopathology and their submission forms are delivered to the histopathology laboratory, where the samples are prepared for processing in the light of information supplied on the request from. The technical staff inspect the samples to ensure there is adequate tissue for processing, that the tissue is properly fixed and decide whether additional treatments, e.g., decalcification of bone, are required prior to routine processing. The majority of tissues are trimmed by the technical staff to fit into processing cassettes and processed into paraffin blocks. For unusual samples, or where technicians may be unsure, the tissues are trimmed by or under the supervision of a pathologist. Multiple tissue

types may be placed into a single paraffin block which is labelled with the *labref.* If multiple blocks are required, each block is labelled sequentially with the *labref* and number of the block, for example two blocks for *labref* 123456 are labelled 123456/1 and 123456/2. The number of blocks, their tissue content and any information regarding sample quality is recorded on the submission form by the technical staff. The forms are then separated from the samples and delivered to the administration staff for data entry into the computer system.

5.2.2.3 Data entry

The administration staff initiate the computerised pathology record for each submission on receipt of the original submission forms. Details are transcribed from the form to the first and second data entry screens. The first data entry screen (Figure 5-1) contains fields for demographic data. Barcode scanning technology is used to enter the *labref* and practice code, while manual keyboard entry is required for all other fields on this screen. The second data entry screen (Figure 5-2) is used to record which tests have been requested on the submission form for the sample(s). Test code formats are used and can be found by using the simple search facility to access this information held in the laboratory test database (Section 5.2.1.1). Once data entry for histopathology requests is complete, the administration staff generate a daily paper-based worklist, listing all histopathology sample *labref*s, owner and animal name and the date the samples were booked in. This list, together with the submission forms, is then delivered to the histopathology laboratory.

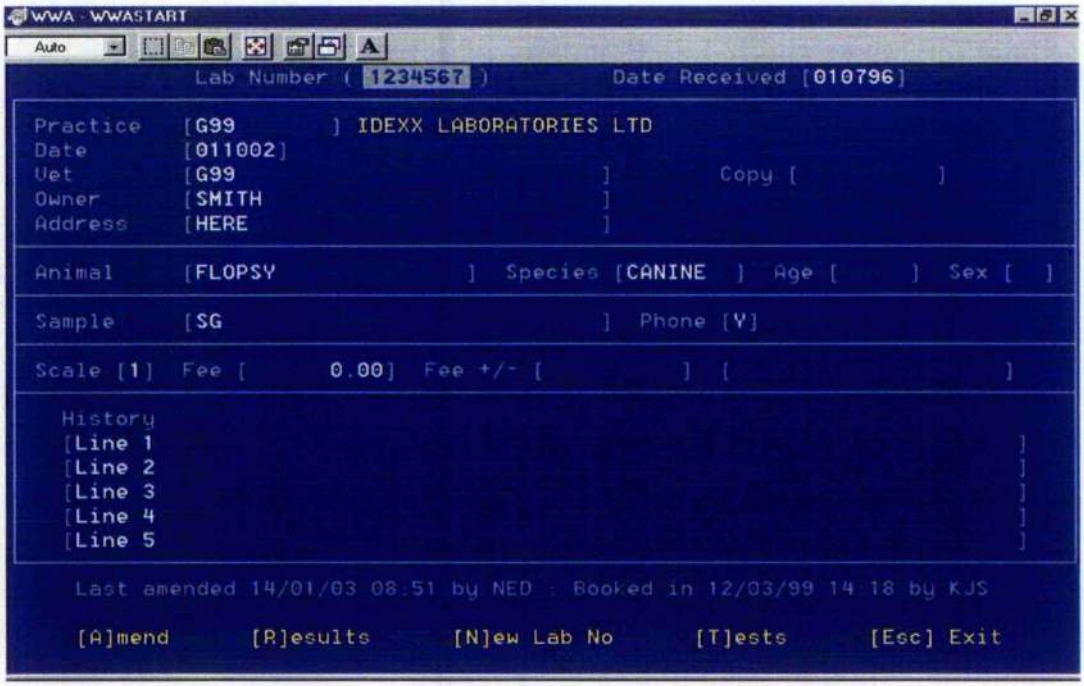


Figure 5-1 Demographic data entry screen for the commercial histopathology database. Data entry for the Lab Number and Practice fields is performed by barcode scanning technology. All other fields are free text fields with data manually transcribed from the biopsy submission form by administrative personnel.

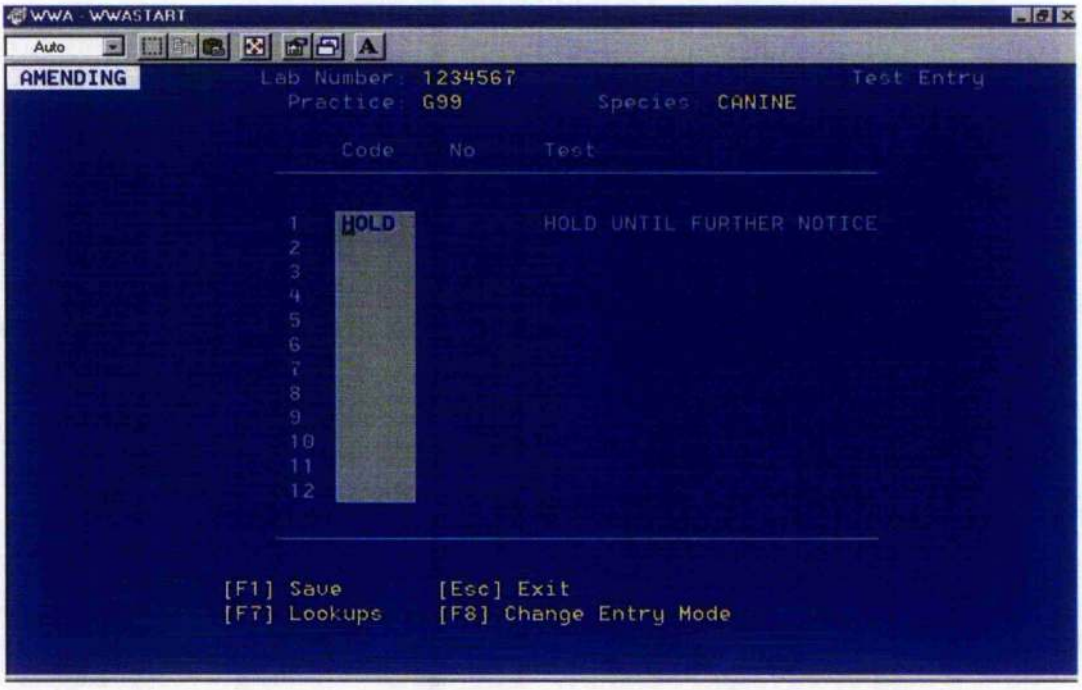


Figure 5-2 Test data entry screen for the commercial histopathology database. Test codes may be entered manually or can be retrieved from a lookup menu accessed using the F7 key.

5.2.2.4 Slide preparation

The processed blocks are sectioned on a microtome, the sections mounted and stained. The prepared slides are labelled with the *labref*, and placed with the original submission form ready for collection by the duty histopathologists. The technical staff manually enter information pertaining to each *labref* on the worklist, being the number of blocks and the date the slides were prepared. On collection of submission forms and slides, the pathologist initials the appropriate *labref* entries on the worklist. The paper-based worklist is checked regularly by the technical staff to ensure that no samples have been missed from the day's submissions. Generally, prepared slides are available for the pathologists within 24 hours of receipt of the sample by the laboratory, unless prior procedures such as decalcification have been required.

5.2.2.5 Microscopic examination

The reporting pathologist reviews the clinical details given on the submission form and following examination of the prepared slides, writes a full report describing the microscopic findings (Chapter 1, Section 1.5.2), making if possible a specific diagnosis and providing prognostic or behavioural information.

If further procedures are deemed necessary, such as special staining or immunohistochemistry (Chapter 1, Section 1.5.4), a provisional report may be written and sent to the veterinary practice with a final report pending the results of further procedures.

5.2.2.6 Report generation

The report is entered directly into the laboratory database. The accuracy of the data entry on the first screen (Figure 5-1), containing the demographic data entered by the administration staff, is checked against the original submission form by the pathologist. A third screen (Figure 5-3) is used for report generation. Field headings include "diagnosis", "histology", "comment" and "prognosis". All fields are for free text entry and the number of fields completed is at the discretion of the pathologist.

Figure 5-3 Report generation screen for the commercial histopathology database. All fields are free text entry fields and all data entry on this screen is performed by the diagnostic histopathologist.

5.2.2.7 Report delivery to practitioner

The laboratory delivers the majority of reports to veterinary practices via an automated facsimile system integrated into the William Woodards software framework (Section 5.2.1.1.). Once the pathologist has completed his/her provisional or final report, he/she authorises the report on-screen, after which the data enters a "fax queue" and is automatically faxed to the practice. For practices requiring hard copy, a paper copy of the report is also generated and is posted to the practice within 24 hours of the report's generation. Some practices receive e-mail reports that can be transferred into practice-held computer records.

5.2.3 Data source

The study dataset described below was derived from a commercially owned veterinary laboratory database. The histopathology service run by this commercial laboratory receives

samples from locations throughout the UK. Over 99% of samples originate from privately owned or charity veterinary practices, with very few being submitted from veterinary schools, other diagnostic laboratories, or other non-practice sources. All samples for histological examination received in a single submission (single or multiple biopsies) are recorded under the same identification number (*labref*). Each record is initiated using information supplied by veterinary personnel on a purpose-designed submission form (Appendix 5), one of which should accompany every submission.

5.2.4 Data extraction

Unlike the procedure for data extraction from the CIDRU database (Chapter 2, Section 2.5.1.1) which was carried out by the author, data extraction from the laboratory database was performed by the dedicated database manager at the laboratory. The author requested all relevant demographic and pathological data pertaining to canine biopsies submitted to the laboratory from January 1993 to August 1999. The data were exported from the database in text format onto the database manager's hard drive, then recorded onto a compact disc and posted to the author.

Data from the following 12 fields were received:

1. Lab number.
2. Requesting practice code.
3. Vet's name.
4. Date received.
5. Animal's name.
6. Animal's age.
7. Animal's sex.
8. Data entered against OURR test format code (Our reference)
9. Data entered against DIAG test format code (Diagnosis)
10. Data entered against HIST test format code (Histology findings)
11. Data entered against COM4 test format code (Comment)
12. Data entered against PROG test format code (Prognosis)

Owner information was not extracted because the laboratory wished to preserve client confidentiality, in compliance with the Data Protection Act (1998) (Data Protection Act, 1998). The data were provided in one large file, as well as in files designated by the year of tissue submission. The data were imported into Microsoft Excel using the "Text Import Wizard" and were initially saved in comma separated value (.csv) format to allow the most efficient use of hard drive disk space. The smallest of these files, containing information for all canine histology samples submitted between January and August 1999, was chosen for initial data inspection and interrogation. This choice was made because the smallest file was considered to be the most manageable dataset with which to investigate the database content, and to determine what was required to prepare the data for statistical analysis. This file is henceforth referred to as "Histo99".

5.2.5 Data preparation and cleaning

Microsoft Excel 97 and Microsoft Access 97 (Microsoft Corporation) were used to prepare the Histo99 file for epidemiological analysis. As for the CIDRU dataset, the outcome of interest was the presence or absence of neoplasia in a tissue biopsy. The risk factors of interest for the study were identical to those examined for the CIDRU data with the exception of breed of dog, because this was not electronically recorded in the database during the study time period. Another risk factor, site of biopsy, was not recorded in a specific field in the database, but this was considered for inclusion because the free text fields, "diagnosis", "histology", "comment" and "prognosis" were available for perusal and could potentially identify the site of biopsy.

Using spreadsheet functions, additional columns were created in the Histo99 file to allow recording of site of biopsy ("site") as a free text field, and to allow recording of an "outcome" code for the diagnosis of the biopsy, as non-neoplastic (coded "0"), neoplastic (coded "1") or indeterminate (coded "2"). Where multiple biopsies were submitted under one *labref*, this feature was recorded in the "site" field. To speed the data preparation process, data filter functions in Microsoft Excel were used to examine the "diagnosis" field and to enter "site" and "outcome" codes as efficiently as possible. For example, a lesion with the diagnosis "histiocytoma" is neoplastic and must originate from the integument, thus the speed of data entry for outcome and site could be considerably enhanced for such

lesions. Where no diagnosis was present in the "diagnosis" field, the free text fields "histology", "comment" and "prognosis" were fully inspected to obtain this information if possible. Text-string searches using neoplasia suffixes such as "-umour" or "-oma" of the free text fields did not significantly reduce the time required for this data preparation stage, because of the large overlap between neoplastic and non-neoplastic diagnoses identified with this technique. For example, searching using "-oma" revealed many free text fields containing words such as "granulomatous" or "granuloma", being terms used to describe non-neoplastic features of a tissue biopsy, as well as many neoplastic terms, e.g., "histiocytoma", "mastocytoma" and "carcinoma". Searching using the non-neoplastic suffix "-itis" was also performed, though again there was limited gain expediting this stage. Inspection of the free text fields was a particularly labour-intensive and time-consuming task. Data preparation was performed by the author and one assistant for the Histo99 file, with the author reviewing all prepared records prior to further analysis.

Following completion of site identification and outcome coding, the following fields were selected for export and conversion to Microsoft Access:

Lab number.
Requesting practice code
Age of animal
Gender of animal
Diagnosis
Site
Outcome
Date received

Prior to conversion to Microsoft Access format, the "date received" file was converted from its numerical format to a date format (dd/mm/yy) using spreadsheet functions. The "age" field was exported in its original format, containing both numerical and text data (e.g. 11 YRS, 11 YR, 11YR, 11MTH, 11 MTHS, 11WKS). Once converted to a Microsoft Access table, systematic querying of the field was used to create an exclusively numerical field for age of dog, so these data could be used in subsequent statistical analyses. Microsoft Access queries were also used to create site codes identical to those used for the CIDRU dataset, based upon the newly-created free text field "site".

Following data manipulation of the dataset, a hierarchical and iterative approach to the data cleaning procedure similar to that used for the CIDRU dataset was developed (Chapter 3, Section 3.2.3), focusing on the main host-related variables of age and gender of dog, and site of biopsy. The practice code for each record was noted. However, further assessment of type and location of the submitting veterinary source was not undertaken because the relevant details were not made available by the laboratory for this study for reasons of client confidentiality. Incremental corrections were made to the study dataset to facilitate progress.

There were two stages to the data cleaning procedure. First, the data underwent preliminary inspection to locate obvious inconsistencies, e.g., missing or nonsense values. Data were then stratified to conduct logic checks e.g., the use of valid gender and site codes, correct gender recorded for biopsies originating from gender-specific sites and correct gender and site recorded for gender-specific diagnoses. The "histology", "comment" and "prognosis" fields as described above were again consulted during this process to further correct discrepancies.

5.2.6 Research design

As for the biopsy population compiled by the CIDRU diagnostic histopathology service described in Chapter 4, a retrospective case-control study design was chosen for the analysis of the commercial histopathology dataset.

5.2.6.1 Definition of cases and controls

The host-related variables selected for initial investigation in this study were identical to those selected from the CIDRU dataset (Chapter 3, Section 3.2.2) where possible. The selections of age and gender ("sex") of dog from which a biopsy was taken, and site of biopsy were based, as previously stated, on results from previous studies (Dorn *et al.*, 1968; Cohen *et al.*, 1974; MacVean *et al.*, 1978).

The unit of investigation was the biopsy, not the animal. Criteria for a biopsy to be selected as a case were known age and gender of dog, known biopsy site, and a histopathological

confirmation of neoplasia. The same criteria as for the cases were used for the selection of a control population from the study population, though with a histopathological confirmation of a non-neoplastic diagnosis (Chapter 3, Section 3.2.4.1). Selection of cases and controls was assisted by the "outcome" code created during the data preparation process (Section 5.2.5). Records containing biopsies from multiple sites, or with biologically distinct diagnoses if multiple biopsies were obtained from the same site, were excluded.

5.2.6.2 Coding of independent variables

The categorical independent variables, gender and site, were subdivided for further analysis. Two coding formats were used to examine gender. The first used the most common entries present in the database, with a total of six categories: male entire (ME), male neutered (MN), male unknown (M), female entire (FE), female neutered (FN) and female unknown (F). The second coding system considered gender separately from neuter status i.e., as two independent variables, the first being male/female (M/F), the second being "neuter status" (entire (E), neutered (N), unknown (U)).

Site of biopsy was coded using the same numerical system as described in Chapter 3, Section 3.2.5. Thus, site was grouped according to the same eight represented biopsy sites as in the CIDRU database analysis (Chapters 3 and 4), with all biopsies from other sites being placed into a separate category.

5.2.7 Statistical analysis

The statistical analysis of the study dataset was undertaken in a similar way to the CIDRU dataset, described in Chapter 4. Initially, the effects of host-related risk factors, and then the influence of practice were assessed upon the outcome of a histopathological diagnosis of neoplasia in a biopsy.

5.2.7.1 Univariable analysis of host-related variables

An initial screening univariable logistic regression was performed to identify those variables that had little or no association with the outcome of neoplasia. All variables significant at P 0.25 were considered eligible for inclusion in a multivariable analysis (Hosmer and Lemeshow, 1989).

5.2.7.2 Multivariable analysis of host-related variables

Following the screening univariable logistic regression, dummy variables were generated for any categorical variable with more than two levels. Because there were two code structures describing gender and neutering status, two models were created. The first assessed the effect of gender when subdivided into six gender categories (Model 1), and the second considered gender and neutering status as separate independent variables in the model building process (Model 2). In each case, the model building process followed the method described in Chapter 4, Section 4.2.2.2.

The goodness-of-fit of the final models were assessed by the Hosmer-Lemeshow statistic (Lemeshow and Hosmer, 1982), and outliers were assessed by inspection of the standardized residual, the deviance, the leverage, the predicted probability and the influence diagnostic, delta beta (Chapter 4, Section 4.2.2.2). Outliers with the most influential covariate patterns were then removed and the significance of the variables included in the models re-evaluated. The criteria adopted for intervention was a change of greater than 20% in any of the variable coefficients and/or evidence of statistical instability.

5.2.7.3 Univariable analysis of practice

Univariable logistic regression was performed to assess the association of practice with the outcome of neoplasia. Cross tabulation of the outcome and practice produced many zero cells. It was decided to adopt the same inclusion criterion for practice as described for the CIDRU dataset (Chapter 4, Section 4.2.2.3.1), i.e., only those contributing more than five neoplastic and more than five non-neoplastic biopsies were eligible for inclusion in a

multivariable analysis which contained a variable representing practice. Data relevant to these practices and their associated biopsies underwent a univariable screening procedure of the host-related risk factors as described in Section 5.2.7.1.

5.2.7.4 Multivariable analysis with practice

Because the dataset exhibited the same feature of grouping by practice as the CIDRU dataset, a similar approach to dealing with extra-binomial variation (EBV) was adopted (Chapter 4, Section 4.2.2.3.2). A fixed-effects model was generated with fixed-effects terms for each eligible practice being added to the main (host-related) effects models. All analyses to this stage were performed using Minitab 12.21 (Minitab Inc., State College, PA). A logistic-binomial models was then created with practice being included as a random-effects term, treating practice data as indistinguishable. This second model was generated using EGRET for Windows, version 2.0.3 (Cytel Software Corporation, 1999).

5.3 RESULTS

5.3.1 General

During the time period represented in the Histo99 file (January to August 1999), a total of 12046 canine biopsies were submitted to the histopathology service at the laboratory, from 694 veterinary practices located throughout the United Kingdom.

5.3.2 Data cleaning

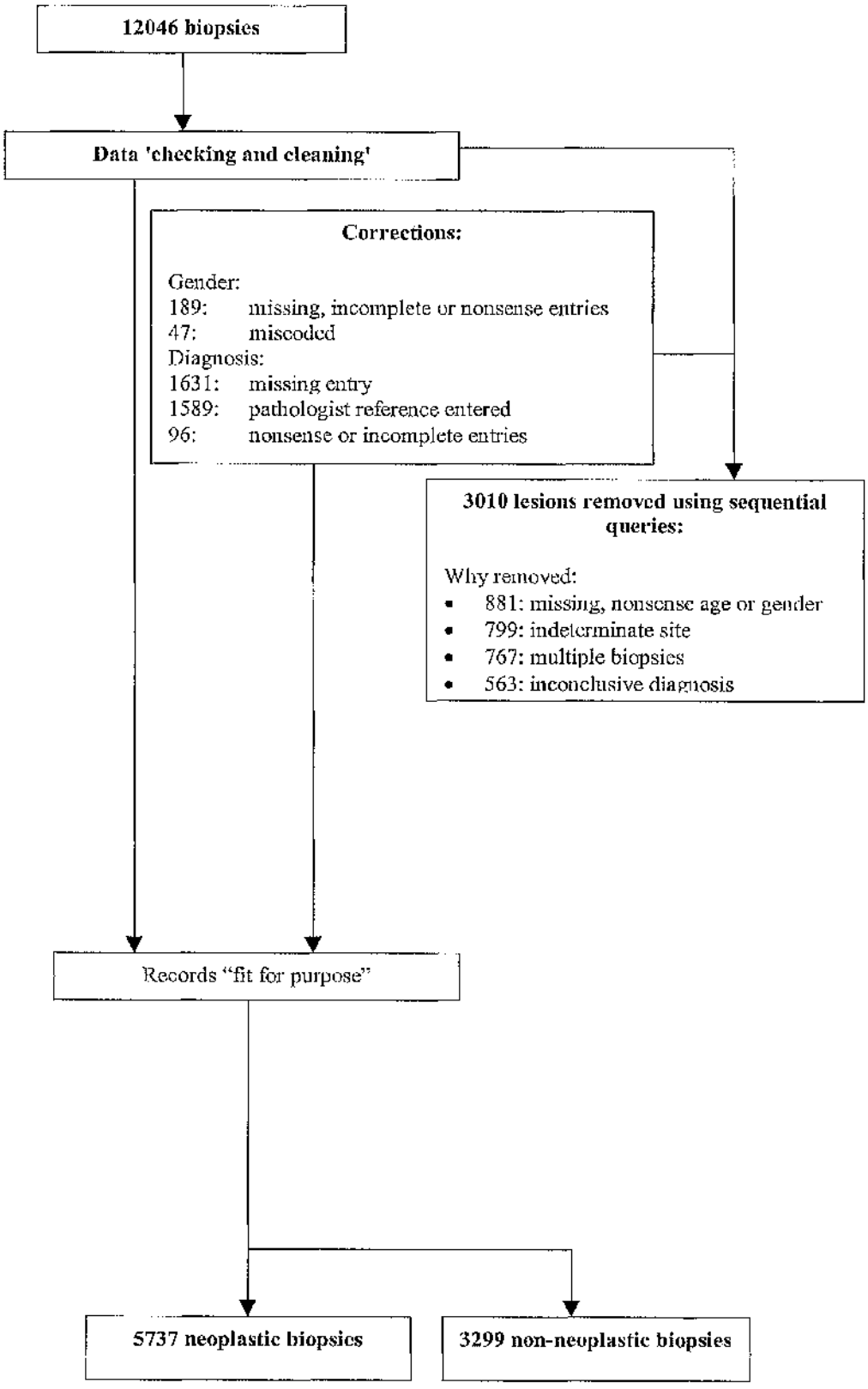
The percentage of records found to be ineligible for inclusion in the statistical analysis was 25.0% (3010/12046). The causes of record exclusion were missing or nonsense values (881/3010, 29.3%), indeterminate site (799/3010, 26.5%), multiple biopsies submitted under one *labref* (767/3010, 25.5%), and an "inconclusive" diagnosis being recorded for the biopsy (563/3010, 18.7%).

For each host-related variable of interest, i.e., age and gender of dog, and site of biopsy, data checks were performed during the data cleaning process and corrections made by

inspecting the full histopathology report where this was necessary. Initial inspection of the "sex" field showed that for those records with an entry, there was a significant proportion which described sex as male or female, but did not clarify neutering status (4660/12046, 38.7%). Due to the number of records involved, it was decided to keep those with known gender but unknown neutering status, and examine this within statistical models. Neutering status was determined by inspection of the "diagnosis", "histology", "comment" and "prognosis" fields, where these fields provided gender-specific data. Coding errors in the gender field, which could be corrected to allow record inclusion in the cleaned dataset, were detected in 2.6% (236/9036) of records. Of these, 189 (80.1%) contained missing, incomplete or nonsense entries and 47 (19.9%) contained miscoding of gender status, i.e., MN rather than ME; FN rather than FE.

Following construction of the study dataset, the original "diagnosis" field was compared to the "diagnosis" field in the cleaned dataset, using Microsoft Access queries to quantify the changes required to produce records that were fit for epidemiological analysis. The number of records where no diagnosis was originally entered was 1631/9036 (18%). Data errors in the "diagnosis" field were also significant, the most common being the entry of the pathologist reference (OURREF) (1589/9036, 17.6%), followed by other illegal or incomplete entries (96/9036, 1.1%).

Following data cleaning, 5737 neoplastic biopsies were eligible as cases and 3299 controls were selected for further analysis. Figure 5-4 summarizes the handling of the Histo99 dataset.


12046 biopsies
Data 'checking and cleaning'
Corrections:
Gender:
189: missing, incomplete or nonsense entries
47: miscoded
Diagnosis:
1631: missing entry
1589: pathologist reference entered
96: nonsense or incomplete entries
3010 lesions removed using sequential queries:
Why removed:
881: missing, nonsense age or gender
799: indeterminate site
767: multiple biopsies
563: inconclusive diagnosis
Records "fit for purpose"
5737 neoplastic biopsies
3299 non-neoplastic biopsies

5.3.3 Qualitative observations on data quality

The setup of the data fields produced a unique set of problems when preparing the data for statistical analysis. Although there was a vast quantity of information provided in the files and the percentage of supplied demographic information for each record was high, the lack of coding systems for data of epidemiological interest resulted in many hours being spent converting data into formats that were suitable for statistical software. Specific issues are detailed below.

The setup of the "age" field, using both numerical and text characters, was problematic. The number of illegal entries was few, but there were many ways that the same age could be recorded, e.g., 18 months of age could be expressed using 12 different combinations of keyboard strokes. The loose definition of data criteria for this field was suitable for the immediate purposes of the record, but led to considerable data manipulation being necessary to convert its content to a statistically-friendly format.

The gender field was less troublesome, with the vast majority of records having one of six major gender codes. The lack of in-built data safety checks for this field is likely to have contributed to data error, and justified the time committed to cross-referencing gender with the free text report fields for gender-specific diagnoses, to increase data accuracy. There was a large number of records with only gender and not neutering status, which is likely to reflect partially the quality and completeness of data supplied by the veterinarian.

Of the free text fields, only data in the "diagnosis" field were insufficient for its designated purpose, being to provide a succinct description of the examined lesion. When data in this field of the original and the cleaned dataset were compared, 36.7% (3316/9036) of records were found to be lacking information. Other fields contained sufficient information regarding biological behaviour and prognosis for the lesion, allowing clinical decisions to be made with confidence. Although data were adequate for the immediate purpose of providing the practitioner with a comprehensive report, the absence of a dedicated "site" field was the major factor limiting the quantity of data that could be studied by the author using the chosen study design and risk factors for analysis. The Histo99 file was chosen because of its smaller size compared to data files from other years. However, the level of

preparation time required for the file was approximately five months for a single operator. Prior knowledge of disease processes and histopathological terminology was also essential because the inspection of many histopathological descriptions was necessary to extract information relevant to further analysis. The requisite level of data preparation precluded the examination and preparation of other files, each of which covered 12 month time periods, compared to the eight month time period of the file selected for initial inspection. Thus, each potentially contained 33% more data than the Histo99 file and would have required a time frame beyond that available for the study.

5.3.4 Statistical analysis

5.3.4.1 Host-related variables

5.3.4.1.1 Univariable analysis

The summary statistics for the cases and controls by age and by subcategory of the categorical independent variables identified for investigation are presented in Table 5-1. Age and gender of dog and site of biopsy were significantly associated with a histopathological diagnosis of neoplasia. Neutering status was not significantly associated with the outcome when considered as a separate variable. The results of the univariable screening of independent variables are presented in Tables 5-2 and 5-3.

	Controls		Cases	
Number	3299	(100%)	5737	(100%)
Age (years)				
1st quartile	3		6	
Median	7		9	
3rd quartile	10		11	
Gender (6 categories)				
Male entire	652	(19.76%)	1001	(17.45%)
Male neutered	420	(12.73%	673	(11.73%)
Male unknown	612	(18.55%	985	(17.17%)
Female entire	315	(9.55%	649	(11.31%)
Female neutered	683	(20.70%	1266	(22.07%)
Female unknown	617	(18.70%	1163	(20.27%)
Gender				
Male	1684	(51.0%)	2659	(46.35%)
Female	1615	(49.0%)	3078	(53.65%)
Neuter status				
Entire	967	(29.31%)	1650	(28.76%)
Neutered	1103	(33.43%)	1939	(33.80%)
Unknown	1229	(37.25%)	2148	(37.44%)
Site of biopsy				
Skin[a]	2309	(69.99%)	3566	(62.16%)
Mouth/pharynx	165	(5.00%)	469	(8.18%)
Mammary gland	58	(1.76%)	831	(14.48%)
Lymphatic system	97	(2.94%)	277	(4.83%)
Liver	91	(2.76%)	33	(0.58%)
Spleen	67	(2.03%)	73	(1.27%)
Reproductive system	67	(2.03%)	218	(3.80%)
Intestine	135	(4.09%)	34	(0.59%)
Other	310	(9.40%)	236	(4.11%)

[a]Includes biopsies of epithelial, mesenchymal and melanocytic cell origin

Table 5-1 Summary statistics for case-control data for the age, gender, neuter status and site of origin of canine biopsies submitted to a commercial histopathology database during 1/1/1999 – 30/8/1999.

	Controls	Cases	P-value	Odds Ratio	95% CI Lower	95% CI Upper
Number	3299	5737				
Constant						
Age (years)			**<0.00**	1.13	1.11	1.14

Table 5-2 Univariable analysis of the continuous variable, age, investigated for association with an outcome of neoplasia being diagnosed in a canine biopsy submitted to a commercial diagnostic histopathology service.

	Controls	Cases	P-value	Odds Ratio	95% CI Lower	95% CI Upper
Gender (6 categories)			**<0.00**			
Male entire[1]	652	1001		1		
Male neutered	420	673		1.04	0.89	1.22
Male unknown	612	985		1.05	0.91	1.21
Female entire	315	649		1.34	1.14	1.59
Female neutered	683	1266		1.21	1.05	1.38
Female unknown	617	1163		1.23	1.07	1.41
Gender			**<0.00**			
Male[1]	1684	2659		1		
Female	1615	3078		1.21	1.11	1.32
Neuter status			**0.85**			
Entire[1]	967	1650		1		
Neutered	1103	1939		1.03	0.92	1.15
Unknown	1229	2148		1.02	0.92	1.14
Site of biopsy			**<0.00**			
Skin[1a]	2309	3566		1		
Mouth/pharynx	165	469		1.84	1.53	2.21
Mammary gland	58	831		9.28	7.07	12.17
Lymphatic system	97	277		1.85	1.46	2.34
Liver	91	33		0.23	0.16	0.35
Spleen	67	73		0.71	0.50	0.99
Reproductive system	67	218		2.11	1.59	2.78
Intestine	135	34		0.16	0.11	0.24

[1]referent category
[a]includes biopsies of epithelial, mesenchymal and melanocytic cell origin

Table 5-3 Univariable analysis of categorical variables investigated for association with an outcome of neoplasia being diagnosed in a canine biopsy submitted to a commercial diagnostic histopathology service.

5.3.4.1.2 Multivariable analysis

The final host-related effects models containing age and gender of dog and site of biopsy are shown in Tables 5-4 and 5-5. Increasing age was associated with an increased risk (odds) of neoplasia in a biopsy in both model 1 (OR=1.11; 95%CI 1.10, 1.12) and model 2 (OR=1.11; 95%CI 1.10, 1.13). In model 1, biopsies from the oropharynx (OR=1.48; 95%CI 1.22), mammary gland (OR=7.97; 95%CI 6.02, 10.54), lymphatic system (OR=1.64; 95%CI 1.29, 2.09) and reproductive systems (OR=1.72; 95%CI 1.28, 2.30) had an increased risk of a neoplastic diagnosis compared to biopsies from the skin and connective tissues (the referent category for site), when controlled for age and gender of dog. Biopsies from the liver (OR=0.19; 95%CI 0.13, 0.29), spleen (OR=0.51; 95%CI 0.36, 0.72) intestine (OR=0.16; 95%CI 0.11, 0.23) and other non-specified sites (OR=0.43; 95%CI 0.36, 0.51) showed decreased risk of a neoplastic diagnosis when controlled for age and gender of dog. Model 2 showed the same directions of risk (Table 5-5). These results were similar to the results of the CIDRU multivariable analysis of host-related variables (Chapter 4, Section 4.3.2.1, Table 4-4). The Hosmer-Lemeshow statistic of model 1 was 15.4 with eight degrees of freedom (P = 0.05), indicating poor model fit. The Hosmer-Lemeshow statistic of model 2 was 18.4 with eight degrees of freedom (P = 0.02).

Variable	Coefficient	SD	P-values	Odds Ratio	95%CI Lower	95%CI Upper
Constant	-0.29					
Age	0.10	0.01	0.00	1.11	1.10	1.12
Gender						
Male entire[1]	0.00			1		
Male neutered	0.09	0.08	0.31	1.09	0.92	1.29
Male unknown	0.10	0.08	0.19	1.11	0.95	1.29
Female entire	-0.08	0.1	0.4	0.92	0.77	1.11
Female neutered	0.04	0.07	0.61	1.04	0.90	1.20
Female unknown	-0.06	0.08	0.45	0.94	0.81	1.10
Site of biopsy						
Skin[1a]	0.00			1		
Mouth/pharynx	0.39	0.1	0.00	1.48	1.22	1.79
Mammary gland	2.08	0.14	0.00	7.97	6.02	10.54
Lymphatic system	0.50	0.12	0.00	1.64	1.29	2.09
Liver	-1.66	0.21	0.00	0.19	0.13	0.29
Spleen	-0.67	0.17	0.00	0.51	0.36	0.72
Reproductive system	0.54	0.15	0.00	1.72	1.28	2.30
Intestine	-1.85	0.2	0.00	0.16	0.11	0.23
Other	-0.85	0.09	0.00	0.43	0.36	0.51

[1]referent category
[a]Includes biopsies of epithelial, mesenchymal and melanocytic cell origin

Table 5-4 Coefficients, odds ratios and 95% confidence intervals from the multivariable logistic regression model containing the host-related main effects for factors associated with the risk of neoplasia being diagnosed from a canine biopsy submitted to a commercial diagnostic histopathology service: Gender being assessed as 6 categories (Model 1).

Variable	Coefficient	SD	P-value	Odds Ratio	95%CI Lower	95%CI Upper
Constant	-0.24	0.05	0.00			
Age	0.11	0.01	0.00	1.11	1.10	1.13
Gender						
Male[1]	0.00			1		
Female	-0.08	0.05	0.09	0.92	0.84	1.01
Site of biopsy						
Skin[1a]	0.00			1		
Mouth/pharynx	0.39	0.1	0.00	1.48	1.22	1.78
Mammary gland	2.05	0.14	0.00	7.75	5.88	10.23
Lymphatic system	0.49	0.12	0.00	1.64	1.29	2.08
Liver	-1.65	0.21	0.00	0.19	0.13	0.29
Spleen	-0.67	0.17	0.00	0.51	0.36	0.72
Reproductive system	0.48	0.15	0.00	1.62	1.22	2.16
Intestine	-1.85	0.2	0.00	0.16	0.11	0.23
Other	-0.85	0.93	0.00	0.43	0.36	0.51

[1]referent category
[a]Includes biopsies of epithelial, mesenchymal and melanocytic cell origin

Table 5-5 Coefficients, odds ratios and 95% confidence intervals from the multivariable logistic regression model containing the host-related main effects for factors associated with the risk of neoplasia being diagnosed from a canine biopsy submitted to a commercial diagnostic histopathology service (Model 2).

5.3.4.2 Univariable analysis of practice

Visual inspection and summary statistics of the distribution of biopsy submission by practice showed moderate variation in the numbers of biopsies contributed from individual practices (1 to 85 biopsies, median 8, interquartile range 2 to 18). There were 505 practices contributing less than 5 neoplastic and less than 5 non-neoplastic lesions (505/694, 72.8%); these practices contributed 34.3% (3099/9036) of biopsies in the study dataset. Because fitting these practices to a logistic regression model would cause instability in the parameter estimates, it was decided to exclude these practices and their biopsies from any analyses which included a practice variable. Univariable logistic regression of practice in the reduced dataset (5937 biopsies from 189 practices) showed practice was still significantly associated with the outcome of neoplasia (Appendix 6).

5.3.4.3 Multivariable analysis with practice as a fixed-effect

The structure of model 1 was chosen to include practice as a fixed-effect, in keeping with the gender classifications used in the statistical analyses described in Chapter 4. The main-effects model of the host-related variables for the reduced dataset described in Section 5.2.8.4 is presented in Table 5-6. As in the full main-effects model, increasing age was shown to be associated with an increased odds of neoplasia in a biopsy (OR=1.11, 95%CI 1.10, 1.13) when controlled for gender and site of biopsy. The directions of odds for neoplasia in biopsies from specific genders of dog, and sites of biopsy, were identical to those shown in the full main-effects model. Specifically, biopsies from the oropharynx (OR=1.39; 95%CI 1.11, 1.76), mammary gland (OR=8.76; 95%CI 6.13, 12.52), lymphatic system (OR=1.37; 95%CI 1.02, 1.83) and reproductive systems (OR=1.61; 95%CI 1.13, 2.30) had an increased risk of a neoplastic diagnosis compared to biopsies from the skin and connective tissues (the referent category for site), when controlled for age and gender of dog. Biopsies from the liver (OR=0.16; 95%CI 0.10, 0.26), spleen (OR=0.59; 95%CI 0.40, 0.87) intestine (OR=0.15; 95%CI 0.09, 0.24) and other non-specified sites (OR=0.41; 95%CI 0.33, 0.52) showed decreased risk of a neoplastic diagnosis when controlled for age and gender of dog.

Variable	Coefficient	SD	P-value	Odds Ratio	95%CI Lower	Upper
Constant	0.61	0.08	0.00			
Age	0.11	0.01	0.00	1.11	1.1	1.13
Gender						
Male entire[1]	0.00			1		
Male neutered	0.18	0.10	0.09	1.19	0.97	1.46
Male unknown	0.15	0.09	0.12	1.16	0.96	1.39
Female entire	-0.04	0.12	0.71	0.96	0.76	1.21
Female neutered	0.08	0.09	0.36	1.09	0.91	1.3
Female unknown	0.06	0.1	0.5	1.07	0.88	1.29
Site of biopsy						
Skin[1a]	0.00			1		
Mouth/pharynx	0.33	0.12	0.01	1.39	1.11	1.76
Mammary gland	2.17	0.18	0.00	8.76	6.13	12.52
Lymphatic system	0.31	0.15	0.04	1.37	1.02	1.83
Liver	-1.83	0.25	0.00	0.16	0.1	0.26
Spleen	-0.53	-.2	0.01	0.59	0.4	0.87
Reproductive system	0.47	0.18	0.01	1.61	1.13	2.3
Intestine	-1.89	0.24	0.00	0.15	0.09	0.24
Other	-0.88	0.11	0.00	0.41	0.33	0.52

[1]referent category
[a]Includes biopsies of epithelial, mesenchymal and melanocytic cell origin

Table 5-6 Coefficients, odds ratios and 95% confidence intervals from the multivariable logistic regression model containing the host-related main effects for factors associated with the risk of neoplasia being diagnosed from a canine biopsy submitted to a diagnostic histopathology service (reduced dataset).

Practice was then added as a fixed-effect to the model. The statistics associated with the covariates of prime interest, being age and gender of dog, and site of biopsy are presented in Table 5-7. The full model, including statistics associated with practice, is presented in Appendix 7. The addition of practice as a fixed-effect caused a significant improvement in model fit indicated by a reduction in the Hosmer-Lemeshow statistic (2.97, P = 0.94), indicating that a degree of extra-binomial variation was accounted for by the practice term. Examination of the coefficients and 95% CIs in the model including practice as a fixed-effect revealed similar results for the host-related covariates as were apparent in the reduced dataset main-effects model (Table 5-6).

Variable	Coefficient	SD	P-value	Odds Ratio	95%CI Lower	95%CI Upper
Constant	0.60	0.39	0.12			
Age	0.11	0.01	0.00	1.12	1.0	1.14
Gender						
Male entire[1]	0.00			1		
Male neutered	0.18	0.11	0.1	1.2	0.96	1.48
Male unknown	0.13	0.1	0.21	1.13	0.93	1.38
Female entire	-0.07	0.12	0.59	0.94	0.73	1.19
Female neutered	0.07	0.1	0.47	1.07	0.89	1.3
Female unknown	0.05	0.1	0.61	1.05	0.86	1.29
Site of biopsy						
Skin[1a]	0.00			1		
Mouth/pharynx	0.33	0.12	0.01	1.39	1.09	1.77
Mammary gland	2.26	0.19	0.00	9.54	6.62	13.75
Lymphatic system	0.36	0.16	0.02	1.43	1.05	1.95
Liver	-1.97	0.26	0.00	0.14	0.08	0.23
Spleen	-0.61	-0.21	0.00	0.55	0.36	0.82
Reproductive system	0.43	0.19	0.02	1.54	1.06	2.23
Intestine	-1.94	0.25	0.00	0.14	0.09	0.23
Other	-0.94	0.12	0.00	0.39	0.31	0.49

[1]referent category
[a]Includes biopsies of epithelial, mesenchymal and melanocytic cell origin

Table 5-7 Coefficients, odds ratios and 95% confidence intervals from the multivariable logistic regression model containing the host-related main effects and practice as a fixed-effect for factors associated with the risk of neoplasia being diagnosed from a canine biopsy submitted to a diagnostic histopathology service (Practice data shown in Appendix 7).

5.3.4.4 Multivariable analysis with practice as a random-effect

Logistic-binomial regression, assuming indistinguishable data, was used to fit a random-effects term for practice to the reduced dataset main-effects model, based upon the *a priori* knowledge that biopsies were grouped, or clustered by practice. Results are presented in Table 5-8. Inclusion of the random-effects term caused a reduction in model deviance, indicating a better model fit when EBV was taken into account using this method. Interestingly, there were several changes in the parameter estimates for site of biopsy, with the odds of neoplasia for biopsies from the oropharynx (OR=1.05; 95%CI 0.92, 1.18), lymphatic system (OR=1.14; 95%CI 0.97, 1.34) and reproductive systems (OR=1.12; 95%CI 0.92, 1.35) becoming insignificant. The odds of neoplasia in biopsies from the mammary gland reduced from over eight times that of the odds of neoplasia in a biopsy

from the referent category, to only 1.4 times that of the odds of neoplasia in a biopsy from the skin/connective tissues (OR=1.42; 95%CI 1.28, 1.57), when controlled for age and gender. The odds of neoplasia in a splenic biopsy (OR=0.86; 95%CI 0.66, 1.12) also became insignificant in this model.

Variable	Coefficient	SD	P-value	Odds Ratio	95%CI Lower	95%CI Upper
Constant	-5.42	0.01	0.00			
Age	0.05	0.00	0.00	1.05	1.04	1.06
Gender						
Male entire[1]	0.00			1		
Male neutered	0.05	0.07	0.44	1.05	0.93	1.19
Male unknown	-0.01	0.06	0.85	0.99	0.88	1.11
Female entire	-0.03	0.07	0.61	0.97	0.84	1.11
Female neutered	0.01	0.06	0.89	1.01	0.9	1.13
Female unknown	-0.02	0.06	0.68	0.98	0.87	1.1
Site of biopsy						
Skin[1a]	0.00			1		
Mouth/pharynx	0.04	0.06	0.48	1.05	0.92	1.18
Mammary gland	0.35	0.05	0.00	1.42	1.28	1.57
Lymphatic system	0.13	0.08	0.12	1.14	0.97	1.34
Liver	-1.04	0.21	0.00	0.35	0.23	0.54
Spleen	-0.15	0.13	0.26	0.86	0.66	1.12
Reproductive system	0.11	0.09	0.26	1.12	0.92	1.35
Intestine	-1.19	0.21	0.00	0.3	0.2	0.46
Other	-0.43	0.08	0.00	0.65	0.55	0.77
Random-effect term						
%SCL	0.00	0.02				

[1]referent category
[a]Includes biopsies of epithelial, mesenchymal and melanocytic cell origin

Table 5-8 Coefficients, odds ratios and 95% confidence intervals from the multivariable logistic regression model containing the host-related main effects and practice as a random-effect for factors associated with the risk of neoplasia being diagnosed from a canine biopsy submitted to a diagnostic histopathology service (reduced dataset).

5.4 DISCUSSION

The primary aim of the laboratory is to provide exemplary service to veterinary practitioners in all disciplines of laboratory diagnostics. The computing hardware and software systems are specifically designed for seamless integration of data from administrative, technical and diagnostic departments. The laboratory's database represents

one of the major potential data sources for canine neoplasia in the UK. However with this laboratory being a non-academic and commercial pathology service, there is no requirement for members of the department to conduct research, and research carried out must not compromise the efficacy and timely provision of diagnostic information to clients. Consequently, there have been few studies conducted using data from the database, and this was the first to examine its potential as a secondary data source for the study of canine neoplasia.

Interrogation of the histopathology records revealed a number of issues concerning the current file structure and data types which limited their immediate usefulness for epidemiological analyses. Standardisation of data entry is of prime importance if datasets are to be easily adapted for epidemiological analyses. Regarding this database, the age field was informative but was only suitable for analysis following a number of data manoeuvres. The use of drop-down menus in this field to regulate the way in which age data were entered by database users could provide data in a format immediately suitable for secondary purposes.

Similar modification to the gender field (providing drop-down menus) could be helpful to reduce significant number of keystroke errors in this field. Without limiting the content of the gender field on the first data entry screen, the current alternative is for a second data quality check at the stage of report generation to be carried out by the pathologist responsible for the biopsy examination. For gender-specific diagnoses, a check at this stage may allow the data in this field to be easily corrected by the pathologist. A disadvantage to this procedure is the need to move between data entry screens. This increases the time taken to produce each report, and this may reduce the pathologist's motivation to correct data in a database field which are indirectly corrected by their diagnosis and report. For many records, there may also be issues with the quality and completeness of gender data provided by the submitting practitioner. A review of a random sample of the original submission forms would be helpful to investigate further this hypothesis and highlight which levels of data procurement could be targeted to improve the completeness of information in this field.

The pricing structure policy of encouraging multiple biopsy submissions caused particular problems for the current study because there was no provision in the file structure to extract easily the data for individual biopsies contained within these records. Due to time constraints, these records were not fully investigated, leading to a significant number of file omissions even though the records were complete and suitable for analysis in all other respects. Restructuring of the database to allow separate entries for each biopsy, linked to the animal by a unique identifier, would allow for increased data inclusion and greater understanding of the contents of the database. The structuring of data on related levels may also allow for more meaningful epidemiological analyses with multi-level modelling techniques.

Owner data was unavailable for this study due to reasons of client confidentiality. The lack of these data may have introduced some bias because there was no method of checking for repeat biopsy submissions from the same animal, thus keeping the unit of investigation as the biopsy, not the animal. A possible solution would be the creation of a unique animal identifier as well as a unique biopsy identifier, which would ensure owner data confidentiality while improving further the suitability of the database content for epidemiological analyses.

In any study using veterinary diagnostic information, accuracy and consistency in the use of nomenclature are vital (Bonnett *et al.*, 1997). The database showed the difficulties of conducting epidemiological analyses on secondary sources of data where standardised nomenclature and coding systems are not implemented. This was an issue with the CIDRU dataset because although coding systems were used to some extent, their uniqueness disallowed accurate comparisons of the results of the study with other studies of canine neoplasia. For this database, the CIDRU coding system for site of biopsy was chosen because of its simplicity and the author's familiarity with its use in the CIDRU database. Its use also allowed direct comparison of results from the two studies.

For the study of host factors of pathological lesions, the site of the lesion is paramount to thorough epidemiological analysis. This laboratory's submission form is designed for this information to be supplied as free text (Appendix 5). An improvement to the form's design would be to provide an animal silhouette diagram to encourage a graphical presentation of

the site of the lesion to be supplied by the practitioner. The site of lesion was commonly referred to by the pathologists in their description of findings and where a specific site was verified, this information was highlighted and coded to be suitable for inclusion in further analyses. This part of the data preparation process caused the most concern because of the need to balance the benefit (increasing the number of records eligible for the study) versus the time needed to complete the task. The introduction of a text field for site would partially address this deficiency. Realistically, this should only be a short-term option while a coding system was introduced and implemented. As mentioned in Chapter 3, Section 3.4, coding and nomenclature systems can involve a high initial costs due to training and software requirements (Feigl *et al.*, 1981; Moss, Smith and Nicholas, 1997). However, their long-term advantages are enormous for epidemiological research, as proven by cancer research in the medical field.

Many of the problems encountered in analysis of this database, and this would equally apply to similar datasets in other similar commercial laboratories, arise from it having been set up to allow rapid and accurate reporting of information to requesting veterinary practitioners. The potential for its use in epidemiological research was not originally anticipated. Alterations or additions to data input requirements in long established databases to increase their potential use by academic researchers would, undoubtedly, entail significant costs, both actual in terms of the costs of changing software and request forms, and in time for training and need for additional data input (McCandlish, 2002). Within a commercial laboratory, a persuasive argument would need to be made for such alterations of no direct or immediate benefit to the organisation, to be undertaken.

The data cleaning approach was effective in producing a study dataset suitable for further analyses, and identified a number of important potential quality assurance issues. There remain concerns associated with the cost of alterations to the database software to prevent coding errors, and with the development of a cleaning procedure with defined levels of allowable coding errors, so that further epidemiological analyses will not be invalidated. The "diagnosis" field was problematic and discussion of its use among users of the database could be considered to increase the accuracy of its content. Lack of accurate entry in the "diagnosis" field was a prime factor necessitating the examination the free text field of each record. Examination of histopathology reports also requires the data user to have

some prior knowledge of histopathological terminology, or direct communication with a pathologist to assist with terminology queries.

Data cleaning preceding epidemiological analysis and *ad hoc* studies are best at detecting registration problems in human cancer registries. Cancer registration data need "to be seen to be of high quality" to be valued and inspire confidence (Brewster, 1995). However, no large-scale database can be perfect (Skeet, 1991). Canine cancer registries remain rare although the VMDB (Chapter 1, Section 1.6.3.2.1) is routinely used to source data for epidemiological studies of canine neoplasia. The quality of data obtained from the VMDB is the responsibility of the institution maintaining the database, so researchers using its data cannot make direct comment. To the author's knowledge, there have been no recent published reports regarding the data quality of the VMDB, though it is assumed that internal data quality assessments are performed for the benefit of those institutions contributing data. A report describing data quality of the VMDB in the literature may nevertheless be beneficial to increase awareness of data quality issues in not only research areas, but also in the veterinary clinical domain.

The aim of the study in this chapter involved the host factors of age and gender of dog, and the site from which a biopsy was taken. The need for prioritisation of these data enabled a systematic approach to the data cleaning process to be developed, in order to retain the maximum number of records for statistical analyses. The hierarchical and iterative data cleaning method gave consideration to which missing data could be most easily retrieved from the content of the database, albeit that this process could still be time-consuming and labour-intensive due to the database structure. Searching of the electronic record was infinitely more time-efficient than searching stored paper records, or making contact with the submitting practice to retrieve or correct data. However, the removal of records with incomplete demographic and/or diagnostic data was inevitable with this approach. Inspection of the electronic histopathology report was essential for the data-retrieval process from this database and as described above, modification to the database structure would be likely to improve greatly the usefulness of the database for detailed epidemiological analyses.

The results of the screening univariable analysis for age of dog and site of biopsy were consistent with previous clinical descriptive and other univariable investigations of canine neoplasia (Dorn *et al.*, 1968b; Cohen *et al.*, 1974; Yamagami, 1996). The univariable analysis of neutering status, when considered as a separate variable to gender (being male/female), was insignificant. This is inconsistent with the results of other studies where neutering status has been found to be highly significant on the outcome of neoplasia. However in studies where neutering status has been significant, a crucial piece of known information has been the age of the animal when neutered, and could therefore be controlled for within the analyses (Schneider, Dorn and Taylor, 1969).

Results from the host-related effects models (Tables 5-4 and 5-5) show that four of the selected sites, *viz* mouth/pharynx, mammary gland, lymphatic system, and the reproductive system, were significantly associated with an increase in the odds of a biopsy sample having a neoplastic diagnosis, when compared to a biopsy from the skin. Age was also significantly associated with an increased odds of a neoplastic diagnosis. The increase in odds of neoplasia with increasing age is in agreement with other studies (Priester and Mantel, 1971; Moulton, 1990; Dobson *et al.*, 2002).

In this study it was elected to treat gender in two ways, producing two models for comparison, in order to include as many records as possible in the analysis. The initial approach of considering gender in six categories was based upon the most common codes present in the dataset, which reflected the terminology used by practitioners to describe this demographic feature of their samples. However, for this dataset, the alternative approach of fitting gender and neutering status as two effects was also considered because of the large number of biopsies for which neutering status was unknown. Comparison of the two models showed little difference in variable coefficients or model fit, thus further analyses were undertaken using the first gender coding approach, to allow more detailed comparisons to be made with the results from the CIDRU dataset, presented in Chapter 4.

Although gender, when considered as six categories, was shown to be significant to the outcome of neoplasia in the univariable analysis, it became insignificant once the other host-related factors of age and site of biopsy were controlled for in a multivariable model (Model 1, Table 5-4). This may in part be due to the number of records for which neutering

status was unknown. It is well known that neutering decreases the risk of gender-specific neoplasms, particularly that of mammary neoplasia if the procedure is carried out prior to two years of age (Dorn *et al.*, 1968b; Schneider, Dorn and Taylor, 1969). In this study, the age of neutering was not recorded, so the effects of neutering at different ages could not be taken into account in the analysis.

The skin was chosen as the referent site for the logistic regression analyses and biopsies of epithelial, mesenchymal and melanocytic cell origin were included in this category. In the final host-related effects models (Tables 5-4 and 5-5), biopsies from the mammary gland were seven times more likely to be neoplastic than those from the skin, and there were significant increases in odds ratios for neoplasia in biopsies from the oropharynx, lymphatic system and reproductive system, when controlled for age and gender. Nearly 75% of biopsies included in the study population originated from the skin or mammary gland, which is likely to reflect the relative ease of detection of abnormalities and subsequent sampling of tissues from these superficial sites. This figure is similar to the proportions of biopsies submitted from these sites in other datasets presented by Dorn *et al.* (1968b), MacVean *et al.* (1978), Nordstoga *et al.* (1997) and the CIDRU dataset (Chapter 4).

The priorities of this histopathology laboratory are to provide rapid, efficient service to their clients. The use of the database for epidemiological analyses of canine neoplasia was not immediately anticipated when the laboratory was established. This is reflected in the file structure of the database, particularly as a systematized coding and nomenclature system for demographic and diagnostic data is not incorporated into the file structure. The coding system used in the CIDRU database to identify site of biopsy (Chapter 2, Table 2.2) was chosen to prepare the dataset for epidemiological analyses because of its relative simplicity and the author's familiarity with the system. As mentioned in Chapter 4, Section 4.4, the use of this system disallows direct comparison of the study results with other studies of canine neoplasia. However, comparisons with the results from the CIDRU dataset analysis can be made.

Analyses of the host-related risk factors in both datasets showed an increase in odds of neoplasia with increasing age. Also in agreement were the increased risk of neoplasia in

biopsies from the oropharynx, mammary gland, lymphatic system and reproductive system, and decreased risk in the other sites selected for detailed study (Chapter 4, Table 4-4; Tables 5-4 and 5-5). Analysis of the reduced datasets, based upon the inclusion criteria stipulated for practice, showed that the inclusion of practice in the regression model as a fixed effect had a significant effect upon the model fit in both cases. This verifies that practice-related factors have a significant effect upon the data and further study of these factors is warranted. Of note is the effect of the inclusion of practice upon the odds ratio of neoplasia in biopsies from the mammary gland, which was reduced in the CIDRU dataset, but significantly increased in the commercial laboratory's dataset. However, it must be remembered that the practice inclusion criteria (>5 neoplastic and >5 non-neoplastic biopsies submitted) caused a greater percentage of records to be excluded from the laboratory dataset compared with the CIDRU dataset (34.3% and 10% respectively). This might in part reflect the much shorter time period covered by the commercial dataset (eight months) compared with the time period covered by the CIDRU dataset (12.5 years). There may also be factors associated with those practices whose data were excluded from each dataset which could influence the outcome of neoplasia in mammary gland biopsies submitted by those practices. Such factors could include a veterinarian's clinical judgement of the likelihood of a particular mammary gland biopsy being malignant, whether therapy and management decisions would be altered based upon histopathological findings and the financial status of the owner.

The inclusion of practice as a random-effect showed that there was strong evidence of the presence of EBV in the model. Host-related covariates were significantly altered from those seen in the models controlling for only host-related effects (Models 1 and 2, Tables 5-4 and 5-5). Alterations in the covariates in a model with improved fit indicates that these results are only associated with certain practices, and that EBV is significant in the data. Conducting practice-based research to identify practice-related factors for more detailed study could provide further insight into EBV in the dataset, using a multi-level modelling approach. The potential for this route of research is a subject for further enquiry.

The advantages of using a data source derived mainly from first-opinion practices have been previously discussed (Chapter 4, Section 4.4). This is the first epidemiological study of canine neoplasia using a commercial histopathology laboratory database, rather than

data from an academic institution, such as those presented in Chapters 3 and 4. There are obvious similarities between the biopsy caseloads from the two histopathology services, particularly with regard to types of biopsy submissions. The major difference lies with the quantity of biopsy submissions and thus, the amount of data generated by the histopathology services. The Histo99 dataset, prior to data cleaning, contained 12046 biopsy records, submitted over an eight month time period. In contrast, 21371 biopsies were submitted to the CIDRU histopathology service during a 12.5 year time period. Large datasets are required to give power to statistical analyses and are more likely to allow detailed analyses because there is less likelihood of the "sparse data problem" causing invalidation of analyses, particularly where subsets of the dataset, such as biopsies submitted from a certain site, may be of interest. The data resource created by this commercial histopathology service has vast potential for epidemiological research due to the sheer volume of biopsy submissions.

This study describes a data cleaning approach applied to a dataset derived from a population of canine biopsies examined by a commercial histopathology laboratory. Suggestions for ways to modify the database structure to improve its usefulness for epidemiological analyses are made based upon the findings from the data cleaning procedure. Univariable and multivariable analyses confirm many of the perceived risks for neoplasia based on previous clinical descriptive and univariable investigations. The results are consistent with those from another univariable and multivariable analysis of a canine biopsy population (Chapters 3 and 4). The increased statistical stability of models including practice as a variable suggests that further investigation of practice-related factors upon canine biopsy submission would be useful to identify other risk factors influencing the content of readily available data sources suitable for investigation of canine neoplasia in the UK.

CHAPTER 6

SPATIAL AND SPATIO-TEMPORAL ANALYSIS OF A CANINE BIOPSY POPULATION COMPILED BY A DIAGNOSTIC HISTOPATHOLOGY SERVICE

6.1 INTRODUCTION

The discipline of epidemiology involves the study of the distribution and determinants of disease or other health-related issues. Historically, the focus in epidemiologic research has been on person/subject or time, with little regard for the implications of place or space (Moore and Carpenter, 1999). The usefulness of maps to investigate the aetiology of disease incidence has been recognised for over a century, with possibly the best-known example being the map of the 1855 cholera outbreak around the Broad Street pump in Soho by John Snow (Snow, 1855). Although the recognition that disease has a spatial component is not new, it has taken the development of powerful computer software with the ability to handle geographical data to make spatial analysis a common procedure for the epidemiologist.

As discussed in Chapter 1, Section 1.10, the combination of geographical information systems (GIS), spatially referenced data and statistical analyses of spatial data have led to numerous publications describing studies of aggregations or "clusters" of cancer occurrence in the medical literature (Hjalmars *et al.*, 1996; Gatrell and Bailey, 1996; Kulldorff *et al.*, 1998). The clustering of disease in space, time or space-time may be examined. However, the problem of distinguishing between clusters occurring due to chance alone, and those occurring due to some underlying spatial or temporal risk factor, must be addressed. The epidemiologist is typically interested in clusters of disease cases only after having adjusted for spatial variations in the density of the background population itself. On a map representing the cases as a spatial point pattern, an apparent disease cluster in a particular area could be misleading because it may be explained simply by a clustering of the population itself in that area (Kulldorff and Nagarwalla, 1995).

In animal populations, clusters of disease can occur at various levels and for different reasons. Examples of this include clustering of disease within a litter or herd. It has become common practice in epidemiological research to expect such clustering and to adjust for it when performing a statistical analysis (Carpenter, 2001). In relation to the studies presented in Chapters 4 and 5, clustering of data by submitting practice was expected in both datasets examined. Thus, fixed- and random-effects logistic regression models were used to adjust for this feature (Chapter 4, Section 4.2.2.3.2 and Chapter 5, Section 5.2.8.4). However, detection of possible geographical clusters of disease requires spatially referenced data and careful consideration of which spatial data analysis method may be suited to the dataset.

The choice of cluster detection method is dependent upon the type of data to be analysed, whether it be represented as discrete points in space, point data, or as an aggregation of individual events, such as counts of the number of cases, within fixed areal units (Bailey and Gatrell, 1995). Several reviews of spatial analytical techniques have been published in the past 5 years (Waller and Jacquez, 1995; Gatrell and Bailey, 1996; Jacquez, 1996a; Moore and Carpenter, 1999; Ward and Carpenter, 2000a; Carpenter, 2001).

In the case of point data analyses, the focus of interest is on the distribution of the event locations themselves, e.g., the distribution of disease cases or controls. Briefly, the analysis of point data has two main approaches, broadly defined as distance-based, or quadrat-based methods. Distance-based methods use a test statistic based on measuring distances between disease cases, while the other is based on studying the variability of case counts in certain subsets of the study region, often called quadrats (Kulldorff and Nagarwalla, 1995). It is important to know exactly what a particular method can detect. There are those that are able to detect whether clustering is present in the data but cannot detect their location, e.g., Moran (Moran, 1948), Cuzick-Edwards' (Cuzick and Edwards, 1990), *k*-function analysis (Diggle *et al.*, 1995), and Whittemore (Whittemore *et al.*, 1987).

There are many situations where the location of the clusters is as important a question to answer as whether or not the detected clusters are significant. Tests for the detection of clusters which also graphically identify them include those methods described by Openshaw *et al.* (1987, 1988) and Besag and Newell (1991). Both methods graphically

identify possible clusters by using a multitude of overlapping circles as quadrats. The method by Openshaw and co-workers, called the geographical analysis machine (GAM), uses multiple overlapping circles of variable size that centre on points at regular intervals within an area of interest. The method compares the disease risk within the circle to the disease risk of the whole population. With the GAM, a criticism of the method is that a separate significance test is made for each circle, leading to the problem of multiple testing. This may identify a number of false positive clusters, i.e., those that have occurred due to chance alone (McKenzie, 1999).

The spatial scan statistic, and the space-time scan statistic, were developed with the purpose of testing for the significance of clustering within space and/or time within a geographical data set, while simultaneously identifying the location of the clusters (Kulldorff and Nagarwalla, 1995; Kulldorff *et al.*, 1997). The method uses a moving circular window, the diameter of which varies at each point location up to a maximum set in the analysis. This has the advantage of being useful in a dataset that contains points with a variable density, so the test can be applied to aggregated as well as non-aggregated data. The problem of pre-selection bias when testing for purported clusters is also overcome because it tests for clustering within areas of all sizes at each location. The method compares the risk of disease within each circle to that outside the circle, using a hypothesis testing technique that overcomes the multiple testing problem (Kulldorff and Nagarwalla, 1995). Another advantage of this technique is that it lists the geographical coordinates of the points included within significant clusters, allowing the location of the clusters to be mapped. The analysis may be performed using a Poisson distribution if the underlying population at risk is known, or a Bernoulli distribution if the data are in case-control format. The spatial scan statistic and space-time scan statistic have been used to investigate clusters of childhood leukaemia (Hjalmars *et al.*, 1996), childhood astrocytoma (Hjalmars *et al.*, 1999), breast cancer (Kulldorff *et al.*, 1997), and to evaluate a brain cancer cluster alarm (Kulldorff *et al.*, 1998).

As discussed in Chapter 1, Section 1.10, the only spatial data analytical studies of canine neoplasia reported in the literature are those by O'Brien *et al.* (1999, 2000). The paucity of studies of the spatial and temporal components of canine neoplasia occurrence are likely to reflect the lack of readily available, spatially-referenced canine neoplasia data, and the

well-recognised lack of accurate, up-to-date background population data for this species (Chapter 1, Section 1.6.1). Epidemiological investigation of the geographical distribution of canine neoplasia could lead to further understanding of the aetiology and pathogenesis of canine tumour types, with direct benefits for the veterinary care of the species. Useful information could also be provided for the human medical field, given the suitability of the dog as a sentinel for disease occurrence (Chapter 1, Section 1.8).

The objectives of the current study were to apply the spatial scan statistic and the space-time scan statistic to the CIDRU canine biopsy dataset described in Chapters 3 and 4. Issues concerning the spatial references available for the data, and subsequent interpretation of the results are also discussed.

6.2 MATERIALS AND METHODS

6.2.1 Data Source

The study dataset was derived from the CIDRU database, previously described in Chapter 2. Univariable and multivariable analyses had shown that practice was a significant variable on the outcome of neoplasia in a canine biopsy submitted to the CIDRU diagnostic histopathology service. As explained in Chapter 4, Section 4.2.2.3, due to the instability that would be caused by inclusion of practices submitting very low numbers of neoplastic or non-neoplastic biopsies (less than or equal to five of either type) in statistical analyses, it was decided to include only those biopsies submitted from practices contributing more than five neoplastic and more than five non-neoplastic biopsies in the spatial and temporal analyses, as was decided for analyses including practice as a fixed or random effect (Chapter 4, Section 4.2.2.3).

6.2.2 Data preparation

As previously described in Chapter 3, Section 3.1.2.3, address details for each practice had been verified by checking the data in the CIDRU database "Vet" file against entries in the Royal Mail Postcode Directory. Where practices were not listed, they were contacted by telephone using the telephone number listed in the CIDRU database to obtain correct

details, particularly practice postcode. Microsoft Access was used to create a table containing the unique veterinary practice identifier, *vetref* and relevant postcode. This table was linked by *vetref* to a table containing biopsy data pertinent to those practices eligible for inclusion in further analyses (107 practices contributing 16536 biopsies). Because temporal analyses were also planned, Microsoft Excel functions were used to create a field for each biopsy indicating the year of biopsy submission. This field, linked by the unique biopsy identifier *histref*, was imported into Microsoft Access in preparation for the creation of files suitable for analysis by spatial statistical software.

6.2.2.1 Geocoding

To give each biopsy a spatial reference, the postcode of the submitting practice was used. Postcodes in the UK are precise point locators, corresponding to grid references and have been used for many years by the health services to perform area-based analyses. The Royal College of Veterinary Surgeons (RCVS) produce an annual publication, the Registry of Veterinary Practices, which contains information on the practices' services and specialties, as well as current address and contact details. In 1999, the RCVS Registry was made available in electronic format, and a copy of the practice database was requested by the Comparative Epidemiology and Informatics Research Group based at GUVS. When received, the data were transferred to a Microsoft Access database and a file containing practice names and postcodes was created. This file was saved onto floppy disk and mailed to Graphical Data Capture Limited for geocoding (Chapter 1, Section 1.10).

The geocoded data were received by return of post and imported into MapInfo, a GIS software program. Using the customized Standard Query Language (SQL) in MapInfo, a table was created containing the geographical coordinates and address details for all practices in the RCVS Directory, and this table then imported into Microsoft Access. Postcode data in the RCVS practice file and the CIDRU practice file were used to link the tables, enabling the spatially-referenced data in the RCVS file to be allocated to the CIDRU practices, in preparation for spatial analyses.

6.2.3 Statistical analysis

The designation of the practice data set as containing cases (neoplastic biopsies) and controls (non-neoplastic biopsies) and its aggregated nature (biopsies aggregated by practice) led to the choice of the spatial analysis method developed by Kulldorff *et al.* (1995), described in Section 6.1. The method uses point data, which in this study was the location of each of the 107 practices contributing biopsies to the study dataset (16536 biopsies).

The spatial scan statistic was applied to the data assuming a Bernoulli distribution with uniform risk over space under the null hypothesis of absence of spatial clustering. The scan statistic applies a moving circular window that centres on each point in turn. As the window moves over the study area it defines a collection Z of zones $Z \subset G$, where G is the total number of points. At each position, the radius of the window varies so that the window includes a minimum of zero neighbouring points up to a maximum number set in the analysis. For each window the method tests the null hypothesis against the alternative hypothesis that there is an elevated risk of neoplastic biopsies within, compared with outside, the window. For each window, the numbers of neoplastic biopsies inside and outside the window are noted. On the basis of these numbers, the likelihood function for the Bernoulli model is calculated for each circular window, expressed as:

$$L(Z,p,q) = p^{n_Z}(1-p)^{\mu(Z)-n_Z} q^{n_G-n_Z}(1-q)^{(\mu(G)-\mu(Z))-(n_G-n_Z)} \qquad \text{equation 6-1.}$$

where n_Z is the number of observed points in zone Z, n_G is the total number of observed points, p is the probability of an individual point being within zone Z, with the probability for individuals outside the zone is q. The measure corresponds to an individual in either one of two states, in this example, with a neoplastic or non-neoplastic diagnosis. The window with the maximum likelihood is denoted the "most likely cluster". The expected distribution of the likelihood ratio test statistic under the null hypothesis is obtained by repeating the same analytical exercise for a large number (999) of random replications of the dataset through a Monte Carlo simulation process. It is then used to estimate a P-value, expressing the probability of obtaining the observed cluster under the null hypothesis. The

null hypothesis is rejected at an alpha level of 0.05 if the simulated P is less than or equal to 0.05 for the most likely cluster.

In addition to the most likely cluster, the spatial scan statistic identifies secondary clusters in the dataset and can order them according to their likelihood ratio. Secondary clusters are reported if their likelihood ratio is larger than the likelihood ratio for the most likely cluster for at least one dataset simulated under the null hypothesis (i.e., p <1.0). The software does not report any sets of practices that partly overlap the most likely cluster and that have a likelihood almost as high, because most of them provide little additional information. As a result, the results of the analysis must be considered as representing the approximate location of a cluster whose exact boundaries are uncertain (Kulldorff *et al.*, 1997).

6.2.3.1 Spatial analysis

Spatial analyses were conducted at two different spatial scales using the software implementation of the spatial scan statistic, SaTScan. The first analysis was conducted using the maximum size of a window as recommended by the developers of SaTScan, to include up to 50 per cent of the study population. So, for each practice, the moving circular window was centred at its coordinate point, and the radius increased continuously until 50 per cent of the total population of biopsies was included. The setting of a maximum size is recommended to avoid the situation whereby the spatial window becomes so large that it is inappropriate to refer to clusters in that zone as representing any mechanisms of epidemiological relevance. During the analysis, the algorithm generates an infinite number of distinct circular windows, with different sets of neighbouring practices within them, and each being a possible candidate for a cluster (Kulldorff *et al.*, 1998).

It was of interest to see whether any significant clusters, if found, could be decomposed into multiple overlapping subclusters, each of which would allow rejection of the null hypothesis of complete randomness on their own strength (Kulldorff *et al.*, 1997). To examine this possibility, the maximum extent of the spatial window was limited to include 10 per cent of the study population in a second analysis. The location of clusters found in this way would identify practices in close proximity with biopsy data which showed a tendency to clustering.

6.2.3.2 Temporal analysis

Temporal analysis was performed using the temporal scan statistic. This has a window that moves in one dimension (time), defined in the same way as the height of the cylinder used by the space-time scan statistic (Section 6.2.3.3). This means that it is flexible in both location and size, covering anything from the length of a pre-defined time interval to the maximum temporal cluster size specified in the analysis (Kulldorff *et al.*, 1998). Temporal analyses were run first using the maximum recommended time window being 90 per cent of the total study period (11 years) then to include 16 per cent of the total study period (2 years).

6.2.3.3 Space-time analysis

Space-time clustering of annual neoplastic biopsy submission from practices was evaluated using the space-time scan statistic. The space-time scan statistic is similar to the spatial scan statistic except that it is defined by a cylindrical window with a circular geographic base and with height corresponding to time. The base is centred around each of the practices located throughout the study region, with the radius varying continuously in size. The height reflects any possible time interval of less than or equal to half the total study period, as well as the study period as a whole. The window is then moved in space and time so that for each possible geographical location and size it also considers each possible time interval. In effect, an infinite number of overlapping cylinders of different size and shape is obtained, jointly covering the entire study region. Each cylinder reflects a possible cluster (Kulldorff *et al.*, 1998). Cases were assumed to be Bernoulli distributed with uniform risk over space and time under the null hypothesis, and with different risks inside and outside at least one of the cylinders under the alternative hypothesis. The maximum settings as recommended for both space and time windows were used in the first analysis, being 50 per cent of the study population and 90 per cent of the study years (11 years). A second analysis was then performed, with the maximum space window set to include 10 per cent of the study population and the maximum time window set to include 16 per cent of the total study period (2 years). The location of these clusters would thus identify practices in close proximity with biopsy data which showed a tendency to cluster in consecutive years.

All spatial, space-time and temporal clustering calculations were performed using SaTScan, which has been designed specifically to implement the spatial scan and the space-time scan statistic (Kulldorff *et al.*, 1998). SaTScan produces a test file that lists the geographical coordinates and the identification of each practice included in each of the significant clusters. These data were then used in a GIS to map the locations of the clusters. Maps were produced using MapInfo (MapInfo Corporation).

6.3 RESULTS

The initial study population consisted of 18366 biopsies submitted from 436 veterinary practices, to the CIDRU diagnostic histopathology service at GUVS during January 1986 - June 1998. Once the inclusion criteria for practice had been fulfilled, a study population of 10905 neoplastic biopsies (cases) and 5631 non-neoplastic biopsies (controls) submitted from 107 practices located throughout the UK, was eligible for inclusion in the spatial and space-time analyses. The geographical distribution of the 107 eligible practices from the original 436 practices is presented in Figure 6-1.

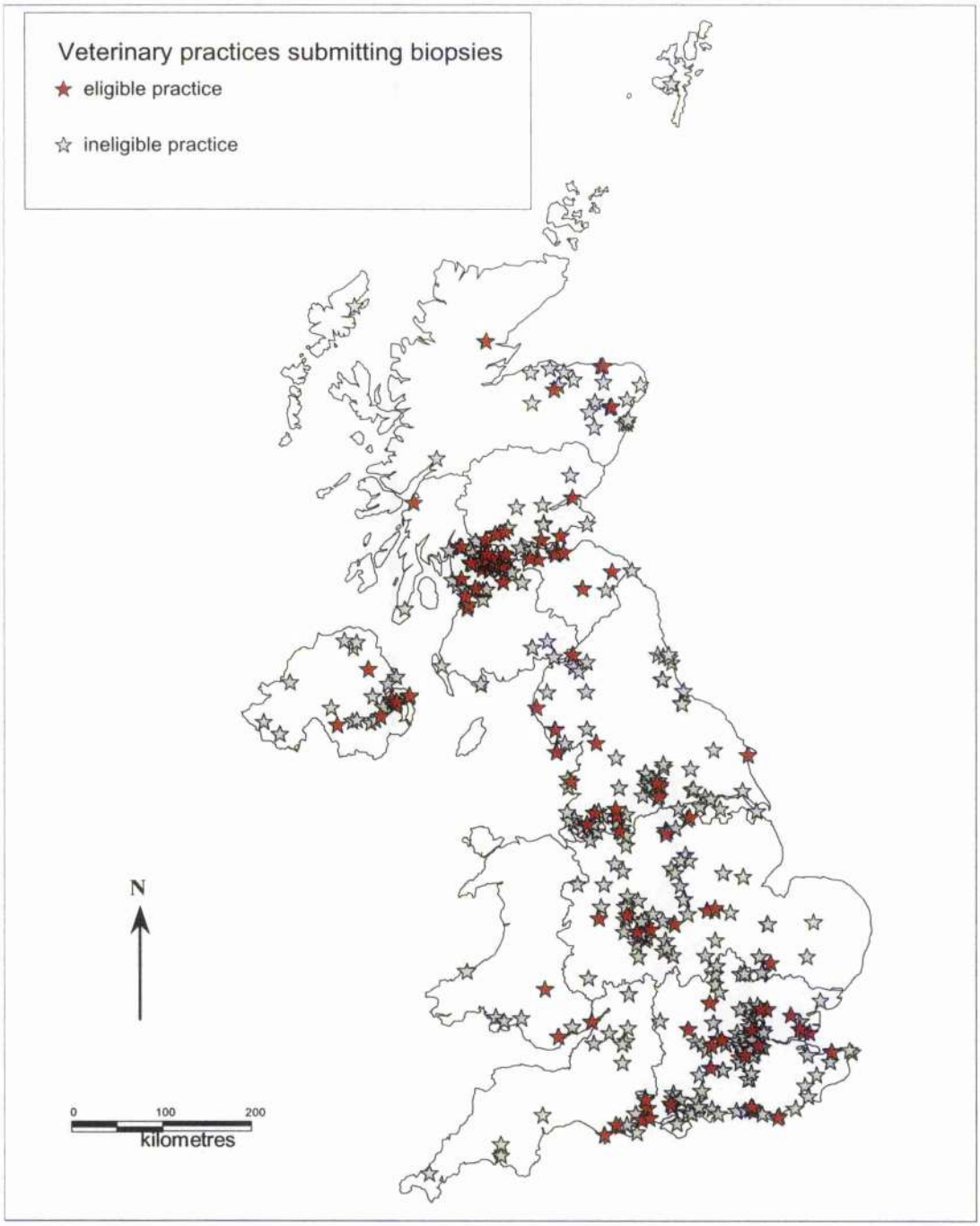


Figure 6-1 Map showing the location of practices contributing biopsies to the study dataset. Eligible practices for inclusion in spatial and spatio-temporal analyses were those contributing greater than five neoplastic and greater than five non-neoplastic biopsies during the study time period.

6.3.1 Spatial analysis

The most likely clusters and significant secondary clusters from each analysis are presented in Tables 6-1 to 6-4. The most likely high-rate spatial cluster from the analysis with the maximum spatial window of 50% included 32 practices located in central-southern England. The relative risk (RR) for neoplastic biopsies being submitted from these practices was 1.067 (p=0.001). The most likely low-rate spatial cluster with the maximum spatial window of 50% included 35 practices located in Scotland, Northern Ireland and northern England, with a RR for neoplastic biopsies from these practices being 0.915 (p=0.001). A secondary low-rate spatial cluster identified a single practice located in northern England with a RR of 0.264 (p=0.001). These clusters are displayed in table 6-1 and figure 6-2.

Cluster ID	Relative Risk	P-value	Number of practices in cluster
1	0.915	0.001	35
2	1.067	0.001	32
3	0.264	0.049	1

Table 6-1 Relative risk, P-value and number of practices for all significant spatial clusters with high or low rates and maximum spatial window (50%), shown in Figure 6-2.

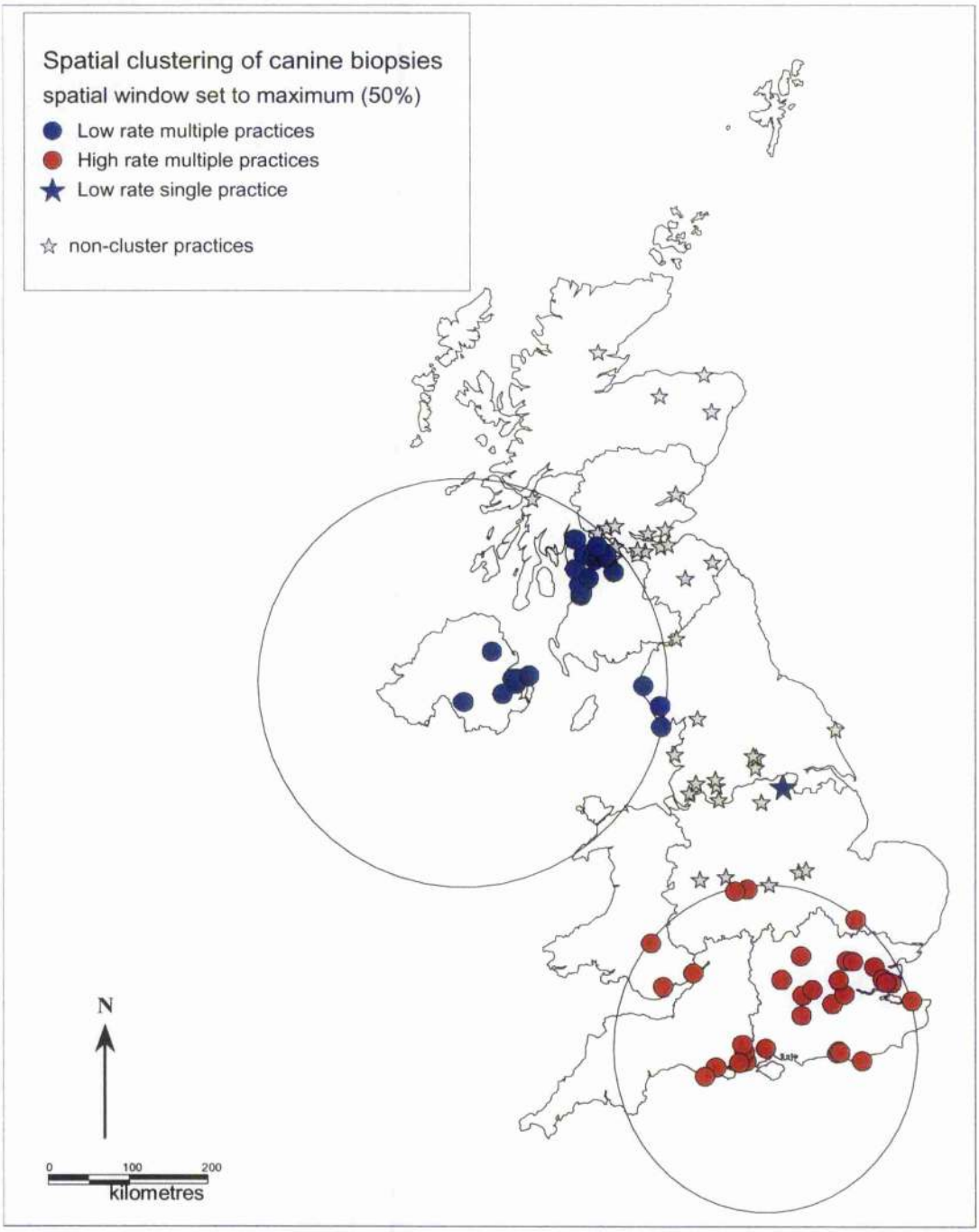


Figure 6-2 Location and radii of all significant spatial high and low rate clusters in the analysis with a maximum spatial window to include up to 50% of the study population.

6.3.2 Temporal analysis

Temporal analyses including 90 percent of the study time period showed the most likely cluster to include all 107 practices during 1988 - 1993 (RR 0.97, p=0.001). When the time window was reduced to include up to 16 percent of the study period, the most likely cluster again included all 107 practices, during 1989 - 1990 (RR 0.95, p=0.003). Because all practices were included, the results were considered meaningless in terms of significant temporal clustering.

6.3.3 Space-time analysis

Table 6-2 and Figure 6-3 show the most likely spatio-temporal clusters from the analysis with the maximum spatial window of 50% and maximum temporal window of 90%. The most likely low-rate spatio-temporal cluster included the same 35 practices in the north of the UK as for the spatial only analysis, with a RR of 0.899 (p=0.001) during the 1988 - 1998 time period. A secondary low-rate spatio-temporal cluster included two practices located in central England, with a RR of 0.181 (p=0.001), one of which was also identified as a cluster by the spatial only analysis. The most likely high-rate spatio-temporal cluster included 23 practices located in Wales and central/southern England (RR 1.122, p=0.001).

Cluster ID	Cluster period	Relative Risk	P-value	Number of practices in cluster
1	1988 - 1998	0.899	0.001	35
2	1990 - 1998	1.122	0.001	23
3	1988 - 1992	0.181	0.001	2

Table 6-2 Time period, relative risk, P-value and number of practices for all significant space-time clusters with high or low rates and maximum space (50%) and time window (90%), shown in Figure 6-3.

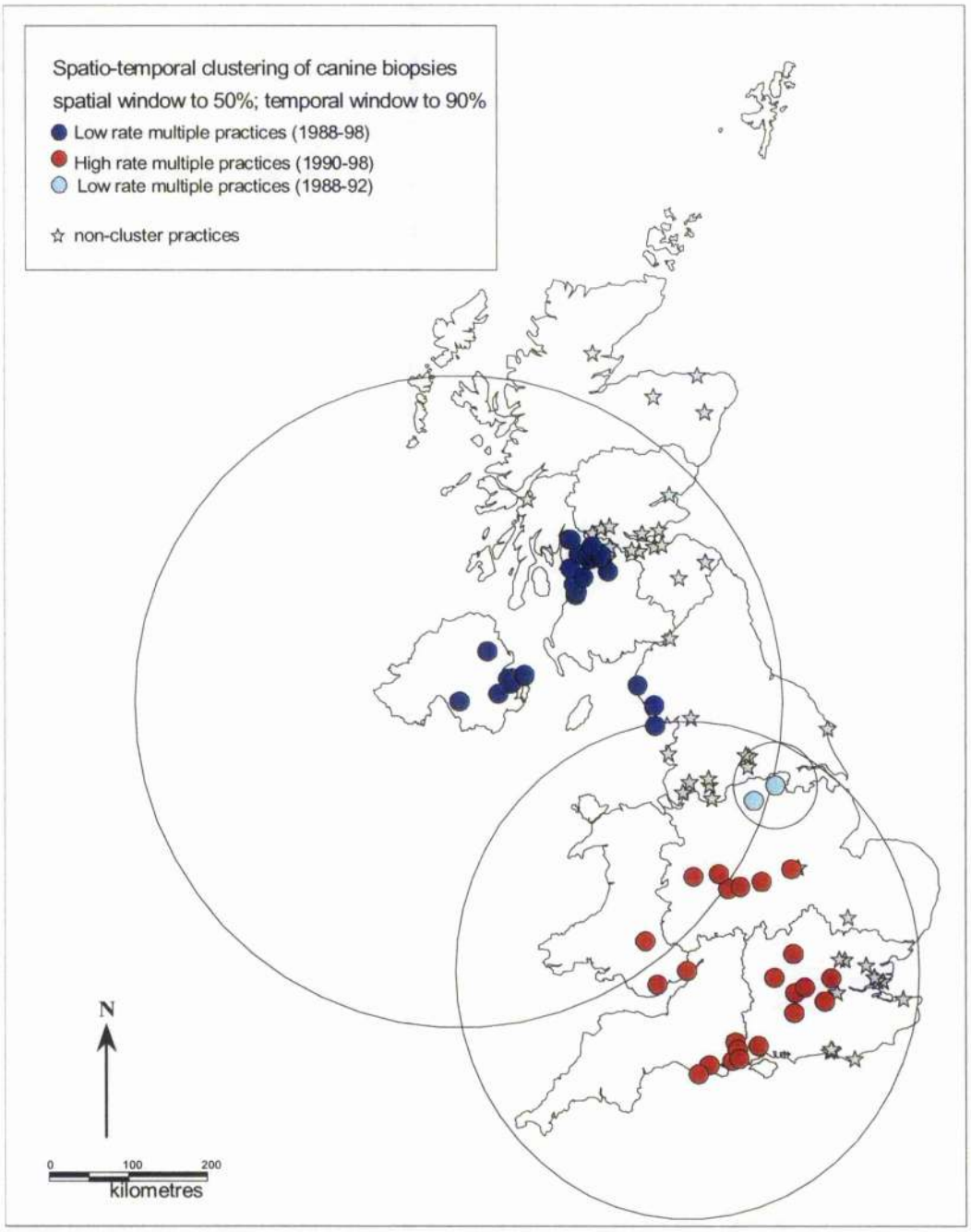


Figure 6-3 Location and radii of all significant high and low rate space-time clusters in the analysis with maximum space window to include up to 50% of the study population and time window to include up to 90% of years in the study time period.

When the above spatial and space-time clusters were further examined by reducing the size of the spatial and temporal scanning windows, three or more smaller subclusters of varying size emerged from each of the main clusters. In addition, new clusters were identified, including between one and six practices. These are presented in tables 6-3 and 6-4, and figures 6-4 and 6-5. When the spatial analysis was conducted with a space window restricted to 10%, the most likely high-rate cluster emerged from the original high-rate cluster, including six practices of the original 32 practices located in southern England.(RR 1.154, p=0.001). A secondary high-rate cluster including three practices (RR 1.154, p=0.006) and interestingly, two secondary low-rate clusters, each including a single practice emerged from the same group of 32 practices with a RR of 0.317 (p=0.001) and RR 0.315 (p=0.001), respectively. The original most likely low-rate spatial cluster which included 35 practices located in Scotland, Northern Ireland and northern England decomposed into three low-rate clusters, located in the Glasgow region (four practices, RR 0.537, p=0.001), north-west England (three practices, RR 0.746, p=0.001) and Northern Ireland (one practice, RR 0.484, p=0.001). Two new high-rate clusters were identified, one in southern Scotland (one practice, RR 1.183, p=0.001) and the other in north-west England (four practices, RR 1.084, p=0.018). Two new low-rate clusters emerged in Scotland, one in the Edinburgh region (two practices, RR 0.633, p=0.001), the other in the north (one practice, RR 0.491, p=0.034).

Cluster ID	Relative Risk	P-value	Number of practices in cluster
1	1.154	0.001	6
2	0.537	0.001	4
3	0.264	0.001	1
4	0.317	0.001	1
5	0.315	0.001	1
6	0.746	0.001	3
7	0.484	0.001	1
8	1.183	0.001	1
9	0.633	0.001	2
10	1.154	0.006	3
11	1.084	0.018	4
12	0.491	0.034	1

Table 6-3 Relative risk, P-value and number of practices for all significant spatial clusters with high or low rates and maximum spatial window of 10%, shown in Figure 6-4.

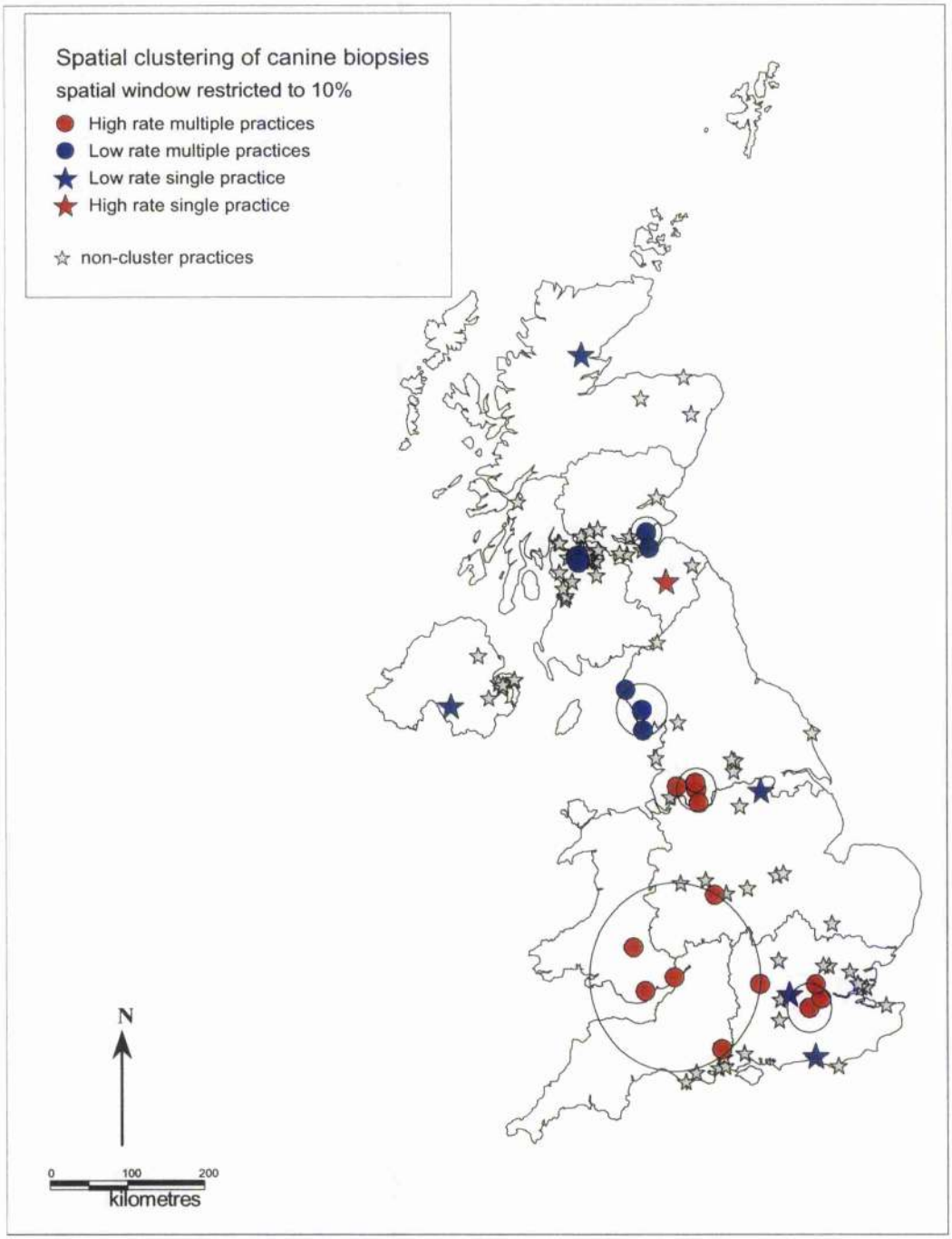


Figure 6-4 Location and radii of all significant high and low rate spatial clusters in the analysis with the spatial window restricted to include up to 10% of the study population.

When the spatial window was limited to 10% and the temporal window to 16% in the spatio-temporal analysis, three low-rate clusters emerged from the original most likely low-rate cluster, which included 35 practices in the north of the UK, during 1988 - 1998. These three clusters were located in the Glasgow region during 1991 - 1992 (three practices, RR 0.342, p=0.001), north-west England during 1989 - 1990 (three practices, RR 0.53, p=0.001) and Northern Ireland during 1992 - 1993 (one practice, RR 0.000, p=0.003). From the most likely high-rate spatio-temporal cluster including 23 practices in Wales and mid-southern England during 1990 - 1998, emerged one high-rate cluster located in Wales and mid-England during 1991 - 1992 (6 practices, RR 1.202, p=0.003) and two low-rate clusters, located in the London region during 1988 - 1989 (four practices, RR 0.152, p=0.001) and a single practice in mid-England during 1992 - 1993 (RR 0.00, p=0.008). A further three low-rate clusters also emerged, located in eastern Scotland during 1989 - 1990 (two practices, RR 0.00, p=0.001), southern England during 1990 - 1991 (one practice, RR 0.00, p=0.001) and the Edinburgh region during 1992 - 1993 (RR 0.515, p=0.015), and two high-rate clusters, located in southern Scotland during 1989 - 1990 (RR 1.348, p=0.039) and north-west England during 1995 - 1996 (RR 1.225, p=0.049).

Cluster ID	Cluster period	Relative Risk	P-value	Number of practices in cluster
1	1988 - 1989	0.213	0.001	2
2	1989 - 1990	0.000	0.001	2
3	1989 - 1990	0.530	0.001	3
4	1991 - 1992	0.342	0.001	3
5	1990 - 1991	0.000	0.001	1
6	1988 - 1989	0.152	0.001	4
7	1992 - 1993	0.000	0.003	1
8	1991 - 1992	1.202	0.003	6
9	1992 - 1993	0.000	0.008	1
10	1992 - 1993	0.515	0.015	3
11	1989 - 1990	1.348	0.039	1
12	1995 - 1996	1.225	0.049	4

Table 6-4 Time period, relative risk, P-value and number of practices for all significant space-time clusters with high or low rates and maximum spatial window of 10% and time window of 16%, shown in Figure 6-5.

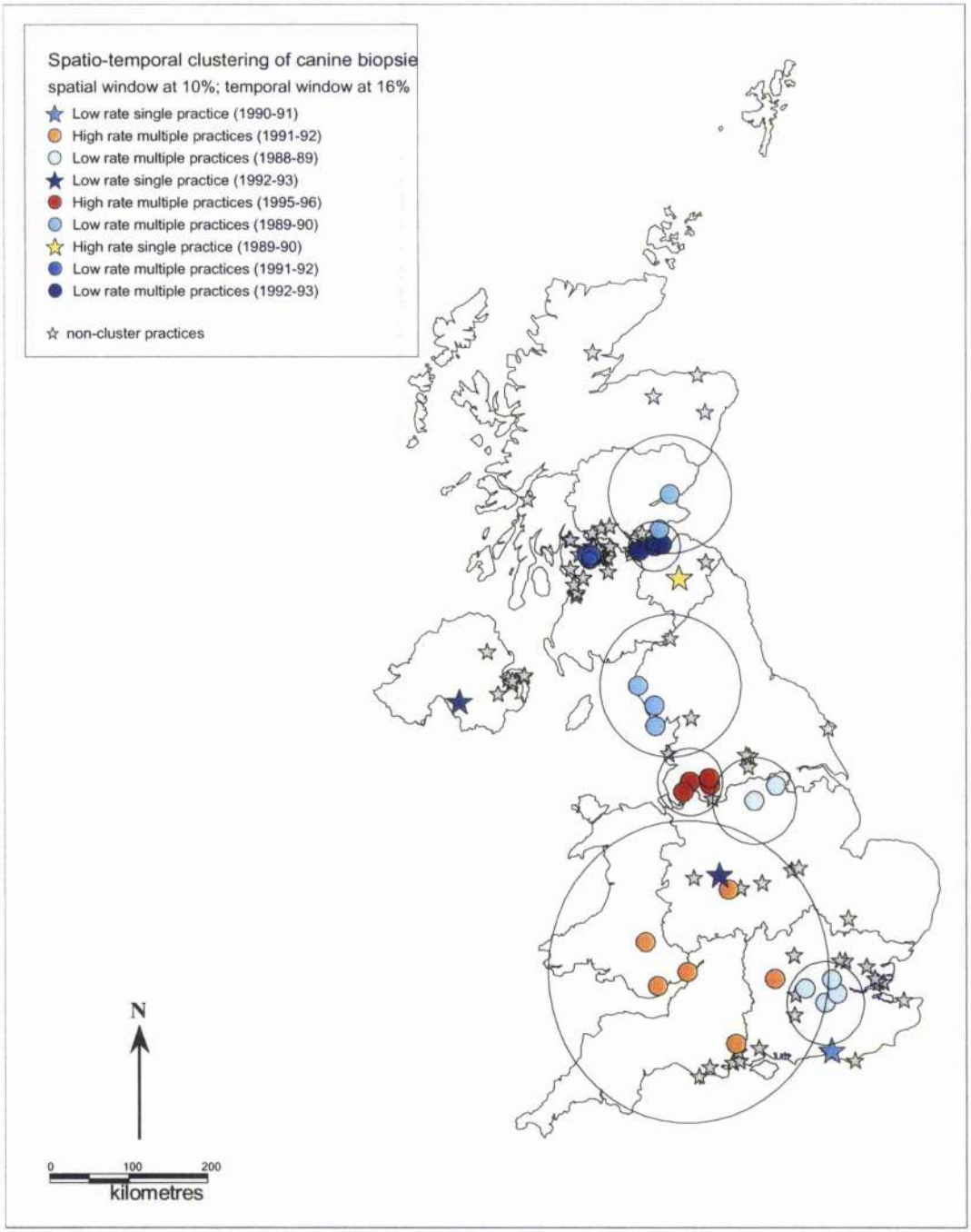


Figure 6-5 Location and radii of all significant high and low rate space-time clusters in the analysis with the maximum space window restricted to include up to 10% of the study population and the maximum time window restricted to include up to 16% of years in the study time period.

The majority of the significant clusters found with the maximum space and time windows covered large areas of the study region, illustrated in figures 6-2 and 6-3. It was interesting to note that the overall most likely clusters for both spatial and spatio-temporal analyses were low-rate clusters. The secondary clusters were of variable rate and size. Restricting the space and time windows resulted in the emergence of subclusters from the main clusters and new clusters. These analyses identified a number of single practices as being locations which allowed rejection of the null hypothesis of complete spatial randomness on their own strength.

6.4 DISCUSSION

For the purpose of this analysis, the biopsy population was spatially referenced according to the location of the submitting practice, because residential addresses for each dog from which a biopsy sample was taken, were unavailable. Ideally, spatial analyses of disease are conducted using residential data for optimal accuracy of disease occurrence within the designated study region (Gatrell and Bailey, 1996). However, there are strict controls governing data use to preserve client confidentiality in studies where subjects, or as in this study, the owners or guardians of the study subjects, have not been directly approached regarding use of information emanating from their association with the owner of the data source, being in this instance, the Department of Veterinary Pathology at GUVS.

The use of the practice postcode as the spatial reference for the biopsy data led to a high level of data aggregation at designated point sources, i.e., practice locations. The spatial and space-time statistics were specifically chosen as the statistical method to analyse this dataset because they have been developed to apply to aggregated as well as non-aggregated point data. A second level of aggregation was also present as practices are geographically aggregated in areas of high human, and therefore pet-owning, population density. The possibility of urban versus rural differences in the occurrence of neoplasia in dogs was mentioned in early reports of canine tonsillar squamous cell carcinoma which was described as a cancer type affecting dogs primarily from urban areas (Withers, 1939; Cohen, Brodey and Chen, 1964) (Chapter 1, Section 1.7.3.4.6). These early studies did not mention the possibility that apparent increase in cancer frequency in these geographical regions could be attributable to a relatively higher pet population density in areas of high

human population. The presence of an urban versus rural phenomenon, with significant aggregation of selected canine cancers in an area of high population density, was described by O'Brien *et al.* (1999) as a possible explanation for the aggregation seen in the study. Although the space and space-time scan statistics are designed to deal with aggregated data, it is pertinent to consider that the full effects of prior spatial aggregation of data by practice may not be fully accounted for by the statistical technique.

The spatial and space-time analyses with maximum space and time windows identified large clusters with both high- and low-rates of neoplasia occurrence within the study region. The most likely low-rate clusters were situated in the north of the UK, and the high-rate clusters occurred in the south of the country. It could be hypothesized that there may be geographical or environmental influences that caused such distinct trends in neoplasia risk, however this assumption is foolhardy due to the complexity of factors affecting the creation of this dataset. As discussed in Chapter 4, Section 4.4, owner-related, veterinarian-related and laboratory-related factors are all likely to have influenced data collection, and further investigations of these factors are required to quantify their effect, where possible, on the probability of a neoplastic diagnosis in a canine tissue biopsy submitted to the CIDRU diagnostic histopathology service. It was interesting to discover that by restricting the spatial and temporal scale criteria of the analysis the large clusters were partitioned into smaller clusters, with some high-rate clusters emerging from the main low-rate clusters and conversely for the main high-rate clusters. This process identified single or groups of practices in locations where the likelihood ratio for the increased or decreased risk of neoplasia within compared with outside the restricted window centred on these locations, was sufficient to cause rejection of the null hypothesis on their own strength. This leads to the formulation of a hypothesis that there may be factors relating to the practices identified by this method which influence the outcome of neoplasia in biopsies submitted from them and those practices in their close vicinity.

In the space-time analysis with restricted space and time windows, the reason for the emergence of four low-rate clusters was obvious, in that no neoplastic biopsies had been submitted from the identified practices during particular time periods. It would be highly implausible for there to have been no neoplasia occurrence. One explanation could be that

the veterinarians working at these practices had personal preference for one of the alternative histopathology services during those time periods.

Previous epidemiological studies of canine neoplasia have identified host-related and environmental risk factors for its occurrence (Chapter 1, Section 1.7). Chapter 4 of this thesis described an investigation of the host-related risk factors of age, gender and breed of dog, and site of biopsy, on the outcome of neoplasia in the biopsy population under study in these analyses, as well as investigating the role of submitting practice as both a fixed and random effect. Ideally, these covariates need to be accounted for in any spatial or spatio-temporal analysis as well. However, the form of the dataset, being case-control format with more than half of the study population being cases, precluded the use of the Poisson model in SaTScan, which is only statistically appropriate for a case-control study population if it contains a small numbers of cases compared to controls. The Poisson model in SatScan, primarily designed for studies where the background population-at risk is known, allows for covariates to be included in the analysis (Kulldorff *et al.*, 1998). The Bernoulli model, statistically appropriate for the data used in this study, does not allow adjustments to be made for covariates in the analysis. Many researchers have shown that canine neoplasia is generally a disease of older individuals (Priester and Mantel, 1971; Moulton, 1990; Dobson *et al.*, 2002), and age of dog was verified as a risk factor for neoplasia in a biopsy in the current study population in Chapter 4. Adjustment for known risk factors is essential to determine accurately the overall risk of neoplasia occurrence within the study population, so without this being possible, additional caution must be exercised regarding the interpretation of the results of these spatial analyses. For example, there may be substantial geographical heterogeneity with respect to the age of the dog population, with particular areas containing predominantly older dogs. If this was the situation, neoplastic cases could appear clustered in those areas and age-adjusted analyses would be needed to assess accurately the significance of such clusters (O'Brien *et al.*, 1999).

At present, there is no readily available, up-to-date information on the size and demography of the canine population in the UK (Chapter 1, Section 1.6.1). It could be argued that the canine background population could be estimated based upon human population Census data and data pertaining to levels of pet ownership in the UK. However,

the level by which the geographical density of the canine population differs from, or is similar to the human population has not been explored on a countrywide basis and results using such assumed data may be highly inaccurate and misleading. Human cancer epidemiology is a vast research area based upon data collected by population-based cancer registries throughout the world, with networks of regional cancer registries existing in many countries to facilitate data collection and analysis. With Census data providing background population data, the publication of cancer incidence data has become commonplace (Black *et al.*, 1997; Devesa *et al.*, 1999; Bray *et al.*, 2002). Currently, the paucity of background canine population data limits the potential uses of data collected by histopathology services, particularly for analyses where spatially-referenced data is essential. The major population-based canine cancer registries, the CANR and Tulsa Registries (Chapter 1, Section 1.6.3) were created prior to the availability of powerful computer hardware and software which are necessary for geocoding of data and computation-intensive spatial analyses. Due to the financial and personnel demands of such a registry, the establishment of a similar data source is unlikely in the immediate future in the UK. However acknowledgement of the existing data sources for canine neoplasia and recognition of their potential as tools for the investigation of geographical occurrence of canine neoplasia are required to expand upon the exploratory analyses presented here.

From this study, hypotheses may be generated that practice locations included within significant clusters may have practice-related factors associated with neoplasia occurrence in their patient population. The practices identified by reducing the study windows as being part of significant subclusters within the main clusters are a logical starting point for enquiry into practice-related factors. Further investigation in the form of questionnaire surveys could be considered to unravel the complex relationships within the practice environment that lead to biopsy procurement and submission to a diagnostic histopathology service. In addition to the factors described in Chapter 4, Section 4.4, the number of veterinarians working at each practice is likely to influence the absolute numbers of biopsy submissions per practice, which will impact the spatial analysis of data that is geocoded by practice location. This analysis included only biopsies from practices which fulfilled a predetermined inclusion criteria relating to the number and type of biopsies submitted. The vast variation in biopsy submissions, described in Chapter 4,

Section 4.3.1.2, was a direct indication of the significance of practice-related factors upon the dataset. Larger practices are likely to service larger geographical areas than smaller practices, however this aspect is likely to show great variation between urban and rural areas, which in itself would need to be considered as a potential risk factor. Distances travelled by owners with their dogs to the point location of practice may vary substantially. Urban practices are likely to serve a client base residing within a smaller radius, compared to their rural counterparts, even though practice size, in terms of number of veterinarians and client population, may be similar. Within that radius may be differing environmental exposures between the rural and urban populations, the study of which requires far greater resources than those available for the current study. The possibility exists that the spatial aggregation of cancer cases is not directly related to any population or demographic factor, but is actually determined by local environmental risk factors that play significant causal roles in the pathogenesis of the diseases.

The presence of a socioeconomic gradient in health and neoplasia has been documented in humans (Smith *et al.*, 1992; MacKie and Hole, 1996) and the role of socioeconomic differentials in pet mortality has been debated (Smith and Bonnett, 1998). Socioeconomic factors may influence biopsy submission at the veterinarian and owner level and could be responsible partly for the pattern of results seen in this analysis, of low-rate clusters in the north of the country, and high-rate clusters in the south. In general, the northern regions of the UK are recognised as having a lower socioeconomic status than the south. In geographical areas with low socioeconomic status, a veterinarian may be less inclined to submit a biopsy from a lesion that clinically, based upon known biological behaviour and appearance, is considered to be benign. Additionally, where a lesion is considered to be highly malignant and the owner is financially compromised, the most practical outcome for the dog may be euthanasia, rather than histopathological confirmation of a fatal malignancy. Thus, less neoplastic lesions may be submitted, reducing the likelihood of a high-rate cluster being detected.

It is reiterated that identification and quantification of practice-related factors is an important area for further investigation. There may be further benefits, with the results of such analyses potentially being used to study the use of practices by owners, such as has been implemented to examine population movements to key services in the NHS (Bullen,

Moon and Jones, 1996). Patterns of owner to veterinarian or veterinary practice allegiance could be determined, and the number and geographical distribution of those owners with tendencies to use more than one veterinary practice may be revealed. This could lead to investigations of the causes for such behaviour and allow the development of management practices designed specifically to reduce the phenomenon of multiple practice usage in that sector of the pet owner population.

This analysis suggests overall trends for canine biopsy data in the UK, with statistically significant differences between the north and south of the country, with main low-rate and high-rate clusters, respectively. It is interesting to note that there are distinct geographical regions of the UK from which there were no practices submitting sufficient biopsies to allow their inclusion in the analysis. However, it must be remembered that this data source is only one of several data sources for canine biopsy data, and is comparatively small next to a similar dataset obtained from a commercial source (Chapter 5). Only biopsy data from 107 practices, submitting tissue samples to one diagnostic histopathology service, have been examined in this study. During the last year of data used (1998), 3073 veterinary practices were registered with the RCVS throughout the UK. The actual number of veterinary practices was undoubtedly higher because registration with the RCVS was, and still is, an optional procedure. Available to these practices were 17 diagnostic histopathology services, with one at each of the six UK veterinary schools and 11 in the private sector. Thus, the dataset presented here represents a miniscule proportion of the canine biopsy population present in the UK during the study time period (1986 - June 1998) and interpretation of the results can only apply to this biopsy population. Nevertheless, were there no spatial or temporal factors affecting the dataset, there would have been no significant clusters identified. Whilst these clusters may not be definitive, they do point towards underlying processes which are spatial and temporal in nature.

This exploratory analysis achieves the aim of demonstrating the application of spatial techniques to a secondary data source where data collection has been undertaken for a purpose other than that for which the study was designed. This is similar to the spatio-temporal studies of canine neoplasia conducted by O'Brien *et al.* (1999, 2000), so far as the canine data was obtained from another veterinary records database, the VMDB, described in Chapter 1, Section 1.6.4.1. However, the data sources are fundamentally

different because the VMDB contains data originating from referral hospitals, and thus referral bias and other factors associated with case selection are present. To the author's knowledge, the current study is the only example of spatial analytical techniques being applied to canine neoplasia data obtained from first-opinion veterinary practices, which allows the results to be considered with direct reference to the veterinary environment that impacts most strongly on the general canine population.

The principal aim of the current study was to apply spatial analytical techniques to a canine biopsy population for the purpose of identifying whether aggregation of neoplasia was present, using a method which also provided the location of any significant clusters. The analysis enabled the generation of hypotheses concerning the geographical occurrence of canine neoplasia in the UK. Significant spatial and space-time clustering was found in this biopsy population across the study region, however, the results need to be interpreted with caution in light of unquantified veterinary practice-related, or sociological and economic influences upon the submission of biopsies to the study population. Their effects upon the likelihood of diagnosis of neoplasia in canine biopsies require further investigation.

CHAPTER 7

GENERAL DISCUSSION AND CONCLUSIONS

The studies presented in this thesis describe the use of canine biopsy data collected by diagnostic histopathology services for epidemiological investigations of the risk factors associated with neoplasia being diagnosed in the biopsy populations. In addition to exploring host-related risk factors, mixed-effect logistic regression modelling was used to account for any extra-binomial variation due to the group effect of submitting practice in the study population. Finally, the geographical distribution of canine biopsy submission and the occurrence of neoplasia within the study population was examined using spatial and spatio-temporal statistical methods. Examination of database structures and preparation of the data sources revealed the importance of data quality assurance for epidemiological studies and confirmed many of the previously identified risk factors for neoplasia. The results also suggest differing geographical trends in canine neoplastic biopsy occurrence. To further examine possible environmental factors and their role in the aetiology of canine neoplasia, first there is a need for examination of practice-based risk factors for their effects upon the investigation and diagnosis of canine neoplasia in the UK.

Detailed investigation of risk factors for canine neoplasia began with two population-based studies, the CANR (Dorn *et al.*, 1968a; Dorn *et al.*, 1968b) and Tulsa Registry (MacVean *et al.*, 1978), described in Chapter 1, Section 1.6.3.1. Prior to collecting data for risk factor analysis, both studies questioned all veterinary practices in a defined study area about the size of the animal population serviced by the practice, to obtain a background population-at-risk for further studies. Such census studies are costly and time-consuming to perform. As a result, these two studies remain the largest of their type and are still quoted widely, even though it is now over 25 years since their inception. It is interesting to note that no comment was made regarding the level of practitioner, and more importantly, owner compliance to the proposed aim of submission of all suspected neoplasms.

Because of the paucity of canine population data, the most common study design for epidemiological studies of canine neoplasia has become the case-control study. This study

design requires the selection of a control (non-diseased) group in an unbiased manner from those subjects which would have been included in the case (diseased) series had they developed the disease under study (Kelsey, Moore and Glickman, 1998). From the CANR studies, the Veterinary Medical Database (VMDB) was established (Chapter 1, Section 1.6.3.2.1) (Anonymous, 2002b) and has provided the data for many case-control studies of canine neoplasia. Referral bias is inherent in all studies utilising VMDB data because all medical records within the database originate from referral veterinary teaching hospitals. In order to remove this source of bias from epidemiological studies of canine disease, some researchers have explored alternative data sources, such as veterinary practice records (Reid-Smith *et al.*, 2000), pet insurance company databases (Wood *et al.*, 2000a; Dobson *et al.*, 2002) and veterinary pathology department records (Goldschmidt and Shofer, 1992; Yamagami, 1996). The use of these first-opinion data sources has been viewed as necessary to generate study results with direct relevance to the general canine population, though it is well recognised that non-veterinary-attending segments of the canine population, such as roaming strays and dogs cared for by owners who do not use veterinary services, remain unaccounted for in all studies using the data collected by members of the above categories. The exclusion of these same segments of the dog population from the current studies is acknowledged.

The aims of the research presented in this thesis included examination of the structure and function of two data sources derived from first-opinion veterinary practices, initially to determine their suitability for epidemiological studies. The databases of two diagnostic histopathology services which receive biopsies from dogs attending veterinary practices throughout the UK were selected for interrogation. Although similar in function, i.e., providing histopathological assessment of tissue biopsies, one database was owned by an academic institution and the other by a commercial organisation. The data recording structure of each was designed according to their respective requirements, with the academically-owned database containing features beyond those required solely for data entry and report generation. The relational database structure of the CIDRU database, described in Chapter 2, and the linear structure of the commercial database, described in Chapter 5, Section 5.2.1, were crucially different in the areas of data entry checks, data detail, and ease of data extraction. The database structures themselves differed primarily due to the commercial organisation's requirements for maximum efficiency in the use of

disk space and their decision to adapt an existing laboratory management system for their own purposes. The creators of the CIDRU database chose a relational database design for its flexibility and relatively efficient use of disk space, as well as integrating a user-friendly query language which could be used by personnel who possessed little prior knowledge of software programming. This last feature was in direct contrast to the commercial database, which required the writing of specialised computer code to facilitate data extraction for the studies presented in Chapter 5.

There were differences in database content, the most significant being the inability to record breed data in the commercial database. For the study of canine disease, breed has been recognised for many years as being a risk factor for certain conditions, including particular types of neoplasia, for example, bone tumours (Tjalma, 1966; Ru, Terracini and Glickman, 1998) (Chapter 1, Section 1.7.2). The presence of breed data in the CIDRU database allowed breed to be controlled for in the multivariable analyses presented in Chapter 4 and identified breed predispositions to neoplasms from particular anatomical sites in the study dataset.

The foresight of the developers of the CIDRU database for its potential as a data source for future epidemiological analyses is also to be noted due to the development of coding systems for particular data fields to facilitate searching and extracting data. Coding and nomenclature systems in human medicine, particularly in cancer registries, are paramount to epidemiological studies and are constantly being reviewed and updated to improve data accuracy for researchers and healthcare professionals (Anonymous, 2002a). The coding system for the CIDRU database (Chapter 2, Tables 2-2 and 2-3) although simplistic in comparison to others such as SNVDO, SNOMED and ICD (Chapter 1, Section 1.6.2) greatly assisted the data cleaning process of the CIDRU data even though there were significant numbers of data errors present (Chapter 3). By using a hierarchical and iterative data cleaning approach, data entry errors were able to be identified qualitatively and quantitatively. Thus, the strengths and weaknesses of the CIDRU database regarding data input were exposed, which allowed specific suggestions to be made for ways to improve overall data quality. The introduction of a breed data field in the commercial database would significantly improve the usefulness of the database content for epidemiological studies of canine disease. Full comparisons between results of the studies using this data

source, presented in Chapter 5, and the risk factor studies presented in Chapter 4, would also have been possible. However, without a universally accepted and implemented coding and nomenclature system for veterinary data, there will inevitably be limitations to making detailed comparisons of risk factor studies conducted from different data sources.

In-built data quality checks were limited in both databases, and resulted in varying degrees of data incligibility and extensive time commitment in preparing the data for statistical analysis. Although there were coding systems for some data fields in the CIDRU database, the lack of programmed range checks, cross-referencing and safety checks resulted in data entry errors such as nonsense values and incorrect data type for the field. Certain data fields in the commercial database caused particular problems, because all demographic data fields had free text entry. Data cleaning is essential for all datasets prior to statistical analyses, but is generally underreported (Maudsley and Williams, 1999). For both databases interrogated for this thesis, the data cleaning processes were informative for quality assurance and highlighted problem data entry areas. Reporting of these findings was important to illustrate the processes required prior to data analysis when utilising data collected for purposes other than the intended epidemiological study.

Case-control studies are designed to compare events or exposures common to both the diseased group (cases) and non-diseased group (controls) to identify any predisposing characteristics or risk factors for the disease in question (Lilienfield and Stolley, 1994; Kelsey, Moore and Glickman, 1998). Of primary interest for the case-control studies conducted in Chapters 4 and 5 was the assessment of the effect of previously identified host-related risk factors on the outcome of neoplasia in these first-opinion-derived, biopsy populations. Because records within the databases under scrutiny were generated by biopsy submissions, it is important to note that the results of these studies are directly referable to a biopsy, not dog, population. The relational database design of the CIDRU database contained a unique identifier field for the dog from which the biopsy was taken. This field allowed checks to be made for multiple biopsy submission from the same dog. The results of this procedure showed that during the 12.5 year study period, a small percentage (6.9%, 1158/16877) of dogs had contributed multiple biopsies, and of those biopsies, only 0.01% (213/18366) were not of a biologically distinct nature. This extremely low number of biopsies was considered to have negligible effect on further analyses, so were retained in

the study dataset. The commercial database did not contain a unique identifier field for dog, but recorded owner name and address to assist accurate biopsy report delivery to the practitioner. However, due to data confidentiality issues, the organisation did not release owner-related information for this study. Consequently, cross-referencing of biopsy submission by owner and practice was not possible to identify multiple biopsy submissions at different times during the study time period, although multiple biopsies under the same "*labref*" number were easily identified by inspection of the record. These records were excluded to allow analyses to be conducted and results interpreted knowing that each record represented one tissue biopsy, thus preserving the unit of investigation as a biopsy. The assumption was made that the probability of submission of biologically similar biopsies from the same dog during the eight month study period (January - August 1999) of the commercial dataset was likely to be extremely low because the number of samples matching these criteria was negligible in the CIDRU dataset which spanned 12.5 years.

Results of multivariable analyses of the host-related variables in Chapters 4 and 5 unsurprisingly confirmed age as being associated with an increased odds of neoplasia in a canine biopsy. This was in agreement with other studies (Priester and Mantel, 1971; Moulton, 1990; Dobson *et al.*, 2002).

Female gender and age at time of neutering were first proved to be significantly associated with the occurrence of mammary neoplasia in the dog in the 1960s (Dorn *et al.*, 1968b; Schneider, Dorn and Taylor, 1969). Age at time of neutering was not known for the current studies. The association of gender and neutering status with the outcome of neoplasia differed between the two datasets under study, with neutered females showing a decreased odds of neoplasia in samples when controlled for age, breed and site in the CIDRU dataset, and gender being insignificant to the outcome in the commercial dataset. Over one-third of biopsy records in the commercial dataset contained gender data with no clarification of neuter status, in contrast to the CIDRU dataset, where the number of records with this feature was low enough to make their exclusion an appropriate method of removing this uncertainty from the data. The reasons for such a high proportion of records in the commercial dataset with incomplete gender and neuter status data are unknown, but may be due to a combination of factors including the level of detail supplied on submission

forms by practitioners, incomplete data entry, and lack of correction of gender data at the point of report generation.

Differing odds of neoplasia in a biopsy were found affecting the breeds selected for more detailed analysis, when controlled for age, gender and site of biopsy in the multivariable analyses presented in Chapter 4. Previous studies have identified certain breeds as being at greater risk for developing particular types of neoplasia though only one breed, the Boxer breed, has been singled out for being predisposed to neoplasia generally (Priester, 1967; Cohen *et al.*, 1974; Priester, Goodman and Theilen, 1977; Misdorp, 1996; Arnesen *et al.*, 2001). This predisposition was also found in the CIDRU biopsy population. Because there are likely to be genetic and heredity traits of each breed influencing the pathogenesis of neoplasia, breed must be regarded as an important variable in any study of canine neoplasia. There are 196 breeds recognised by the Kennel Club in the UK (Kennel Club, 2002b), of which 156 were represented in the CIDRU dataset. To maintain statistical stability in analyses including breed, it was necessary to collapse these many breed groupings, but this procedure was still insufficient to allow the effect of interaction terms to be assessed in the main effects model. The stratification of the data by site and construction of site-specific models partially addressed this problem. However, despite the large size of the dataset, and restriction of breed-specific analyses to the top eight represented purebreeds, sparsity of data in the covariate matrix resulted in non-convergence of two of the site-specific models. The results of these analyses illustrate the need for very large datasets to allow optimal exploratory data analysis. The introduction of breed recording to the commercial dataset would greatly increase its value to epidemiological study, particularly due to the far greater accumulation of data over time compared to the academic database. At the time of writing, approximately 1500 biopsies per month are being submitted to the commercial histopathology service, compared to 100 biopsies per month to the CIDRU histopathology service, leaving no doubt regarding the commercial service's potential as a data source.

Results of multivariable analyses of both datasets showed there was an increased odds of neoplasia in a biopsy from the mouth/pharynx, mammary gland, lymphatic and reproductive systems, when controlling for other selected host-related variables. An extension of these studies would be to examine the benign and malignant biopsy

populations independently. The CIDRU database contained a coding system for diagnosis, allowing rapid identification of biopsies with a neoplastic diagnosis. However, the only method available for differentiating benign from malignant tumours was direct examination of the diagnosis field. Accurate assessment of sample morphology is the role of the histopathologist, who makes a diagnosis of malignancy based upon information regarding biological behaviour of the mass from which the biopsy was taken, and the presence or absence of specific criteria that identify the biopsy as being malignant (Henderson and D'Andrea, 1993; Stone, 1995). Diagnostic terminology may indicate malignancy in many cases. However, using only this information would be likely to lead to misclassification in an indeterminate number of cases unless review of the tissue sample themselves, by a histopathologist, was performed in those cases where diagnostic terminology did not provide a definite answer to the biological behaviour of the biopsy.

The commercial dataset presented an even greater challenge regarding the differentiation of benign from malignant neoplasms. There was no coding system to assist in identification of neoplasms from other classes of biopsies, so direct examination of the diagnosis field was essential for this purpose. Data cleaning revealed absent or nonsense data entries in the diagnosis field of 36.7% of records that were included the final dataset. For these records the pathologists' reports had been reviewed to obtain the diagnostic information necessary for coding of biopsies as cases or controls. Further coding for biological behaviour in both datasets was considered beyond the scope of the current study both due to the additional time required for the task and the likely need for review of tissue biopsies by histopathologists who were already committed to providing the high level of service expected by their clients.

The histopathology data examined in this thesis was generated by veterinary practitioners working in first-opinion practice. Statistical analyses are generally conducted under the assumptions that individual subjects in the study population are independent of each and identically distributed in the population. However, the data under study violated these assumptions because they were derived from discrete practices, causing them to be grouped by practice. Veterinary practitioners will undergo a decision-making process for every canine biopsy submission, involving clinical assessment of the dog and consideration of the owner's wishes. Thus, it is likely that biopsy submission will be influenced by

practitioner, and therefore practice-related factors, leading to within-practice correlation in the data.

In multivariable logistic regression analyses, the presence of a group effect is likely to introduce extra-binomial variation, or overdispersion into the model. To account for this group effect, practice was introduced as an independent variable into the host-related main effects models. Prior to its inclusion, however, univariable analysis of practice in both datasets showed that there was wide variation in the number of biopsies submitted from each practice. In an attempt to achieve statistical stability in analyses including practice, only data from those practices contributing an *a priori* determined number of biopsies were considered eligible for further analyses. Mixed-effect models are well recognised as one method of accounting for extra-binomial variation (Curtis *et al.*, 1993; McDermott and Schukken, 1994; McDermott, Schukken and Shoukri, 1994). Inclusion of practice as a fixed-effect, the most common statistical method used, resulted in improved model fit in both multivariable analyses, indicating that extra-binomial variation was present in the data. However, the use of random-effects logistic regression was considered more appropriate to assess the effect of grouping by practice because the number of practices contributing to each dataset was large. The use of random-effects modelling requires definition of the data in the study as distinguishable or indistinguishable (Corcoran, Coull and Patel, 1999). This definition proved critical in the generation of random-effects models, and it can be debated as to which definition best fitted the data. To classify the data as distinguishable assumes that different practices have different policies regarding which biopsies to send for analysis, and that these differences are based upon the covariates of interest, i.e., age, gender and breed of dog, where data for breed are available, and site of biopsy. This may appear intuitively to be the case, however can only be determined by analysis of these covariates at the practice level, which was outwith the scope of the current study. Therefore, the classification of the data as indistinguishable, and the resultant random-effects models, may be correct.

The large population-based studies of canine neoplasia (CANR, Tulsa Registry, Chapter 1, Section 1.6.3.1) relied upon practitioners to submit all lesions suspicious of being neoplastic to the study, and offered free histopathology examination to overcome owner reluctance to allow biopsy due to the costs they would otherwise have incurred. The

number of suspected neoplasms that were not submitted to the service was impossible to report, and it is likely that practice-related factors, such as an individual veterinary practitioner's assessment of a lesion as non-neoplastic, or their judgement that procurement of the biopsy, which was most likely to involve general anaesthesia, would carry undue risk to the animal's life and so was not performed. Thus, a group effect due to practice was likely to be present in the registry data, but was not accounted for in subsequent statistical analyses. The development of powerful statistical software since these studies were conducted allowed for mixed-effect logistic regression to be used in the current studies, to account for a possible group effect (Chapter 4, Section 4.3.3 and Chapter 5, Section 5.2.8.4). However, in both the population-based studies and the biopsy population studies, unquantified practice-related factors are likely to have influenced data collection.

The CIDRU dataset was selected for the spatial and spatio-temporal analysis presented in Chapter 6, because of the presence of postcode data for practices contributing samples to the database, which allowed spatial referencing of the biopsy data. Spatial analytical methods were combined with a geographic information system (GIS) to describe and illustrate the spatial patterns present in the data. The results indicated geographical differences in the risk of the occurrence of neoplasia in a biopsy, depending upon the location of the submitting practice. Geographical factors that may play a causal role in the pathogenesis of disease include physical and environmental factors, socio-economic factors, and genetic factors. Ideally, the host-related covariates of age, gender and breed of dog, and site of biopsy, would have been controlled for in the spatial analyses. However, the lack of background canine population data restricted the type of spatial data analysis to one that did not allow covariates to be entered (Chapter 6, Section 6.4). Despite this deficiency, the analyses were considered useful due to their identification of overall geographical trends in the occurrence of neoplastic biopsies and the identification of a number of practices which influenced the results of the analysis when the study area was restricted in size in an attempt to identify localised trends.

When the time period for the space-time analyses was restricted, obvious differences between practices due to their number and types of biopsy submissions became apparent. This exploratory data analysis resulted in a number of questions being raised regarding practice-related factors, which as already referred to in multivariable analyses of the

dataset in Chapter 4, were considered to be important to the outcome of interest. The spatial and space-time analyses enabled identification of individual practices with high influence, which led to the proposal of targeting those practices for detailed examination in an attempt to identify practice-specific factors that affect the outcome of neoplasia in biopsies submitted from those practices. Both factors relating to individual veterinarians' investigative approach towards canine neoplasia, and data pertaining to the type of owner population utilising the practice, such as socioeconomic factors and distance between owner residence and practice location, would need to be considered. Socioeconomic factors are recognised to vary widely across the UK and are routinely considered in epidemiological studies of human neoplasia (Carstairs, 1995; Coleman *et al.*, 1999). At the owner level, the level of human-animal bond experienced by the owner may correlate with their pet being presented to a veterinarian, and their choice of the degree of investigation. The socioeconomic status of the owner may influence decisions which ultimately lead to biopsy submission, for which costs of the biopsy procurement procedure and its subsequent inspection by a histopathology service must be covered. In the last ten years, the emergence of a competitive pet insurance market in the UK is likely to have increased client demand for more expensive investigations, and its influence on the data used for investigations such as that presented here must also be considered.

In conclusion, further work is required to determine the spectrum of risk factors affecting the occurrence of canine neoplasia in the UK. Data quality assurance assessments of the available data sources and investigation of practice-related factors influencing these data sources would be invaluable to increase the understanding of fundamental aspects of canine neoplasia research.

APPENDIX 1

BIOPSY SUBMISSION FORM FOR THE CANINE INFECTIOUS DISEASES RESEARCH UNIT DIAGNOSTIC HISTOPATHOLOGY SERVICE (1986 - 1998)

Date sent:

Date received:

CANINE INFECTIOUS DISEASE RESEARCH UNIT
DEPARTMENT OF VETERINARY PATHOLOGY
UNIVERSITY OF GLASGOW VETERINARY SCHOOL
BEARSDEN, GLASGOW G61 1QH.
Telephone: 0141-330 5776

Lab Ref. No

Veterinary Surgeon's name and address:

Tel:

Owner:

Address:

DOG NAME/REF Breed: Age: Sex:

SAMPLES (please tick)

Clotted blood ☐ Swab ☐ Carcase

Fixed tissues ☐ Faeces ☐ Others (specify)

CASE HISTORY/VACCINATION RECORD

TREATMENT

EXAMINATION REQUIRED

Lab. Results: By:

REPORT/DIAGNOSIS

Date report

Charge

APPENDIX 2

BIOPSY SUBMISSION FORM FOR THE CANINE INFECTIOUS DISEASES RESEARCH UNIT DIAGNOSTIC HISTOPATHOLOGY SERVICE (1998)

Companion Animal Diagnostics

University of Glasgow, Veterinary Diagnostic Services, Bearsden, Glasgow G61 1QH
Tel: 0141 330 5777 Fax: 0141 330 5748 e-mail: companion@vet.gla.ac.uk
www.gla.ac.uk/companion

UNIVERSITY of GLASGOW

Canine Infectious Disease Research Unit Request Form

Please complete all shaded sections

Vet reference number

VETERINARY SURGEON'S NAME AND ADDRESS

OWNER'S NAME AND ADDRESS

Post Code

e-mail address

Tel. Number

Send results by: Fax Phone e-mail

Fax Number

Send forms Send transport medium (number)

DATE SAMPLE TAKEN

REASON FOR SAMPLING

Sick Healthy Pre-mating Pre-vaccine check Yes No Vaccinated? Number of dogs in house

SAMPLES SENT

Heparin blood EDTA blood Plasma Serum Faeces Urine Biopsy

Other: please specify:

DESCRIPTION OF DOGS DATE RECEIVED

No	Name	Age/date of birth	Sex	Breed	Our reference
1					
2					
3					

CLINICAL SIGNS AND HISTORY INCLUDING DRUGS AT TIME OF SAMPLING

√ Please tick shaded box to indicate which tests are required

SCREENING HEALTHY DOGS

Antibody screen for the demonstration of immunity

Rabies antibody (FAVN)	A separate submission form is required. Please call 0141 330 5777 for advice
Canine parvovirus (CPV) (HAI)	1ml heparin blood or serum
Canine distemper(CDV) (neutralisation)	1ml heparin blood or serum
Canine adenovirus (CAV-2) (neutralisation)	1ml heparin blood or serum
Canine parainfluenza virus (CpiV) (neutralisation)	1ml heparin blood or serum
Canine herpesvirus (CHV) (neutralisation)	1ml heparin blood or serum

Does the dog need a vaccine booster?

Canine parvovirus (CPV) and canine distemper virus (CVD) antibodies	2ml heparin blood or serum
CPV, CDV and canine adenovirus antibodies*	2ml heparin blood or serum

**Screens for canine parainfluenza virus or Leptospira antibodies are not included because vaccines do not generally induce detectable antibodies*

DIAGNOSTIC TESTING OF SICK DOGS

INFECTIOUS DISEASES

**viral transport medium sent free of charge on request*

Canine parvovirus Detection of faecal virus (PCR) and'or antibody in blood	Faecal swab (not in charcoal)or 1ml heparin blood or serum
Canine coronavirus antibodies	1ml serum
Kennel cough profile CpiV, *Bordetella bronchiseptica*, viral/bacterial isolation and CPiV antibody screen	Tonsillar swabs: 1) in viral transport medium*, 2) in charcoal. Two samples of 1ml heparin blood or serum at 3 week interval
Canine parainfluenza antibody Virus neutralisation test	1ml serum
Canine parainfluenza virus isolation	Throat swab in viral transport medium
Canine distemper Virus neutralisation test	a. 1ml heparin blood/serum: paired samples at 3 week interval b. Serum plus CSF
Canine adenovirus 1, CAV 2 isolation	Throat swab in viral transport medium
Canine adenovirus antibodies/blue eye check For differentiating CAV-1 and CAV-2	1ml heparin blood or serum
Canine herpesvirus isolation	Ocular/tonsil/genital swab in our viral transport medium
Leptospira antibodies *L. icterohaemorrhagiae* and *L. canicola*	1ml serum

PATHOLOGY/BIOPSY

Samples in 10% formalin unless otherwise stated **Phone Duty Pathologist before sampling*

Canine parvovirus	3 samples of small intestine from different levels
Canine distemper	Gastric mucosa, lymph node, bladder, lung and brain
Canine adenovirus	Intestine, liver, lymph node
Canine herpesvirus (neonatal puppy death)	Kidneys, liver
Faded puppy	Whole puppy
Needle aspirates or effusions.	Please put effusions into EDTA
Other, please specify*	

HAEMATOLOGY AND CLINICAL PATHOLOGY

**Supplied free of charge on request*

Complete blood cell count and smear	1ml EDTA blood and 2 air dried blood smears
Coagulation screen	1ml blood in sodium citrate*
Direct Coomb's test	1ml EDTA blood and 2 air dried blood smears
Routine biochemistry	2ml serum or plasma
Other, please specify:	

BACTERIOLOGY

Isolation (includes antibody sensitivity) Swab in transport medium or faeces Origin of swab:

Other, please specify:

PARASITOLOGY

Faecal sample

Giardia **Protozoa screen (Giardia, Isospora)** **Cryptosporidium** **Nematode egg screen**

***Toxoplasma gondii* antibodies**	1ml heparin blood
***Neospora caninum* antibodies**	1ml heparin blood
Other, please specify:	

MYCOLOGY

Ringworm	Plucked hair sample or toothbrush combing
***Aspergillus* antibodies**	1ml heparin blood
Other, please specify:	

ENDOCRINOLOGY

Canine T4 and TSH	1ml blood or serum
T4 (state time post pill)	1ml blood or serum
Dexamethasone suppression test	1ml blood or serum
ACTH stimulation test	1ml blood or serum

July 2001 JMcC/13171

APPENDIX 3

OUTPUT OF UNIVARIABLE LOGISTIC REGRESSION ANALYSIS OF PRACTICE (CIDRU DATASET)

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count	
NEO	1	10905	(Event)
	0	5631	
	Total	16536	

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	0.54788	0.04715	11.62	0.000			
CORRVETC							
2	0.40295	0.08156	4.94	0.000	1.50	1.28	1.76
3	-0.6202	0.2247	-2.76	0.006	0.54	0.35	0.84
5	-0.0036	0.1893	-0.02	0.985	1.00	0.69	1.44
8	0.24722	0.09911	2.49	0.013	1.28	1.05	1.55
11	-0.7457	0.2431	-3.07	0.002	0.47	0.29	0.76
12	0.4235	0.1680	2.52	0.012	1.53	1.10	2.12
18	0.5343	0.1868	2.86	0.004	1.71	1.18	2.46
22	0.0583	0.5097	0.11	0.909	1.06	0.39	2.88
23	0.5082	0.4132	1.23	0.219	1.66	0.74	3.74
26	-0.0592	0.1560	-0.38	0.704	0.94	0.69	1.28
27	0.6759	0.2130	3.17	0.002	1.97	1.29	2.98
29	0.1089	0.2376	0.46	0.647	1.12	0.70	1.78
30	0.0340	0.1365	0.25	0.803	1.03	0.79	1.35
31	0.0807	0.1402	0.58	0.565	1.08	0.82	1.43
33	-2.1060	0.3211	-6.56	0.000	0.12	0.06	0.23
54	-0.4662	0.2076	-2.25	0.025	0.63	0.42	0.94
55	0.2506	0.1589	1.58	0.115	1.28	0.94	1.75
56	0.8674	0.2745	3.16	0.002	2.38	1.39	4.08
57	-0.3161	0.2832	-1.12	0.264	0.73	0.42	1.27
60	0.7178	0.1296	5.54	0.000	2.05	1.59	2.64
61	-0.00542	0.08126	-0.07	0.947	0.99	0.85	1.17
65	-0.1150	0.2663	-0.43	0.666	0.89	0.53	1.50
66	-1.8877	0.3420	-5.52	0.000	0.15	0.08	0.30
68	0.0763	0.3234	0.24	0.814	1.08	0.57	2.03
72	0.61070	0.09800	6.23	0.000	1.84	1.52	2.23
75	0.2308	0.2331	0.99	0.322	1.26	0.80	1.99
76	0.2406	0.2092	1.15	0.250	1.27	0.84	1.92
81	-0.31970	0.09884	-3.23	0.001	0.73	0.60	0.88
82	0.0047	0.1230	0.04	0.970	1.00	0.79	1.28
85	-1.2855	0.3696	-3.48	0.001	0.28	0.13	0.57
86	0.0152	0.2013	0.08	0.940	1.02	0.68	1.51
87	0.9117	0.2362	3.86	0.000	2.49	1.57	3.95
88	0.6222	0.3845	1.62	0.106	1.86	0.88	3.96
89	0.0683	0.2434	0.28	0.779	1.07	0.66	1.73
92	-0.0853	0.3132	-0.27	0.785	0.92	0.50	1.70
100	-0.1540	0.2449	-0.63	0.529	0.86	0.53	1.39
108	-1.2071	0.3215	-3.75	0.000	0.30	0.16	0.56
112	0.3540	0.2290	1.55	0.122	1.42	0.91	2.23
121	0.5806	0.1314	4.42	0.000	1.79	1.38	2.31
132	0.5019	0.4417	1.14	0.256	1.65	0.70	3.93
133	0.2515	0.1650	1.52	0.127	1.29	0.93	1.78
134	-1.1945	0.3752	-3.18	0.001	0.30	0.15	0.63

135	-0.2304	0.1456	-1.58	0.114	0.79	0.60	1.06
137	0.4495	0.1112	4.04	0.000	1.57	1.26	1.95
147	0.4433	0.2008	2.21	0.027	1.56	1.05	2.31
166	0.6308	0.2897	2.18	0.029	1.88	1.06	3.32
182	0.1923	0.1415	1.36	0.174	1.21	0.92	1.60
196	-0.6012	0.1700	-3.54	0.000	0.55	0.39	0.76
197	0.6204	0.1254	4.95	0.000	1.86	1.45	2.38
199	0.2269	0.1825	1.24	0.214	1.25	0.88	1.79
207	-1.8791	0.3042	-6.18	0.000	0.15	0.08	0.28
230	0.0652	0.3476	0.19	0.851	1.07	0.54	2.11
232	-0.5308	0.1908	-2.78	0.005	0.59	0.40	0.85
237	-0.3537	0.3640	-0.97	0.331	0.70	0.34	1.43
247	-0.9183	0.2460	-3.73	0.000	0.40	0.25	0.65
259	0.8590	0.3255	2.64	0.008	2.36	1.25	4.47
260	-0.0883	0.2650	-0.33	0.739	0.92	0.54	1.54
270	-0.2602	0.4435	-0.59	0.557	0.77	0.32	1.84
274	-1.2410	0.5022	-2.47	0.013	0.29	0.11	0.77
297	-0.4902	0.1367	-3.59	0.000	0.61	0.47	0.80
301	0.0275	0.4193	0.07	0.948	1.03	0.45	2.34
322	-0.2966	0.5062	-0.59	0.558	0.74	0.28	2.00
337	0.6843	0.1590	4.30	0.000	1.98	1.45	2.71
361	0.3882	0.2680	1.45	0.147	1.47	0.87	2.49
378	0.3892	0.1971	1.97	0.048	1.48	1.00	2.17
388	0.2850	0.3817	0.75	0.455	1.33	0.63	2.81
437	0.0807	0.3131	0.26	0.797	1.08	0.59	2.00
447	-0.8663	0.4670	-1.85	0.064	0.42	0.17	1.05
542	-0.1424	0.5292	-0.27	0.788	0.87	0.31	2.45
546	0.2194	0.3389	0.65	0.517	1.25	0.64	2.42
557	0.1206	0.1978	0.61	0.542	1.13	0.77	1.66
625	0.3904	0.2820	1.38	0.166	1.48	0.85	2.57
690	-0.4301	0.4882	-0.88	0.378	0.65	0.25	1.69
694	-0.9533	0.4110	-2.32	0.020	0.39	0.17	0.86
772	0.8762	0.3125	2.80	0.005	2.40	1.30	4.43
774	0.3761	0.1705	2.21	0.027	1.46	1.04	2.03
848	-0.0871	0.2049	-0.42	0.671	0.92	0.61	1.37
903	0.0881	0.2426	0.36	0.716	1.09	0.68	1.76
905	0.3904	0.3960	0.99	0.324	1.48	0.68	3.21
923	0.0583	0.3620	0.16	0.872	1.06	0.52	2.15
995	0.3966	0.4479	0.89	0.376	1.49	0.62	3.58
998	0.6561	0.4679	1.40	0.161	1.93	0.77	4.82
1015	-1.1075	0.4457	-2.48	0.013	0.33	0.14	0.79
1018	0.8204	0.3030	2.71	0.007	2.27	1.25	4.11
1087	-0.0978	0.1395	-0.70	0.483	0.91	0.69	1.19
1090	0.4967	0.1611	3.08	0.002	1.64	1.20	2.25
1133	-0.7992	0.5062	-1.58	0.114	0.45	0.17	1.21
1191	0.6238	0.2485	2.51	0.012	1.87	1.15	3.04
1239	-1.6312	0.2511	-6.50	0.000	0.20	0.12	0.32
1269	0.5130	0.2775	1.85	0.065	1.67	0.97	2.88
1271	0.3056	0.2534	1.21	0.228	1.36	0.83	2.23
1290	-0.0371	0.3682	-0.10	0.920	0.96	0.47	1.98
1301	0.5507	0.2160	2.55	0.011	1.73	1.14	2.65
1359	-0.0959	0.3451	-0.28	0.781	0.91	0.46	1.79
1364	0.5156	0.2205	2.34	0.019	1.67	1.09	2.58
1370	0.3684	0.1404	2.62	0.009	1.45	1.10	1.90
1403	-0.0883	0.3717	-0.24	0.812	0.92	0.44	1.90
1432	-0.3937	0.5583	-0.71	0.481	0.67	0.23	2.02
1542	-0.0246	0.3190	-0.08	0.938	0.98	0.52	1.82
1662	-1.1569	0.2281	-5.07	0.000	0.31	0.20	0.49
1717	0.6587	0.1908	3.45	0.001	1.93	1.33	2.81
1815	-0.2114	0.2123	-1.00	0.319	0.81	0.53	1.23

1843	-1.3042	0.2571	-5.07	0.000	0.27	0.16	0.45
2004	-0.3656	0.3528	-1.04	0.300	0.69	0.35	1.39
2018	-0.7149	0.2935	-2.44	0.015	0.49	0.28	0.87
2094	-0.2602	0.3448	-0.75	0.450	0.77	0.39	1.52

Log-Likelihood = -10194.240
Test that all slopes are zero: G = 823.445, DF = 106, P-Value = 0.000

* NOTE * No goodness of fit tests performed.
* The model uses all degrees of freedom.

APPENDIX 4

OUTPUT OF MULTIVARIABLE LOGISTIC REGRESSION ANALYSIS WITH PRACTICE AS A FIXED-EFFECT (CIDRU DATASET)

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count	
NEO	1	10905	(Event)
	0	5631	
	Total	16536	

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-0.19240	0.08436	-2.28	0.023			
AGE	0.119380	0.005205	22.94	0.000	1.13	1.12	1.14
SEXC							
2	-0.16566	0.09738	-1.70	0.089	0.85	0.70	1.03
3	0.05956	0.04099	1.45	0.146	1.06	0.98	1.15
4	-0.18538	0.05796	-3.20	0.001	0.83	0.74	0.93
PBREED							
2	-0.46966	0.08458	-5.55	0.000	0.63	0.53	0.74
3	-0.40396	0.09635	-4.19	0.000	0.67	0.55	0.81
4	0.2608	0.1037	2.52	0.012	1.30	1.06	1.59
5	-0.0929	0.1071	-0.87	0.386	0.91	0.74	1.12
6	-0.4335	0.1113	-3.89	0.000	0.65	0.52	0.81
7	-0.6719	0.1160	-5.79	0.000	0.51	0.41	0.64
8	-0.26495	0.06756	-3.92	0.000	0.77	0.67	0.88
100	-0.40197	0.05943	-6.76	0.000	0.67	0.60	0.75
SITEC							
3	0.30786	0.07421	4.15	0.000	1.36	1.18	1.57
4	1.35165	0.07291	18.54	0.000	3.86	3.35	4.46
5	0.33591	0.09139	3.68	0.000	1.40	1.17	1.67
6	-1.8494	0.1737	-10.64	0.000	0.16	0.11	0.22
7	-0.6066	0.1436	-4.22	0.000	0.55	0.41	0.72
11	0.36091	0.09418	3.83	0.000	1.43	1.19	1.73
13	-0.4647	0.1347	-3.45	0.001	0.63	0.48	0.82
100	-0.42282	0.07608	-5.56	0.000	0.66	0.56	0.76
CORRVETC							
2	0.38150	0.08700	4.39	0.000	1.46	1.23	1.74
3	-0.3193	0.2371	-1.35	0.178	0.73	0.46	1.16
5	0.0478	0.1970	0.24	0.808	1.05	0.71	1.54
8	0.1925	0.1042	1.85	0.065	1.21	0.99	1.49
11	-0.7065	0.2545	-2.78	0.006	0.49	0.30	0.81
12	0.2717	0.1797	1.51	0.131	1.31	0.92	1.87
18	0.6428	0.1975	3.25	0.001	1.90	1.29	2.80
22	0.1987	0.5465	0.36	0.716	1.22	0.42	3.56
23	0.5738	0.4404	1.30	0.193	1.77	0.75	4.21
26	-0.1867	0.1668	-1.12	0.263	0.83	0.60	1.15
27	0.3753	0.2234	1.68	0.093	1.46	0.94	2.26
29	-0.0140	0.2547	-0.05	0.956	0.99	0.60	1.62
30	0.0917	0.1435	0.64	0.523	1.10	0.83	1.45
31	0.2510	0.1463	1.72	0.086	1.29	0.96	1.71
33	-1.6680	0.3274	-5.09	0.000	0.19	0.10	0.36
54	-0.3537	0.2204	-1.61	0.108	0.70	0.46	1.08
55	0.2150	0.1672	1.29	0.199	1.24	0.89	1.72
56	0.9648	0.2836	3.40	0.001	2.62	1.51	4.57
57	-0.2200	0.2950	-0.75	0.456	0.80	0.45	1.43

60	0.6663	0.1363	4.89	0.000	1.95	1.49	2.54
61	-0.05386	0.08667	-0.62	0.534	0.95	0.80	1.12
65	-0.0939	0.2790	-0.34	0.736	0.91	0.53	1.57
66	-1.7620	0.3601	-4.89	0.000	0.17	0.08	0.35
68	0.0815	0.3428	0.24	0.812	1.08	0.55	2.12
72	0.4095	0.1040	3.94	0.000	1.51	1.23	1.85
75	0.1590	0.2448	0.65	0.516	1.17	0.73	1.89
76	0.2279	0.2218	1.03	0.304	1.26	0.81	1.94
81	-0.3329	0.1047	-3.18	0.001	0.72	0.58	0.88
82	0.0077	0.1308	0.06	0.953	1.01	0.78	1.30
85	-1.1717	0.3952	-2.96	0.003	0.31	0.14	0.67
86	0.0809	0.2110	0.38	0.701	1.08	0.72	1.64
87	0.6528	0.2472	2.64	0.008	1.92	1.18	3.12
88	0.5747	0.4014	1.43	0.152	1.78	0.81	3.90
89	0.1087	0.2578	0.42	0.673	1.11	0.67	1.85
92	-0.2338	0.3341	-0.70	0.484	0.79	0.41	1.52
100	-0.0581	0.2524	-0.23	0.818	0.94	0.58	1.55
108	-1.0129	0.3357	-3.02	0.003	0.36	0.19	0.70
112	0.3008	0.2424	1.24	0.215	1.35	0.84	2.17
121	0.4359	0.1380	3.16	0.002	1.55	1.18	2.03
132	0.5940	0.4541	1.31	0.191	1.81	0.74	4.41
133	0.2625	0.1719	1.53	0.127	1.30	0.93	1.82
134	-0.9560	0.3864	-2.47	0.013	0.38	0.18	0.82
135	-0.0135	0.1539	-0.09	0.930	0.99	0.73	1.33
137	0.2365	0.1183	2.00	0.046	1.27	1.00	1.60
147	0.4556	0.2103	2.17	0.030	1.58	1.04	2.38
166	0.6712	0.3021	2.22	0.026	1.96	1.08	3.54
182	0.1363	0.1484	0.92	0.359	1.15	0.86	1.53
196	-0.5984	0.1803	-3.32	0.001	0.55	0.39	0.78
197	0.4822	0.1320	3.65	0.000	1.62	1.25	2.10
199	0.3382	0.1902	1.78	0.075	1.40	0.97	2.04
207	-1.8233	0.3154	-5.78	0.000	0.16	0.09	0.30
230	0.0351	0.3638	0.10	0.923	1.04	0.51	2.11
232	-0.5500	0.2010	-2.74	0.006	0.58	0.39	0.86
237	-0.1691	0.3886	-0.44	0.664	0.84	0.39	1.81
247	-0.8547	0.2587	-3.30	0.001	0.43	0.26	0.71
259	0.9900	0.3352	2.95	0.003	2.69	1.40	5.19
260	-0.0242	0.2818	-0.09	0.932	0.98	0.56	1.70
270	-0.0287	0.5159	-0.06	0.956	0.97	0.35	2.67
274	-1.1882	0.5357	-2.22	0.027	0.30	0.11	0.87
297	-0.4317	0.1446	-2.98	0.003	0.65	0.49	0.86
301	0.0141	0.4500	0.03	0.975	1.01	0.42	2.45
322	-0.6597	0.5319	-1.24	0.215	0.52	0.18	1.47
337	0.6297	0.1675	3.76	0.000	1.88	1.35	2.61
361	0.3933	0.2762	1.42	0.155	1.48	0.86	2.55
378	0.4652	0.2054	2.26	0.024	1.59	1.06	2.38
388	0.2452	0.3980	0.62	0.538	1.28	0.59	2.79
437	-0.0100	0.3359	-0.03	0.976	0.99	0.51	1.91
447	-0.4819	0.4921	-0.98	0.327	0.62	0.24	1.62
542	0.1705	0.5546	0.31	0.759	1.19	0.40	3.52
546	0.0760	0.3605	0.21	0.833	1.08	0.53	2.19
557	0.1967	0.2079	0.95	0.344	1.22	0.81	1.83
625	0.3926	0.2975	1.32	0.187	1.48	0.83	2.65
690	-0.5593	0.5139	-1.09	0.276	0.57	0.21	1.56
694	-0.9216	0.4340	-2.12	0.034	0.40	0.17	0.93
772	0.9020	0.3276	2.75	0.006	2.46	1.30	4.68
774	0.2769	0.1791	1.55	0.122	1.32	0.93	1.87
848	-0.2384	0.2187	-1.09	0.276	0.79	0.51	1.21
903	0.0686	0.2541	0.27	0.787	1.07	0.65	1.76
905	0.3702	0.4161	0.89	0.374	1.45	0.64	3.27

923	0.2318	0.3799	0.61	0.542	1.26	0.60	2.65
995	-0.0998	0.4809	-0.21	0.836	0.91	0.35	2.32
998	0.5942	0.4880	1.22	0.223	1.81	0.70	4.71
1015	-1.2219	0.4691	-2.60	0.009	0.29	0.12	0.74
1018	0.9190	0.3132	2.93	0.003	2.51	1.36	4.63
1087	0.0636	0.1460	0.44	0.663	1.07	0.80	1.42
1090	0.6293	0.1678	3.75	0.000	1.88	1.35	2.61
1133	-0.2821	0.5173	-0.55	0.586	0.75	0.27	2.08
1191	0.4062	0.2608	1.56	0.119	1.50	0.90	2.50
1239	-1.2477	0.2586	-4.82	0.000	0.29	0.17	0.48
1269	0.2637	0.2966	0.89	0.374	1.30	0.73	2.33
1271	0.2541	0.2609	0.97	0.330	1.29	0.77	2.15
1290	0.1281	0.3843	0.33	0.739	1.14	0.54	2.41
1301	0.6337	0.2224	2.85	0.004	1.88	1.22	2.91
1359	0.1156	0.3665	0.32	0.752	1.12	0.55	2.30
1364	0.4885	0.2314	2.11	0.035	1.63	1.04	2.57
1370	0.3660	0.1480	2.47	0.013	1.44	1.08	1.93
1403	0.0225	0.3887	0.06	0.954	1.02	0.48	2.19
1432	-0.4737	0.5842	-0.81	0.417	0.62	0.20	1.96
1542	0.3028	0.3359	0.90	0.367	1.35	0.70	2.61
1662	-0.9575	0.2388	-4.01	0.000	0.38	0.24	0.61
1717	0.6291	0.1996	3.15	0.002	1.88	1.27	2.77
1815	-0.1561	0.2233	-0.70	0.484	0.86	0.55	1.33
1843	-1.2417	0.2746	-4.52	0.000	0.29	0.17	0.49
2004	-0.4531	0.3783	-1.20	0.231	0.64	0.30	1.33
2018	-0.4606	0.3046	-1.51	0.131	0.63	0.35	1.15
2094	-0.2084	0.3623	-0.58	0.565	0.81	0.40	1.65

Log-Likelihood = -9392.320
Test that all slopes are zero: G = 2427.286, DF = 126, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	12940.439	12798	0.186
Deviance	14731.376	12798	0.000
Hosmer-Lemeshow	5.642	8	0.687

APPENDIX 5

BIOPSY SUBMISSION FORM FOR THE COMMERCIAL HISTOPATHOLOGY SERVICE

IDEXX LABORATORIES LTD

PO Box 4, Wetherby, West Yorkshire LS22 4ZR Telephone: (01937) 544000 Fax: (01937) 544001

LAB NUMBER	DATE:	Name & Address of Owner:	FOR LAB USE ONL
	Veterinary Surgeon:		Extra Slides ☐
			Extra Blocks ☐
VET CODE	Address Stamp:	Name of Animal:	Special Stains ☐
		Species & Breed:	
		Age:	FAX YES / NO
		Sex: Neutered or Entire	U T

HISTOLOGY REQUEST FORM

Tissues to be submitted in 10% formol-saline. Polypropylene tubes with fixative are available from the laboratory. Please do not push large tissues into small pots. Wrap pots in sufficient absorbent material to contain any leakage in transit.

History:

Provisional diagnosis:

Recent treatment:

Excision site:

Tissue description:

Has whole tumour been submitted? Yes / No
IF NOT, PLEASE GIVE DETAILS: (This information is essential for prognosis).

Cut surface appearance:

Please feel free to ring Histology Department for advice

APPENDIX 6

OUTPUT OF UNIVARIABLE LOGISTIC REGRESSION ANALYSIS OF PRACTICE (COMMERCIAL HISTOPATHOLOGY SERVICE)

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count	
neo	1	3661	(Event)
	0	2276	
	Total	5937	

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	1.2528	0.3586	3.49	0.000			
pracID							
2	-1.1575	0.5652	-2.05	0.041	0.31	0.10	0.95
3	-1.1350	0.6039	-1.88	0.060	0.32	0.10	1.05
4	-0.6397	0.4334	-1.48	0.140	0.53	0.23	1.23
5	-0.6581	0.4749	-1.39	0.166	0.52	0.20	1.31
6	-0.9163	0.4928	-1.86	0.063	0.40	0.15	1.05
7	-1.4351	0.5008	-2.87	0.004	0.24	0.09	0.64
8	-0.5055	0.5407	-0.94	0.350	0.60	0.21	1.74
9	-1.1474	0.5828	-1.97	0.049	0.32	0.10	0.99
10	-0.6337	0.5902	-1.07	0.283	0.53	0.17	1.69
11	-0.9871	0.4532	-2.18	0.029	0.37	0.15	0.91
12	-0.4990	0.5589	-0.89	0.372	0.61	0.20	1.82
13	-1.4069	0.5323	-2.64	0.008	0.24	0.09	0.70
14	-1.3959	0.5217	-2.68	0.007	0.25	0.09	0.69
15	-0.5960	0.4869	-1.22	0.221	0.55	0.21	1.43
16	-1.0986	0.6619	-1.66	0.097	0.33	0.09	1.22
17	-0.8008	0.6019	-1.33	0.183	0.45	0.14	1.46
18	-1.5712	0.5869	-2.68	0.007	0.21	0.07	0.66
19	-1.2528	0.6153	-2.04	0.042	0.29	0.09	0.95
20	-0.4906	0.4830	-1.02	0.310	0.61	0.24	1.58
21	-0.8781	0.5310	-1.65	0.098	0.42	0.15	1.18
22	-1.2528	0.5732	-2.19	0.029	0.29	0.09	0.88
23	-1.7868	0.4710	-3.79	0.000	0.17	0.07	0.42
24	-0.5596	0.6153	-0.91	0.363	0.57	0.17	1.91
25	-0.2167	0.5017	-0.43	0.666	0.81	0.30	2.15
26	-0.3185	0.5052	-0.63	0.528	0.73	0.27	1.96
27	-0.2113	0.5950	-0.36	0.722	0.81	0.25	2.60
28	-0.8109	0.5578	-1.45	0.146	0.44	0.15	1.33
29	-1.5041	0.6185	-2.43	0.015	0.22	0.07	0.75
30	-0.1542	0.5923	-0.26	0.795	0.86	0.27	2.74
31	-0.5108	0.5255	-0.97	0.331	0.60	0.21	1.68
32	-0.2167	0.5017	-0.43	0.666	0.81	0.30	2.15
33	-0.7419	0.6287	-1.18	0.238	0.48	0.14	1.63
34	-0.7221	0.5361	-1.35	0.178	0.49	0.17	1.39
35	-0.9651	0.6483	-1.49	0.137	0.38	0.11	1.36
36	-0.6466	0.6214	-1.04	0.298	0.52	0.15	1.77
37	-0.9045	0.5203	-1.74	0.082	0.40	0.15	1.12
38	-0.6466	0.6214	-1.04	0.298	0.52	0.15	1.77
39	-0.9845	0.5141	-1.91	0.056	0.37	0.14	1.02
40	-1.0986	0.6619	-1.66	0.097	0.33	0.09	1.22
41	-1.2528	0.6796	-1.84	0.065	0.29	0.08	1.08
42	-0.6774	0.5497	-1.23	0.218	0.51	0.17	1.49
43	-0.8109	0.5578	-1.45	0.146	0.44	0.15	1.33

44	-0.4224	0.4616	-0.92	0.360	0.66	0.27	1.62
45	-1.0296	0.5946	-1.73	0.083	0.36	0.11	1.15
46	-0.8109	0.5578	-1.45	0.146	0.44	0.15	1.33
47	-1.0986	0.6619	-1.66	0.097	0.33	0.09	1.22
48	-1.0704	0.5585	-1.92	0.055	0.34	0.11	1.02
49	-1.6582	0.6375	-2.60	0.009	0.19	0.05	0.66
50	-0.8473	0.5804	-1.46	0.144	0.43	0.14	1.34
51	-0.0827	0.4487	-0.18	0.854	0.92	0.38	2.22
52	-0.6466	0.6214	-1.04	0.298	0.52	0.15	1.77
53	-0.9474	0.5026	-1.88	0.059	0.39	0.14	1.04
54	-1.0296	0.5278	-1.95	0.051	0.36	0.13	1.00
55	0.0910	0.5820	0.16	0.876	1.10	0.35	3.43
56	-1.4351	0.5585	-2.57	0.010	0.24	0.08	0.71
57	-0.2113	0.5950	-0.36	0.722	0.81	0.25	2.60
58	-1.0521	0.5750	-1.83	0.067	0.35	0.11	1.08
59	-1.0116	0.5394	-1.88	0.061	0.36	0.13	1.05
60	-0.3365	0.4781	-0.70	0.482	0.71	0.28	1.82
61	-1.3863	0.6296	-2.20	0.028	0.25	0.07	0.86
62	0.1335	0.5804	0.23	0.818	1.14	0.37	3.57
63	-1.1192	0.5123	-2.18	0.029	0.33	0.12	0.89
64	-0.5849	0.4519	-1.29	0.196	0.56	0.23	1.35
65	-1.2528	0.6437	-1.95	0.052	0.29	0.08	1.01
66	-1.0296	0.4910	-2.10	0.036	0.36	0.14	0.93
67	-0.1542	0.5648	-0.27	0.785	0.86	0.28	2.59
68	-0.1542	0.5923	-0.26	0.795	0.86	0.27	2.74
69	-0.4726	0.5110	-0.92	0.355	0.62	0.23	1.70
70	-0.8350	0.4359	-1.92	0.055	0.43	0.18	1.02
71	-0.9904	0.5527	-1.79	0.073	0.37	0.13	1.10
72	-0.2412	0.5468	-0.44	0.659	0.79	0.27	2.29
73	-1.4131	0.4573	-3.09	0.002	0.24	0.10	0.60
74	-1.3068	0.4866	-2.69	0.007	0.27	0.10	0.70
75	-1.1575	0.5652	-2.05	0.041	0.31	0.10	0.95
76	-1.8814	0.5659	-3.32	0.001	0.15	0.05	0.46
77	-0.9651	0.6483	-1.49	0.137	0.38	0.11	1.36
78	-0.0632	0.5612	-0.11	0.910	0.94	0.31	2.82
79	-2.0260	0.6101	-3.32	0.001	0.13	0.04	0.44
80	-0.5596	0.5622	-1.00	0.320	0.57	0.19	1.72
81	-1.2528	0.6437	-1.95	0.052	0.29	0.08	1.01
82	-0.2495	0.5028	-0.50	0.620	0.78	0.29	2.09
83	-1.6582	0.6375	-2.60	0.009	0.19	0.05	0.66
84	-0.5596	0.4656	-1.20	0.229	0.57	0.23	1.42
85	-1.1192	0.6296	-1.78	0.075	0.33	0.10	1.12
86	-0.8733	0.5081	-1.72	0.086	0.42	0.15	1.13
87	-0.1306	0.4725	-0.28	0.782	0.88	0.35	2.22
88	-1.3754	0.4588	-3.00	0.003	0.25	0.10	0.62
89	0.1335	0.4741	0.28	0.778	1.14	0.45	2.89
90	-0.8299	0.4644	-1.79	0.074	0.44	0.18	1.08
91	-0.9651	0.6483	-1.49	0.137	0.38	0.11	1.36
92	-1.4351	0.5585	-2.57	0.010	0.24	0.08	0.71
93	-0.5596	0.5434	-1.03	0.303	0.57	0.20	1.66
94	-1.1192	0.4422	-2.53	0.011	0.33	0.14	0.78
95	-0.4055	0.6055	-0.67	0.503	0.67	0.20	2.18
96	0.2353	0.5511	0.43	0.669	1.27	0.43	3.73
97	-1.2528	0.6153	-2.04	0.042	0.29	0.09	0.95
98	-0.7980	0.4220	-1.89	0.059	0.45	0.20	1.03
99	-0.7563	0.4934	-1.53	0.125	0.47	0.18	1.23
100	-0.6597	0.4500	-1.47	0.143	0.52	0.21	1.25
101	-0.4906	0.5815	-0.84	0.399	0.61	0.20	1.91
102	-1.1474	0.5828	-1.97	0.049	0.32	0.10	0.99
103	-1.0986	0.6619	-1.66	0.097	0.33	0.09	1.22

104	-0.7221	0.4561	-1.58	0.113	0.49	0.20	1.19
105	-0.3704	0.4964	-0.75	0.456	0.69	0.26	1.83
106	-0.0632	0.5612	-0.11	0.910	0.94	0.31	2.82
107	-0.3520	0.4866	-0.72	0.469	0.70	0.27	1.83
108	-0.1542	0.5434	-0.28	0.777	0.86	0.30	2.49
109	0.2191	0.4652	0.47	0.638	1.24	0.50	3.10
110	-1.4271	0.4650	-3.07	0.002	0.24	0.10	0.60
111	-1.6582	0.5804	-2.86	0.004	0.19	0.06	0.59
112	-0.8850	0.5627	-1.57	0.116	0.41	0.14	1.24
113	-1.4198	0.5444	-2.61	0.009	0.24	0.08	0.70
114	-0.6022	0.5054	-1.19	0.233	0.55	0.20	1.47
115	-0.8755	0.4459	-1.96	0.050	0.42	0.17	1.00
116	-1.3328	0.5374	-2.48	0.013	0.26	0.09	0.76
117	-0.4318	0.5094	-0.85	0.397	0.65	0.24	1.76
118	-0.3365	0.5182	-0.65	0.516	0.71	0.26	1.97
119	-1.5041	0.6185	-2.43	0.015	0.22	0.07	0.75
120	-1.0222	0.4416	-2.31	0.021	0.36	0.15	0.85
121	-1.4171	0.4411	-3.21	0.001	0.24	0.10	0.58
122	-1.2528	0.5923	-2.12	0.034	0.29	0.09	0.91
123	-0.8473	0.4824	-1.76	0.079	0.43	0.17	1.10
124	-0.7419	0.6287	-1.18	0.238	0.48	0.14	1.63
125	-0.8473	0.5172	-1.64	0.101	0.43	0.16	1.18
126	-0.8473	0.5172	-1.64	0.101	0.43	0.16	1.18
127	-0.8961	0.6094	-1.47	0.141	0.41	0.12	1.35
128	-0.7419	0.6287	-1.18	0.238	0.48	0.14	1.63
129	-0.9651	0.5683	-1.70	0.089	0.38	0.13	1.16
130	-1.2528	0.6153	-2.04	0.042	0.29	0.09	0.95
131	-0.9343	0.4863	-1.92	0.055	0.39	0.15	1.02
132	-0.1001	0.5898	-0.17	0.865	0.90	0.28	2.87
133	-1.3173	0.5077	-2.59	0.009	0.27	0.10	0.72
134	-0.6242	0.5659	-1.10	0.270	0.54	0.18	1.62
135	-1.1838	0.5164	-2.29	0.022	0.31	0.11	0.84
136	-0.8473	0.6375	-1.33	0.184	0.43	0.12	1.49
137	-0.9079	0.4786	-1.90	0.058	0.40	0.16	1.03
138	-0.7758	0.4676	-1.66	0.097	0.46	0.18	1.15
139	-0.7221	0.5361	-1.35	0.178	0.49	0.17	1.39
140	-1.3134	0.4999	-2.63	0.009	0.27	0.10	0.72
141	-1.4641	0.4848	-3.02	0.003	0.23	0.09	0.60
142	-1.6776	0.4752	-3.53	0.000	0.19	0.07	0.47
143	-1.1192	0.6296	-1.78	0.075	0.33	0.10	1.12
144	-0.7138	0.5956	-1.20	0.231	0.49	0.15	1.57
145	-0.1919	0.4510	-0.43	0.670	0.83	0.34	2.00
146	-0.8008	0.4954	-1.62	0.106	0.45	0.17	1.19
147	0.0594	0.5567	0.11	0.915	1.06	0.36	3.16
148	-0.1542	0.4603	-0.33	0.738	0.86	0.35	2.11
149	-0.9904	0.5527	-1.79	0.073	0.37	0.13	1.10
150	-1.1350	0.6039	-1.88	0.060	0.32	0.10	1.05
151	-0.4055	0.6055	-0.67	0.503	0.67	0.20	2.18
152	-0.8174	0.5275	-1.55	0.121	0.44	0.16	1.24
153	-1.0986	0.6619	-1.66	0.097	0.33	0.09	1.22
154	-0.8473	0.5804	-1.46	0.144	0.43	0.14	1.34
155	-1.0857	0.4610	-2.36	0.019	0.34	0.14	0.83
156	-1.0296	0.4910	-2.10	0.036	0.36	0.14	0.93
157	-0.6650	0.5330	-1.25	0.212	0.51	0.18	1.46
158	-0.8473	0.5804	-1.46	0.144	0.43	0.14	1.34
159	-0.2029	0.5669	-0.36	0.720	0.82	0.27	2.48
160	-0.8850	0.5627	-1.57	0.116	0.41	0.14	1.24
161	-0.4726	0.5110	-0.92	0.355	0.62	0.23	1.70
162	-1.0902	0.4872	-2.24	0.025	0.34	0.13	0.87
163	-0.9163	0.4928	-1.86	0.063	0.40	0.15	1.05

164	-0.3878	0.5534	-0.70	0.483	0.68	0.23	2.01
165	-0.8698	0.4906	-1.77	0.076	0.42	0.16	1.10
166	-0.9426	0.4554	-2.07	0.038	0.39	0.16	0.95
167	-0.7062	0.5216	-1.35	0.176	0.49	0.18	1.37
168	-0.0206	0.5595	-0.04	0.971	0.98	0.33	2.93
169	-0.5596	0.5622	-1.00	0.320	0.57	0.19	1.72
170	-0.2542	0.5693	-0.45	0.655	0.78	0.25	2.37
171	-2.3514	0.5923	-3.97	0.000	0.10	0.03	0.30
172	-1.3218	0.5164	-2.56	0.010	0.27	0.10	0.73
173	-0.4318	0.5094	-0.85	0.397	0.65	0.24	1.76
174	-0.7758	0.4676	-1.66	0.097	0.46	0.18	1.15
175	-1.0415	0.4848	-2.15	0.032	0.35	0.14	0.91
176	-1.0986	0.6619	-1.66	0.097	0.33	0.09	1.22
177	-0.7419	0.4663	-1.59	0.112	0.48	0.19	1.19
178	-0.3365	0.4952	-0.68	0.497	0.71	0.27	1.89
179	-0.5268	0.4770	-1.10	0.269	0.59	0.23	1.50
180	-0.0690	0.4644	-0.15	0.882	0.93	0.38	2.32
181	-0.2793	0.4578	-0.61	0.542	0.76	0.31	1.86
182	-0.8473	0.6375	-1.33	0.184	0.43	0.12	1.49
183	-2.2336	0.5981	-3.73	0.000	0.11	0.03	0.35
184	-0.8065	0.4807	-1.68	0.093	0.45	0.17	1.15
185	-1.5404	0.6483	-2.38	0.017	0.21	0.06	0.76
186	-0.2113	0.5950	-0.36	0.722	0.81	0.25	2.60
187	-1.1097	0.5217	-2.13	0.033	0.33	0.12	0.92
188	-0.3365	0.6016	-0.56	0.576	0.71	0.22	2.32
189	-1.0296	0.5278	-1.95	0.051	0.36	0.13	1.00

Log-Likelihood = -3799.731
Test that all slopes are zero: G = 304.876, DF = 188, P-Value = 0.000

* NOTE * No goodness of fit tests performed.
* NOTE * The model uses all degrees of freedom.

APPENDIX 7

OUTPUT OF MULTIVARIABLE LOGISTIC REGRESSION ANALYSIS WITH PRACTICE AS A FIXED-EFFECT (COMMERCIAL HISTOPATHOLOGY DATASET)

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count	
neo	1	3661	(Event)
	0	2276	
	Total	5937	

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	0.6016	0.3877	1.55	0.121			
ageyrs	0.111139	0.008177	13.59	0.000	1.12	1.10	1.14
sexc							
2	0.1791	0.1099	1.63	0.103	1.20	0.96	1.48
3	0.12641	0.09996	1.26	0.206	1.13	0.93	1.38
4	-0.0666	0.1247	-0.53	0.593	0.94	0.73	1.19
5	0.07014	0.09647	0.73	0.467	1.07	0.89	1.30
6	0.0520	0.1025	0.51	0.612	1.05	0.86	1.29
sitec							
3	0.3296	0.1236	2.67	0.008	1.39	1.09	1.77
4	2.2560	0.1864	12.11	0.000	9.54	6.62	13.75
5	0.3604	0.1576	2.29	0.022	1.43	1.05	1.95
6	-1.9683	0.2582	-7.62	0.000	0.14	0.08	0.23
7	-0.6067	0.2106	-2.88	0.004	0.55	0.36	0.82
11	0.4307	0.1905	2.26	0.024	1.54	1.06	2.23
13	-1.9385	0.2474	-7.83	0.000	0.14	0.09	0.23
100	-0.9394	0.1201	-7.82	0.000	0.39	0.31	0.49
pracID							
2	-1.5293	0.6029	-2.54	0.011	0.22	0.07	0.71
3	-1.3594	0.6350	-2.14	0.032	0.26	0.07	0.89
4	-0.6738	0.4606	-1.46	0.144	0.51	0.21	1.26
5	-0.8363	0.5051	-1.66	0.098	0.43	0.16	1.17
6	-1.4085	0.5256	-2.68	0.007	0.24	0.09	0.69
7	-1.6129	0.5352	-3.01	0.003	0.20	0.07	0.57
8	-0.6150	0.5712	-1.08	0.282	0.54	0.18	1.66
9	-1.6096	0.6319	-2.55	0.011	0.20	0.06	0.69
10	-1.0166	0.6516	-1.56	0.119	0.36	0.10	1.30
11	-1.5783	0.4870	-3.24	0.001	0.21	0.08	0.54
12	-0.9369	0.5853	-1.60	0.109	0.39	0.12	1.23
13	-1.6450	0.5609	-2.93	0.003	0.19	0.06	0.58
14	-1.7537	0.5477	-3.20	0.001	0.17	0.06	0.51
15	-0.5281	0.5264	-1.00	0.316	0.59	0.21	1.65
16	-1.5454	0.7289	-2.12	0.034	0.21	0.05	0.89
17	-1.1339	0.6362	-1.78	0.075	0.32	0.09	1.12
18	-2.0735	0.6172	-3.36	0.001	0.13	0.04	0.42
19	-1.2369	0.6794	-1.82	0.069	0.29	0.08	1.10
20	-0.8022	0.5242	-1.53	0.126	0.45	0.16	1.25
21	-1.2796	0.5717	-2.24	0.025	0.28	0.09	0.85
22	-1.4737	0.6100	-2.42	0.016	0.23	0.07	0.76
23	-2.1898	0.5079	-4.31	0.000	0.11	0.04	0.30
24	-0.9493	0.6472	-1.47	0.142	0.39	0.11	1.38
25	-0.5299	0.5284	-1.00	0.316	0.59	0.21	1.66
26	-0.6164	0.5289	-1.17	0.244	0.54	0.19	1.52
27	-0.5669	0.6281	-0.90	0.367	0.57	0.17	1.94

28	-0.9788	0.5813	-1.68	0.092	0.38	0.12	1.17
29	-1.5788	0.6425	-2.46	0.014	0.21	0.06	0.73
30	-0.3784	0.6223	-0.61	0.543	0.68	0.20	2.32
31	-0.8498	0.5568	-1.53	0.127	0.43	0.14	1.27
32	-0.3327	0.5375	-0.62	0.536	0.72	0.25	2.06
33	-0.8796	0.6614	-1.33	0.184	0.41	0.11	1.52
34	-0.9760	0.5724	-1.71	0.088	0.38	0.12	1.16
35	-1.0131	0.7372	-1.37	0.169	0.36	0.09	1.54
36	-1.1159	0.6531	-1.71	0.088	0.33	0.09	1.18
37	-1.1956	0.5540	-2.16	0.031	0.30	0.10	0.90
38	-0.9445	0.7040	-1.34	0.180	0.39	0.10	1.55
39	-1.2311	0.5465	-2.25	0.024	0.29	0.10	0.85
40	-0.9750	0.7124	-1.37	0.171	0.38	0.09	1.52
41	-1.6437	0.7337	-2.24	0.025	0.19	0.05	0.81
42	-0.9569	0.5752	-1.66	0.096	0.38	0.12	1.19
43	-0.8146	0.5918	-1.38	0.169	0.44	0.14	1.41
44	-0.7346	0.4935	-1.49	0.137	0.48	0.18	1.26
45	-1.4882	0.6599	-2.26	0.024	0.23	0.06	0.82
46	-0.9239	0.5890	-1.57	0.117	0.40	0.13	1.26
47	-1.2682	0.6980	-1.82	0.069	0.28	0.07	1.11
48	-1.2268	0.5957	-2.06	0.039	0.29	0.09	0.94
49	-1.8991	0.6594	-2.88	0.004	0.15	0.04	0.55
50	-1.2099	0.6284	-1.93	0.054	0.30	0.09	1.02
51	-0.5481	0.4748	-1.15	0.248	0.58	0.23	1.47
52	-1.1551	0.6589	-1.75	0.080	0.32	0.09	1.15
53	-1.3772	0.5366	-2.57	0.010	0.25	0.09	0.72
54	-1.5498	0.5812	-2.67	0.008	0.21	0.07	0.66
55	-0.1829	0.6054	-0.30	0.763	0.83	0.25	2.73
56	-1.7984	0.6051	-2.97	0.003	0.17	0.05	0.54
57	-0.3186	0.6270	-0.51	0.611	0.73	0.21	2.48
58	-1.3578	0.6236	-2.18	0.029	0.26	0.08	0.87
59	-1.1591	0.5778	-2.01	0.045	0.31	0.10	0.97
60	-0.6570	0.5045	-1.30	0.193	0.52	0.19	1.39
61	-1.5122	0.6844	-2.21	0.027	0.22	0.06	0.84
62	0.1112	0.6097	0.18	0.855	1.12	0.34	3.69
63	-1.3129	0.5411	-2.43	0.015	0.27	0.09	0.78
64	-0.8575	0.4817	-1.78	0.075	0.42	0.17	1.09
65	-1.9727	0.6857	-2.88	0.004	0.14	0.04	0.53
66	-1.5551	0.5301	-2.93	0.003	0.21	0.07	0.60
67	-0.4438	0.6117	-0.73	0.468	0.64	0.19	2.13
68	-0.4507	0.6286	-0.72	0.473	0.64	0.19	2.18
69	-0.9169	0.5436	-1.69	0.092	0.40	0.14	1.16
70	-1.2510	0.4656	-2.69	0.007	0.29	0.11	0.71
71	-1.2868	0.5856	-2.20	0.028	0.28	0.09	0.87
72	-0.5146	0.5829	-0.88	0.377	0.60	0.19	1.87
73	-1.8464	0.4916	-3.76	0.000	0.16	0.06	0.41
74	-1.7035	0.5184	-3.29	0.001	0.18	0.07	0.50
75	-1.5363	0.6008	-2.56	0.011	0.22	0.07	0.70
76	-2.4095	0.6097	-3.95	0.000	0.09	0.03	0.30
77	-1.4990	0.7089	-2.11	0.034	0.22	0.06	0.90
78	0.0132	0.6020	0.02	0.983	1.01	0.31	3.30
79	-2.2553	0.6568	-3.43	0.001	0.10	0.03	0.38
80	-0.8434	0.5881	-1.43	0.152	0.43	0.14	1.36
81	-1.6188	0.6779	-2.39	0.017	0.20	0.05	0.75
82	-0.3482	0.5362	-0.65	0.516	0.71	0.25	2.02
83	-1.9454	0.7169	-2.71	0.007	0.14	0.04	0.58
84	-0.9798	0.5008	-1.96	0.050	0.38	0.14	1.00
85	-1.5442	0.7007	-2.20	0.028	0.21	0.05	0.84
86	-1.1904	0.5410	-2.20	0.028	0.30	0.11	0.88
87	-0.7205	0.5024	-1.43	0.152	0.49	0.18	1.30

88	-1.5502	0.4955	-3.13	0.002	0.21	0.08	0.56
89	-0.1965	0.5015	-0.39	0.695	0.82	0.31	2.20
90	-0.9388	0.4909	-1.91	0.056	0.39	0.15	1.02
91	-1.2582	0.6907	-1.82	0.068	0.28	0.07	1.10
92	-2.1611	0.6334	-3.41	0.001	0.12	0.03	0.40
93	-0.6991	0.5706	-1.23	0.220	0.50	0.16	1.52
94	-1.2992	0.4727	-2.75	0.006	0.27	0.11	0.69
95	-1.0016	0.6508	-1.54	0.124	0.37	0.10	1.32
96	-0.1355	0.5819	-0.23	0.816	0.87	0.28	2.73
97	-1.6072	0.7009	-2.29	0.022	0.20	0.05	0.79
98	-0.9802	0.4519	-2.17	0.030	0.38	0.15	0.91
99	-1.0811	0.5306	-2.04	0.042	0.34	0.12	0.96
100	-0.8322	0.4774	-1.74	0.081	0.44	0.17	1.11
101	-0.9443	0.6109	-1.55	0.122	0.39	0.12	1.29
102	-1.5746	0.6298	-2.50	0.012	0.21	0.06	0.71
103	-1.1128	0.7079	-1.57	0.116	0.33	0.08	1.32
104	-0.7753	0.4874	-1.59	0.112	0.46	0.18	1.20
105	-0.5923	0.5242	-1.13	0.259	0.55	0.20	1.55
106	-0.5537	0.5914	-0.94	0.349	0.57	0.18	1.83
107	-0.4134	0.5250	-0.79	0.431	0.66	0.24	1.85
108	-0.7684	0.5778	-1.33	0.184	0.46	0.15	1.44
109	0.0195	0.4932	0.04	0.968	1.02	0.39	2.68
110	-1.4186	0.4937	-2.87	0.004	0.24	0.09	0.64
111	-1.8762	0.6193	-3.03	0.002	0.15	0.05	0.52
112	-1.6990	0.6051	-2.81	0.005	0.18	0.06	0.60
113	-1.6017	0.5756	-2.78	0.005	0.20	0.07	0.62
114	-0.8546	0.5531	-1.55	0.122	0.43	0.14	1.26
115	-1.0032	0.4726	-2.12	0.034	0.37	0.15	0.93
116	-1.4200	0.5734	-2.48	0.013	0.24	0.08	0.74
117	-0.4853	0.5409	-0.90	0.370	0.62	0.21	1.78
118	-0.6979	0.5386	-1.30	0.195	0.50	0.17	1.43
119	-1.3698	0.6538	-2.10	0.036	0.25	0.07	0.92
120	-1.4391	0.4723	-3.05	0.002	0.24	0.09	0.60
121	-1.5762	0.4702	-3.35	0.001	0.21	0.08	0.52
122	-1.2131	0.6339	-1.91	0.056	0.30	0.09	1.03
123	-0.8131	0.5093	-1.60	0.110	0.44	0.16	1.20
124	-0.8425	0.7144	-1.18	0.238	0.43	0.11	1.75
125	-1.1676	0.5512	-2.12	0.034	0.31	0.11	0.92
126	-1.0518	0.5479	-1.92	0.055	0.35	0.12	1.02
127	-0.6574	0.6656	-0.99	0.323	0.52	0.14	1.91
128	-1.1485	0.6640	-1.73	0.084	0.32	0.09	1.17
129	-1.0058	0.5980	-1.68	0.093	0.37	0.11	1.18
130	-2.0817	0.6735	-3.09	0.002	0.12	0.03	0.47
131	-1.1845	0.5208	-2.27	0.023	0.31	0.11	0.85
132	-0.2000	0.6219	-0.32	0.748	0.82	0.24	2.77
133	-1.3641	0.5518	-2.47	0.013	0.26	0.09	0.75
134	-1.2358	0.6056	-2.04	0.041	0.29	0.09	0.95
135	-1.9004	0.5566	-3.41	0.001	0.15	0.05	0.45
136	-1.0725	0.6921	-1.55	0.121	0.34	0.09	1.33
137	-1.2216	0.5116	-2.39	0.017	0.29	0.11	0.80
138	-0.9969	0.4950	-2.01	0.044	0.37	0.14	0.97
139	-1.1171	0.5730	-1.95	0.051	0.33	0.11	1.01
140	-1.4463	0.5402	-2.68	0.007	0.24	0.08	0.68
141	-1.6399	0.5150	-3.18	0.001	0.19	0.07	0.53
142	-1.7047	0.5032	-3.39	0.001	0.18	0.07	0.49
143	-0.8894	0.6940	-1.28	0.200	0.41	0.11	1.60
144	-0.7816	0.6436	-1.21	0.225	0.46	0.13	1.62
145	-0.6438	0.4758	-1.35	0.176	0.53	0.21	1.33
146	-0.9842	0.5417	-1.82	0.069	0.37	0.13	1.08
147	-0.1910	0.5918	-0.32	0.747	0.83	0.26	2.64

148	-0.4868	0.4904	-0.99	0.321	0.61	0.24	1.61
149	-1.1793	0.5810	-2.03	0.042	0.31	0.10	0.96
150	-1.4406	0.6429	-2.24	0.025	0.24	0.07	0.83
151	-0.5307	0.6374	-0.83	0.405	0.59	0.17	2.05
152	-1.0427	0.5593	-1.86	0.062	0.35	0.12	1.05
153	-1.1587	0.6845	-1.69	0.090	0.31	0.08	1.20
154	-1.0864	0.6145	-1.77	0.077	0.34	0.10	1.13
155	-1.2970	0.4926	-2.63	0.008	0.27	0.10	0.72
156	-1.4179	0.5318	-2.67	0.008	0.24	0.09	0.69
157	-0.8091	0.5606	-1.44	0.149	0.45	0.15	1.34
158	-1.3898	0.6358	-2.19	0.029	0.25	0.07	0.87
159	-0.3207	0.5884	-0.54	0.586	0.73	0.23	2.30
160	-1.3413	0.6009	-2.23	0.026	0.26	0.08	0.85
161	-0.7718	0.5496	-1.40	0.160	0.46	0.16	1.36
162	-1.2376	0.5150	-2.40	0.016	0.29	0.11	0.80
163	-1.4137	0.5230	-2.70	0.007	0.24	0.09	0.68
164	-0.7628	0.6011	-1.27	0.204	0.47	0.14	1.51
165	-1.0909	0.5231	-2.09	0.037	0.34	0.12	0.94
166	-1.1712	0.4828	-2.43	0.015	0.31	0.12	0.80
167	-0.8322	0.5493	-1.51	0.130	0.44	0.15	1.28
168	-0.4719	0.5872	-0.80	0.422	0.62	0.20	1.97
169	-0.6890	0.5890	-1.17	0.242	0.50	0.16	1.59
170	-0.9578	0.6011	-1.61	0.107	0.38	0.12	1.23
171	-2.5399	0.6525	-3.89	0.000	0.08	0.02	0.28
172	-1.3458	0.5622	-2.39	0.017	0.26	0.09	0.78
173	-0.7058	0.5418	-1.30	0.193	0.49	0.17	1.43
174	-1.1462	0.5040	-2.27	0.023	0.32	0.12	0.85
175	-1.3382	0.5229	-2.56	0.010	0.26	0.09	0.73
176	-1.4283	0.7251	-1.97	0.049	0.24	0.06	0.99
177	-1.1108	0.5008	-2.22	0.027	0.33	0.12	0.88
178	-0.9378	0.5296	-1.77	0.077	0.39	0.14	1.11
179	-0.7877	0.5041	-1.56	0.118	0.45	0.17	1.22
180	-0.3451	0.4961	-0.70	0.487	0.71	0.27	1.87
181	-0.6201	0.4879	-1.27	0.204	0.54	0.21	1.40
182	-1.1946	0.6832	-1.75	0.080	0.30	0.08	1.16
183	-2.2187	0.6361	-3.49	0.000	0.11	0.03	0.38
184	-0.9990	0.5138	-1.94	0.052	0.37	0.13	1.01
185	-1.6749	0.6678	-2.51	0.012	0.19	0.05	0.69
186	-0.2022	0.6304	-0.32	0.748	0.82	0.24	2.81
187	-1.2874	0.5625	-2.29	0.022	0.28	0.09	0.83
188	-0.6038	0.6408	-0.94	0.346	0.55	0.16	1.92
189	-1.3139	0.5720	-2.30	0.022	0.27	0.09	0.82

Log-Likelihood = -3403.822
Test that all slopes are zero: G = 1096.692, DF = 202, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	5249.667	5088	0.056
Deviance	6045.958	5088	0.000
Hosmer-Lemeshow	2.973	8	0.936

REFERENCES

Anonymous (2002a) SNOMED International. http://www.snomed.org

Anonymous (2002b) Veterinary Medical Database. http://www.vmdb.org/vmdbinfo.html

Aldrich, T. and Sinks, T. (2002) Things to know and do about cancer clusters. *Cancer Investigation* **20**, 810-816.

Alevizopoulos, K., Vlach, J., Hennecke, S. and Amati, B. (1997) Cyclin E and c-Myc promote cell proliferation in the presence of p16INK4a and of hypophosphorylated retinoblastoma family proteins. *European Molecular Biology Organ Journal* **16**, 5322-5333.

Aquilina, J.W., McKinney, L., Pacelli, A., Richman, L.K., Waters, D.J., Thompson, I., Burghardt, W.F.J. and Bostwick, D.G. (1998) High grade prostatic intraepithelial neoplasia in military working dogs with and without prostate cancer. *Prostate* **36**, 189-193.

Arnesen, K., Gamlem, H., Glattre, E., Grondalen, J., Moe, L. and Nordstoga, K. (2001) The Norwegian Canine Cancer Register 1990-1998. Report from the project "Cancer in the Dog". *European Journal of Companion Animal Practice* **11**, 159-169.

Ash, E.C. (1927) *Dogs: Their history and development*, London: Ernest Benn Ltd.

Auerbach, O., Hammond, E.C., Kirman, D. and Garfinkel, L. (1970) Effects of cigarette smoking on dogs II. Pulmonary neoplasms. *Archives of Environmental Health* **21**, 754-768.

Bailey, T.C. and Gatrell, A.C. (1995) *Interactive spatial data analysis*, New York: Wiley.

Barrett, J.C. (1993) Mechanisms of multistep carcinogenesis and carcinogen risk assessment. *Environmental Health Perspectives* **100**, 9.

Batamuzi, E.K., Kassuku, A.A. and Agger, J.F. (1992) Risk factors associated with canine transmissible venereal tumour in Tanzania. *Preventive Veterinary Medicine* **13**, 13-17.

Beckmann, M.W., Niederacher, D., Schnurch, H.G., Gusterson, B.A. and Bender, H.G. (1997) Multistep carcinogenesis of breast cancer and tumour heterogeneity. *Journal of Molecular Medicine* **75**, 429-439.

Bell, C.M., Lawrence, G., Pheby, D.F., Smith, J. and Coleman, M.P. (1995) The role of cancer registries. *Clinical Oncology (Royal College of Radiologists)* **7**, 143-144.

Bender, A.P., Bender, G.P., Dorn, C.R. and Schneider, R. (1982) Associations between canine benign and malignant neoplasms. *Preventive Veterinary Medicine* **1**, 77-87.

Bender, A.P., Dorn, C.R. and Schneider, R. (1983) Quantitative index for diagnostic surveillance of neoplasms in dogs. *American Journal of Veterinary Research* **44**, 395-398.

Bender, A.P., Dorn, C.R. and Schneider, R. (1984) An epidemiologic study of canine multiple primary neoplasia involving the female and male reproductive systems. *Preventive Veterinary Medicine* **2**, 715-731.

Benecke, N. (1987) Studies on early dog remains from northern Europe. *Journal of Archaeological Science* **14**, 31-49.

Benfield, J.R., Hammond, W.G., Paladugu, R.R., Pak, H.Y., Azumi, N. and Teplitz, R.L. (1986) Endobronchial carcinogenesis in dogs. *Journal of Thoracic and Cardiovascular Surgery* **92**, 880-889.

Benfield, J.R., Shors, E.C., Hammond, W.G., Paladugu, R.R., Cohen, A.H., Jensen, T., Fu, P.C., Pak, H.Y. and Teplitz, R.L. (1981) A clinically relevant canine lung cancer model. *Annals of Thoracic Surgery* **32**, 592-601.

Besag, J. and Newell, J. (1991) The detection of clusters in rare diseases. *Journal of the Royal Statistical Society, Series A* **154**, 143-155.

Black, R.J., Bray, F., Ferlay, J. and Parkin, D.M. (1997) Cancer incidence and mortality in the European Union: cancer registry data and estimates of national incidence for 1990. *European Journal of Cancer* **33**, 1075-1107.

Bodson, L. (2000) Motivations for pet-keeping in Ancient Greece and Rome: a preliminary survey. In: Podberscek, A.L., Paul, E.S. and Serpell, J.A., (Eds.) *Companion Animals and Us: exploring the relationships between people and pets*, pp. 27-41. Cambridge: Cambridge University Press.

Bonfiglio, T.A. and Terry, R. (1983) The pathology of cancer. In: Rubin, P., (Ed.) *Clinical Oncology*, 6th edn. pp. 20-28. New York: American Cancer Society.

Bonnett, B.N., Egenvall, A., Olson, P. and Hedhammar, A. (1997) Mortality in insured Swedish dogs: rates and causes of death in various breeds. *Veterinary Record* **141**, 40-44.

Borjesson, D.L., Christopher, M.M. and Ling, G.V. (1999) Detection of canine transitional cell carcinoma using a bladder tumor antigen urine dipstick test. *Veterinary Clinical Pathology* **28**, 33-38.

Bray, F., Sankila, R., Ferlay, J. and Parkin, D.M. (2002) Estimates of cancer incidence and mortality in Europe in 1995. *European Journal of Cancer* **38**, 99-166.

Bregman, C.L., Hirth, R.S., Sundberg, J.P. and Christensen, E.F. (1987) Cutaneous neoplasms in dogs associated with canine oral papillomavirus vaccine. *Veterinary Pathology* **24** , 477-487.

Breslow, N. (1982) Design and analysis of case-control studies. *Annual Review of Public Health* **3**, 29-54.

Breslow, N.E. and Day, N.E. (1980) Statistical methods in cancer research. Volume I - The analysis of case-control studies. *IARC Scientific Publications (Lyon)* **32**, 5-338.

Brewster, D. (1995) Improving the quality of cancer registration data. *Journal of the Royal Society of Medicine* **88**, 268-271.

Buck, W.B. (1979) Animals as monitors of environmental quality. *Veterinary and Human Toxicology* **21**, 277-284.

Bukowski, J.A. and Wartenberg, D. (1997) An alternative approach for investigating the carcinogenicity of indoor air pollution: Pets as sentinels of environmental cancer risk. *Environmental Health Perspectives* **105**, 1312-1319.

Bullen, N., Moon, G. and Jones, K. (1996) Defining localities for health planning:a GIS approach. *Social Science and Medicine* **42**, 801-806.

Calvert, C.A. (1990) Canine viral papillomatosis. In: Greene, C.E., (Ed.) *Clinical microbiology and infectious diseases of the dog and cat*, 2nd edn. pp. 288-290. Philadelphia, USA: W.B. Saunders.

Cangul, I.T., Teske, E. and Ingh, T.S.G.A. (1998) Immunohistochemical staining results of frozen and paraffin embedded sections of 14 canine malignant lymphoma cases: a comparative study. *European Journal of Veterinary Pathology* **4**, 17-22.

Carlo, G.L., Cole, P., Miller, A.B., Munro, I.C., Solomon, K.R. and Squire, R.A. (1992) Review of a study reporting an association between 2,4-dichlorophenoxyacetic acid and canine malignant lymphoma: report of an expert panel. *Regulatory Toxicology and Pharmacology* **16**, 245-252.

Carpenter, T.E. (2001) Methods to investigate spatial and temporal clustering in veterinary epidemiology. *Preventive Veterinary Medicine* **48**, 303-320.

Carpinito, G.A., Stadler, W.M., Church, P.A., Lamm, D.L., Lange, P.H., Messing, E.M., Pasciak, R.M., Reservitz, G.B., Ross, R.N., Rukstalis, D.B., Sarosdy, M.F., Soloway, M.S., Thiel, R.P., Vogelzang, N. and Hayden, C.L. (1996) Urinary nuclear matrix protein as a marker for transitional cell carcinoma of the urinary tract. *Journal of Urology* **156**, 1280-1285.

Carstairs, V. (1995) Deprivation indices: their interpretation and use in relation to health. *Journal of Epidemiology and Community Health* **49 Suppl 2**, S3-S8.

Case, J. T. (1998) Status of health information standards in veterinary medicine. http://www.usaha.org/speeches/speech98/s98case.html. United States Animal Health Association. 1998 Proceedings.

Cesano, A., Visonneau, S., Jeglum, K.A., Owen, J., Wilkinson, K., Carner, K., Reese, L. and Santoli, D. (1996) Phase I clinical trial with a human major histocompatibility complex nonrestricted cytotoxic T-cell line (TALL-104) in dogs with advanced tumors. *Cancer Research* **56**, 3021-3029.

Chaganti, S.R., Mitra, J. and LoBue, J. (1992) Detection of canine homologs of human MYC, BCL2, IGH, and TCRB genes by Southern blot analysis. *Cancer Genetics and Cytogenetics* **62**, 9-14.

Chu, L.L., Rutteman, G.R., Kong, J.M.C., Ghahremani, M., Schmeing, M., Misdorp, W., Garderen, E.v. and Pelletier, J. (1998) Genomic organization of the canine p53 gene and its mutational status in canine mammary neoplasia. *Breast Cancer Research and Treatment* **50**, 11-25.

Clutton-Brock, J. (1994) The Unnatural World:behavioural aspects of humans and animals in the process of domestication. In: Manning, A. and Serpell, J., (Eds.) *Animals and Human Society*, pp. 23-35. London: Routledge.

Clutton-Brock, J. and Jewell, P. (1993) Origin and domestication of the dog. In: Evans, H.E. *Miller's anatomy of the dog*, pp. 21-31. Philadelphia: W.B. Saunders.

Cohen, D. (1985) The canine transmissible venereal tumor: A unique result of tumor progression. *Advances in Veterinary Science and Comparative Medicine* **43**, 75-112.

Cohen, D., Booth, S. and Sussman, O. (1959) An epidemiological study of canine lymphoma and its public health significance. *American Journal of Veterinary Research* **20**, 1026-1031.

Cohen, D., Brodey, R.S. and Chen, A.B. (1964) Epidemiologic aspects of oral and pharyngeal neoplasms of the dog. *American Journal of Veterinary Research* **25**, 1776-1779.

Cohen, D., Reif, J.S., Brodey, R.S. and Keiser, H. (1974) Epidemiological analysis of the most prevalent sites and types of canine neoplasia observed in a veterinary hospital. *Cancer Research* **34**, 2859-2868.

Coiera, E. (1997) *Guide to medical informatics, the internet and telemedicine*, 1st edn. London and NewYork: Chapman and Hall.

Coleman, M.P., Babb, P., Damiecki, P., Grosclaude, P., Honjo, S. and Jones, J. (1999) *Cancer Survival Trends in England and Wales 1971-1995:Deprivation and NHS Region Series SMPS No. 61*, London: The Stationery Office.

Cook, A.K., Breitschwerdt, E.B., Levine, J.F., Bunch, S.E. and Linn, L.O. (1993) Risk factors associated with acute pancreatitis in dogs: 101 cases (1985-1990). *Journal of the American Veterinary Medical Association* **203**, 673-679.

Corcoran, C., Coull, B. and Patel, A. (1999) Logistic regression with random effects. In: *Egret for Windows user manual*, pp. 173-196. Boston: Department of Biostatistics, Harvard School of Public Health.

Cotchin, E. (1951) Neoplasms in small animals. *Veterinary Record* **63**, 67-78.

Cotchin, E. (1954) Further observations on neoplasms in dogs, with particular reference to site of origin and malignancy Part I: Cutaneous, female genital and alimentary systems. *British Veterinary Journal* **110**, 218-230.

Cote, R.A., Rothwell, D.J., Palotay, J.L., Beckett, R.S. and Brochu, L. (1993) *The Systematized Nomenclature of Medicine. SNOMED International*, Northfield, Illinois: College of American Pathologists.

Cotran, R.S., Kumar, V. and Collins, T. (1999) Neoplasia. In: Cotran, R.S., Kumar, V. and Collins, T., (Eds.) *Robbins pathologic basis of disease*, 6th edn. pp. 260-327. Philadelphia: W.B. Saunders.

Council for Science and Society (1988) *Companion Animals in Society: report of a working party: Council for Science and Society*, Oxford: Oxford University Press.

Cullen, J.M., Page, R. and Misdorp, W. (2002) An overview of cancer pathogenesis, diagnosis and management. In: Meuten, D.J., (Ed.) *Tumors in Domestic Animals*, pp. 3-44. Ames: Iowa State Press.

Curtis, C.R., Mauritsen, R.H., Kass, P.H., Salman, M.D. and Erb, H.N. (1993) Ordinary versus random-effects logistic regression for analyzing herd-level calf morbidity and mortality data. *Preventive Veterinary Medicine* **16**, 207-222.

Curtis, C.R., Mauritsen, R.H., Salman, M.D. and Erb, H.N. (1988) The enigma of herd: a comparison of different models to account for group effects in multiple logistic regression analysis. *Acta Veterinaria Scandinavica* **Suppl 84**, 462-465.

Cuzick, J. and Edwards, R. (1990) Spatial clustering for inhomogeneous populations. *Journal of the Royal Statistical Society, Series B* **52**, 73-104.

Data Protection Act. (1998) http://www.hmso.gov.uk/acts/acts1998/19980029.htm

Davis, J.M. and Valla, F.R. (1978) Evidence for the domestication of the dog 12,000 years ago in the Natufian of Israel. *Nature* **276**, 608-610.

Devesa, S.S., Grauman, D.J., Blot, W.J., Pennello, G.A., Hoover, R.N. and Fraumeni, J.F. (1999) *Atlas of Cancer Mortality in the United States, 1950-94*, Washington: National Cancer Institute.

Devilee, P., van Leeuwen, I.S., Voesten, A., Rutteman, G.R., Vos, J.H. and Cornelisse, C.J. (1994) The canine p53 gene is subject to somatic mutations in thyroid carcinoma. *Anticancer Research* **14**, 2039-2046.

Dickinson, P.J., McEntee, M.C., Lipsitz, D., Keel, K. and LeCouteur, R.A. (2001) Radiation induced vertebral osteosarcoma following treatment of an intradural extramedullary spinal cord tumor in a dog. *Veterinary Radiology and Ultrasound* **42**, 463-470.

Diggle, P.J., Chetwynd, A.G., Haggkvist, R. and Morris, S.E. (1995) Second-order analysis of space-time clustering. *Statistical Methods in Medical Research* **4**, 124-136.

Dobson, J.M., Samuel, S., Milstein, H., Rogers, K. and Wood, J.L.N. (2002) Canine neoplasia in the UK: estimates of incidence rates from a population of insured dogs. *Journal of Small Animal Practice* **43**, 240-246.

Dohoo, I.R. and Martin, S.W. (1984) Disease, production and culling in Holstein-Friesian cows. V. Survivorship. *Preventive Veterinary Medicine* **2**, 771-784.

Dorn, C.R. (1968) Comparative studies of cancer in dogs and cats. In: Schwabe, C.W., (Ed.) *Veterinary Medicine and and Human Health*, Baltimore: Williams and Wilkins.

Dorn, C.R. and Chaulk, L.E. (1964) A survey of practice characteristics and animal admissions in two California counties. *Journal of the American Veterinary Medical Association* **145**, 358-360.

Dorn, C.R. and Priester, W.A. (1976a) Epidemiologic analysis of oral and pharyngeal cancer in dogs, cats, horses, and cattle. *Journal of the American Veterinary Medical Association* **169**, 1202-1206.

Dorn, C.R. and Schneider, R. (1976b) Inbreeding and canine mammary cancer: a retrospective study. *Journal of the National Cancer Institute* **57**, 545-548.

Dorn, C.R., Taylor, D.O., Chaulk, L.E. and Hibbard, H.H. (1966) The prevalence of spontaneous neoplasms in a defined canine population. *American Journal of Public Health and the Nations Health* **56**, 254-265.

Dorn, C.R., Taylor, D.O., Frye, F.L. and Hibbard, H.H. (1968a) Survey of animal neoplasms in Alameda and Contra Costa Counties, California. I. Methodology and description of cases. *Journal of the National Cancer Institute* **40**, 295-305.

Dorn, C.R., Taylor, D.O. and Hibbard, H.H. (1967) Epizootiologic characteristics of canine and feline leukemia and lymphoma. *American Journal of Veterinary Research* **28**, 993-1001.

Dorn, C.R., Taylor, D.O. and Schneider, R. (1971) Sunlight exposure and risk of developing cutaneous and oral squamous cell carcinomas in white cats. *Journal of the National Cancer Institute* **46**, 1073-1078.

Dorn, C.R., Taylor, D.O., Schneider, R., Hibbard, H.H. and Klauber, M.R. (1968b) Survey of animal neoplasms in Alameda and Contra Costa Counties, California. II. Cancer morbidity in dogs and cats from Alameda County. *Journal of the National Cancer Institute* **40**, 307-318.

Duhaime, R.A., Norden, D., Corso, B., Mallonee, S. and Salman, M.D. (1998) Injuries and illnesses in working dogs used during the disaster response after the bombing in Oklahoma City. *Journal of the American Veterinary Medical Association* **212**, 1202-1207.

Dumonceaux, G.A. and Beasley, V.R. (1990) Emergency treatments for police dogs used for illicit drug detection. *Journal of the American Veterinary Medical Association* **197**, 185-187.

Egenvall, A., Bonnett, B.N., Olson, P. and Hedhammar, A. (1998) Validation of computerized Swedish dog and cat insurance data against veterinary practice records. *Preventive Veterinary Medicine* **36**, 51-65.

Elliott, P., Martuzzi, M. and Shaddick, G. (1995) Spatial statistical methods in environmental epidemiology: a critique. *Statistical Methods in Medical Research* **4**, 137-159.

Ershler, W.B. and Longo, D.L. (1997) Aging and cancer: issues of basic and clinical science. *Journal of the National Cancer Institute* **89**, 1489-1497.

Farber, E. (1984) The multistep nature of cancer development. *Cancer Research* **44**, 4217

Fears, T.R., Bird, C.C., Guerry, D., Sagebiel, R.W., Gail, M.H., Elder, D.E., Halpern, A., Holly, E.A., Hartge, P. and Tucker, M.A. (2002) Average midrange ultraviolet radiation flux and time outdoors predict melanoma risk. *Cancer Research* **62**, 3992-3996.

Feigl, P., Breslow, N.E., Laszlo, J., Priore, R.L. and Taylor, W.F. (1981) U.S. centralized cancer patient data system for uniform communication among cancer centers. *Journal of the National Cancer Institute* **67**, 1017-1024.

Fisher, D.J., Naydan, D., Werner, L.L. and Moore, P.F. (1995) Immunophenotyping lymphomas in dogs: comparison of results from fine needle aspirate and needle biopsy samples. *Veterinary Clinical Pathology* **24**, 118-123.

Fleege, J.C., van Diest, P.J. and Baak, J.P.A. (1992) Quality Control Methods for Data Entry in Pathology using a Computerized Data Management System Based on an Extended Data Dictionary. *Human Pathology* **23**, 91-97.

Flynn, J.N., Hanlon, L. and Jarrett, O. (2000) Feline leukaemia virus: protective immunity is mediated by virus-specific cytotoxic T lymphocytes. *Immunology* **101**, 120-125.

Friedmann, E., Katcher, A.H., Lynch, J.J. and Thomas, S.A. (1980) Animal companions and one-year survival of patients after discharge from a coronary unit. *Public Health Reports* **95**, 307-312.

Friedmann, E. and Thomas, S.A. (1995) Pet ownership, social support, and one-year survival after acute myocardial infarction in the Cardiac Arrhythmia Suppression Trial (CAST). *American Journal of Cardiology* **76**, 1213-1217.

Frye, F.L., Dorn, C.R. and Taylor, D.O. (1967) Characteristics of canine mammary gland tumor cases. *Journal of the American Animal Hospital Association* **3**, 1-12.

Garbe, P.L. (1988) The companion animal as a sentinel for environmentally related human diseases. *Acta Veterinaria Scandinavica* **Suppl 84**, 290-292.

Gardner, D.G. (1996) Spontaneous squamous cell carcinomas of the oral region in domestic animals: a review and consideration of their relevance to human research. *Oral Diseases* **2**, 148-154.

Gatrell, A.C. and Bailey, T.C. (1996) Interactive spatial data analysis in medical geography. *Social Science and Medicine* **42**, 843-855.

Gillette, S.M., Gillette, E.L., Powers, B.E. and Withrow, S.J. (1990) Radiation-induced osteosarcoma in dogs after external beam or intraoperative radiation therapy. *Cancer Research* **50**, 54-57.

Ginn, P.E., Fox, L.E., Brower, J.C., Gaskin, A., Kurzman, I.D. and Kubilis, P.S. (2000) Immunohistochemical detection of p53 tumor-suppressor protein is a poor indicator of prognosis for canine cutaneous mast cell tumors. *Veterinary Pathology* **37**, 33-39.

Glickman, L., Emerick, T., Glickman, N., Glickman, S., Lantz, G., Perez, C., Schellenberg, D., Widmer, W. and Yi, Q.L. (1996) Radiological assessment of the relationship between thoracic conformation and the risk of gastric dilatation-volvulus in dogs. *Veterinary Radiology and Ultrasound* **37**, 174-180.

Glickman, L.T., Domanski, L.M., Maguire, T.G., Dubielzig, R.R. and Churg, A. (1983) Mesothelioma in pet dogs associated with exposure of their owners to asbestos. *Environmental Research* **32**, 305-313.

Glickman, L.T., Schofer, F.S., McKee, L.J., Reif, J.S. and Goldschmidt, M.H. (1989) Epidemiologic study of insecticide exposures, obesity, and risk of bladder cancer in household dogs. *Journal of Toxicology and Environmental Health* **28**, 407-414.

Goldschmidt, M.H. (1993) Biopsies from the pathologist's point of view. *Seminars in Veterinary Medicine and Surgery (Small Animal)* **8**, 284-288.

Goldschmidt, M.H. and Shofer, F.S. (1992) *Skin tumors of the dog and cat*. Oxford: Pergamon Press Ltd.

Greger, V., Debus, N., Lohmann, D., Hopping, W., Passarge, E. and Horsthemke, B. (1994) Frequency and parental origin of hypermethylated RB1 alleles in retinoblastoma. *Human Genetics* **94**, 491-496.

Gregory, F. (2000) The demographics and epidemiology of pet ownership and canine relinquishment. University of Glasgow. PhD thesis.

Griffey, S.M., Kraegel, S.A. and Madewell, B.R. (1998) Rapid detection of K-ras gene mutations in canine lung cancer using single-strand conformational polymorphism analysis. *Carcinogenesis* **19**, 959-963.

Griffey, S.M., Madewell, B.R., Dairkee, S.H., Hunt, J.E., Naydan, D.K. and Higgins, R.J. (1993) Immunohistochemical reactivity of basal and luminal epithelium-specific cytokeratin antibodies within normal and neoplastic canine mammary glands. *Veterinary Pathology* **30**, 155-161.

Griffey, S.M., Verstraete, F.J., Kraegel, S.A., Lucroy, M.D. and Madewell, B.R. (1998) Computer-assisted image analysis of intratumoral vessel density in mammary tumors from dogs. *American Journal of Veterinary Research* **59**, 1238-1242.

Hahn, K.A. (1998) Chromosomal changes associated with cancer. In: Morrison, W.B., (Ed.) *Cancer in dogs and cats: medical and surgical management*, pp. 53-59.

Hahn, K.A. and Anderson, T.E. (2000) Tumors of the respiratory tract. In: Bonagura, J.D., (Ed.) *Kirk's current veterinary therapy XIII: small animal practice*, pp. 500-505. Philadelphia, USA: W.B. Saunders.

Hahn, K.A., Bravo, L., Adams, W.H. and Frazier, D.L. (1994) Naturally occurring tumors in dogs as comparative models for cancer therapy research. *In Vivo* **8**, 133-143.

Harari, R., Shahar, R., Zuckerman, A. and Nyska, A. (1995) Silver stained nucleolar organizer regions (AgNORs) in canine mammary tumors measured by image analysis and direct counting: correlation to histology and clinical behavior. *Israel Journal of Veterinary Medicine* **50**, 147-155.

Hargis, A.M. (1981) A review of solar-induced lesions in domestic animals. *Compendium on Continuing Education for the Practicing Veterinarian* **3**, 287-294.

Hayes, H.M., Tarone, R.E., Cantor, K.P., Jessen, C.R., McCurnin, D.M. and Richardson, R.C. (1991) Case-control study of canine malignant lymphoma: positive association with dog owner's use of 2,4-dichlorophenoxyacetic acid herbicides. *Journal of the National Cancer Institute* **83**, 1226-1231.

Hayes, H.M.J., Hoover, R. and Tarone, R.E. (1981) Bladder cancer in pet dogs: a sentinel for environmental cancer? *American Journal of Epidemiology* **114**, 229-233.

Hayes, H.M.J. and Pendergrass, T.W. (1976) Canine testicular tumors: epidemiologic features of 410 dogs. *International Journal of Cancer* **18**, 482-487.

Hayes, H.M.J., Wilson, G.P. and Fraumeni, H.F., Jr. (1982) Carcinoma of the nasal cavity and paranasal sinuses in dogs: descriptive epidemiology. *Cornell Veterinarian* **72**, 168-179.

Hayes, H.M.J., Wilson, G.P., Pendergrass, T.W. and Cox, V.S. (1985) Canine cryptorchism and subsequent testicular neoplasia: case-control study with epidemiologic update. *Teratology* **32**, 51-56.

Heederik, D.J.J. (1994) Toxic Terror: The Truth Behind the Cancer Scares. *British Medical Journal* **308**, 545

Henderson, R.A. and D'Andrea, G.H. (1993) Interpreting the pathologist's report. *Seminars in Veterinary Medicine and Surgery (Small Animal)* **8**, 289-293.

Hirao, K., Matsumura, K., Imagawa, A., Enomoto, Y. and Hosogi, Y. (1974) Primary neoplasms in dog liver induced by diethylnitrosamine. *Cancer Research* **34**, 1870-1882.

Hjalmars, U., Kulldorff, M., Gustafsson, G. and Nagarwalla, N. (1996) Childhood leukaemia in Sweden: using GIS and a spatial scan statistic for cluster detection. *Statistics in Medicine* **15**, 707-715.

Hjalmars, U., Kulldorff, M., Wahlqvist, Y. and Lannering, B. (1999) Increased incidence rates but no space-time clustering of childhood astrocytoma in Sweden, 1973-1992: a population-based study of pediatric brain tumors. *Cancer* **85**, 2077-2090.

Hosmer, D.W. and Lemeshow, S. (1989) *Applied Logistic Regression*, New York: Chichester Wiley.

Howard, E.B. and Neilson, S.W. (1965) Neoplasia of the Boxer dog. *American Journal of Veterinary Research* **26**, 1121-1131.

Hughes, A.W., Sherlock, D.A., Hamblen, D.L. and Reid, R. (1987) Sarcoma at the site of a single hip screw. A case report. *Journal of Bone and Joint Surgery - British Volume* **69**, 470-472.

Inskip, P.D., Hartshorne, M.F., Tekkel, M., Rahu, M., Veidebaum, T., Auvinen, A., Crooks, L.A., Littlefield, L.G., McFee, A.F., Salomaa, S., Makinen, S., Tucker, J.D., Sorensen, K.J., Bigbee, W.L. and Boice, J.D., Jr. (1997) Thyroid nodularity and cancer among Chernobyl cleanup workers from Estonia. *Radiation Research* **147**, 225-235.

Ivanov, V.K., Tsyb, A.F., Nilova, E.V., Efendiev, V.F., Gorsky, A.I., Pitkevich, V.A., Leshakov, S.Y. and Shiryaev, V.I. (1997) Cancer risks in the Kaluga oblast of the Russian Federation 10 years after the Chernobyl accident. *Radiation and Environmental Biophysics* **36**, 161-167.

Jacquez, G.M. (1996a) A k nearest neighbour test for space-time interaction. *Statistics in Medicine* **15**, 1935-1949.

Jacquez, G.M. (1996b) Disease cluster statistics for imprecise space-time locations. *Statistics in Medicine* **15**, 873-885.

Jaffe, M.H., Hosgood, G., Taylor, H.W., Kerwin, S.C., Hedlund, C.S., Lopez, M.K., Davidson, J.R., Miller, D.M. and Paranjpe, M. (2000) Immunohistochemical and clinical evaluation of p53 in canine cutaneous mast cell tumors. *Veterinary Pathology* **37**, 40-46.

Johnson, A.S., Couto, C.G. and Weghorst, C.M. (1998) Mutation of the p53 tumor suppressor gene in spontaneously occurring osteosarcomas of the dog. *Carcinogenesis* **19**, 213-217.

Johnson, R.C. (1992) Canine spirocercosis and associated esophageal sarcoma. *Compendium on Continuing Education for the Practicing Veterinarian* **14**, 577-580.

Joslin, C.A. (1990) National cancer control and cancer registration. *British Journal of Cancer* **62**, 882.

Keller, E.T. and Madewell, B.R. (1992) Locations and types of neoplasms in immature dogs: 69 cases (1964-1989). *Journal of the American Veterinary Medical Association* **200**, 1530-1532.

Kelsey, J.L., Moore, A.S. and Glickman, L.T. (1998) Epidemiologic studies of risk factors for cancer in pet dogs. *Epidemiologic Reviews* **20**, 204-217.

Kennel Club (2002a) History of the Kennel Club. http://www.the-kennel-club.org.uk/

Kennel Club (2002b) The Kennel Club breed standards. http://www.the-kennel-club.org.uk/

Kent, P., O'Donoghue, J.M. and O'Hanlon, D.M. (1995) Linkage analysis and the susceptibility gene (BRCA-1) in familial breast cancer. *European Journal of Surgical Oncology* **21**, 240.

King, R.J.B. (2000) *Cancer biology*, Harlow: Pearson Education.

Kitchell, B.K., Orenberg, E.K., Brown, D.M., Hutson, C., Ray, K., Woods, L. and Luck, E. (1995) Intralesional sustained-release chemotherapy with therapeutic implants for treatment of canine sun-induced squamous cell carcinoma. *European Journal of Cancer* **31**, 2093-2098.

Klein, M.K. (1996) Tumors of the female reproductive system. In: Withrow, S.J. and MacEwen, E.G., (Eds.) *Small Animal Clinical Oncology*, pp. 347-345. Philadelphia: W.B. Saunders.

Knapp, D.W., Glickman, N.W., Widmer, W.R., DeNicola, D.B., Adams, L.G., Kuczek, T., Bonney, P.L., DeGortari, A.E., Han, C. and Glickman, L.T. (2000) Cisplatin versus cisplatin combined with piroxicam in a canine model of human invasive urinary bladder cancer. *Cancer Chemotherapy and Pharmacology* **46**, 221-226.

Knapp, D.W. and Waters, D.J. (1997) Naturally occurring cancer in pet dogs: important models for developing improved cancer therapy for humans. *Molecular Medicine Today* **3**, 8-11.

Knight, G.C. and Douglas, S.W. (1943) The incidence of cancer in dogs. *Veterinary Record* **55**, 195-196.

Koss, L.G., Woyke, S. and Olszewski, W. (1984) *Aspiration Biopsy. Cytologic Interpretation and Histologic Bases*, New York: Igaku-Shoin.

Kulldorff, M., Athas, W.F., Feurer, E.J., Miller, B.A. and Key, C.R. (1998) Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *American Journal of Public Health* **88**, 1377-1380.

Kulldorff, M., Feuer, E.J., Miller, B.A. and Freedman, L.S. (1997) Breast cancer clusters in the northeast United States: a geographic analysis. *American Journal of Epidemiology* **146**, 161-170.

Kulldorff, M. and Nagarwalla, N. (1995) Spatial disease clusters: detection and inference. *Statistics in Medicine* **14**, 799-810.

Kulldorff, M., Rand, K., Gherman, G., Williams, K. G., and DeFrancesco, D (1998) SatScan, version 2.1, software for calculating the spatial, temporal and space-time scan statistics. http://www3.cancer.gov/prevention/bb/satscan.html. National Cancer Institute.

Kwapien, R.P., Giles, R.C., Geil, R.G. and Casey, H.W. (1980) Malignant mammary tumors in beagle dogs dosed with investigational oral contraceptive steroids. *Journal of the National Cancer Institute* **65**, 137-144.

Landi, M.T., Baccarelli, A., Tarone, R.E., Pesatori, A., Tucker, M.A., Hedayati, M. and Grossman, L. (2002) DNA repair, dysplastic nevi, and sunlight sensitivity in the development of cutaneous malignant melanoma. *Journal of the National Cancer Institute* **94**, 94-101.

LaRue, S.M., Fox, M.H., Withrow, S.J., Powers, B.E., Straw, R.C., Cote, IM and Gillette, E.L. (1994) Impact of heterogeneity in the predictive value of kinetic parameters in canine osteosarcoma. *Cancer Research* **54**, 3916-3921.

Laurier, D. and Bard, D. (1999) Epidemiologic studies of leukemia among persons under 25 years of age living near nuclear sites. *Epidemiologic Reviews* **21**, 188-206.

Leach, M. (1961) *God had a dog:folklore of the dog*, New Brunswick: Rutgers University Press.

Lemeshow, S. and Hosmer, D.W. (1982) The use of goodness-of-fit statistics in the development of logistic regression models. *American Journal of Epidemiology* **115**, 92-106.

Lengerich, E.J., Teclaw, R.F., Mendlein, J.M., Mariolis, P. and Garbe, P.L. (1992) Pet populations in the catchment area of the Purdue comparative oncology program. *Journal of the American Veterinary Medical Association* **200**, 51-56.

Levine, R.A. and Fleischli, M.A. (2000) Inactivation of p53 and retinoblastoma family pathways in canine osteosarcoma cell lines. *Veterinary Pathology* **37**, 54-61.

Li, X.Q., Hom, D.L., Black, J. and Stevenson, S. (1993) Relationship between metallic implants and cancer: a case-control study in a canine population. *Veterinary and Comparative Orthopaedics and Traumatology* **6**, 70-74.

Lilienfield, D.E. and Stolley, P.D. (1994) Observational studies:II Case-control studies. In: Anonymous *Foundations of Epidemiology*, 3rd edn. pp. 226-251. Oxford: Oxford University Press.

London, C. (2000) Tumor biology. In: Ettinger, S.J., (Ed.) *Textbook of veterinary internal medicine: diseases of the dog and cat*, 5th edn. pp. 478-483. Philadelphia: W.B. Saunders.

London, C.A., Kisseberth, W.C., Galli, S.J., Geissler, E.N. and Helfand, S.C. (1996) Expression of stem cell factor receptor (c-kit) by the malignant mast cells from spontaneous canine mast cell tumours. *Journal of Comparative Pathology* **115**, 399-414.

Lorenz, K. (1954) *Man Meets Dog*, London: Methuen.

MacEwen, E.G. (1990) Spontaneous tumors in dogs and cats: models for the study of cancer biology and treatment. *Cancer and Metastasis Reviews* **9**, 125-136.

MacEwen, E.G., Khanna, C. and Radinsky, R. (2001) Cancer biology and metastasis. In: Withrow, S.J. and MacEwen, E.G., (Eds.) *Small Animal Clinical Oncology*, 3rd edn. pp. 18-34. Philadelphia: W.B. Saunders.

MacKie, R.M. and Hole, D.J. (1996) Incidence and thickness of primary tumours and survival of patients with cutaneous malignant melanoma in relation to socioeconomic status. *British Medical Journal* **312**, 1125-1128.

MacVean, D.W., Monlux, A.W., Anderson, P.S., Jr., Silberg, S.L. and Roszel, J.F. (1978) Frequency of canine and feline tumors in a defined population. *Veterinary Pathology* **15**, 700-715.

Madewell, B.R. (1987) Cancer diagnosis. In: Theilen, G.H. and Madewell, B.R., (Eds.) *Veterinary Cancer Medicine*, pp. 3-12. Philadelphia: Lea and Febiger.

Madewell, B.R. (1997) Tumor markers. In: Anonymous *Clinical biochemistry of domestic animals*, pp. 761-784. San Diego: Academic Press.

Madewell, B.R., Deitch, A.D., Higgins, R.J., Marks, S.L. and deVere, W.R. (1991) DNA flow cytometric study of the hyperplastic and neoplastic canine prostate. *Prostate* **18**, 173-179.

Madewell, B.R. and Griffey, S.M. (2000) Modern diagnostic strategies for cancer sampling guidelines. In: Bonagura, J.D., (Ed.) *Kirk's current veterinary therapy XIII: small animal practice*, pp. 452-458. Philadelphia: W.B. Saunders.

Madewell, B.R., Griffey, S.M. and Munn, R.J. (1992) Ultrastructure of canine vasoformative tumors. *Journal of Vascular Research* **29**, 50-55.

Mantel, N. and Haenszel, W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* **22**, 719-748.

Martin, P.M., Cotard, M., Mialot, J.P., Andre, F. and Raynaud, J.P. (1984) Animal models for hormone-dependent human breast cancer. Relationship between steroid receptor profiles in canine and feline mammary tumors and survival rate. *Cancer Chemotherapy and Pharmacology* **12**, 13-17.

Martin, R.A. (1993) Incisional versus excisional biopsy. *Seminars in Veterinary Medicine and Surgery (Small Animal)* **8**, 228-234.

Maudsley, G. and Williams, E.M. (1999) What lessons can be learned for cancer registration quality assurance from data users? Skin cancer as an example. *International Journal of Epidemiology* **28**, 809-815.

Mayr, B., Dressler, A., Reifinger, M. and Feil, C. (1998) Cytogenetic alterations in eight mammary tumors and tumor-suppressor gene p53 mutation in one mammary tumor from dogs. *American Journal of Veterinary Research* **59**, 69-78.

Mayr, B., Schaffner, W., Botto, I., Reifinger, M. and Loupal, G. (1997) Canine tumour suppressor gene p53 - mutation in a case of adenoma of circumanal glands. *Veterinary Research Communications* **21**, 369-373.

McCandlish, I.A.P. (2002) Personal communication.

McDermott, J.J. and Schukken, Y.H. (1994) A review of methods used to adjust for cluster effects in explanatory epidemiological studies of animal populations. *Preventive Veterinary Medicine* **18**, 155-173.

McDermott, J.J., Schukken, Y.H. and Shoukri, M.M. (1994) Study design and analytic methods for data collected from clusters of animals. *Preventive Veterinary Medicine* **18**, 175-191.

McKenzie, J.S. (1999) The use of habitat analysis in the control of wildlife tuberculosis in New Zealand. Massey University. PhD thesis.

Meehan, T.P., Hennen, J., Radford, M.J., Petrillo, M.K., Elstein, P. and Ballard, D.J. (1995) Process and outcome of care for acute myocardial infarction among medicare beneficiaries in Connecticut: A quality improvement demonstration project. *Annals of Internal Medicine* **122**, 928-936.

Menache, S. (2000) Hunting and attachment to dogs in the pre-Modern period. In: Podberscek, A.L., Paul, E.S. and Serpell, J.A., (Eds.) *Companion Animals & Us: exploring the relationships between people and pets*, pp. 42-60. Cambridge: Cambridge University Press.

Mendoza, S., Konishi, T., Dernell, W.S., Withrow, S.J. and Miller, C.W. (1998) Status of the p53, Rb and MDM2 genes in canine osteosarcoma. *Anticancer Research* **18**, 4449-4453.

Michaelis, J., Haaf, H.G., Zollner, J., Kaatsch, P., Krummenauer and Berthold, F. (1996) Case control study of neuroblastoma in west-Germany after the Chernobyl accident. *Klinische Padiatrie* **208**, 172-178.

Michell, A.R. (1999) Longevity of British breeds of dog and its relationships with sex, size, cardiovascular variables and disease. *Veterinary Record* **145**, 625-629.

Misdorp, W. (1987) Incomplete surgery, local immunostimulation, and recurrence of some tumour types in dogs and cats. *Veterinary Quarterly* **9**, 279-286.

Misdorp, W. (1996) Veterinary cancer epidemiology. *Veterinary Quarterly* **18**, 32-36.

Miyazawa, T. and Jarrett, O. (1997) Feline leukaemia virus proviral DNA detected by polymerase chain reaction in antigenaemic but non-viraemic ('discordant') cats. *Archives of Virology* **142**, 323-332.

Mol, J.A., van Garderen, E., Rutteman, G.R. and Rijnberk, A. (1996) New insights in the molecular mechanism of progestin-induced proliferation of mammary epithelium: induction of the local biosynthesis of growth hormone (GH) in the mammary glands of dogs, cats and humans. *Journal of Steroid Biochemistry and Molecular Biology* **57**, 67-71.

Monlux, A.W., Roszel, J.F., MacVean, D.W. and Palmer, T.W. (1977) Classification of Epithelial Canine Mammary Tumors in a Defined Population. *Veterinary Pathology* **14**, 194-217.

Moore, D.A. and Carpenter, T.E. (1999) Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiologic Reviews* **21**, 143-161.

Moran, P.A.P. (1948) The interpretation of statistical maps. *Journal of the Royal Statistical Society, Series B* **10**, 243-251.

Morris, J. and Dobson, J. (2001a) Diagnosis and staging. In: *Small Animal Oncology*, pp. 15-30. Blackwell Science Ltd.

Morris, J. and Dobson, J. (2001b) Pathogenesis and tumour biology. In: *Small Animal Oncology*, pp. 4-14. Blackwell Science Ltd.

Morris, J.S., Bostock, D.E., McInnes, E.F., Hoather, T.M. and Dobson, J.M. (2000) Histopathological survey of neoplasms in flat-coated retrievers, 1990 to 1998. *Veterinary Record* **147**, 291-295.

Morris, J.S., McInnes, E.F., Bostock, D.E., Hoather, T.M. and Dobson, J.M. (2002) Immunohistochemical and histopathologic features of 14 malignant fibrous histiocytomas from Flat-Coated Retrievers. *Veterinary Pathology* **39**, 473-479.

Morrison, W.B. (1998) Instruments and techniques for nonsurgical biopsy. In: Morrison, W.B., (Ed.) *Cancer in dogs and cats: medical and surgical management*, pp. 123-129. Baltimore: Williams and Wilkins.

Morrison, W.B. and De Nicola, D.B. (1993) Advantages and disadvantages of cytology and histopathology for the diagnosis of cancer. *Seminars in Veterinary Medicine and Surgery (Small Animal)* **8**, 222-227.

Moss, S.M., Smith, J.A. and Nicholas, D.S. (1997) The quality of histopathology data in a computerised cancer registration system: implications for future audit of care. *Public Health* **111**, 101-106.

Moulton, J.E. (1990) *Tumours in Domestic Animals*, 3rd edn. Berkeley and Los Angeles: University of California press.

Moulton, J.E., Rosenblatt, L.S. and Goldman, M. (1986) Mammary tumors in a colony of beagle dogs. *Veterinary Pathology* **23**, 741-749.

Mousing, J., Lybye, H., Barfod, K., Meyling, A., Ronsholt, L. and Willeberg, P. (1990) Chronic pleuritis in pigs for slaughter: an epidemiological study of infectious and rearing system-related risk factors. *Preventive Veterinary Medicine* **9**, 107-119.

Muir, C.S., Demaret, E. and Boyle, P. (1985) The Cancer Registry in cancer control: an overview. In: Parkin, D.M., Wagner, G. and Muir, C.S., (Eds.) *The Role of the Registry in Cancer Control: IARC Scientific Publications (Lyon)*, **66**, 13-26

Mulvihill, J.J. and Priester, W.A. (1978) Tumours in young domestic animals: epidemiologic comparisons with man. In: Severi, L., (Ed.) *Tumors of Early Life in Man and Animals*, Montelucc, Italy: Perugia University, Division of Cancer Research.

Nikula, K.J., Benjamin, S.A., Angleton, G.M., Saunders, W.J. and Lee, A.C. (1992) Ultraviolet radiation, solar dermatosis, and cutaneous neoplasia in Beagle dogs. *Radiation Research* **129**, 11-18.

Nordstoga, K., Arnesen, K., Gamlem, H., Glattre, E., Grondalen, J. and Moe, L. (1997) Cancer in dogs in Norway. *European Journal of Companion Animal Practice* **7**, 41-47.

O'Brien, D.J., Kaneene, J.B., Getis, A., Lloyd, J.W., Rip, M.R. and Leader, R.W. (1999) Spatial and temporal distribution of selected canine cancers in Michigan, USA, 1964-1994. *Preventive Veterinary Medicine* **42**, 1-15.

O'Brien, D.J., Kaneene, J.B., Getis, A., Lloyd, J.W., Swanson, G.M. and Leader, R.W. (2000) Spatial and temporal comparison of selected cancers in dogs and humans, Michigan, USA, 1964-1994. *Preventive Veterinary Medicine* **47**, 187-204.

O'Brien, D.J., Kaneene, J.B. and Poppenga, R.H. (1993) The use of mammals as sentinels for human exposure to toxic contaminants in the environment. *Environmental Health Perspectives* **99**, 351-368.

Ogilvie, G.K., Fettman, M.J., Mallinckrodt, C.H., Walton, J.A., Hansen, R.A., Davenport, D.J., Gross, K.L., Richardson, K.L., Rogers, Q. and Hand, M.S. (2000) Effect of fish oil, arginine, and doxorubicin chemotherapy on remission and survival time for dogs with lymphoma. A double-blind, randomized placebo-controlled study. *Cancer* **88** , 1916-1928.

Olsen, S.F., Martuzzi, M. and Elliott, P. (1996) Cluster analysis and disease mapping--why, when, and how? A step by step guide. *British Medical Journal* **313**, 863-866.

Olsen, S.J. (1985) *Origins of the Domestic Dog: The Fossil Record*, Tucson: The University of Arizona Press.

Onions, D.E. (1980) RNA-dependent DNA polymerase activity in canine lymphosarcoma. *European Journal of Cancer* **16**, 345-350.

Onions, D.E. (1984) A prospective survey of familial canine lymphosarcoma. *Journal of the National Cancer Institute* **72**, 909-912.

Openshaw, S., Charlton, M., Wymer, C. and Craft, A.W. (1987) A mark 1 analysis machine for the automated analysis of point data sets. *International Journal of Geographical Information Systems* **1**, 358.

Openshaw, S., Craft, A.W., Charlton, M. and Birch, J.M. (1988) Investigation of leukaemia clusters by use of a Geographical Analysis Machine. *Lancet* **1**, 272-273.

Padgett, G.A., Madewell, B.R., Keller, E.T., Jodar, L. and Packard, M. (1995) Inheritance of histiocytosis in Bernese mountain dogs. *Journal of Small Animal Practice* **36**, 93-98.

Palotay, J.L. (1983) SNOMED-SNOVET: an information system for comparative medicine. *Medical Informatics* **8**, 17-21.

Peaston, A.E. and Watson, A.J. (1995) A survey on veterinary clinical oncology. *Australian Veterinary Practitioner* **25**, 2-7.

Pectasides, D., Bafaloucos, D., Antoniou, F., Gogou, L., Economides, N., Varthalitis, J., Dimitriades, M., Kosmidis, P. and Athanassiou, A. (1996) TPA, TATI, CEA, AFP, beta-HCG, PSA, SCC, and CA 19-9 for monitoring transitional cell carcinoma of the bladder. *American Journal of Clinical Oncology* **19**, 271-277.

Perez, A.D., Rutteman, G.R., Pena, L., Beynen, A.C. and Cuesta (1998) Relation between habitual diet and canine mammary tumors in a case-control study. *Journal of Veterinary Internal Medicine* **12**, 132-139.

Pet Food Manufacturers Association (2002) Pet ownership. http://www.pfma.com/petownership.htm

Pfaffenberger, C.J., Scott, J.P., Fuller, J.L., Ginsburg, B.E. and Biefelt, S.W. (1976) *Guide dogs for the blind: their selection, development, and training*, Amsterdam, Oxford, New York: Elsevier Scientific Publishing Company.

Pfeiffer, J.D. and Wick, M.R. (1991) The pathologic evaluation of neoplastic diseases. In: Holleb, A.I., Fink, D.J. and Murphy, G.P., (Eds.) *American Cancer Society Textbook of Clinical Oncology*, 1st edn. pp. 7. Atlanta: American Cancer Society.

Pierrepoint, C.G. (1985) Possible benefits to veterinary medicine of considering the dog as a model for human cancer. *Journal of Small Animal Practice* **26**, 43-47.

Pitts, M. and Roberts, M. (1997) *Fairweather Eden: life in Britain half a million years ago as revealed by the excavations at Boxgrove*, London: Century.

Pollari, F.L., Bonnett, B.N., Allen, D.G., Bamsey, S.C. and Martin, S.W. (1996a) Quality of computerized medical record abstract data at a veterinary teaching hospital. *Preventive Veterinary Medicine* **27**, 141-154.

Pollari, F.L., Bonnett, B.N. and Bamsey, S.C. (1996b) Electronic medical records at North American veterinary colleges - 1995. *Journal of Veterinary Medical Education* **23**, 21-24.

Powers, B.E. (1996) The pathology of neoplasia. In: Withrow, S.J. and MacEwen, E.G., (Eds.) *Small Animal Clinical Oncology* , 2nd edn. pp. 4-15. Philadelphia: W.B. Saunders.

Powers, B.E. (2001) The pathology of neoplasia. In: Withrow, S.J. and MacEwen, E.G., (Eds.) *Small Animal Clinical Oncology*, 3rd edn. pp. 4-17. Philadelphia: W.B. Saunders.

Priester, W.A. (1964) *Standard Nomenclature of Veterinary Diseases and Operations*, Washington D.C.: U.S. Govt. Print Off.

Priester, W.A. (1967) Canine lymphoma: relative risk in the boxer breed. *Journal of the National Cancer Institute* **39**, 833-845.

Priester, W.A. (1975) Collecting and using veterinary clinical data. In: Ingram, D.G., Mitchell, W.B. and Martin, S.W., (Eds.) *Animal Disease Monitoring*, pp. 119-128. Illinois: Charles C Thomas.

Priester, W.A., Goodman, D.G. and Theilen, G.H. (1977) Nine simultaneous primary tumors in a Boxer dog. *Journal of the American Veterinary Medical Association* **170**, 823-826.

Priester, W.A. and Mantel, N. (1971) Occurrence of tumors in domestic animals. Data from 12 United States and Canadian colleges of veterinary medicine. *Journal of the National Cancer Institute* **47**, 1333-1344.

Priester, W.A. and McKay, F.W. (1980) The occurrence of tumors in domestic animals. *National Cancer Institute Monographs* **No. 54**, National Cancer Institute.

Priester, W.A., Wade, W.M. and McKay, F.W. (1977) *Coding Supplement to the Standard Nomenclature of Veterinary Diseases and Operations*, Washington,D.C: U.S. Govt. Print Off.

Prokocimer, M. and Rotter, V. (1994) Structure and function of p53 in normal cells and their aberrations in cancer cells: projection on the hematologic cell lineages. *Blood* **84** , 2391-2411.

Prymak, C (1999) Principles of surgical oncology. Small Animal Oncology into the new millenium. http://members.aol.com/jcrammond/saonm.pdf

Ragland III, W.L. and Gorham, J.R. (1967) Tonsillar carcinoma in rural dogs. *Nature* **214**, 925-926.

Rajack, L.S. (1997) Pets and human health: the influence of pets on cardiovascular and other aspects of owners' health. University of Cambridge. PhD thesis.

Ray, W.A., Griffin, M.R., Fought, R.L. and Adams, M.L. (1992) Identification of fractures from computerized medicare files. *Journal of Clinical Epidemiology* **45**, 703-714.

Rebar, A.H., Hahn, F.F., Halliwell, W.H., DeNicola, D.B. and Benjamin, S.A. (1980) Microangiopathic hemolytic anemia associated with radiation-induced hemangiosarcomas. *Veterinary Pathology* **17**, 443-454.

Reid-Smith, R.J. (1999) The incidence of neoplasia in the canine and feline populations of private veterinary practices in southern Ontario. Faculty of Graduate Studies, University of Guelph. Doctor of Veterinary Science thesis.

Reid-Smith, R.J., Bonnett, B.N., Martin, S.W., Kruth, S.A., Abramms-Ogg, A. and Hazlett, M.J. (2000) The incidence of neoplasia in the canine and feline patient populations of private veterinary practices in southern Ontario. *Proceedings of the 9th International Symposium on Veterinary Epidemiology and Economics, Breckenridge, Colorado, USA.*

Reif, J.S. and Brodey, R.S. (1969) The relationship between cryptorchidism and canine testicular neoplasia. *Journal of the American Veterinary Medical Association* **155**, 2005-2010.

Reif, J.S., Bruns, C. and Lower, K.S. (1998) Cancer of the nasal cavity and paranasal sinuses and exposure to environmental tobacco smoke in pet dogs. *American Journal of Epidemiology* **147**, 488-492.

Reif, J.S. and Cohen, D. (1971) The environmental distribution of canine respiratory tract neoplasms. *Archives of Environmental Health* **22**, 136-140.

Reif, J.S., Dunn, K., Ogilvie, G.K. and Harris, C.K. (1992) Passive smoking and canine lung cancer risk. *American Journal of Epidemiology* **135**, 234-239.

Reif, J.S., Lower, K.S. and Ogilvie, G.K. (1995) Residential exposure to magnetic fields and risk of canine lymphoma. *American Journal of Epidemiology* **141**, 352-359.

Reif, J.S., Maguire, T.G., Kenney, R.M. and Brodey, R.S. (1979) A cohort study of canine testicular neoplasia. *Journal of the American Veterinary Medical Association* **175**, 719-723.

Restucci, B., Vico, G.d. and Maiolino, P. (2000) Evaluation of angiogenesis in canine mammary tumors by quantitative platelet endothelial cell adhesion molecule immunohistochemistry. *Veterinary Pathology* **37**, 297-301.

Reynolds, P.M., Reif, J.S., Ramsdell, H.S. and Tessari, J.D. (1994) Canine exposure to herbicide-treated lawns and urinary excretion of 2,4- dichlorophenoxyacetic acid. *Cancer Epidemiology, Biomarkers and Prevention* **3**, 233-237.

Ritvo, H. (1994) Animals in nineteenth-century Britain: complicated attitudes and competing categories. In: Manning, A. and Serpell, J.A., (Eds.) *Animals and Human Society:changing perspectives*, pp. 106-126. London: Routledge.

Robinson, G.W. (1968) A comparison of licensed and hospital dog populations. *Journal of the American Veterinary Medical Association* **152**, 1383-1389.

Rosenberg, Z.F., Crowther, R.L., Essex, M., Jarrett, O. and Hasceltine, W.A. (1981) Isolation via transfection of feline leukemia viruses from DNA of naturally occurring feline lymphomas. *Virology* **115**, 203-210.

Rothman, K.J. (1990) A sobering start for the cluster busters' conference. *American Journal of Epidemiology* **132**, S6-13.

Rothman, K.J. and Greenland, S. (1998) *Modern Epidemiology*, 2nd edn. Philadelphia: Lippincott-Raven.

Ru, G., Terracini, B. and Glickman, L.T. (1998) Host related risk factors for canine osteosarcoma. *Veterinary Journal* **156**, 31-39.

Rutteman, G.R. and Misdorp, W. (1993) Hormonal background of canine and feline mammary tumours. *Journal of Reproduction and Fertility - Supplement* **47**, 483-487.

Sagartz, J.E., Bodley, W.L., Gamblin, R.M., Couto, C.G., Tierney, L.A. and Capen, C.C. (1996) p53 tumor suppressor protein overexpression in osteogenic tumors of dogs. *Veterinary Pathology* **33**, 213-221.

Salisbury, S.K. (1998) Principles of the surgical biopsy of cancer. In: Morrison, W.B., (Ed.) *Cancer in dogs and cats: medical and surgical management*, pp. 131-137. Baltimore: Williams and Wilkins.

Sandusky, G.E., Carlton, W.W. and Wrightman, K.A. (1987) Diagnostic immunohistochemistry of canine round cell tumours. *Veterinary Pathology* **24**, 495-499.

Sanz, C., Marcos, B., Rivas, C., Accituno, E., Pena, L., Garcia, M.J. and Echezarreta, G. (1998) Image analysis in canine lymphomas: a comparative study with flow cytometry and proliferation with MIB1 and PC10. *European Journal of Veterinary Pathology* **4**, 73-79.

Sapp, W.J. and Adams, E.W. (1970) C-type viral particles in canine venereal tumor cell cultures. *American Journal of Veterinary Research* **31**, 1321-1323.

Sargartz, J.E., Bodley, W.L., Gamblin, R.M., Couto, C.G., Tierney, L.A. and Capen, C.C. (1996) p53 tumor suppressor protein overexpression in osteogenic tumors of dogs. *Veterinary Pathology* **33**, 213-221.

Savishinsky, J.S. (1983) Pet Ideas: the domestication of animals, human behaviour, and human emotions. In: Katcher, A.H. and Beck, A.M., (Eds.) *New Perspectives on Our Lives with Companion Animals*, pp. 112-131. Philadelphia: University of Pennsylvania Press.

Savolainen, P., Zhang, Y., Luo, J., Lundeberg, J. and Leitner, T. (2002) Genetic Evidence for an East Asian Origin of Domestic Dogs. *Science* **298**, 1610-1613.

Schafer, K.A., Kelly, G., Schrader, R., Griffith, W.C., Muggenburg, B.A., Tierney, L.A., Lechner, J.F., Janovitz, E.B. and Hahn, F.F. (1998) A canine model of familial mammary gland neoplasia. *Veterinary Pathology* **35**, 168-177.

Schneider, R. (1970a) Comparison of age, sex, and incidence rates in human and canine breast cancer. *Cancer* **26**, 419-426.

Schneider, R. (1970b) The natural history of malignant lymphoma and sarcoma in cats and their associations with cancer in man and dog. *Journal of the American Veterinary Medical Association* **157**, 1753-1758.

Schneider, R. (1975) A Population-Based Animal Tumor Registry. In: Ingram, D.G., Mitchell, W.B. and Martin, S.W., (Eds.) *Animal Disease Monitoring*, pp. 162-172. Illinois: Charles C Thomas.

Schneider, R., Dorn, C.R. and Klauber, M.R. (1968) Cancer in households. A human-canine retrospective study. *Journal of the National Cancer Institute* **41**, 1285-1292.

Schneider, R., Dorn, C.R. and Taylor, D.O. (1969) Factors influencing canine mammary cancer development and postsurgical survival. *Journal of the National Cancer Institute* **43**, 1249-1261.

Schneider, R. and Vaida, M.L. (1975) Survey of canine and feline populations: Alameda and Contra Costa Counties, California, 1970. *Journal of the American Veterinary Medical Association* **166**, 481-486.

Schwabe, C.W. (1994) Animals in the Ancient World. In: Manning, A. and Serpell, J., (Eds.) *Animals and Human Society*, 1st edn. pp. 36-58. London: Routledge.

Serpell, J.A. (1986) *In the company of animals*, Oxford: Basil Blackwell.

Serpell, J.A. and Paul, E.S. (1994) Pets and the development of positive attitudes towards animals. In: Manning, A. and Serpell, J., (Eds.) *Animals and Human Society. Changing Perspectives*, pp. 127-144. London and New York: Routledge.

Shimada, A., Shinya, K. and Awakura, T. (1993) Cutaneous papillomatosis associated with papillomavirus infection in a dog. *Journal of Comparative Pathology* **108**, 103

Shofer, F.S., Sonnenschein, E.G., Goldschmidt, M.H., Laster, L.L. and Glickman, L.T. (1989) Histopathologic and dietary prognostic factors for canine mammary carcinoma. *Breast Cancer Research and Treatment* **13**, 49-60.

Skeet, R.G. (1991) Quality and quality control. In: Jensen, O.M., Parkin, D.M., MacLennan, R., Muir, C.S. and Skeet, R.G., (Eds.) *Cancer Registration Principles and Methods. IARC Scientific Publications, Lyon* **95**, pp. 101-107.

Smith, G.D. and Bonnett, B. (1998) Socioeconomic differentials in the mortality of pets. Probably reflect the same differences in material circumstances as in their owners. *British Medical Journal (Clinical Research edition)* **317**, 1671-1673.

Smith, G.D., Carroll, D., Rankin, S. and Rowan, D. (1992) Socioeconomic differentials in mortality: evidence from Glasgow graveyards. *British Medical Journal* **305**, 1554-1557.

Snow, J. (1855) *On the Mode of Communicating Cholera*, 2nd edn. London: John Churchill.

Sonnenschein, E.G., Glickman, L.T., Goldschmidt, M.H. and McKee, L.J. (1991) Body conformation, diet, and risk of breast cancer in pet dogs: a case-control study. *American Journal of Epidemiology* **133**, 694-703.

Spangler, W.L. and Culbertson, M.R. (1992) Prevalence, type, and importance of splenic diseases in dogs: 1,480 cases (1985-1989). *Journal of the American Veterinary Medical Association* **200**, 829-834.

Stedman, T.L. (2000) *Stedman's medical dictionary*, 27th edn. Baltimore: Lippincott Williams and Wilkins.

Stevenson, S. (1991) Fracture-associated sarcomas. *Veterinary Clinics of North America - Small Animal Practice* **21**, 859-872.

Stevenson, S., Hohn, R.B., Pohler, O.E.M., Fetter, A.W., Olmstead, M.L. and Wind, A.P. (1982) Fracture-associated sarcoma in the dog. *Journal of the American Veterinary Medical Association* **180**, 1189-1196.

Stone, E.A. (1995) Biopsy:priciples, technical considerations, and pitfalls. *Veterinary Clinics of North America, Small Animal Practice* **25**, 33-45.

Straw, R.C. (1996) Tumors of the skeletal system. In: Withrow, S.J. and MacEwen, E.G., (Eds.) *Small Animal Clinical Oncology*, 2nd edn. pp. 287-315. Philadelphia: W.B. Saunders.

Stula, E.F., Barnes, J.R., Sherman, H., Reinhardt, C.F. and Zapp, J.A., Jr. (1978) Liver and urinary bladder tumors in dogs from 3,3'-dichlorobenzidine. *Journal of Environmental Pathology and Toxicology* **1**, 475-490.

Sykes, G.P., King, J.M. and Cooper, B.C. (1985) Retrovirus-like particles associated with myeloproliferative disease in the dog. *Journal of Comparative Pathology* **95**, 559-564.

Taylor, D.O. (1965) California's Four-Phase Cancer Field Research Program. *Journal of the American Veterinary Medical Association* **147**, 1492-1497.

Taylor, G.N., Lloyd, R.D., Miller, S.C. and Muggenburg, B.A. (2000) Radium-induced eye melanomas in dogs. *Health Physics* **79**, 196-198.

Taylor, G.N., Thurman, G.B., Mays, C.W., Shabestari, L., Angus, W. and Atherton, D.R. (1981) Plutonium-induced osteosarcomas in the St. Bernard. *Radiation Research* **88**, 180-186.

Tchernov, E. and Valla, F.F. (1997) Two new dogs, and other Natufian dogs, from the southern Levant. *Journal of Archaeological Science* **24**, 65-95.

Teclaw, R.F., Mendlein, J.M., Garbe, P.L. and Mariolis, P. (1992) Characteristics of pet populations and households in the Purdue Comparative Oncology Program catchment area, 1988. *Journal of the American Veterinary Medical Association* **201**, 1725-1729.

Teske, E. (1994) Canine malignant lymphoma: a review and comparison with human non-Hodgkin's lymphoma. *Veterinary Quarterly* **16**, 209-219.

Theilen, G. (1997) Possible relationships between canine hematopoietic neoplasia, other malignancies and immune mediated diseases. *Leukemia* **11 Suppl 3**, 187-188.

Thompson, E.T. and Hayden, A.C. (1961) *Standard Nomenclature of Diseases and Operations*, 5th edn. Chicago: American Medical Association.

Thornburg, L. P. (2001) Cancer Registry. http://www.caninecancer.net

Thrusfield, M.V. (1989) Demographic characteristics of the canine and feline populations of the UK in 1986. *Journal of Small Animal Practice* **30**, 76-80.

Tjalma, R.A. (1966) Canine bone sarcoma: estimation of relative risk as a function of body size. *Journal of the National Cancer Institute* **36**, 1137-1150.

Tomley, F.M., Armstrong, S.J., Mahy, B.W. and Owen, L.N. (1983) Reverse transcriptase activity and particles of retroviral density in cultured canine lymphosarcoma supernatants. *British Journal of Cancer* **47**, 277-284.

Tondel, M., Carlsson, G., Hardell, L., Eriksson, M., Jakobsson, S., Flodin, U., Skoldestig, A. and Axelson, O. (1996) Incidence of neoplasms in ages 0-19 y in parts of Sweden with high 137Cs fallout after the Chernobyl accident. *Health Physics* **71**, 947-950.

Tuan, Y-F. (1984) *Dominance and Affection: The Making of Pets*, New Haven: Yale University Press.

van Leeuwen, I.S., Cornelisse, C.J., Misdorp, W., Goedegebuure, S.A., Kirpensteijn, J. and Rutteman, G.R. (1997) P53 gene mutations in osteosarcomas in the dog. *Cancer Letters* **111**, 173-178.

van Leeuwen, I.S., Hellmen, E., Cornelisse, C.J., Van den Burgh, B. and Rutteman, G.R. (1996) P53 mutations in mammary tumor cell lines and corresponding tumor tissues in the dog. *Anticancer Research* **16**, 3737-3744.

Vastag, B. (1999) Giving new meaning to the word "Watchdog"? *Journal of the National Cancer Institute* **91**, 112-114.

Veldhoen, N., Stewart, J., Brown, R. and Milner, J. (1998) Mutations of the p53 gene in canine lymphoma and evidence for germ line p53 mutations in the dog. *Oncogene* **16**, 249-255.

Veldhoen, N., Watterson, J., Brash, M. and Milner, J. (1999) Identification of tumour-associated and germ line p53 mutations in canine mammary cancer. *British Journal of Cancer* **81**, 409-415.

Vila, C., Savolainen, P., Maldonado, J.E., Amorim, I.R., Rice, J.E., Honeycutt, R.L., Crandall, K.A., Lundeberg, J. and Wayne, R.K. (1997) Multiple and ancient origins of the domestic dog. *Science* **276**, 1687-1689.

Waller, L.A. and Jacquez, G.M. (1995) Disease models implicit in statistical tests of disease clustering. *Epidemiology* **6**, 584-590.

Ward, M.P. and Carpenter, T.E. (2000a) Analysis of time-space clustering in veterinary epidemiology. *Preventive Veterinary Medicine* **43**, 225-237.

Ward, M.P. and Carpenter, T.E. (2000b) Techniques for analysis of disease clustering in space and in time in veterinary epidemiology. *Preventive Veterinary Medicine* **45**, 257-284.

Waters, D.J. (1999) High-grade prostatic intraepithelial neoplasia in dogs. *European Urology* **35** , 456-458.

Waters, D.J. and Bostwick, D.G. (1997a) The canine prostate is a spontaneous model of intraepithelial neoplasia and prostate cancer progression. *Anticancer Research* **17**, 1467-1470.

Waters, D.J., Hayden, D.W., Bell, F.W., Klausner, J.S., Qian, J. and Bostwick, D.G. (1997b) Prostatic intraepithelial neoplasia in dogs with spontaneous prostate cancer. *Prostate* **30**, 92-97.

Wayne, R.K., Nash, W.G. and O'Brien, S.J. (1987) Chromosomal evolution of the Canidae. I. Species with high diploid numbers. *Cytogenetics and Cell Genetics* **44**, 134-141.

Wayne, R.K., Van Valkenburgh, B., Kat, P.W., Fuller, T.K., Johnson, W.E. and O'Brien, S.J. (1989) Genetic and morphologic divergences among sympatric canids (Mammalia: Carnivora). *Journal of Heredity* **80**, 447-454.

Weiss, E. and Greenberg, G. (1997) Service dog selection tests: Effectiveness for dogs from animal shelters. *Applied Animal Behaviour Science* **53**, 297-308.

Whittemore, A.S., Friend, N., Brown, J.B.W. and Holly, E.A. (1987) A test to detect clusters of disease. *Biometrika* **74**, 635

Wilcke, J. R., Livesay, P., and Zimmerman, K. (2002) SNOMED record abstracting guidelines. http://snomed.vetmed.vt.edu/guidelines/default.htm

Withers, F.W. (1939) Squamous-celled carcinoma of the tonsil in the dog. *Journal of Pathology and Bacteriology* **49**, 429-432.

Withrow, S.J. (1996) Why worry about cancer in pet animals? In: Withrow, S.J. and MacEwen, E.G., (Eds.) *Small Animal Clinical Oncology*, 2nd edn. pp. 1-3. Philadelphia: W.B. Saunders.

Withrow, S.J. (2001) Biopsy principles. In: Withrow, S.J. and MacEwen, E.G., (Eds.) *Small Animal Clinical Oncology*, 3rd edn. pp. 63-69. Philadelphia: W.B. Saunders.

Wood, J.L.N., Dobson, J.M. and Rogers, K. (2000a) An epidemiological study of cancer in British dogs based on insurance company data. *Proceedings of the 9th Symposium of the International Society for Veterinary Epidemiology and Economics, Breckenridge, Colorado.*

Wood, J.L.N., Samuel, S., Milstein, I.H., Rogers, K. and Dobson, J.M. (2000b) The effects of breed of dog in the epidemiology of canine neoplasia in the UK. *BSAVA Congress 2000, Birmingham, UK*

Yamagami, T. (1996) A retrospective and prospective study on canine and feline neoplasm biopsies. *Bulletin of the Nippon Veterinary and Animal Science University* **45**, 105-107.

Yates, J.M. and Davidge, M.G. (1984) Can you measure performance? Perspectives in NHS management. *British Medical Journal* **288**, 1935-1936.

Zemann, B.I., Moore, A.S., Rand, W.M., Mason, G., Ruslander, D.M., Frimberger, A.E., Wood, C.A., L'Heureux, D.A., Gliatto, J. and Cotter, S.M. (1998) A combination chemotherapy protocol (VELCAP-L) for dogs with lymphoma. *Journal of Veterinary Internal Medicine* **12**, 465-470.

Zeuner, F.E. (1963) *A History of Domesticated Animals*, London: Hutchinson.