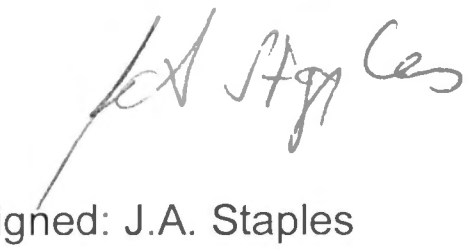
ENVIRONMENTAL INFLUENCES ON MULTIPLE SCLEROSIS AND OTHER AUTOIMMUNE DISORDERS

Judith Ann Staples

A thesis submitted for the degree of Doctor of Philosophy of the Australian National University

October 2012

This is to confirm that, unless otherwise stated, this thesis is entirely my own original work, conducted through the National Centre for Epidemiology and Population Health of the Australian National University.



Signed: J.A. Staples

ACKNOWLEDGMENTS

I wish to thank the following:

Professor Anne-Louise Ponsonby for involving me in the study and for being a source of expert advice during my candidature.

Associate Professor Lynette Lim for expert biostatistical guidance and overall supervision, and for her unfailing encouragement and friendship throughout.

Dr Katie Glass for additional statistical input and for diligently reading and constructively commenting on the final draft.

Professor Tony McMichael for overall support and encouragement, and for commenting on parts of the manuscript.

Emeritus Professor James McLeod for MS Survey data and for providing personal access to original clinical patient files at the University of Sydney.

Dr Lilia Deschamps, BMRC Australia, and Dr Peter Gies, ARPANSA, for Australian UV data.

To Derek and D.J.C. – for ongoing support and inspiration

This thesis was edited by Elite Editing, and editorial intervention was restricted to Standards D and E of the *Australian Standards for Editing Practice*.

ABSTRACT

The risk of autoimmune disorders such as multiple sclerosis (MS) may be influenced by environmental factors early in life. This epidemiological thesis provides new knowledge on the role of early life environmental factors in MS in particular, for which the specific aetiology is unknown.

Two main potential environmental determinants of MS, ultraviolet radiation (UVR) and infections, are explored through an analysis of their timing of action in the life course and a consideration of their possible protective effects. Other organ-specific autoimmune disorders whose aetiology is also unknown, including type 1 diabetes and rheumatoid arthritis (RA), are also assessed for links with UVR for comparison with MS.

Two existing national Australian datasets provided the outcome data for the analyses. The 1995 Australian National Health Survey (NHS) provided summary prevalence estimates for ecological (population-level) analysis of four immune disorders other than MS. The 1981 Australian MS Survey, used for the largest part of the thesis, provided individual-level MS-case data for 1981 throughout Australia, and was further modified to construct a longitudinal MS-rates study dataset for analysis of timing of birth.

Ecological analysis of the 1995 NHS data showed that geographic latitude was positively associated, and regional ambient UVR inversely associated, with the prevalence of type 1 diabetes in Australia. Ambient UVR exposure may thus be a protective factor against such disease at the population level. The association supports previous ecological findings for MS in Australia and adds to the evidence that UVR exposure might be a modifiable determinant for autoimmune disease generally.

Longitudinal analysis of the reconstructed 1981 Australian MS Survey dataset showed that increased MS risk of around 30% was evident in Australians born in November to December (southern hemisphere early summer) compared with those born in May to June (early winter). This MS-risk pattern, indicating environmental factor(s) acting around the time of birth, mirrored (seasonally) that seen for MS in the northern hemisphere, suggesting globally similar perinatal environmental determinants modifying the risk of MS onset. Most importantly, this Australian pattern was found to be fully accounted for by individual, regional (state) and seasonal ambient UVR levels specific to the prenatal period seven to eight months before birth. Low ambient (maternal) UVR exposure in the first trimester of pregnancy thus appears to be associated with a higher risk of MS in the offspring.

Birth-order analysis of cases in the 1981 MS Survey further showed that early birth order was independently associated with MS risk, MS cases being more likely to be one of the *older* siblings in their sibships. Consistent with the hygiene hypothesis, this result suggests that a lack of microbial exposure in early childhood may increase MS risk later in life.

Population health implications of these findings are discussed. In particular, safe sun exposure and/or vitamin D supplementation during early pregnancy may help prevent subsequent onset of high-morbidity, long-duration and presently incurable autoimmune disorders such as MS in the offspring.

PUBLICATIONS ARISING FROM THIS THESIS

Staples J.A, A-L. Ponsonby, L.L-Y. Lim and A.J. McMichael (2003). Ecologic analysis of some immune-related disorders, including type 1 diabetes, in Australia: latitude, regional ultraviolet radiation, and disease prevalence. *Environ Health Perspect* **111**(4): 518-523.

This paper (see Appendix I) is based on the contents in Chapter 4.

Staples, J., A-L. Ponsonby and L. Lim (2010). Low maternal exposure to ultraviolet radiation in pregnancy, month of birth, and risk of multiple sclerosis in offspring: longitudinal analysis. *BMJ* **340**: c1640.

This paper (see Appendix II) is based on the contents in Chapters 5 to 7.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABIS	All Babies in southeast Sweden
ABS	Australian Bureau of Statistics
ACT	Australian Capital Territory
AIRE	Autoimmune regulator
APC	Antigen-presenting cell
APS-1	Autoimmune polyendocrine syndrome-1
ARPANSA	Australian Radiation Protection and Nuclear Safety Agency
ASSDA	Australian Social Science Data Archive
BBB	Blood-brain barrier
BDNF	Brain-derived neurotrophic factor
BMRC	Bureau of Meteorology Research Centre
C'	Complement
CCPGSMS	Canadian Collaborative Project on Genetic Susceptibility to Multiple Sclerosis
CCSG	Canadian Collaborative Study Group
CI	Confidence interval
CIA	Collagen-induced arthritis
CLT	Central limit theorem
CNS	Central nervous system
CSF	Cerebrospinal fluid
CTLA	Cytotoxic T-lymphocyte associated antigen
CUA	*cis*-Urocanic acid
DAISY	Diabetes Autoimmunity Study in the Young
DC	Dendritic cell
d.f.	Degrees of freedom
DHA	Docosahexaenoic acid
DiMe	Childhood Diabetes in Finland Study
DIPP	Diabetes Prediction and Prevention Trial
DLN	Draining lymph node
DSS	Disability status scale
EAE	Experimental allergic/autoimmune encephalomyelitis, 'mouse MS'
EBV	Epstein-Barr virus
EBNA1	EBV nuclear antigen 1
ECP	Extracorporeal photopheresis
EPA	Eicosapentaenoic acid
FDE	First demyelinating events

FFb	Fraction of first born
FFb_2	Fraction of first born relative to last born
FFQ	Food frequency questionnaire
FLb	Fraction of last born
Foxp3	Forkhead Box P3
GA	Glatiramer acetate
HERV	Human endogenous retroviruses
HLA	Human leucocyte antigen
HR	Hazard ratio
HST	Histamine
HHV6	Human herpes virus 6
ICAM	Intercellular adhesion molecule
IDDM	Insulin-dependent diabetes mellitus
IFN	Interferon
Ig	Immunoglobulin
IL	Interleukin
IM	Infectious mononucleosis
IPEX	Immunodysregulation, polyendocrinopathy, enteropathy, X-linked
IRR	Incidence rate ratio
IS	Immune suppression
ISAAC	International Study of Asthma and Allergies in Childhood
iTreg	Induced (Foxp3+) regulatory T cell(s)
IU	International units
K	Keratinocytes
LC	Langerhans cell
LCMV	Lymphocytic choriomeningitis virus
LFA-1	Lymphocyte function-associated antigen 1
LPS	Lipopolysaccharide
MΦ	Macrophage
MBO	Mean birth order
MBP	Myelin basic protein
MED	Minimum erythemal dose
MHC	Major histocompatibility complex
MMP	Matrix metalloproteinase
M/N	Macrophages+neutrophils
MOG	Myelin oligodendrocyte glycoprotein
MON	Monosymptomatic optic neuritis
MOP	Methoxypsoralen

MRI	Magnetic resonance imaging
MS	Multiple sclerosis
MST	Mast cells
NHIS	National Health Interview Survey
NHS	National Health Survey
NK	Natural killer
NKT	Natural Killer T cell
NO	Nitric oxide
NOD	Non-obese diabetic
NSW	New South Wales
NT	Northern Territory
nTreg	Natural regulatory T cell(s)
NZ	New Zealand
OR	Odds ratio
PAF	Platelet-activating factor
PLP	Proteolipid protein
PPMS	Primary progressive multiple sclerosis
PUVA	Psoralen + UVA irradiation
QLD	Queensland
RA	Rheumatoid arthritis
RCT	Randomised controlled trial
RNA	Ribonucleic acid
RR	Relative risk
RRMS	Relapsing-remitting multiple sclerosis
SA	South Australia
s.d.	Standard deviation
SD	Statistical division
SDist	Statistical district
SES	Socio-economic status
SFV	Semliki Forest virus
SLE	Systemic lupus erythematosus
SLO	Secondary lymphoid organs
SPMS	Secondary progressive multiple sclerosis
SSPE	Subacute sclerosing panencephalitis
TAS	Tasmania
TCR	T-cell receptor
TEDDY	The Environmental Determinants of Diabetes in the Young
TGF-β	Transforming growth factor-beta

TNF-α	Tumour necrosis factor-alpha
TLR	Toll-like receptor
TMEV	Theiler's murine encephalomyelitis virus
TOMS	Total ozone mapping spectrometer
TRIGR	Trial to Reduce IDDM in the Genetically at Risk
UK	United Kingdom
US	United States
UV	Ultraviolet
UVA	Ultraviolet A
UVB	Ultraviolet B
UVI	Ultraviolet Index
UVR	Ultraviolet radiation
UVR_{eff}	Effective UVR
VCAM	Vascular cell-adhesion molecule
VDR	Vitamin D receptor
VEGF	Vascular endothelial growth factor
VLA4	Very late antigen 4
VIC	Victoria
WA	Western Australia
WHO	World Health Organization

PROLOGUE

Focus of this thesis

'The ultimate goal of epidemiologic research is the elaboration of causes that can explain patterns of disease occurrence'

Rothman and Greenland (1998) [1]

Organ-specific autoimmune disorders, such as multiple sclerosis (MS), type 1 diabetes mellitus and rheumatoid arthritis (RA), are a significant public health problem, particularly in the Western world, yet their aetiology remains unknown. Some 5 to 20% of the world's population are affected [2] and the morbidity burden contributes markedly to public health costs because of their often long duration; as yet, however, no cures are known. The incidence of autoimmune disorders is increasing worldwide [3], the rate over three decades indicating environmental influences rather than genetic change [4]. Epidemiological studies have indicated that a multifactorial aetiology of MS and other such autoimmune disorders is likely, several putative, possibly sequential, environmental risk factors having been proposed. Further study is required to determine the significance of these candidate risk factors, and in particular their timing of action in the life course before disease onset. Any one such environmental factor may then be amenable to intervention to prevent eventual onset of autoimmune disease [1]. This thesis focuses largely on MS, with additional consideration of type 1 diabetes and RA where applicable.

Autoimmune diseases such as MS have long been observed to be distributed along geographic latitude gradients, particularly among Caucasian populations, with the prevalence of disease increasing with increase in latitude. These gradients are not easily explained by the postulated infectious aetiology for autoimmune disease, wherein autoimmunity is thought to be triggered by early life exposure to viral or bacterial antigens mimicking 'self' tissues [5]. In 1997, McMichael and Hall hypothesised that the latitude gradient may reflect direct suppression of autoimmune activity by ultraviolet radiation (UVR) in sunlight, rather than resulting directly from infective agents, diet or possibly disease susceptibility genes as previously thought. This new hypothesis proposed that, for MS in particular, as latitude was inversely proportional to solar UVR, 'the level of UVR exposure influences aspects of immune functioning involved in the pathogenesis of MS' and that 'UVR-induced suppression of immune function, maximal at low latitudes, attenuates the [infection-related]

autoimmune process that underlies MS' [5]. To test this hypothesis, McMichael and Hall stressed there was a need, first, 'to show that there is an association between UVR and MS at the individual level' and, second, to answer the question 'Does UVR act early in life?', the latter emphasising the probable importance of timing of this (and other such factors) in the aetiology of MS [6]. These questions and this hypothesis are the inspiration and basis for this thesis.

Therefore, the two main, possibly protective, environmental factors explored in this thesis are UVR in sunlight, and infections. Both factors are tested at the individual level for MS, and UVR is also investigated at the population level for some immune disorders other than MS. Moreover, these factors at the individual level are considered particularly with regard to their time of action during the life course (prior to disease onset), because timing can help identify the nature of determinant factors [7]. The timing of infections—for example, whether early or late in childhood/adolescence—can also determine whether they are ultimately protective or directly causal [3, 8].

The overall aim of this thesis is to explore the distribution (in place and time), and thus identify possible environmental determinants, of organ-specific autoimmune disorders in Australia, focusing particularly on MS. As specific aims, the following hypotheses will be explored:

1. Prevalence of autoimmune disorders other than MS, such as type 1 diabetes and RA, exhibits an inverse UVR gradient within Australia (as already shown for MS in Australia in 2001) [9].
2. Immunosuppressive UVR in sunlight acts near the time of birth to protect against subsequent MS risk.
3. Infection or other microbial exposure acts in early childhood to protect against subsequent MS risk.

To test the first hypothesis, an ecological analysis of geographic latitude, regional ambient UVR and prevalence of four immune disorders other than MS, including type 1 diabetes and RA, is conducted at the population level. Such an inverse UVR gradient for autoimmune disorders other than MS was one of McMichael and Hall's specific predictions to test their UVR hypothesis [5]; this information was unknown at the commencement of this thesis, particularly in the southern hemisphere. The broad latitude range and resulting wide range of ambient UVR levels over the Australian continent are a further advantage for this study (results published as Staples et al., 2003, Appendix I).

For the second specific hypothesis, a timing (month)-of-birth (i.e. seasonal) pattern is initially investigated at the individual level, to determine whether an environmental factor such as UVR (or infection or diet or other climatic factor) might be influencing MS risk at this possibly critical life-course period. A further question of whether such a timing-of-birth pattern varies over the latitudinal breadth of Australia is also posed, given that the latitude (and UVR) range is ideal for studying this question. Few studies on timing of birth of MS had been published in the northern hemisphere prior to the commencement of this thesis, and such information was unknown in the southern hemisphere. Finally, for this specific hypothesis, a direct regional (birthplace) and seasonal UVR exposure factor during prenatal development is independently and prospectively explored at the individual level, to contribute specific information on the nature of this possible determinant and particularly its prenatal timing, which had not been studied previously. This prenatal UVR study is unique, suggesting important new information for risk of MS (published as Staples et al., 2010, Appendix II).

To test the third specific hypothesis for MS, that early childhood infection(s) may influence MS risk protectively (also known as the 'hygiene hypothesis', see Chapter 3), an independent analysis using the proxy factor of early birth order is conducted at the individual level. Few studies on sibship characteristics of MS cases in the southern hemisphere had been published prior to this thesis and the findings from the northern hemisphere are inconsistent.

The research questions for this thesis are:

1. Are lower UVR (and higher latitude) associated with prevalence of type 1 diabetes and RA (and other selected immune disorders) in Australia?

2a. Do MS cases show a timing (month)-of-birth pattern in Australia?

   b. Does this pattern vary by (individual) birthplace region over Australia?

   c. Is (individual) perinatal ambient UVR associated with MS risk? Can perinatal ambient UVR explain a timing-of-birth pattern?

3. Do MS cases show a birth-order pattern at the individual level in Australia? Specifically, are MS cases more likely to be one of the older siblings (i.e. early born) in their sibships?

Finally, both major candidate factors for MS (UVR and infections) are considered and discussed as part of a possible sequential 'cascade model' of disease determinants in order to identify where (and when) population health interventions to prevent MS might possibly be instigated.

Datasets

Secondary analyses of existing datasets are a well-established method for addressing research questions that are important and difficult to answer directly, particularly for rare disease conditions such as MS with a long latency period, for which new prospective cohort studies would be both time-consuming and expensive [10, 11]. The use of good quality, national datasets can mean less wastage of available, already well-collected information, and can also reduce the likelihood of bias in primary studies due to, for example, recall, non-response and even the effect on the diagnostic process of attention caused by the research question [12]. The relatively large size of national datasets, and the number of relevant measures, can also ensure population representativeness, making statistical inference more straightforward and enabling a generalisable answer to a high-impact question [11].

This thesis makes use of two existing national Australian datasets, one (for the first hypothesis) comprising summary estimates of several health conditions of the Australian population in 1995, as sampled by the Australian Bureau of Statistics (ABS), and the other (for the second and third hypotheses) comprising original individual-level data on most MS cases in Australia in 1981, as sampled comprehensively by noted MS researcher and neurologist Professor J. McLeod, his team at the University of Sydney, Australia, and his Australian colleagues.

The 1995 National Health Survey (NHS) is part of a comprehensive three to five-yearly health survey by the national statistical organisation in Australia, ABS, that is also responsible for conducting the five-yearly Census of Population and Housing. Like the United States (US) National Health Interview Survey (NHIS), the NHS is conducted by personal interview in households using stratified multi-stage probability design and targeting the civilian, non-institutionalised population of all ages residing in Australia. Good supporting documentation and statistical advice for use of this survey and a high response rate achieved by the ABS (see Chapter 4) make this dataset reliable for valid secondary analysis of selected disorders. The standardised design and interview questionnaire, where obvious errors and biases have been noted and rectified, also contributes to validity; for example, partial-response bias has been accounted for in the 1995 NHS by drawing on previous analysis experience [13].

The 1981 MS Survey, kindly made available for this thesis by Professor McLeod, is used for the largest part of this thesis (see Chapters 5 to 8). This population-based cross-sectional survey is based on comprehensive and state-standardised

ascertainment of MS cases in Australia on 30 June 1981, the date chosen to coincide with an ABS national Census of Population and Housing. MS cases were ascertained widely from a number of national and state sources, and potential cases then generally individually interviewed by the study neurologists, diagnosed, and these diagnoses independently validated. The well-coordinated nature of this unique survey contributes to overall validity and usefulness of this dataset. Several internationally recognised peer-reviewed publications on the epidemiology of MS in Australia, based solely on this dataset, resulted between 1987 and 2011 (see Chapter 5, Table 5.1), making this dataset a valuable resource for the purposes of this thesis. Access to original clinical files of the cases further aided this candidate's familiarity with the primary interviews conducted and subsequent treatment of data.

However, challenges in using this 'prevalence' dataset for timing-of-birth analysis included necessary derivation of suitable denominators and adjustment of these to account for original sampling limitations. This existing Australian survey of MS cases was then utilised essentially as a longitudinal dataset suitable for such analysis, enabling a complete comparison of timing of birth between MS cases and a relevant reference population in a 1920 to 1950 Australia-born 'cohort', rather than simply a disease prevalence study.

Thus, there are issues to be noted that are inherent in using existing datasets: for example, for secondary analysis to be appropriate, the dataset must be adequate to answer the particular research question(s). While the secondary analyst does not have control over design and sample size, the sample should both reflect the population of interest and be large enough to detect a clinically meaningful effect [10]. Familiarity with the dataset and access to information related to the purpose, content, population representation, and previously conducted studies relating to the existing data, provide the information required to assess the quality and hence the reliability and validity of the data [14]. Checks for the extent of missing observations, accuracy of coding and appropriateness of measures available to answer the research questions are also needed [10, 12], as is confirmation of data integrity after transference or other mathematical manipulation by the secondary analyst. These issues, where applicable, are discussed in the relevant chapters of this thesis, together with strategies used to overcome inevitable data limitations. Identification of these limitations here should also inform future studies. A range of different methodological and statistical approaches is applied in the analyses of these two datasets for testing the above hypotheses in this thesis.

Chapter structure

Chapter 1 presents an overview of autoimmune disease, including mechanisms involved in the maintenance and breakdown of immunological self-tolerance leading to such disease. The clinical features and immuno-pathogenesis of MS are described.

Chapters 2 and 3 review the epidemiological evidence for autoimmune disease determinants, particularly environmental factors, and mechanisms for their action, focusing mainly on MS. Chapter 2 considers evidence for sunlight as UVR and/or vitamin D and its possible early life timing of action. Chapter 3 reviews evidence for infections and other possible environmental determinants, and then considers the overall life-course timing of these determinants together.

Chapter 4 focuses on the ecological analysis of the 1995 NHS and reports findings on associations between latitude, regional ambient UVR and prevalence of four immune disorders, including type 1 diabetes and RA.

Chapters 5, 6 and 7 cover analysis of prenatal UVR and timing-of-birth risk patterns of MS in Australia, including specific methodology for this purpose. Chapter 5 focuses on the 1981 MS Survey data, to be used as numerators for subsequent analysis, and reports on cleaning, verification and comparison of these data with those of prior publications.

Chapter 6 describes derivation of reference-population denominators from the 1981 census and supplementary births-registration sources for construction of the required dataset, together with necessary final adjustment of these denominators. Preliminary prevalence relationships are reported as a validity check of the constructed longitudinal dataset to be used for subsequent timing-of-birth analysis.

Chapter 7 presents and discusses results of the main statistical analysis of timing of birth of MS cases in Australia. Prenatal UVR as a prospective regional and seasonal exposure factor is also considered as an independent determinant for MS, and the findings discussed in relation to the timing-of-birth pattern.

Chapter 8 focuses on the original unit-record 1981 MS Survey dataset and analyses the association between birth order of MS cases and MS risk in Australia. Specific methodology for this purpose is included.

The final chapter, Chapter 9, summarises the key findings of this thesis and discusses their implications for future research directions and population health policy.

CHAPTER 1

PATHOGENESIS OF MS AND OTHER ORGAN-SPECIFIC AUTOIMMUNE DISORDERS: IMMUNE MECHANISMS

'The organism possesses certain contrivances by means of which the immunity reaction ... is prevented from acting against the organism's own elements and so giving rise to autotoxins ... so that one might be justified in speaking of a 'horror autotoxicus' of the organism'

Dr Paul Ehrlich, Nobel Laureate, the 'father' and founder of modern immunology, 1897 [15]

1.1 Introduction—autoimmunity and disease

When Paul Ehrlich, a leading German medical scientist working on disease toxins (such as diphtheria), antitoxins and immunity at the turn of the last century, published his 'Side-chain Theory of Antibody Formation' in 1897 and 'On Immunity with Special Reference to Cell Life' [16, 17], he helped establish the science of immunology and introduced a radically new way of thinking about the immune system. His original concept was that the immune system was able to recognise foreign antigens, for example, from pathogens, *before* they were actually encountered in the body. This was achieved by having a 'repertoire' of pre-existing cell-surface receptors, ready to specifically bind any foreign antigen. On encountering the antigen, these cells (lymphocytes) would then be stimulated to produce more specific receptors and secrete them into the extracellular fluid as antibodies. Today, Ehrlich's idea is a principal feature of the lymphocyte clonal-selection theory that is the basis of modern immunology; this was further developed by Niels Jerne [18], MacFarlane Burnet [19] and other immunologists some 60 years later [20]. In addition to identifying lymphocytes, the white-blood-cell mediators of immunity, Ehrlich's experiments on specific receptors and antibodies led him to conclude that immunity was exclusively directed to foreign materials or antigens, and that normally there was no reactivity towards 'self'. That is, the body tolerated 'self' and scrupulously avoided an immune attack on itself (that is, 'horror autotoxicus' or avoidance of autoimmunity).

In fact, we now know that some level of self-reactivity does occur and, indeed, is crucial for the immune system to function optimally [2, 21-23]. Response to self-antigens is

required in the system to ensure that pathogenic microbes that have invaded host cells intracellularly, or have been 'internalised' by the host's phagocytic cells (macrophages) during a non-specific, *innate* inflammatory response, are also recognised as foreign and killed. This is part of the less rapid but antigen-specific *adaptive* immune system of higher animals [24], summarised in Figure 1.1 and the accompanying legend, and further discussed in Section 1.3.2.

Therefore, the notion of a controlled response to self-antigens and no over-abiding or persistent pathological self-reactivity is still current. Thus, we should here differentiate between the terms 'autoimmunity' or 'autoimmune responses', both of which define a necessary physiologic level of self-reactivity, and the term 'autoimmune disease', which defines a pathological state [2, 23].

Autoimmune disease may be said to result when healthy self-tolerance fails and the immune system acts inappropriately [25, 26], or persistently [23, 27] or, in certain micro-environments, such as in the presence of inflammatory cytokines [2], attacks self-tissues.

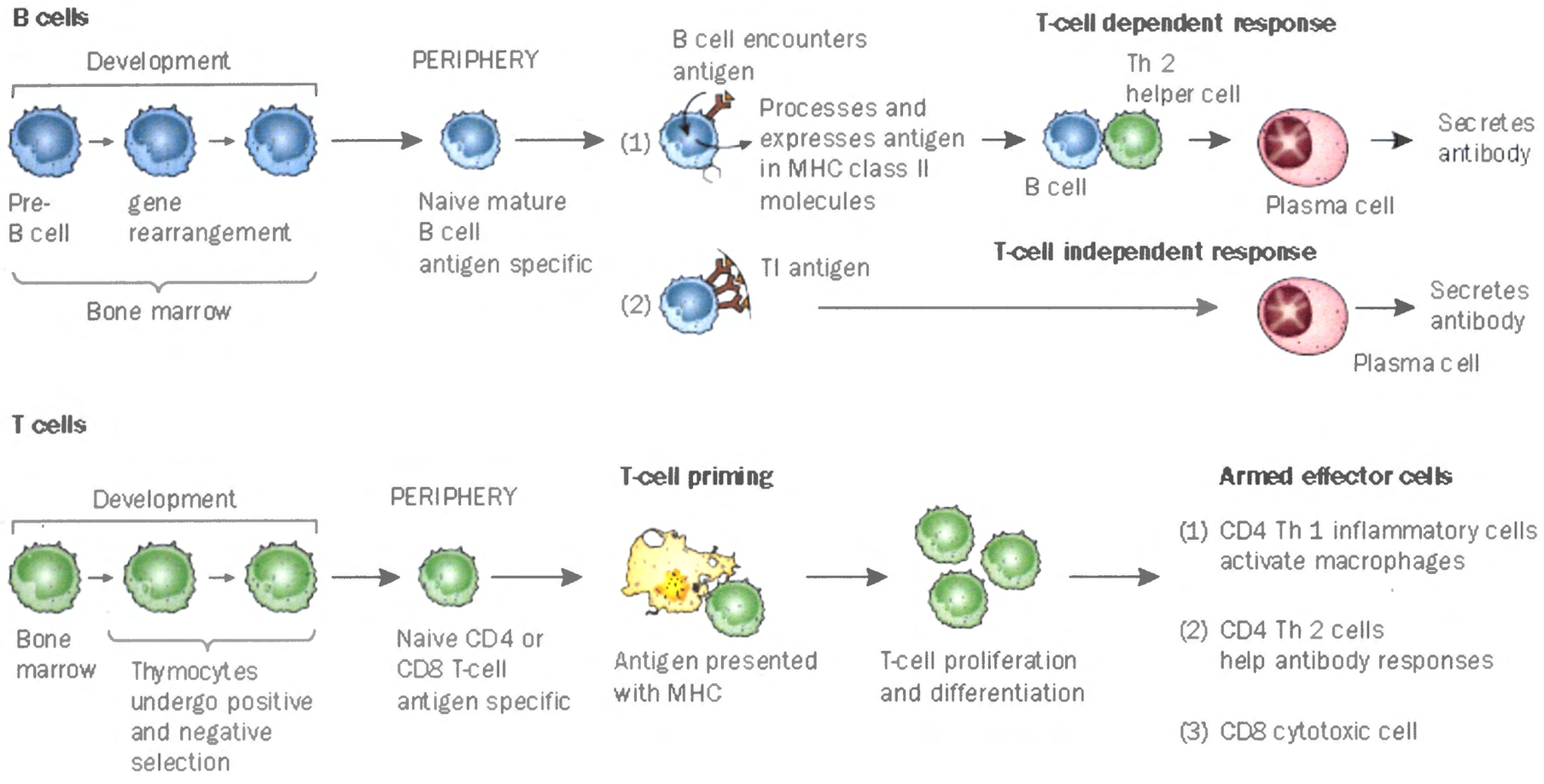


Figure 1.1: Development and role of T and B lymphocytes in antigen-specific (adaptive) immunity; for example, to microbial infection. Both T (thymus) and B (bone marrow) lymphocyte cells originate in bone marrow, but T cells migrate to thymus at early stage as thymocytes. Both cell types undergo random rearrangement of DNA segments, forming a 'repertoire' of antigen-specific receptors on the cell surface that is sufficient for all pathogens likely to be encountered in life. Naïve B cells encounter microbial antigen in secondary lymphoid organs (SLO, e.g. lymph nodes, spleen) and are activated to become plasma cells secreting neutralising/opsonising antibodies, either assisted by, or independently of, T-helper (Th) cells. Naïve T cells encounter antigen-

presenting cells (e.g. dendritic cells) in SLO; antigen is presented as microbial peptide complexed with self-MHC (major histocompatibility complex) molecule, either class I (leading to *CD8+ [cytotoxic] T cells recognising endogenous antigen, e.g. viral) or class II (*CD4+ [helper] T cells recognising exogenous bacteria); co-stimulation, for example, via cytotoxic T-lymphocyte-associated antigen 4 (CTLA4) (on DCs), is also necessary. Antigen-activated T cells proliferate (clonal expansion) and differentiate into effector T-helper cell subsets, producing soluble cytokines that effect both the eradication of pathogen and (downstream) control of tissue damage; most effectors migrate to the site of inflammation but a few remain in SLO as memory T cells for future responses. (Figure from Parkin and Cohen, 2001 [24].) (*CD cell-surface markers defined in footnote 2, Section 1.2.1)

Three such organ-specific autoimmune diseases are MS, type 1 (insulin-dependent) diabetes mellitis (also known as juvenile diabetes or insulin-dependent diabetes mellitis [IDDM]) and RA. The autoimmune targets for these three different disorders are thought to be, respectively, the myelin-producing cells in the central nervous system (MS), the insulin-producing islet beta-cells in the pancreas (type 1 diabetes) and the collagen-producing cells in the synovia of bone joints (RA) [25]. Some 5 to 20% of the world's population is affected by these or other known autoimmune conditions, particularly in Western countries [2]; at least 0.1% of children and adolescents (<20 years) are affected by type 1 diabetes [26, 28]. While 60 to 75% of patients are female in MS and RA, a relatively equal risk between males and females is seen in childhood onset type 1 diabetes. There are also notable differences in the age distribution among autoimmune diseases; for example, the mean or median age of onset for type 1 diabetes, MS and RA is, respectively, 10, 32 and 58 years [28, 301].

Autoimmune diseases are among the leading causes of death among young and middle-aged women (aged <65 years) in the US [29]; almost all autoimmune diseases (except type 1 diabetes) disproportionately affect women [28]. Given that such diseases are typically of long duration, they also constitute a significant morbidity burden [22], and contribute markedly to the costs of public health. For example, type 1 diabetes has been estimated to have average lifetime medical support costs of approximately $1.3 million in the US per person [30]. Yet, despite much intensive research, the aetiology of these diseases remains unknown, their pathogenesis obscure and cures presently unattainable; therapy must therefore be directed at modifying the autoimmune response, since the self-antigens themselves cannot be eliminated from the body.

Autoimmune disease is usually manifested as chronic inflammation, in the absence of ongoing infection or any other discernible cause. The self-targeted autoimmune response is thought to be mediated largely by self-reactive T-cell lymphocytes (i.e. *cell*-mediated immune response), although auto-antibodies produced by B cells are also often involved [2, 27, 31]. The Th1 subtype of T-helper cells[1] was for many years considered to be the self-reactive T cell most important in mediation of autoimmune disease, by virtue of its secreted pro-inflammatory cytokines, including interleukin 2 (IL-2), interferon-γ (IFN-γ) and tumour necrosis factor-α (TNF-α). Polarisation of immune

[1] T-helper cells are effector T cells of the adaptive immune system, assisting cytotoxic T cells to perform their cell-to-cell function by secretion of immunoregulatory cytokines (chemical mediators) [24].

responses towards the production of Th1 cytokines, relative to other T-cell types, was thought to be the essential basis of the autoimmune reaction in MS, type 1 diabetes and RA [25, 26]. However, the 'immune deviation' paradigm of just a decade ago, based on a simple balance between Th1 and Th2 cytokines (the latter secreting anti-inflammatory IL-4, IL-5 and IL-13) [25], has since been superseded by T-cell mediation theories of increasing complexity. For example, pro-inflammatory IL-17-producing T cells have been recently credited with causing and sustaining tissue damage in organ-specific autoimmune disorders in the brain, synovium, intestine and other tissues, and now appear to play a major role in immune-mediated tissue injury as well as Th1 cells [31-35].

In addition, the identification of distinct lineages of immunosuppressive 'regulatory T cells' with wide effects on many immune cells has shifted the emphasis from specific effector T-cell types to less specific, but crucial, regulatory mechanisms to control responses to pathogens and to prevent autoimmune disease. That is, a resurgence of interest in a special subgroup of T cells with 'suppressor' properties has occurred in the last 15 years, and these re-named 'regulatory T cells' have now been found to be of critical importance in limiting inflammation and maintaining a state of dominant self-tolerance in the immunological system [36-39].

Thus, a dynamic homeostatic balance between regulatory and self-reactive T cells is now considered to be the dominant factor in the avoidance of autoimmune disease; these regulatory T cells will be explored further in Section 1.3.1. Therefore, autoimmune disease can be described as a breakdown in homeostatic self-tolerance mechanisms that can occur in a number of different ways [21, 40, 41]. Thus, the T-cell-mediated inflammatory response, if over-expressed or insufficiently regulated, may contribute to autoimmune disease because of a breakdown in tolerance of self-antigens.

In this chapter, the immunological mechanisms for maintaining self-tolerance, and the general conditions possibly leading to the breakdown of this tolerance, will be reviewed by focusing on recent human studies as much as possible, supplemented by studies in various organ-specific disease models in mammals (mainly mice, for example, experimental allergic/autoimmune encephalomyelitis [EAE, 'mouse MS']), in which much experimental work has necessarily been conducted. The history, clinical features and immuno-pathogenesis of MS as a particular example of human organ-specific

autoimmune disease will then be discussed in the final section (see Section 1.5, together with a brief introduction to EAE in Section 1.5.4).

1.2 Immunological self and tolerance

'An extraordinarily precise mechanism of self-recognition holds us back from the brink of autoimmune disease'

Sir Gustav Nossal [22]

Immunological (self-)tolerance may be defined as a functional state in which there is either no apparent immune response to self-antigens found naturally in the body, or in which an anti-self immune response exists but does not lead to tissue damage [27]. The maintenance of immunological tolerance to self-antigens in the body, while still retaining the ability to respond rapidly and effectively to invading non-self potential pathogens, is the crux of the immune system's functional homeostasis. Mechanisms that allow the immune system to mount a protective response towards pathogen-derived foreign antigens, while avoiding a pathological response to self-antigens at the same time, have evolved in a necessarily complex manner.

Immunological tolerance, the system by which the body does not normally over-respond to self-antigens, can be thought of as being generated at two levels, 'central' tolerance, developing primarily in foetal and neonatal life, and 'peripheral' tolerance, developing postnatally for the control of self-reactivity outside the thymus [26, 27].

1.2.1 Central T-cell tolerance

The central lymphoid organs where early life lymphopoiesis takes place comprise the thymus and bone marrow [26]. Although both T (thymus) cells and B (bone marrow, antibody-producing) cells (Figure 1.1) are involved in most immunity processes (and the terms 'central' and 'peripheral' apply equally to both kinds of lymphocytes), T cells appear to play a dominant role in immune tolerance and autoimmunity [21, 42].

The chief mechanism of T-cell self-tolerance is the initial selection, sorting and then deletion by apoptosis (programmed cell death) of most of the self-reactive T cells in the thymus. In the thymus, immature T cells (thymocytes) encounter circulating peptides derived from endogenous (self) proteins that are presented to them bound to major histocompatibility complex (MHC, self) molecules. Depending on the binding affinity between the T-cell antigen receptors (TCRs) and the peptides, only a few of these (self) T-cell complexes eventually survive the process of 'negative selection' or 'clonal deletion', and are then able to migrate to the body's peripheral systems where foreign antigens may be encountered. The exiting naïve (i.e. having not yet encountered their specific foreign antigen) T cells, either T-helpers (bearing the CD4+ cell surface marker)[2] or cytotoxic T cells (CD8+ marker), are self-tolerant but able to recognise foreign antigen when presented to them with self-MHC in the periphery [24] (see Figure 1.1). Self-recognising *regulatory* T cells (mostly CD4+) are now known to also be produced in the thymus and emerge at this early stage [43, 44] (to be discussed in detail in Section 1.3.1). Thus, more than 98% of the self-reacting cell complexes are deleted, or purged, centrally from the immune system in the thymus [20, 45], as was postulated originally by immunologists such as MacFarlane Burnet in 1957 [19].

That is, the centrally located selection mechanisms are not completely effective in removing self-reactive cell complexes, nor are they meant to be (in contrast to earlier thinking). Some, indeed many, auto-reactive cells are necessarily part of the peripheral T-cell repertoire of healthy individuals—the survival of naïve T and B cells in the periphery requires continuous exposure to auto-antigens [2, 22]. For example, both T and B cells recognising insulin or myelin basic protein (MBP) can be isolated from persons without type 1 diabetes or MS [21, 42, 46]. Therefore, rather than central tolerance being thought of as a 'leaky' process [26] with these escapee lymphocytes requiring constant surveillance and elimination, the ongoing presence in the periphery of, for example, myelin-reactive T cells in unaffected persons, is now known to be an essential feature of the immunological tolerance system. However, mechanisms to maintain this state of self-tolerance in the periphery are then required.

[2] Many of the functionally important cell surface molecules and their receptors are described by the cluster of differentiation (CD) nomenclature, based on their identification by characterised monoclonal antibodies [26].

1.2.2 Peripheral T-cell tolerance

A number of different peripheral tolerance mechanisms contributing to the prevention of immune pathology have been known for some years, including passive mechanisms such as immunologic ignorance (T-cell unresponsiveness) and sequestration of antigens (for example, isolating self-antigens behind cellular or vascular barriers), and active mechanisms such as antigen-induced T-cell apoptosis and T-cell anergy (induced unresponsiveness) [47]. Further, antigen-presenting cells, and particularly dendritic cells, have increasingly been recognised as crucial to the overall process of achieving peripheral self-tolerance.

Tolerogenic dendritic cells

Dendritic cells (DCs) can be described as 'professional' specialised antigen-presenting cells (APCs) in that they have a unique ability to first-activate *naïve* T cells, as opposed to 'non-professional' APCs (e.g. B cells, monocytes, macrophages and endothelial cells) capable of stimulating mainly memory (i.e. antigen-experienced) T cells [48]. That is, DCs are particularly specialised for acquiring, processing and presenting both self- and non-self-antigens to naïve T cells so that effector T-cell responses to invading pathogens or tissue damage in peripheral tissues can be instigated (see Figure 1.1). Thus, they have been termed 'the sentinels of the immune system' because they are strategically located at all pathogen entry sites of the body. There, DCs respond to environmental 'danger' signals, including pro-inflammatory cytokines IL-1, TNF-α and IL-6, prostaglandins and pathogen products such as lipopolysaccharide (LPS), and, once activated, migrate from peripheral tissues into SLO (such as lymph nodes) where they mature and most effectively induce and regulate effector T-cell responses to infection [49-51]. Their up-regulated expression of co-stimulatory molecules, and production of soluble pro-inflammatory cytokines such as IL-1, TNF-α and IL-12, then promotes the differentiation of the responder (effector) T cells and amplifies the effector T-cell response [47].

However, DC function has more recently been found to vary with their state of differentiation—that is, whether immature or mature [27]. It is now evident that

immature myeloid[3] DCs have a very different and pivotal role in the induction and maintenance of peripheral tolerance by being able to differentiate into either immunogenic DCs, as described, or into 'tolerogenic' DCs, which can induce T-cell tolerance and thus modulate immune responses [50, 52]. Tolerogenic DCs induce T-cell tolerance by a variety of mechanisms, ranging from production of anti-inflammatory cytokines with broad attenuating effects, to the induction of antigen-specific T-cell responses resulting, on the one hand, in anergy or deletion of effector T cells or, alternatively, stimulation of suppressive regulatory T cells [50] (Figure 1.2).

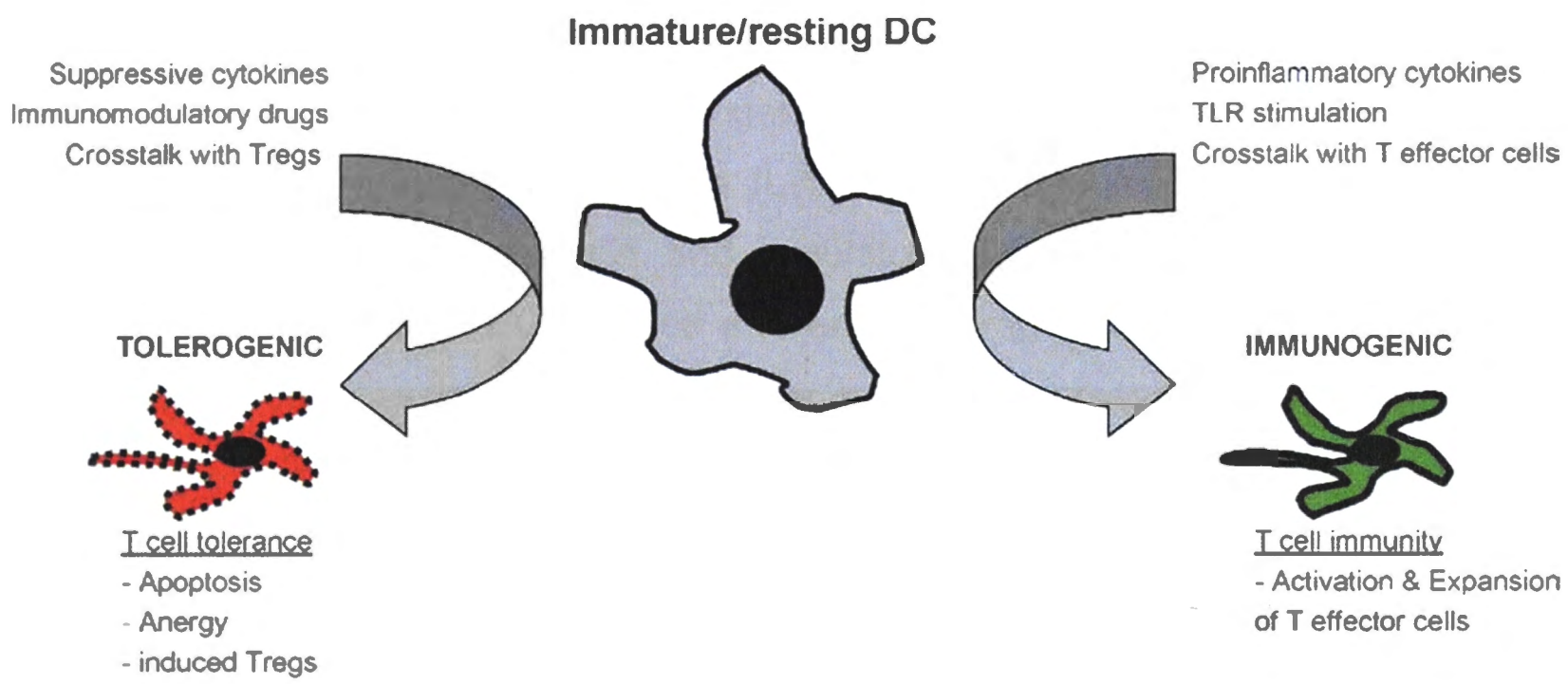


Figure 1.2: Functional plasticity of immature/resting DCs. Highly sensitive to stimulatory and suppressive agents, immature myeloid DCs are accessible to activating and inhibitory factors that favour their immunogenic or tolerogenic function, respectively. Immunosuppressive agents, such as IL-10 or TGF-β, irreversibly convert immature DCs into tolerogenic DCs with locked immunosuppressive function. Tolerogenicity can also be induced in immature DCs by activated regulatory T cells or immunosuppressive drugs. In contrast, pro-inflammatory cytokines (i.e. IL-1, IL-6, TNF-α), microbial products (toll-like receptor [TLR] ligands) and the crosstalk with activated effector T cells induce the terminal differentiation of immature DCs into immunogenic DCs, insusceptible to cellular or soluble tolerating signals. (Figure from Steinbrink et al., 2009 [50].)

[3] DCs are classified into two cell types: myeloid DCs are the most well characterised and are found throughout the body as interstitial cells or as Langerhans cells (LCs) in the skin; they can become either immunogenic (producing IL-12 and efficiently priming naïve T cells) or tolerogenic. Plasmacytoid DCs are less well characterised, are found in the bone marrow and secrete IFN-β in response to viruses and other pathogens; they may be considered naturally occurring tolerogenic DCs [31, 52, 53].

This modulation by tolerogenic DCs occurs primarily constitutively (that is, in the normal, steady, non-inflammatory state when self-antigens are the main antigens being presented in the tissues), but can also occur at the end of an inflammatory process caused by pathogens, in order to control established immune responses and prevent chronic autoimmune disease [27, 49]. That is, in the absence of inflammation and disease, tolerogenic DCs constantly present innocuous self- and non-self-antigens in a way that promotes self-tolerance, and this occurs, in large part, by control of regulatory T cells [49, 54].

In turn, regulatory T cells (considered next section) are able to affect DC development, preventing differentiation into immunogenic DCs and even inducing production of immunosuppressive cytokines such as IL-10 by (tolerogenic) DCs. Thus, there is a mutual interaction between DCs and regulatory T cells for maintaining immune homeostasis: tolerogenic DCs induce and enhance regulatory T cells and, inversely, regulatory T cells assist DCs to become immunosuppressive, in a vital two-way relationship [54]. DCs are instrumental in maintaining a balance between the activities of effector T cells and regulatory T cells in the immune system.

In humans, the incidence of the autoimmune disease systemic lupus erythematosus (SLE) is correlated with a chronically activated DC phenotype [54]. DCs with an altered phenotype, as well as dysfunctional interactions between DCs and regulatory T cells, have been described also in animal models of autoimmune disease (for example, 'mouse MS' [EAE]) and in MS patients (reviewed by Zozulya and Wiendl, 2008 [55]). Conversely, towards tolerance, *in vitro* generated immature (self) antigen-pulsed DCs have been shown to induce antigen-specific T-cell tolerance *in vivo* in human volunteers, by induction of antigen-specific regulatory T cells [50]. More evidence of modulation by DCs has come from autoimmune disease therapy. For example, IFN-β treatment in MS up-regulates a tolerogenic molecule on DCs related to apoptosis (programmed cell death) [56], and reputedly also has a role in inducing regulatory T cells via interactions between them and DCs [55]. Further drug treatments shown to promote induction of tolerogenic DCs include the glucocorticosteroids (discussed with respect to MS therapy in Section 1.5.5); these impair DC maturation and inhibit co-stimulatory expression and pro-inflammatory IL-12 production by DCs, leading to decreased allo- (non-self) stimulatory capacity. Indeed, many anti-inflammatory and immunosuppressive drugs appear to act via induction of such DCs with tolerogenic properties (reviewed by Adorini and Penna, 2009 [52]).

Thus, self-tolerance comprises central tolerance mechanisms, wherein immature lymphocytes recognising self-antigens in bone marrow or thymus are removed by apoptosis, and peripheral (or adaptive) tolerance, where mature lymphocytes encounter self-antigens in peripheral tissues and are killed, shut down or otherwise controlled [57]. Peripheral tolerance mechanisms include anergy or deletion (apoptotic cell death) of effector T cells, modulation by tolerogenic DCs and, most essentially, suppression by regulatory T cells, discussed next.

1.3 Maintenance of tolerance

1.3.1 Regulatory T cells

There has been increasing recognition of specific T-cell populations with suppressive/regulatory properties that are dedicated to the maintenance of tolerance in both humans and experimental animals [36-38, 42, 48, 58-60]. T-cell populations that are able to inhibit the response of other (effector) T cells, and thus suppress the autoimmune response, were described in humans and other mammals several decades ago and have been 'revived' in importance only relatively recently. These regulatory T cells playing a key role in the control of both reactivity to self-antigens and to non-self (such as in graft transplantation) have, appropriately, been receiving much recent attention because they offer hope as use for possible future therapies for autoimmune disorders [61-65].

'Suppressor T cells' were initially defined in the 1970s when Gershon and Kondo described a population of T cells that suppressed the immune response of mice to foreign antigens [66], but the concept was not developed further during the next two decades. It was later found that a variety of autoimmune diseases developed in mice whose thymus was removed soon after birth, clearly indicating the presence of thymic cells that were capable of suppressing autoimmunity. The subsequent discovery of a subpopulation of T cells that could critically control auto-reactive T cells of mice *in vivo* revived the original concept of regulatory cells that were suppressive in action [58]. These now-termed 'regulatory T cells' were found to have potent inhibitory effects on immune responses to foreign antigens and on the development of autoimmune disease. That is, autoimmune disease in multiple organs could be spontaneously produced in normal rodents by removal of these regulatory T cells, but not by general reduction of

T-helper cells. Organ-specific autoimmune disease induction was dependent on the genetic strain, suggesting an autoimmune mechanism that may be common to each disease but is target-modified by genetics [67]. Moreover, in 'adoptive transfer' experiments (discussed further in Section 1.5.4) in mice (reviewed by Sakaguchi et al., 2006 [68] and Roncarolo and Battaglia, 2007 [64]), these regulatory T cells could block the development of autoimmune disease even in thymectomised mice [20].

Further studies identifying the presence of regulatory T cells in human peripheral blood, and their ability to suppress T-cell proliferation *in vitro,* followed [69-71]. Both endogenous 'natural' regulatory T cells (nTreg), produced and fully differentiated in the thymus, and 'induced' regulatory T cells, obtained by administration of exogenous antigen in the periphery, have now been identified, most being of the original CD4+ T-helper type [37] (Figure 1.3). It has been suggested that the thymic-derived regulatory T cells, such as natural nTreg (considered next), have a central homeostatic function to regulate T-cell proliferation in lymphoid organs, while peripheral antigen-induced regulatory T cells, such as Tr1 (considered in the following subsection) have anti-inflammatory capacity and infiltrate injured tissues to control inflammation and tissue destruction locally [72]

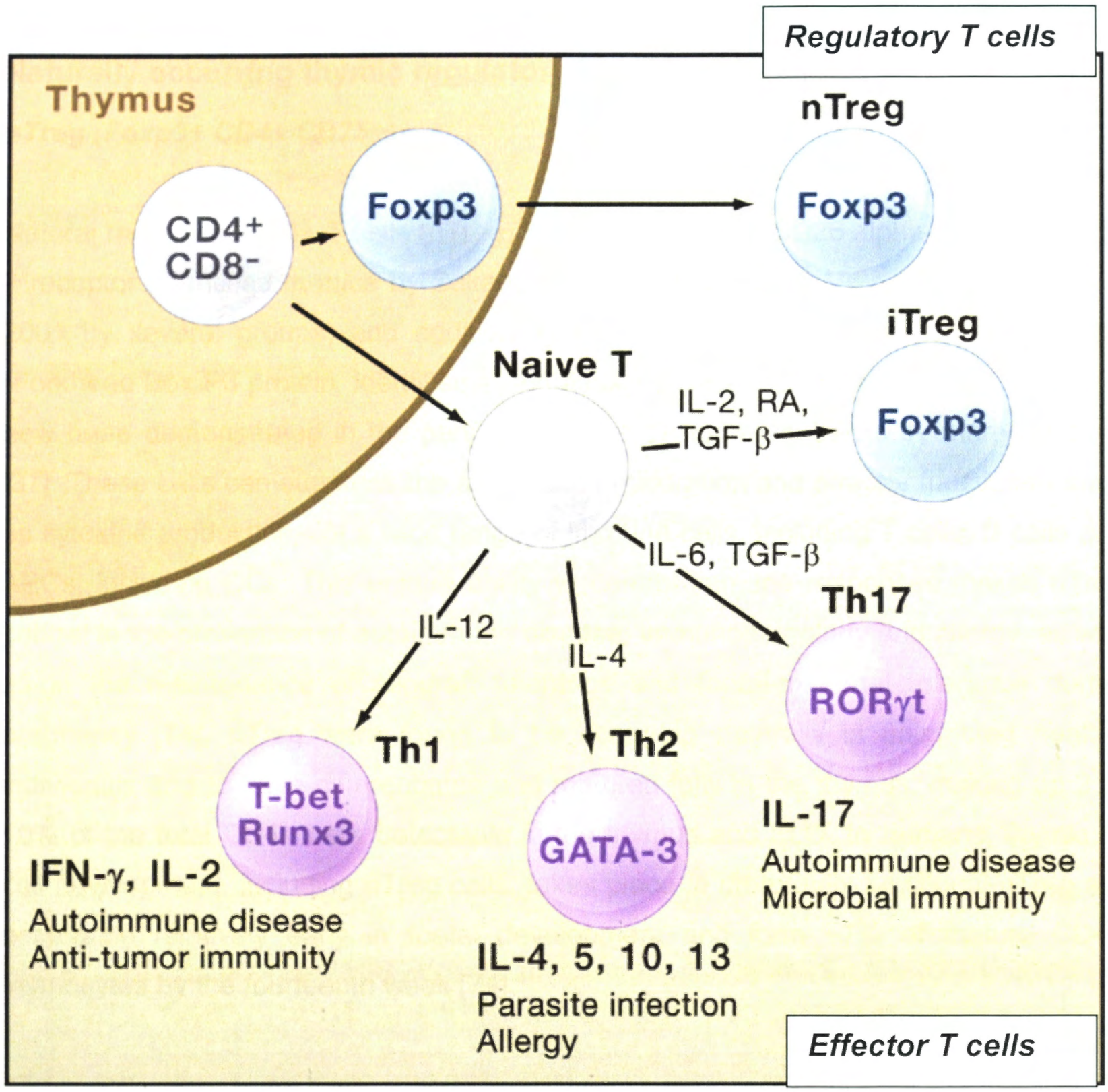


Figure 1.3: Differentiation of CD4+ T cells into regulatory or effector T cells. Cytokines and transcription factors (such as Foxp3) that promote the differentiation of naïve T cells into regulatory (induced iTreg), or effector T cells (Th1, Th2, Th17) for specific conditions, are shown. (Figure adapted from Sakaguchi et al., 2008 [37].)

Naturally occurring thymic regulatory T cells

nTreg (Foxp3+ CD4+ CD25+)

Natural regulatory CD4+ T cells (nTreg)[4] co-expressing the CD25 alpha chain of the IL-2 receptor (identified in mice by Sakaguchi and colleagues in 1995 and in humans in 2001 by several groups) and additionally expressing the Foxp3 transcription factor (Forkhead Box P3 protein, identified in regulatory T cells in 2003, see Figure 1.3) have now been demonstrated in the peripheral blood of human subjects by many workers [37]. These cells can suppress the activation, proliferation and effector functions—such as cytokine production—of a wide range of immune cells, including T cells, B cells and APCs, including DCs. This unique ability to control immune responses makes nTreg central in the prevention of autoimmune disease, immunopathology and allergy, as well as in the maintenance of allograft tolerance and foetal-maternal tolerance during pregnancy [74]. nTreg were found to be occurring naturally in uninfected healthy individuals and to have differentiated and matured fully in the thymus, making up 2 to 10% of the total CD4+ cells detectable in the thymus and SLO. In humans, thymic T-cell development, including nTreg cells, takes place *in utero;* fully functional nTreg are detectable relatively early in foetal development and form ~7% of mature CD4+ thymocytes by the fourteenth week [74].

nTreg cells appear to have a unique ability to suppress neighbouring T cells in an antigen-specific and cell contact-dependent way, inhibiting proliferation and IL-2 production by either CD4+ or CD8+ T cells *in vitro* [48, 59, 64, 72, 75]. Contact-dependent suppression may act via direct cytolytic effects on target cells (leading to apoptosis) or by 'metabolic disruption' (i.e. inducing release of adenosine nucleosides in or from target) [76]. As discussed in the previous section, nTreg may also inhibit by repressing maturation and/or function of DCs and, therefore, the priming of effector T cells (see Figure 1.1) [39, 54].

nTreg may also inhibit via production of soluble anti-inflammatory factors, such as TGF-β (transforming growth factor-β) and IL-10 (these generally inhibitory cytokines discussed in more detail later in this section). For example, IL-10 produced by nTreg is

---

[4] Given the currently accepted importance of naturally-occurring thymic regulatory T cells in immunomodulation, some publications in the literature refer to these regulatory T cells simply as 'Treg'; however, to avoid confusion and to be consistent with major publications, the term 'nTreg' will be used here. The term 'iTreg' will then be reserved for *antigen-induced* Foxp3+ regulatory T cells [37, 73, 74] (as Figure 1.3), and not include other inducible regulatory T cells such as Tr1 and Th3, which are more often, but not always, classified as Foxp3-.

known to be essential for immune regulation at mucosal surfaces such as the intestines, lungs and skin [73]; TGF-β is required for maintenance of Foxp3+ expression [77]. That is, there appear to be a multitude of potential mechanisms constituting the nTreg arsenal—some may be involved in key regulatory functions, such as homeostasis and 'homing' (to particular tissues *in vivo*), and others in, for example, suppression in different tissues (e.g. mucosal versus lymphoid tissue) or in different diseases (e.g. autoimmune versus inflammatory). However, many of the exact mechanisms by which nTreg cells exert suppression *in vitro* and *in vivo* still remain to be determined [74].

Overall, nTreg are vital to the prevention of autoimmune disorders in both humans and experimental animals. The importance of T-cell regulation in human disease is highlighted by the severe inflammation and autoimmune disorders in people with IPEX (immunodysregulation, polyendocrinopathy, enteropathy, X-linked) syndrome, who completely lack nTreg cells as a result of a mutation in the *Foxp3* gene. These individuals develop a broad range of auto-antibodies, insulin-dependent diabetes, thyroiditis, eczema, haemolytic anaemia and irritable bowel disease and, in the absence of a bone marrow transplant, die at an early age [78].

In MS, also, Viglietta and co-workers found that the functional activity of these unique nTreg cells correlated inversely with MS disease activity; importantly, these data demonstrated alterations in function, rather than numbers, of these T cells from patients with the most common form of MS, relapsing-remitting MS (RRMS) [79]. More recently, the same research group confirmed that natural nTreg, and not induced iTreg, were functionally less suppressive in patients with MS [80]. Several groups have now shown that nTreg are functionally impaired or have deficits in their maturation or in their thymic emigration in patients with MS [81-83], while Venken and colleagues demonstrated impaired nTreg function together with decreased Foxp3 expression in RRMS [84]. Huan and colleagues also showed lower levels of Foxp3 messenger ribonucleic acid (RNA) and protein expression in MS patients than in healthy people [85]. Another study by Venken and co-workers confirmed altered generation of nTreg cells, pointing to disturbed thymic development and function in MS patients [86].

However, whether such nTreg cell dysfunction has a causal role in MS, or whether this cell dysfunction represents a more general defect within the immunoregulatory system that can be associated with any autoimmune disorder, is yet to be determined [87]. Cvetanovich and Hafler further caution that many earlier studies have used less defined human regulatory T-cell populations (and therefore impure or heterogeneous

populations, in terms of CD and Foxp3 expression, that might obscure results). Nevertheless, they conclude that nTreg show multiple differences in RRMS patients, compared with healthy controls, that are likely to play roles in the complex pathogenesis of RRMS, including reduced thymic output and various nTreg cell impairments [88].

Defective suppressor function has also been found *in vitro* in these regulatory T cells from patients with type 1 diabetes [89]. A recent study showed that the same regulatory T cells from patients with type 1 diabetes had a reduced capacity to respond to IL-2, this defect in IL-2-induced signalling being correlated with a loss of Foxp3 expression [90]. Other recent studies on type 1 diabetes patients' regulatory T cells also showed ineffective suppression [91, 92]. Indeed, such dysfunction in autoimmune disorders such as type 1 diabetes may be due, in part, to resistance of effector T cells in these type 1 diabetic patients to nTreg suppression [78, 92].

Data demonstrating that Foxp3+ regulatory T cells are functionally compromised in RA patients have also been published [93], the impaired function linked to reduced expression of CTLA-4 [94]. These data further indicated that this suppressive function could be restored by anti-TNF-α, which, rather than reversing the CTLA-4 defect, induced a novel regulatory T-cell population with potent suppressive properties [94] and has been of some therapeutic value in RA [95]. In RA, further, IL-17-producing T-helper cells were resistant to suppression by these regulatory T cells, making this inflammatory condition complex to treat [94].

Interestingly, recent studies (reviewed by Bettini and Vignali, 2009 [96] and Campbell and Koch, 2011 [73]) have shown that different, specialised subsets of nTreg cells are able to restrain the different types of CD4+ effector T-cell responses (i.e. by Th1, Th2, or Th17 T cells) by using the corresponding nuclear transcription factors[5] that are required for initial differentiation of these effector T cells. That is, nTreg utilise different transcriptional 'programmes' depending on the T-helper subset they are attempting to suppress, in order to maintain or restore immune homeostasis. For example, 'T-bet', which is the master transcription factor controlling the differentiation, migration and function of IFNγ-producing Th1 cells (see Figure 1.3), is also expressed by a subset of nTreg cells and is now found to be required for nTreg cell homeostasis and function

[5] The functional specialisation of the various CD4+ T-cell subsets is due to the differential expression of 'master' transcription factors in the cell nucleus, which are activated by signals transduced from the bound antigen on the cell surface to the nucleus and which then turn on distinct programmes of gene expression that control T-cell proliferation, function and migration [24, 73].

during Th1-type pro-inflammatory responses [97]. Possibly T-helper cells and nTreg cells have co-evolved in tandem [96], such that pro-inflammatory immune responses that are necessary in an infection can also be regulated to ensure that host tissues are not harmed after pathogen eradication.

NKT cells

Another unique 'natural', thymus-derived, self-recognising, regulatory T-cell population comprises 'NKT' cells. These are a subset of T cells expressing receptors found on both CD4+ T cells and natural killer (NK)[6] cells. NKT cells can both respond rapidly to invading pathogens (as innate immune cells do) and produce both Th1 and Th2-type cytokines adaptively (Figure 1.4); thus, they are said to bridge the innate and adaptive (i.e. antigen-responding) immune systems. NKT cells have been shown to mediate both protective and regulatory immune functions, including tumour rejection, protection against infectious microbes, maintenance of transplant tolerance and inhibition of autoimmune disease development, their immunomodulatory functions being linked, at least in part, to generation of tolerogenic DCs [98].

NKT cells have been recently functionally linked with amelioration of EAE via Th17 cells in the gut [99]. A selective reduction in NKT cell numbers has also been shown in human patients with autoimmune diseases, including RA, type1 diabetes and MS (reviewed by Balato et al., 2009 [98]). For example, decreased numbers of both NKT and nTreg cells were demonstrated in the peripheral blood of type 1 diabetes patients, suggesting possible thymic dysfunction and/or broad T-cell defects [100].

[6] NK cells are a white blood cell type of the innate immune system with both cytolytic and potent cytokine-producing properties [24].

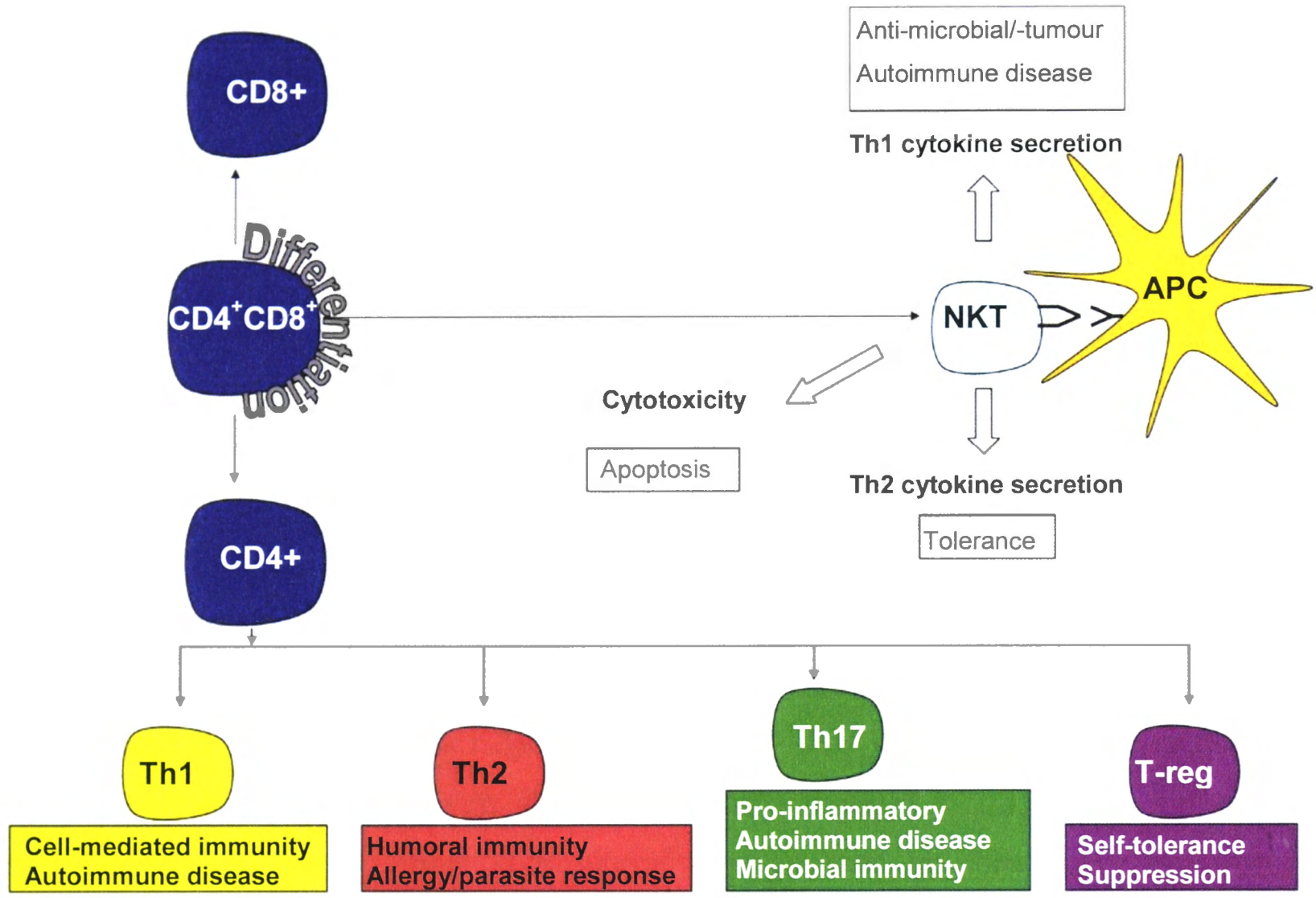


Figure 1.4: Differentiation of (CD4+ CD8+) T cells (in thymus) into CD4+ subsets, CD8+ T cells (CD8+ discussed further in Section 1.5.4), NKT cells and activation of NKT cells by APCs. Activated NKT cells produce Th1 or Th2 cytokines and have direct cytotoxicity activity leading to apoptosis of the target cell. (Figure adapted from Balato et al., 2009 [98].)

Antigen-induced regulatory T cells

Antigen-induced regulatory T cells are adaptive T cells produced in SLO in the periphery of the body (i.e. outside the central thymus and bone marrow, in the lymph nodes and spleen). They are phenotypically diverse and induced by antigenic signals in the presence of immunosuppressive cytokines, and often in special circumstances, such as after mucosal administration of antigen or in chronic viral infections [101]. In older adult humans, the induced regulatory T cells may play a greater role, compared with nTreg cells, in maintaining immune homeostasis because thymic output in humans is decreased following puberty and fewer natural nTreg cells subsequently emerge from the thymus [73].

iTreg (Foxp3+ CD4+ CD25+)

Foxp3+ CD4+ CD25+ regulatory T cells were originally considered to be a homogeneous population of naturally occurring, thymus-derived T cells (i.e. nTreg). However, similar Foxp3+ regulatory T cells ('iTreg') can be induced from CD25- precursors *in vivo*, and *ex vivo* with IL-2 and TGF-β [77]. Both populations express Foxp3 and suppress immune responses through contact-dependent mechanisms and the production of soluble factors, including the cytokines TGF-β, IL-10 and (in mice) IL-35 [78].

iTreg are induced in the periphery from a CD4+ CD25- Foxp3- (naïve) T-cell population following T-cell receptor (TCR) stimulation (by antigen) in the presence of cytokines TGF-β and IL-2 (see Figure 1.3). That is, iTreg differ in their antigen specificity (i.e. TCR repertoire) compared with nTreg (which are thought to be largely self-reactive, see Figure 1.5) and in their TCR signal strength and co-stimulatory requirements needed for their generation; however, whether they have any unique function *in vivo* has been unclear. It is possible that nTreg and iTreg have different roles in the adaptive immune response; for example, whereas nTreg (in mice) can be converted to Th17-type cells by IL-6 under some conditions, induced iTreg are resistant to this cytokine and thus might retain suppressive function at inflammatory sites [77]. Other reviewers take this concept of separate roles further and propose that nTreg and iTreg cells may synergise to achieve optimal regulation. For example, Curotto de Lafaille and Lafaille (2009) consider that there is a division of labour between nTreg and iTreg cells, based on the different TCR repertoires of both regulatory T-cell populations (Figure 1.5). That is, although iTreg seem tailored to respond to foreign antigens and neo-antigens (e.g. tumour antigens), it is likely that iTreg can also be generated in response to self-antigens and can then synergise with nTreg cells in the control of autoimmune inflammation [102].

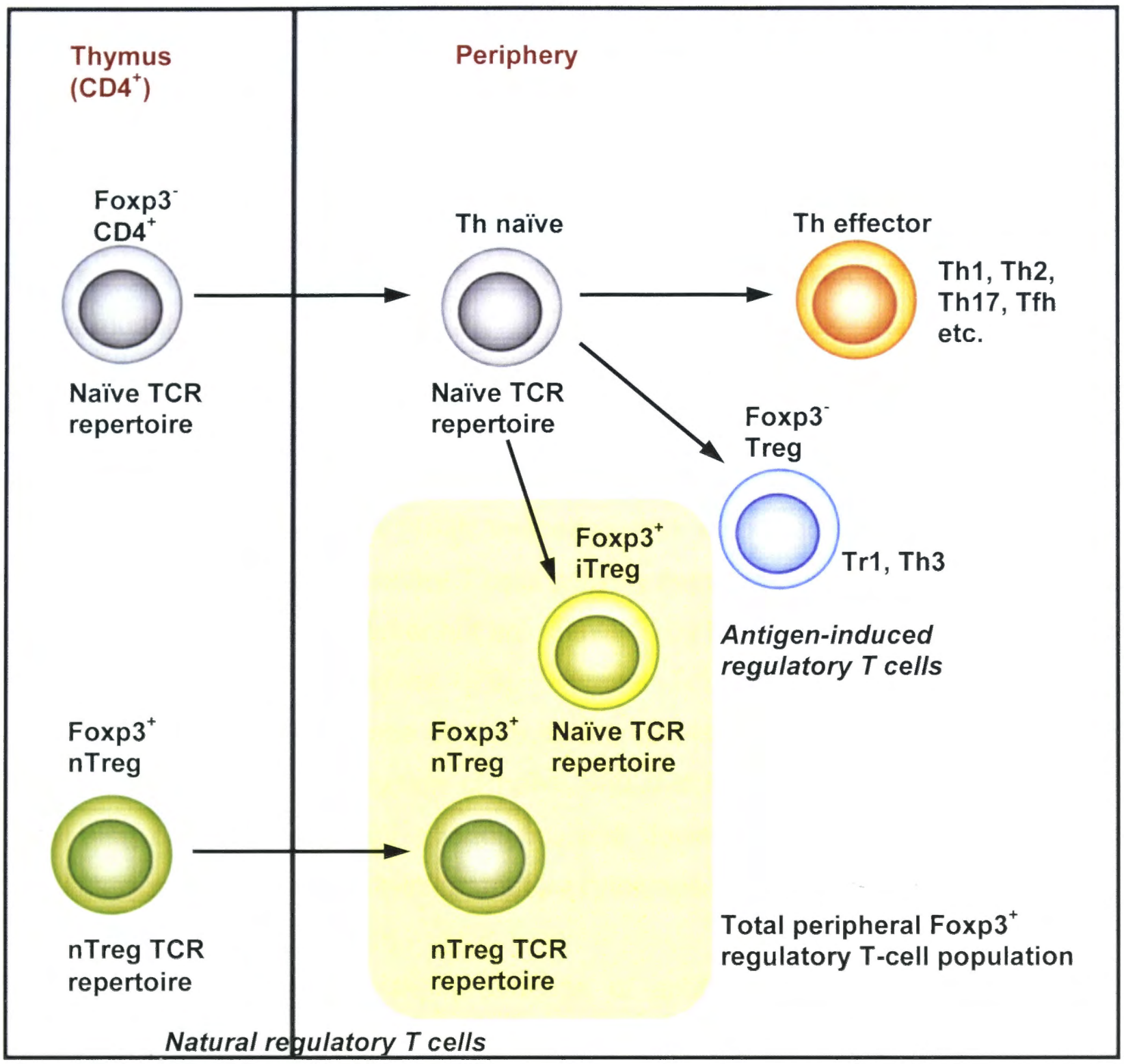


Figure 1.5: Thymic and peripheral generation of Foxp3+ regulatory T cells. Natural nTreg cells differentiate in the thymus and migrate to peripheral tissues. Induced Foxp3+ iTreg cells differentiate in SLO and tissues. The peripheral population of Foxp3+ regulatory T cells comprises both nTreg and iTreg cells. nTreg and Foxp3+ iTreg cells likely differ in their TCR repertoire, because the iTreg cell repertoire is drawn from naïve conventional CD4+ T cells that have survived negative selection (for self-antigens) in the thymus. Conversely, nTreg cells are selected (positively for self-antigens) by high-avidity interactions in the thymus (Treg, Foxp3– regulatory T cell). (Figure adapted from Curotto de Lafaille and Lafaille, 2009 [102].)

Other inducible regulatory T cells

Other inducible regulatory T cells, including Tr1 (T-regulatory type 1) and Th3 (T-helper type 3), appear to derive from the CD25- [48] and Foxp3- [102] T-cell subtype in the periphery (Figure 1.5); Tr1 and Th3 rely on cytokines such as IL-10 and TGF-β to effect suppression. Both these types of induced regulatory T cells and their cytokines have potent modulatory properties.

Type 1 regulatory cells (Tr1) have been more studied and appear to play an important role in self-tolerance induction; self-reactive Tr1 cells have been isolated from healthy human donors [103], indicating that auto-reactive T cells circulate in healthy people and that regulatory T cells actively suppress their function [64]. Tr1 cells do not have a unique cell marker but are identified by their high production of IL-10 and not pro-inflammatory cytokines [78]. Human Tr1 cells down-regulate immune responses through the release of both of the anti-inflammatory cytokines IL-10 and TGF-β, and regulate the function of naïve and memory Th1-type or Th2 responses. They also display a memory phenotype, and down-regulate the expression of co-stimulatory molecules and pro-inflammatory cytokines by APCs [64].

Further, Tr1 cells enhance the production of antibodies IgD, IgA and IgG by B lymphocytes. Importantly, Tr1 cells are inducible, antigen specific and need to be activated through their TCRs to exert their suppressive functions. However, once activated, they mediate suppression in an antigen non-specific manner [104]. For human Tr1-cell generation, IL-10 is necessary but probably not sufficient; other necessary factors include the presence of APCs, which provide, in addition to IL-10, other soluble and surface molecules crucial for Tr1-cell differentiation [64].

Human Tr1 cells appear to migrate to peripheral sites to directly regulate the inflammation of injured tissues and to maintain tolerance at mucosal surfaces. This suggests that the suppressive effect of regulatory Tr1 cells occurs locally and selectively in peripheral inflamed tissues [72]. Interestingly, and in support of this, an EAE experimental model has shown that Tr1 cells prevent encephalomyelitis only if the specific myelin antigen is administered in the brain and not systemically [105], suggesting that regulation of the pathogenic autoimmune response by Tr1 is localised in the inflamed environment. Indeed, in the human immune response to certain infectious agents, pathogen-specific Tr1 cells can be generated *in vivo* during the course of bacterial, viral, fungal or parasitic infections; the major purpose of these cells

here is to control inflammation and collateral tissue damage, thereby preventing chronic infectious states that may otherwise result if this mechanism is subverted by the pathogen [38].

In human autoimmune disorders, impaired Tr1 function in MS patients has been demonstrated, lower amounts of IL-10 being produced by these cells [106, 107]. Further, CD4+ T cells from MS patients were more resistant to suppression by IL-10 because the IL-10 signalling cascade was defective in MS patients [107].

T-helper 3 (Th3) cells are another antigen-specific regulatory cell population that originates in the periphery; Th3 cells mediate suppression through the secretion of TGF-β and, similar to Tr1 cells, they do not have a unique cell-surface marker [78]. Th3 cells were originally propagated from animals that became tolerant to orally administered protein antigen. They provide help for IgA synthesis and display suppressive properties for both Th1 and Th2 cells [108]. Th3 cells specific for myelin antigens have been shown to suppress autoimmunity in animal models of autoimmune encephalitis [109]. However, at least some Th3 cells appear to be TGF-β-induced Foxp3+ iTreg [37]. Relevant to this finding, Carrier and co-workers, using a TGF-β-transgenic mouse model, showed that TGF-β-dependent induction of antigen-specific Th3 cells in the periphery effected, in turn, the generation of antigen-specific Foxp3+ regulatory cells (i.e. iTreg) by producing required TGF-β. Th3 cells may thereby play a crucial role in inducing and maintaining peripheral tolerance, by driving the differentiation of antigen-specific Foxp3+ iTregs in the periphery [110].

Thus, both of these peripherally produced, induced regulatory T-cell types act in a down-regulating, anti-inflammatory way. Interestingly, some peripheral regulatory T cells exhibit antigen specificity while others appear to have lost such specificity. Initiation of regulation appears to be antigen-dependent, whereas, as regulation expands both quantitatively and qualitatively, antigen dependency appears to be lost and the production of more non-specific immunosuppressive cytokines, such as IL-10 and TGF-beta, appears to be the result [111].

CD8+ regulatory T cells of a variety of types, induced in association with MHC class I (self) antigens, have also been reported. One type appears to play a prominent role in protection from MBP-stimulated EAE in mice [38]; another type appears to resemble CD4+ Tr1 cells because their suppressive effects are primarily mediated by IL-10. A further type, characterised as CD28-, also expresses Foxp3 messenger RNA and can

render APCs tolerogenic [38]. Importantly, in humans, CD8+ T cells prevalent in MS lesions may have a regulatory function by direct effects on CD4+ cells [31, 55].

Immunomodulatory cytokines

The interrelations between nTreg, DCs, iTreg, Tr1 and Th3 cells, and the role of cytokines in the suppressive effects mediated by the different populations of regulatory T cells, are yet to be fully defined. Possibly, all types of regulatory T cells co-operate *in vivo* to prevent autoimmune reactions [48]. However, both TGF-β and IL-10 cytokines appear consistently as mediators of these generally immunosuppressive processes.

IL-10

IL-10 is a major immunomodulatory (down-regulating) cytokine that can inhibit the activation and effector function of many kinds of immune cells, playing a crucial role in preventing inflammatory and autoimmune disorders. This cytokine was originally described as a T-helper 2 (Th2)-type cytokine, and later associated with regulatory T cells such as Tr1 [104]. It is now known that IL-10 is much more broadly expressed by many cells of both the adaptive immune system (e.g. Th1, Th2, Th17, Tr1, CD8+ T cells and B cells) and the innate immune system (e.g. DCs, macrophages, mast cells, NK cells, eosinophils and neutrophils), affirming its role as a feedback regulator of diverse immune responses.

IL-10 exhibits broad anti-inflammatory properties by its suppression of both macrophage and DC function, including APC function and suppression of pro-inflammatory cytokine production. Thus, IL-10 has a primary role in limiting the immune and inflammatory responses to pathogens in order to prevent damage to the host [112]. Indeed, although IL-10-deficient mice can clear certain intracellular pathogens more efficiently than normal, this efficient dispensing of pathogens is often accompanied by immunopathology that can be lethal to the host [113].

IL-10 inhibits the development of Th1- and Th17-type responses by acting on DCs and macrophages, but also suppresses Th2 and allergic responses. As well, because IL-10 can be produced by Th1, Th2 and Th17 cells, an additional feedback loop effectively limits the innate effector functions of macrophages and DCs and their subsequent activation of T cells [114]. However, IL-10 also up-regulates the differentiation of IL-10-

secreting regulatory T cells, forming a positive regulatory loop for its induction [104]. This cytokine can also activate mast cells and enhance the functions of CD8+ T cells, NK cells and B cells in some situations; thus, IL-10 has important effects on the development of an immune response, and a central role in the induction of tolerance and maintenance of immunological homeostasis in humans and other mammals.

The key *in vivo* role of IL-10 has been demonstrated in several experimental animal models. For example, IL-10 'knock-out' mice (that is, a genetic strain in which IL-10 production is down-regulated) develop EAE induced by a myelin protein (myelin oligodendrocyte glycoprotein [MOG]) [115] and spontaneously develop inflammatory bowel disease [112, 116]. These genetically modified mice are also more susceptible to RA [117]. IL-10 has also been administered prophylactically to various animal models and mostly shown to be effective in the prevention of inflammatory and organ-specific autoimmune disorders such as EAE, diabetes mellitus and arthritis (reviewed by Asadullah et al., 2003 [118]).

However, some data are conflicting, and the therapeutic action of IL-10 may differ depending on the local micro-environment, the disease stage and the IL-10 concentration. Nevertheless, clinical trials in humans are in progress using recombinant IL-10 in immune-mediated inflammatory diseases such as psoriasis, Crohn's disease and RA, with varying results (reviewed by O'Garra et al., 2008 [112]).

TGF-β

TGF-β has been found to be essential in the adaptive (peripheral) immune system, affecting most immune-cell types as shown by the broad distribution of the relevant receptor. TGF-β initiates 'signalling' events in target cells that affect cell-fate decisions, proliferation and survival by modifying cellular transcriptional programmes. Using more recent genetic targeting approaches, TGF-β signalling has been implicated in the inhibition of cytolytic and Th1-cell differentiation, proliferation and apoptosis in particular. This direct role was confirmed when aggressive autoimmune lesions resulted after abrogation of TGF-β signalling in mice [119]. A further direct role for TGF-β in the differentiation of pro-inflammatory IL-17-producing Th17 cells has also been recently uncovered: TGF-β in the presence of IL-6 determines the generation of this distinct cell lineage rather than iTreg (see Figure 1.3).

The essential role of TGF-β (together with IL-2) in Foxp3+ iTreg development has already been noted; in this role, TGF-β also requires CTLA-4 in order to induce activated CD4+ cells to express Foxp3. In addition, TGF-β is required to maintain Foxp3 expression in both iTreg and nTreg. It has also been reported that Foxp3+ regulatory T cells can secrete active TGF-β and express this cytokine on their cell surface; membrane-bound TGF-β seems to have stronger regulatory effects on other immune cells than soluble TGF-β and may thus have an important role in the functional properties of both activated nTreg and iTreg [77]. Therefore, TGF-β has both a direct role—in determining T effector cell functions—and an indirect role—in inducing and maintaining peripheral Foxp3+ regulatory cells—in the overall maintenance of immune homeostasis [96, 119].

TGF-β derived from Foxp3+ regulatory cells may further be involved in inducing immature DCs to become tolerogenic, since DCs are principal targets of Foxp3+ regulatory cells and, as discussed earlier, are prevented from maturing into immunogenic DC by them. Further, tolerogenic DCs can produce latent TGF-β and convert this precursor molecule to biologically active TGF-β [77]. TGF-β is also produced by both of the peripherally antigen-induced Th3 and Tr1 cells (particularly Th3), and has generally down-regulatory effects.

In humans, T-cell cultures derived from patients with active MS produced less TGF-β compared to T cells from patients with stable disease [120]. MBP-specific T-cell clones from MS patients also produced less TGF-β compared with MBP-specific T-cell clones from healthy individuals, when the frequencies of MBP-specific T cells were similar in patients and non-patients [121]. Thus, TGF-β levels appear to be associated with appearance of symptoms in MS [122].

IL-2

IL-2 is a major cytokine thought to control regulatory T-cell homeostasis, and particularly iTreg generation and homeostasis. Indeed, Foxp3+ regulatory T cells were initially identified based on their constitutive expression of the high-affinity IL-2 receptor component CD25, and IL-2 is required for TGF-β induction of Foxp3 transcription and suppressor activity [102]. IL-2 seems to have an essential and non-redundant role in controlling nTreg and iTreg function in the periphery, as shown by the development of lymphoproliferative disease and colitis in mice that are deficient in either IL-2 or CD25.

However, the precise way in which IL-2 influences regulatory T-cell function is unknown [73].

In summary, both *central* and *peripheral* mechanisms for maintaining tolerance are required for immunologic homeostasis. Now that the importance of the centrally-derived Foxp3+ (CD4+ CD25+) regulatory T cells (nTreg) in several autoimmune diseases in humans has been recognised [79, 89, 95], the thymic central tolerance system can be thought of as producing not one, but two, initial repertoires of tolerogenic cells (see Figure 1.5) that appear to balance each other. One repertoire is formed by deleting potentially aggressive T cells with high specificity for self-antigens—the original clonal deletion model known since the days of Ehrlich and Burnet [19]; the other repertoire is formed by positive selection of Foxp3+ nTreg and NKT cells that do recognise self-antigens, with the express purpose of suppressing T-cell-mediated interactions and maintaining tolerance. Added to this are now the peripheral tolerance systems, including the induced iTreg, Tr1 and Th3 regulatory T cells, and the specialised antigen-presenting DCs that are responsible for much of the 'dialogue' in the system. Therefore, rather than a simple decision of T-cell clonal deletion or survival, as envisaged several decades ago, subsequent research has shown that individual T (and B) lymphocytes must pass a remarkably complex series of self-reactivity checkpoints before they form part of the combined immune repertoire and potentially become large clones of either effector or regulatory immune cells.

So we may ask now two questions: (1) how is such heavily-regulated immune tolerance temporarily overcome in order to fight potentially pathogenic invaders? and (2) how is self-tolerance more permanently 'broken', resulting in overt autoimmune disease?

1.3.2 Overcoming tolerance to respond to infection

In order to be able to mount an effective response to infection in the face of such immune regulation, both the adaptive and innate immune responses work together to overcome the suppressive properties of regulatory T cells. That is, cells of the innate immune system first produce cytokines in response to pathogen recognition and these initiate the inflammatory response, which is then subsequently amplified by products of the adaptive immune response. For example, microbial induction of toll-like receptors

(TLRs) [7] on DCs via microbial LPS can overcome regulatory T-cell-mediated suppression and allow the generation of an antigen-specific adaptive immune response [123, 124]. The suppressive activities of regulatory T cells may also be overcome by enhanced signalling, especially via co-stimulatory CD28, as might arise following the innate response of DCs to infection [125]. In general, high levels of cytokines such as IL-2, IL-4 or IL-15, or IL-6, may cause proliferation of both natural and induced regulatory T cells and temporarily overcome their cell-to-cell or cytokine-mediated suppressive abilities, until the concentration of growth factor falls again and the regulatory T cells stop proliferating. It seems that when the immune system has eradicated the infection, the regulatory T cells then recover their suppressive function [65] as pictured schematically in Figure 1.6.

It has also been recently observed that Foxp3+ regulatory T cells may not initially be activated or induced in highly inflammatory environments in acute infection; that is, Th1-type or Th2-polarising cytokines can interfere with the induction of Foxp3+ regulatory T cells in mice [126]. In addition, cytokines such as IL-6, IL-1 and IL-12 can down-regulate Foxp3 expression and even convert either iTreg or nTreg to conventional pro-inflammatory effector T cells *in vivo* [73].

Thus, it seems that the steady-state regulatory process can be temporarily subverted to allow induction of an antigen-specific effector response to infection; then, mechanisms appear to be in place in order to prevent ongoing induction of this response and to limit pathogenic anti-self immune responses that might occur at inflammatory sites. For example, both TNF-α and IFN-γ, considered archetypal pro-inflammatory cytokines, are now known to trigger regulatory processes, such as production of anti-inflammatory IL-10, at the same time as they carry out their primary effector role. This means that they can trigger key immunoregulatory pathways 'downstream', for control of an established immune response, to prevent ongoing inflammation [27]. Anti-inflammatory TGF-β production is also often associated with the downstream effects of inflammatory responses [126], while other pro-inflammatory cytokines, such as IL-6 and IL-4, can also have both positive and negative effects on regulatory T-cell activity [73].

---

[7] TLRs on innate immune cells are highly evolutionarily conserved receptors that recognize pathogen-associated molecular patterns. Once recognised, the innate cell can process the pathogenic antigen and present as peptide to antigen-specific (i.e. antigen-recognising) T cells for further action (i.e. adaptive response) [24, 31].

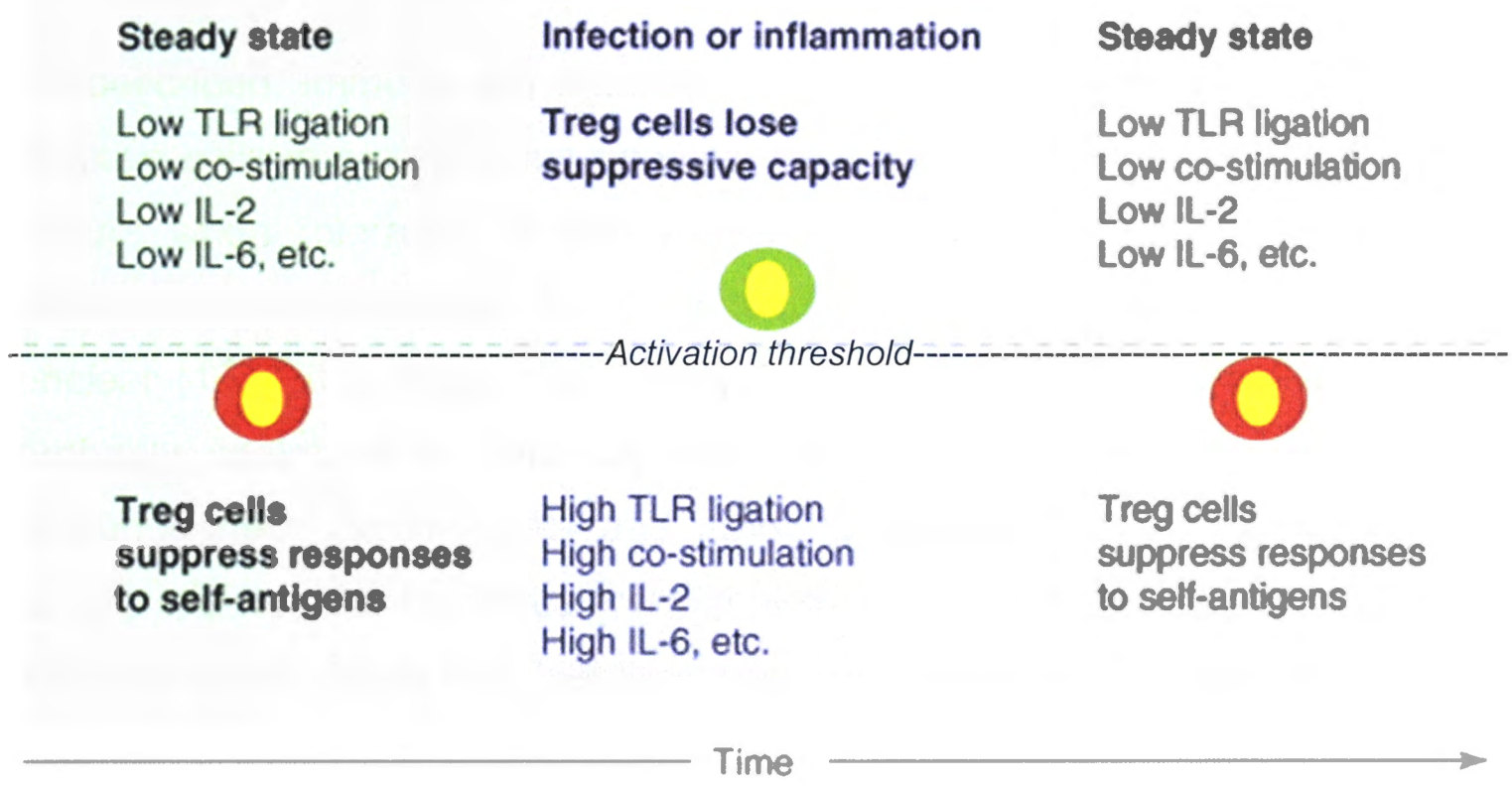


Figure 1.6: Schematic model of response to infection, based on stability of the suppressor phenotype. Regulatory T cells suppress the response of auto-reactive (red) cells to self-antigens in the steady state by ensuring that they do not reach the threshold for activation (hatched line). Their ability to suppress responses can be overcome in the face of infection, as a result of TLR ligation or increasing levels of co-stimulation and cytokines secreted in response to infection. This might permit localised activation of some auto-reactive (green) cells through bystander activation or molecular mimicry (discussed in Section 1.4.2). However, their expansion will be prevented by clonal competition from cells specific for antigens carried by the infectious agent. Available evidence suggests that regulatory T cells retain their suppressive capacity and will thus continue to suppress the response to self-antigens when the infection has been cleared. (Figure adapted from Wraith et al., 2004 [65].)

What, then, might go wrong in the immunoregulatory system, where self-tolerance should be maintained or at least reverted to after temporary infection, to initiate full-scale, persistent autoimmune disease?

1.4 Breakdown of tolerance

As described, immune self-tolerance is a process that eliminates or neutralises self-reactive cells in a continuous ongoing way. The initiation of autoimmune disease then occurs when tolerance to self-antigens is broken—a phenomenon that has long fascinated immunologists—yet the actual initiating immunogens for this process are still unclear [127]. It appears that both intrinsic factors and external precipitating events probably contribute to eventual autoimmune disease. The intrinsic factors involve immunogenetic determinants and initial repertoire formation, and thus are involved largely with central tolerance mechanisms set up in early life (see Section 1.2.1); as discussed next, these may 'set the stage' for susceptibility to autoimmune disease. The eventual precipitating events (introduced subsequently) then appear to relate more to peripheral tolerance mechanisms (see Section 1.2.2), and these events, acting alone or in synergy, may finally lead to clinically established autoimmune disease.

The primary genetic determinants of autoimmune disease are not reviewed here but are discussed briefly in relation to specific environmental factors detailed in Chapter 2.

1.4.1 Immunogenetic factors and faulty tolerance induction

At a first level, the MHC (genetic self) haplotype (e.g. human leucocyte antigen [HLA] subtypes) can influence susceptibility to autoimmune disease. For example, this can be achieved by enhancing the presentation of antigens in the periphery, resulting in increased T-cell activation, or by ineffective presentation of self-antigens in the thymus, leading to more aggressive effector T cells or fewer numbers of regulatory T cells [127].

At a second level, various polymorphisms in genes involved in establishing self-tolerance and immune regulation can occur. For example, expression of the human 'autoimmune regulator' (AIRE) is vital to thymic negative selection and thus to the prevention of organ-specific autoimmune disease [128, 129]; humans deficient in the *AIRE* gene develop a severe multi-organ autoimmune disease characterised by variable combinations of autoimmune endocrine diseases, including type 1 diabetes. AIRE is primarily expressed by epithelial cells in the thymic medulla and induces the transcription of a large number of tissue-specific self-antigens that would otherwise be sequestered in the periphery. Thus, AIRE ensures developing thymocytes are exposed to a comprehensive view of 'self', permitting early deletion of auto-reactive T cells

before acquisition of effector functions (see Section 1.2.1) [128, 130]. It is therefore apparent that in the absence of AIRE expression, the establishment of central self-tolerance is incomplete, resulting in an auto-aggressive reaction towards several self-antigens [127].

Another example of aberrant tolerance induction is seen in the murine EAE model of human MS, in which a genetic variant of the self-protein, proteolipid protein (PLP), expressed in thymic epithelial cells, results in the particular auto-reactive T-cell repertoire responsible for the EAE disorder [131]. This antigen-lacking variant may prevent proper presentation of the self-protein to T cells in the thymus, leading to poor self-tolerance induction and autoimmune disease susceptibility.

Tolerance may also be compromised and autoimmune disease may develop as a consequence of altered balance between regulatory T cells and self-reactive conventional T cells. That is, any genetic anomalies that tip the balance toward self-reactive conventional T cells could cause or predispose to autoimmune disease. As discussed, mutations in the *Foxp3* gene result in IPEX syndrome, which is an nTreg-specific immunodeficiency in humans (see Section 1.3.1). In addition, polymorphisms of several genes controlled by *Foxp3* in nTreg cells, such as those encoding CTLA-4, IL-2 and CD25, are associated with susceptibility to common autoimmune diseases (including type 1 diabetes) in humans and rodents. For example, *Il2* and *Ctla4* are main susceptibility genes for type 1 diabetes in the diabetes-prone 'non-obese diabetic' (NOD) mouse model, and both IL-2 and CTLA-4 molecules seem to be important in maintaining the balance between Foxp3+ T-regulatory cells and self-reactive T cells [39]. In humans, a single nucleotide difference in *Ctla4* alleles has been identified as being differentially expressed between people with autoimmune Grave's disease and those without, and type 1 diabetes may also be linked to the same locus of this *Ctla4* gene [132], further highlighting the importance of this CTLA-4 'key negative immune regulator' in autoimmune disease susceptibility generally.

Other genetic polymorphisms may occur, for example, in apoptosis genes, which may lead to excess in effector T cells over regulatory T cells [27]. The importance of normal apoptotic processes for eradicating self-antigen activated lymphocytes in humans is illustrated by a finding that in type 1 diabetes patients, auto-reactive T cells are highly anergic and resistant to apoptosis, requiring 20 to 100 times lower levels of IL-2 than non-diabetic T cells to escape apoptosis [133]. Genes controlling the production of

cytokines such as TNF-α or IL-1 or IL-10 may also show polymorphisms; these too have been associated with susceptibility to autoimmune diseases [27, 134].

1.4.2 External events—autoimmune disease determinants

Even in a genetically predisposed person, some precipitating event—an environmental exposure or a change in the internal environment—is likely to be required for frank auto-reactivity. However, in the case of most autoimmune diseases the actual factor is unknown [2]. External factors may act at an early stage in the life course and interact with both genetic and immunological mechanisms to set the initial conditions for later precipitation of autoimmune disease. This possibility, and some of the factors potentially involved, will be considered in more detail in Chapters 2 and 3. In the present section, the external factors that may be involved in initial activation of self-antigen-specific T cells will be briefly considered, this being the critical early event in the induction of autoimmune disease [135].

Microbial agents

Microbial antigens have the potential to initiate auto-reactivity through a number of different possible mechanisms. Infections are prime candidates for initiating autoimmune disease in predisposed persons, because they frequently induce strong inflammatory responses in various organs and can attract many potentially auto-aggressive lymphocytes to the site of infection.

Microbial infections can cause massive host inflammatory responses producing cytokines that polyclonally activate 'bystander' T cells at the site of infection in a non-antigen-specific way. In this case, the sum of all non-specifically activated cells, rather than a few specific cells, causes collateral damage [127]. For example, the enhancement of type 1 diabetes in a strain of transgenic NOD mice by a Coxsackie B virus infection (this virus having been associated many times with human autoimmune diabetes [136]) has been shown to be a direct result of local infection leading to inflammation, tissue damage and the release of sequestered islet antigens, in turn resulting in the stimulation or re-stimulation of resting bystander auto-reactive T cells [137].

There are several other examples of microbes triggering autoimmune disease in susceptible humans or animals by similar adjuvant effects (reviewed by Verhasselt and Goldman, 2001 [23] and discussed in more detail in Chapter 3). For example, work by Shevach and colleagues on EAE showed that pre-existing auto-reactive T cells stay in a quiescent state unless microbial products such as LPS are present; the presence of these microbial factors allows the development of effector Th1 cells and the induction of EAE through the stimulation of IL-12 production. DCs have been shown to produce IL-12 in response to various microbial products and to be the major source of this cytokine following infection. LPS might also act by interfering with the induction of peripheral T-cell death [23].

Another study showed that self-tolerance could be broken and spontaneous EAE produced by activation of TLRs on APCs, demonstrating response by the *innate* immune system to microbial products, even in the genetically resistant transgenic murine strain used [138]. Similarly, induction of EAE by (microbial) pertussis toxin was shown to be dependent on innate TLR4 and involved recruitment of leucocytes and T cells to the cerebrovascular endothelium; this suggests that infectious agents may have a role in the interaction between T cells and the blood-brain barrier (BBB)[8] in central nervous system (CNS) disorders such as MS, by up-regulating adhesion molecules important for permeability of the CNS to immune cells [140].

Another pathway by which microbial agents can break self-tolerance may be via 'antigenic (or 'molecular') mimicry' (discussed in more detail in Chapter 3)—this is the hypothesised 'mistaken identity' situation in which auto-reactive cells in the host are activated by microbes presenting antigens that are cross-reactive with (host) 'self'. In this process, the anti-foreign immune response causes collateral autoimmune damage by mistake [22]. In MS patients, for example, T cells react with a peptide from the auto-antigen MBP, but also cross-react with peptides from Epstein-Barr virus (EBV), influenzavirus type A and human papillomavirus [141]. In type 1 diabetes, human T cells recognise both an auto-antigen peptide and an analogous peptide from a Coxsackie B virus protein [142]. RA also appears to have an infectious aetiological association, with *Proteus* bacterial species [143] or with other infectious agents such as viruses.

---

[8] The BBB is an important physical and physiological 'dividing line' between the immune system and the CNS and is also the locale, and conduit, for interactions between these two systems. It is a dynamic system that maintains brain homeostasis and limits CNS penetration via interactions of transmembrane and intracellular proteins. Thus, the BBB restricts unregulated mixing of immune substances in the blood with those in the CNS, directly transports neuroimmune-active substances between the blood and CNS and itself secretes neuroimmune substances [139].

Although microbial agents have been associated with several autoimmune diseases, including MS, type 1 diabetes and RA (discussed further in Chapter 3), attempts to establish a direct epidemiological association between specific microbial infections and various autoimmune disorders have been difficult. Part of the problem in establishing this link is that most humans encounter a multitude of infections during their lifetime and most infections are cleared by the time of autoimmune disease diagnosis. Thus, precise 'footprints' documenting an individual's history of viral and bacterial infections are difficult to find.

In addition, several sequential events may be necessary to precipitate autoimmune disease, each event accelerating an already pre-existing autoimmune condition rather than being an initiator [127]. Thus, microbial events are good candidates as part of a suite of factors for initiating autoimmune disease, given recent information that such events can initiate or strengthen allergic inflammation. For example, pathogen-derived products induce the expression of various members of the IL-1-cytokine family which have been shown to enhance and alter Th17, Th1 and Th2 responses [144].

Finally, the observation that some infections, particularly those in early life, might protect an individual from autoimmune disease rather than enhance it—that is, the 'hygiene hypothesis'—appears paradoxical to an infectious aetiology and further complicates the issue of whether infectious agents can be initiators and/or accelerators of autoimmune disease. For example, multiple infections during the first year of life are associated with a significant reduction in the risk of type 1 diabetes [145]. This means that microbes may be capable not only of triggering autoimmune disease but also may provide protection against such disease under different conditions. This hypothesis will be explored further in Chapters 3 and 8.

Other determinants

As discussed, to break self-tolerance microbial infections could combine a non-specific bystander activation component provided by local inflammation, and a cross-reacting antigen-specific element that directs the activated immune system to specifically target particular host proteins. However, such a 'perfect' combination can also be achieved by exposure to drugs or other chemical agents that modify self-components, resulting in the formation of 'neo-antigens' to which no previous tolerance could have been established by the host [127]. For example, metabolism of drugs such as procainamide

can cause antigenic modification of self-components, alteration of the immune repertoire and induction of the autoimmune 'lupus' disorder SLE. Some heavy metals such as mercury can also be immunotoxic in susceptible individuals.

Other external factors that have been linked to CNS disorders such as MS, perhaps via perturbation of the BBB, include prior head injury and psychological stress (reviewed by Goodin et al., 1999 [146]). In addition, several autoimmune disorders, including MS and RA, are more common in women than in men [147], indicating the importance of the internal environment, including hormones. Foreign substances, such as penicillins, or gliadin (from wheat gluten) may also act as haptens and render auto-antigens immunogenic [2]. Other possible environmental factors, including those related to the physical environment (e.g. climate), may also help determine autoimmune disease. For example, low exposure to sunlight and UVR have been implicated by epidemiological evidence; these will be considered in detail in Chapter 2.

Thus, while internal and external determinants of MS and other organ-specific autoimmune diseases have been briefly introduced here, external factors, in particular, will be reviewed fully in chapters 2 and 3. In addition, the epidemiological parameters of MS, particularly those concerning incidence and prevalence and pertaining to the distribution of disease frequency—that is, variation in 'person, place and time'—will be discussed in detail in Chapter 2, together with those of type 1 diabetes and RA where appropriate.

In conclusion, the breakdown of tolerance and the consequent initiation and final precipitation of autoimmune disease appears to be a multifactorial and complex process, requiring both immunogenetic components and a probable sequence of environmental events. During normal autoimmune responses, a delicate balance between activation and elimination of self-reactive lymphocytes is required to avoid tissue damage, and access of lymphocytes to intact tissues must also be limited [23]. In the words of noted Australian immunologist Sir Gustav Nossal [22]:

> ***'We are all constantly teetering on the brink of autoimmune disease'***
>
> ***'Our lymphocytes are not truly resting; rather they are integrating a wealth of positive and negative signals that arise from their antigen receptors and from many other signal-transducing molecules ... [and]***

minor defects or insults can perturb the equilibrium, with disastrous results'

On the one hand, this 'immunological self' complexity and precise mechanism of self-recognition is to our advantage because it is not actually easy to break tolerance to auto-antigens and override the normal regulatory balances [127]. Conversely, however, the complexity makes the task of unravelling the aetiology of autoimmune disorders, such as MS, all the more difficult, and so the development of more specific therapeutic approaches for such autoimmune disease continues to be an ongoing process.

In this setting of breakdown of immunological tolerance and consequent development of autoimmune disease, the clinical and immuno-pathogenic aspects of the specific disorder of MS, in particular, can now be considered.

1.5 MS (*sclérose en plaques disseminêes*)

1.5.1 General features

MS is the most common chronic disabling disease of the CNS in young adults in Western countries, with 1 in 1000 people affected [148]. Noseworthy and colleagues describe the disease as an 'enigmatic, relapsing, and often eventually progressive disorder ... [that] continues to challenge investigators trying to understand the pathogenesis of the disease and prevent its progression' [149, 150].

MS was described more than a century ago by a French neurologist now honoured as the 'founder of modern neurology', Jean Martin Charcot, while noting the accumulation of inflammatory cells within the brain and spinal cord white matter of patients with intermittent episodes of neurologic dysfunction [135]. This led to the term *sclérose en plaques disseminêes* (patches of scarring), or MS [151]. This name conveys the MS pathological hallmark of the demyelinated 'plaque' (i.e. hardened sclerotic lesion), leading to neurologic dysfunction due to interference in nerve transmission, which then becomes 'disseminated in space and time' at multiple sites within the brain and spinal cord of the CNS [150]. Kabat and colleagues further noted immunoglobulins (antibodies) in the cerebrospinal fluid (CSF) of MS patients in 1948, indicative of an inflammatory disease [152]

1.5.2 Natural history and clinical features

As a chronic disease typically beginning in the second or third decade of life, MS can take different forms. The most frequent form is RRMS, characterised by bouts of disease followed by complete or incomplete remissions; this form has a female predominance of approximately 2:1 [150]. After several years, this relapsing disease can transform into secondary progressive MS (SPMS), defined by a slow clinical deterioration. Up to twenty per cent of MS patients have a primary progressive MS (PPMS) at onset, characterised by gradual progression from onset with no remissions; this form has similar incidence in men and women [150, 301].

MS is a chronic inflammatory disease of the CNS, associated with the formation of large confluent plaques of demyelination in the brain and spinal cord [153]. Nerve axons are also injured and destroyed in the demyelinated areas and the loss of these axons appears to result in permanent neurological deficit [154]. In addition to focal demyelinating plaques, there is diffuse damage in the 'normal' white matter and the cortex; this diffuse, global brain injury is most evident in patients with primary progressive or secondary progressive forms of MS. Functionally, the clinical symptoms of MS reflect disturbances in the electrical signal transmission in the CNS due to damage to the myelin sheath and associated axonal and neuronal grey matter injury [155].

1.5.3 Diagnosis of MS

MS is diagnosed by a careful clinical process to demonstrate findings consistent with MS and to rule out other causes with similar symptoms. The basic criteria for diagnosing MS currently, known as the 'McDonald criteria', are based on recommendations by an international panel convened in 2001 [156], which were revised in 2005 [157]. As in previously used criteria [158, 159], these recommendations focus on the objective demonstration of dissemination of lesions in both time and space, and now include magnetic resonance imaging (MRI) integrated with clinical and other paraclinical diagnostic methods, such as analysis of CSF (for evidence of immune inflammation) and visual evoked potentials (electrical nerve-transmission tests). The clinical examination would include history and tests of function. An extensive review of mental, emotional and language functions, movement and coordination, vision, balance and the functions of the five senses would be covered. Birthplace, family history, age at

first symptoms and sex would also be taken into consideration [156] ('MS Australia' website www.msaustralia.org.au, accessed March 2011).

1.5.4 Pathology and immunopathology

The composition of the inflammatory infiltrate in acute MS lesions, together with the local expression of various cytokines and other immune-associated molecules, suggests that the inflammatory response is based on a T-cell-mediated immunological process [160]. Thus, MS is generally agreed to be essentially an autoimmune disease, at least in the acute phase of RRMS [161, 162].

In the early 1930s, an autoimmune, and at times demyelinating, disease was demonstrated in mammals by repeatedly injecting rabbit brain and spinal cord into primates. This disease, known as experimental 'allergic' (later, 'autoimmune') encephalomyelitis, EAE, and now studied mainly in mice, led to the mostly accepted hypothesis that MS was similarly secondary to an autoimmune response to the (self-) proteins of the myelin nerve sheaths in a genetically susceptible host [135].

For the induction of EAE, injecting a rat or mouse with myelin protein in an adjuvant leads to the development of an inflammatory CNS process manifested by paralysis, after which most of these animals recover. Re-injection of myelin in the recovered animals produces no disease, because they have developed immunological memory and have acquired tolerance to future exposures to the myelin antigens. However, injection of cells isolated from sick animals into naïve animals elicits EAE (a phenomenon known as adoptive, or passive, transfer of the disease). This suggests that EAE, and presumably MS, are largely cell-mediated diseases (rather than antibody mediated) [155], although auto-antibodies are clearly also involved in human MS. It is likely that the human MS disease also exhibits immunological mechanisms additional to those mediated by CD4+ Th1 and Th17 cells in EAE; these include CD8+ T cells found in MS lesions, B cells that can present antigen and interact with T cells and produce both pro-inflammatory and regulatory cytokines [31], as well as a neurodegenerative component in the target neural tissue, particularly in later disease stages [154, 160, 161, 163]. Nevertheless, EAE disease models have been a crucial tool for attempting to understand this devastating human disorder (reviewed by Gold et al., 2006 [163]).

Such heterogeneity in human MS is supported by evidence of MS lesions varying in their patterns of demyelination, even though all share an inflammatory reaction dominated by T cells and macrophages [164]. Further, these patterns have been found to be heterogeneous among patients but homogeneous within active plaques from the same patient [160, 165]. In some lesions, the presence of immunoglobulins (e.g. IgG) and products of complement activation indicate that demyelinating antibodies also have a pathogenic role. In others, oligodendrocytes (the primary myelinating cells of the CNS) appear to be the main target for destruction. Thus, it has been suggested by a number of researchers that MS may be a series of syndromes with different pathogenic mechanisms and causes, characterised by: (1) T-cell-mediated immune injury, or T-cell plus complement- and antibody-mediated injury (both these types being characteristic of EAE animal models) and/or (2) primary oligodendrocyte dystrophy (this type less seen in EAE-type models and reminiscent of virus- or toxin-induced demyelination rather than autoimmunity) [135, 162, 165-167].

Therefore, MS may be a much more complex and heterogeneous disease than was previously thought, and the pathogenetic mechanisms and targets of demyelination may be fundamentally different in distinct subgroups or stages of the disease. A now widely accepted, but relatively recent, concept is that MS is initiated by an autoimmune inflammatory process, but later develops a neurodegenerative component, which might progress independently from inflammation [161, 162].

Possible immuno-pathogenesis and progression of MS

Based on recent findings in humans [31, 55, 135, 150, 155, 160-162, 167-169], the possible immuno-pathogenesis and progression of MS can be described as follows (refer to Figures 1.7 and 1.8):

Breakdown in self-tolerance and lack of regulation

Underlying immunoregulatory defects, such as functional decrease in nTreg cells in the circulation, may lead to disturbance in physiological mechanisms controlling T-cell homeostasis (i.e. faulty tolerance) and allow pathologic activation of auto-reactive T cells in the peripheral blood.

Activation of myelin-reactive T cells

Antigen-presenting DCs are activated, possibly by microbial agents or other metabolites, leading to the consequent activation of myelin- or other 'MS-antigen'-specific, auto-reactive T lymphocytes (Figure 1.7). That is, CD4+ T cells (Th1 and/or Th17) are reactive to one or more of several putative self-myelin and non-myelin 'MS antigens', including antigens of astrocytic and neuronal origin. Activation also triggers secretion of cytokines and chemokines that modify the local micro-environment.

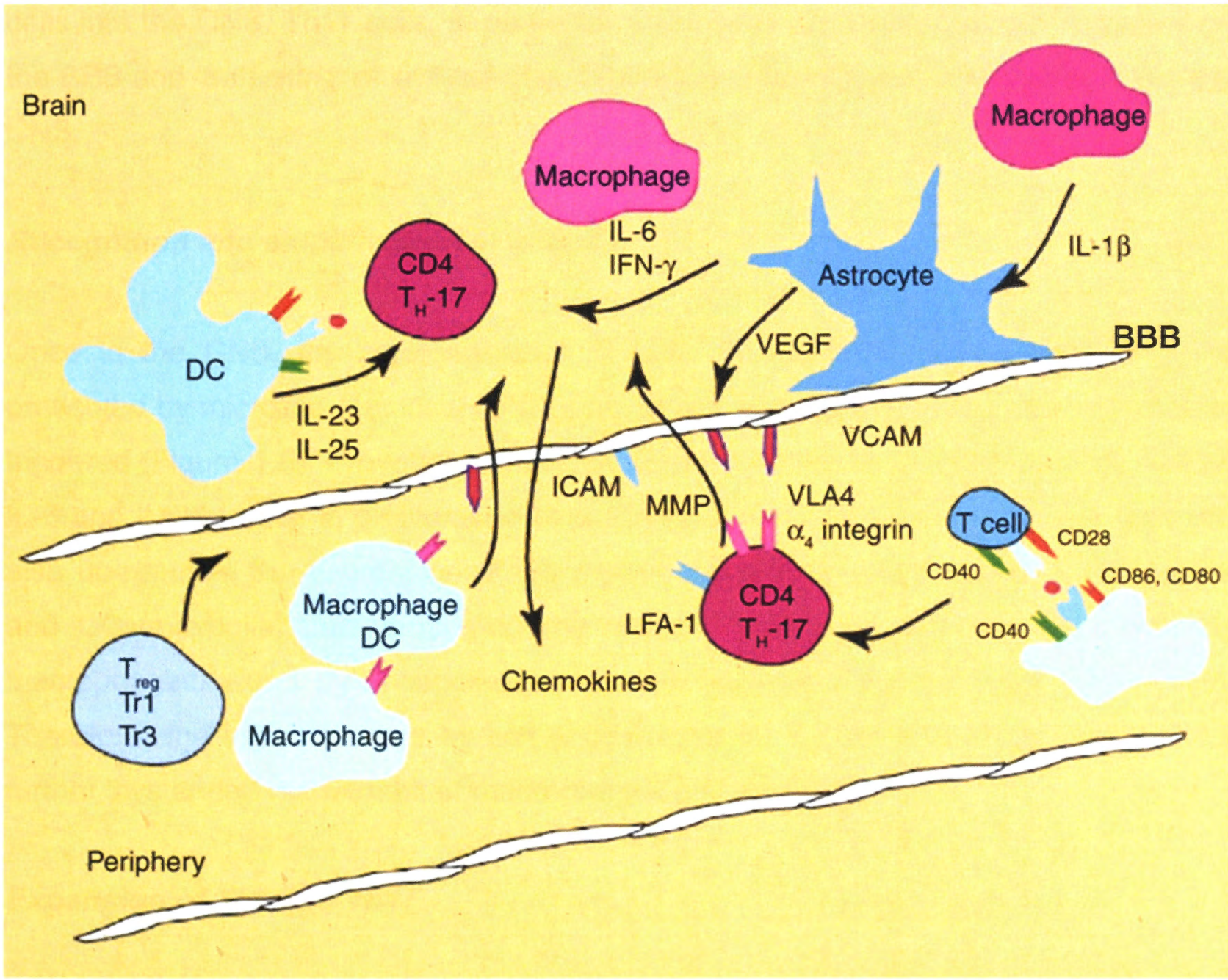


Figure 1.7: Immunological processes during early steps in the development of MS lesions. These processes include opening of the BBB, cell adhesion to cerebrovascular endothelial cells and local reactivation of infiltrating T_H1 or T_H17 cells (T_H17 [CD4] presented here). (Adhesion molecules: ICAM, intercellular adhesion molecule; LFA-1, lymphocyte function-associated antigen 1 [integrin $\alpha_1\beta_2$]; VCAM, vascular cell-adhesion molecule; VLA4, very late antigen 4 [integrin $\alpha_4\beta_1$]. MMP, matrix metalloproteinase; VEGF, vascular endothelial growth factor). (Figure adapted from McFarland and Martin, 2007 [162].)

Disruption of BBB and entry into CNS

Once activated, the CD4+ T cells proliferate, circulate in the peripheral blood and express a variety of cell-surface adhesion molecules, such as VLA4, which links with its ligand, vascular cell-adhesion molecule (VCAM) (Figure 1.7). The activated T cells adhere to the luminal surface of endothelial cells in CNS venules and migrate into the CNS through disruption of the BBB, aided by enzymes such as matrix metalloproteinase (MMP) (Figure 1.7). In the CNS, local factors may up-regulate the expression of various endothelial adhesion molecules, further facilitating the entry of T cells into the CNS. Th17 cells, in particular, have been shown to promote disruption of the BBB and 'trafficking' of self-reactive T cells from the systemic compartment into the CNS.

Recognition and amplification of antigen

Once in the CNS, the myelin-reactive T cells recognise the various MS antigens presented by microglia (i.e. local APCs), are re-activated and an inflammatory cascade triggered (Figure 1.8). Pro-inflammatory cytokines, such as IL-12 and IFN-γ, as well as IL-6 and IL-23, result in proliferation of pro-inflammatory Th1 and Th17 cells, and can also up-regulate the expression of the myelin antigens on neighbouring lymphocytes and DCs/microglia, thus amplifying the immune response. Ingestion of myelin by macrophages, aided by antibodies, can further activate CNS microglia (Figure 1.8). Therefore, the initial damage by self-antigen-specific T cells is a strong stimulus for further 'bystander' recruitment of macrophages and microglia.

Expansion of Th1 and Th17

Pro-inflammatory Th1 and Th17 cells (secreting IFN-γ and IL-2, and IL-17, respectively) and CD8+ cytotoxic T cells proliferate and ultimately result in immune-mediated injury to myelin, oligodendrocytes and axons. The pro-inflammatory cytokine, osteopontin, expressed at high levels within lesions, promotes the survival of activated T cells. Cytokines such as IL-6 may inhibit the suppressive action of regulatory T cells (Figure 1.8).

Auto-antibodies

Anti-myelin or anti-oligodendrocyte antibodies may also gain access to the CNS through the disruption of the BBB as a consequence of the T-cell initiated inflammatory response (Figure 1.8). These antibodies can cause demyelination directly, possibly through the activation of complement, or assist other cell types such as macrophages.

Myelin injury

Activated macrophages and microglia further produce damaging soluble factors, including nitric oxide, reactive oxygen radicals and excito-toxins such as glutamate; sodium- and calcium-ion channels in the axons are also altered deleteriously (Figure 1.8). Thus, multiple mechanisms of immune-mediated injury of myelin have been postulated: (a) cytokine-mediated injury of oligodendrocytes and myelin; (b) digestion of surface myelin antigens by macrophages, assisted by auto-antibodies; (c) complement-mediated injury and (d) direct injury of oligodendrocytes by CD4+ and CD8+ cytotoxic T cells (Figure 1.8).

Immune regulation, remission and myelin repair

The regulation of initiated autoimmune responses may involve both natural and induced regulatory T cells accessing the perivascular space of the CNS (Figure 1.8). Regulatory T cells in the CNS can directly suppress encephalitogenic T cells through anti-inflammatory cytokine secretion (IL-10 and TGF-β) and cell-cell interaction, the latter focusing primarily on Th17 cells, but they can also regulate Th1/Th2 cells. Regulatory T cells also act on local DCs, thereby rendering them tolerogenic; the tolerogenic DCs in turn can enhance regulatory T-cell expansion or suppressive function that contributes to the suppression of ongoing autoimmune CNS inflammation (see Sections 1.2.2 and 1.3.1). Despite the evident existence and importance of such immuno-modulation in MS, it is not yet known whether regulatory T cells can be generated in the brain parenchyma or whether they function primarily in the periphery. CD8+ T cells found in lesions may also have a regulatory function under some conditions.

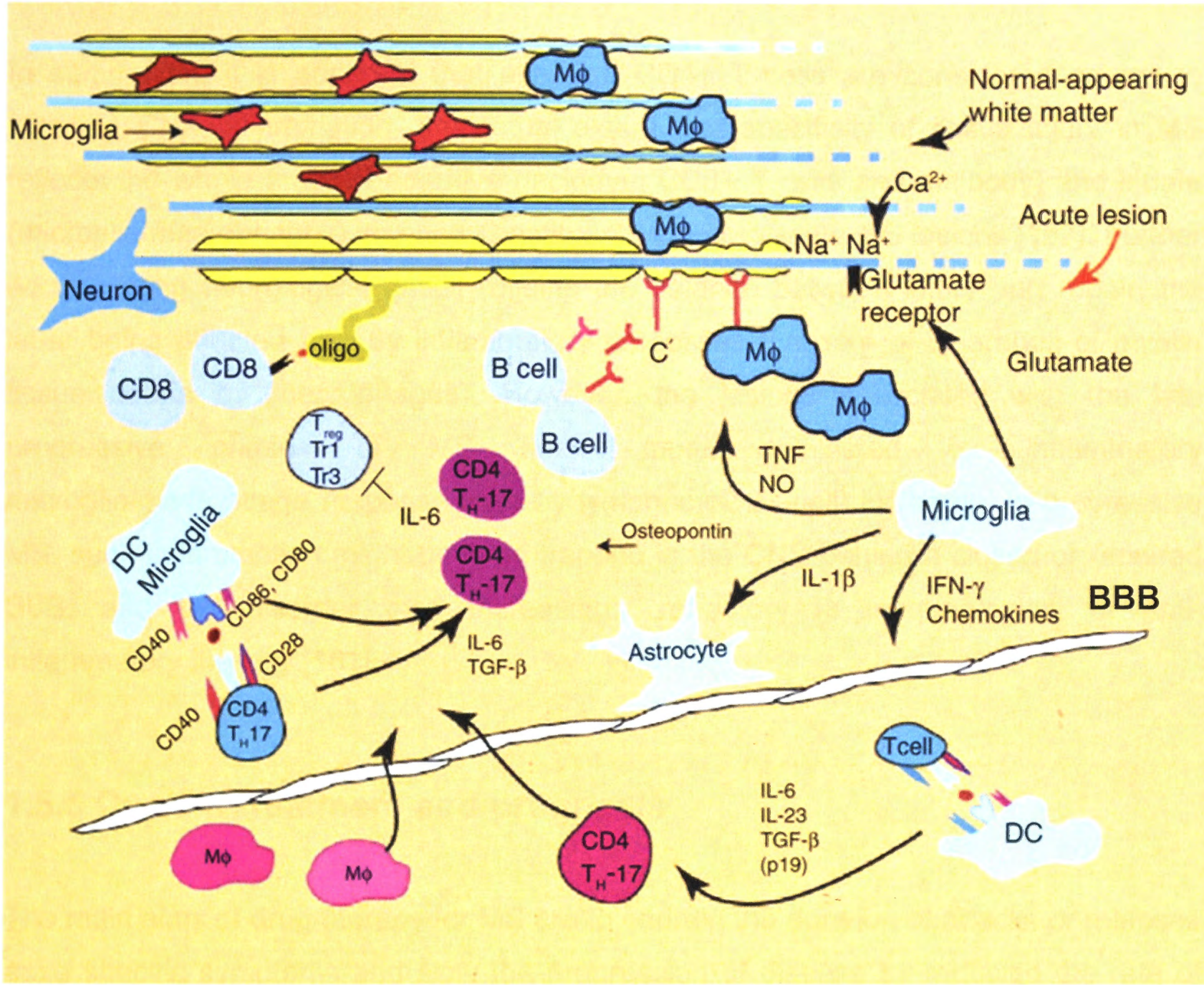


Figure 1.8: Effector mechanisms (T_H17 [CD4] cells shown here) leading to CNS tissue destruction. This includes antibody- and complement-mediated damage, formation of oxygen radicals, glutamate-mediated excito-toxicity, pro-inflammatory cytokine secretion and cell-mediated damage, either through CD8+ T cells or by means of antibody receptors and monocytes, macrophages (MΦ) and microglia (NO, nitric oxide; C', complement). (Figure adapted from McFarland and Martin, 2007 [162].)

The regenerational capacity of the CNS, including axons, has also been appreciated only recently. There are several possible mechanisms of repair of the myelin membrane, including resolution of the inflammatory response followed by spontaneous re-myelination, repair-enhancing antibody-producing B cells and re-myelination resulting from the proliferation, migration and differentiation of resident oligodendrocyte precursor cells. Recent evidence shows that neurotrophins such as brain-derived neurotrophic factor (BDNF), produced by activated T cells, B cells and monocytes, may play a role in re-myelination as well as in protection of axons and neurons.

In summation, it is apparent that although CD4+ T cells are considered central to initiating CNS inflammation, the actual extent and specificity of tissue injury in MS reflects the whole array of adaptive (including CD8+ T cells and antibody) and innate (microglia/macrophages) immune constituents found in acute MS lesions [167]. Further, accumulating neurological deficit reflects the balance between injury and repair, the latter being affected also by inflammatory processes (i.e. rate of clearance of myelin tissue debris by macrophages). However, the lesions associated with the late progressive phases of MS seem more dominated by inflammatory microglia/macrophage response than by lymphocytic (T-cell) infiltrates. In progressive MS, such inflammation might become trapped in the CNS behind a closed or repaired BBB, and the disorder then increasingly refractory to immunological or anti-inflammatory therapy [161].

1.5.5 Current treatment and prognosis

The main aims of drug therapy for MS are to shorten the duration of attacks or relapses, ease specific symptoms and slow the progression of disease by reducing the rate of the attacks. Figure 1.9 provides an overview of basic inductor and effector mechanisms in MS, or EAE, as already described, and indicates possible sites for therapeutic intervention.

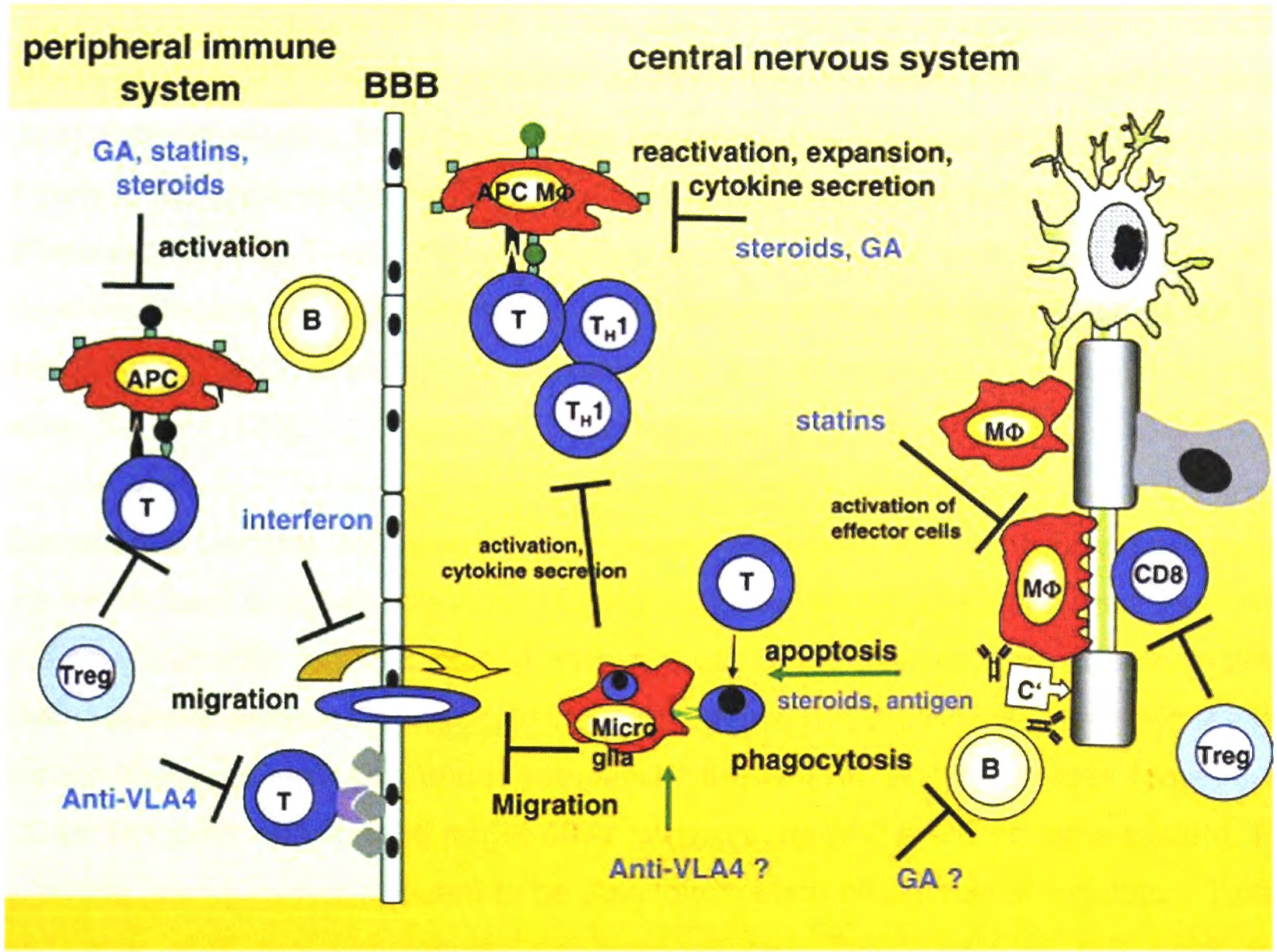


Figure 1.9: Inductor and effector stages of the immune reaction in EAE/MS (T_H1 shown here) and possible sites for therapeutic manipulation. Therapeutic intervention is indicated by black bars (C', complement; GA, glatiramer acetate; MΦ, macrophage; Treg, regulatory T cells). (Figure from Gold et al., 2006 [163].)

Glucocorticosteroids have been used as adjunctive therapy to ease inflammation in RRMS for some years [163], acting at several potential sites (Figure 1.9); it is now known that glucocorticosteroids, in common with several other immunosuppressive and anti-inflammatory drugs, act by inducing tolerogenic DCs [52] (see Section 1.2.2).

Among a number of immuno-therapeutic products, which are used mainly for RRMS, glatiramer acetate (GA, Figure 1.9) is an injected antigen-targeted therapy that appears to mimic the myelin protein and thus block myelin-specific autoimmune responses; it is used for treatment of relapses and reduction of lesions in RRMS [164]. GA is also credited with increasing the suppressive potential of Foxp3+ regulatory T cells, as well as being able to induce CD8+ regulatory T cells with enhanced suppressive ability, in MS patients [55].

Injected beta-interferons (IFN-β1a, b) are used for treatment of relapses and reduction of lesions in RRMS to slow progression and they may also have some cognitive benefit [164]. Beta-interferons have been shown to restore the numbers of Foxp3+ regulatory T cells in MS patients [55], as well as affect BBB 'trafficking' by reducing the migration of monocytes and T cells through human brain endothelial cells [167] (Figure 1.9). However, recent work suggests that beta-interferons may be more effective for Th1 than for Th17 cells, and this might explain the lack of response to beta-interferons in some patients [170].

Compounds blocking adhesion molecules have also been approved for use, including the monoclonal antibody anti-VLA4 (Figure 1.9), also known as natalizumab. However, this drug currently requires careful and very controlled use because of rare fatalities ('MS Australia' website, www.msaustralia.org.au, March 2011) [171]. A number of other immunotherapies are also under evaluation; these now tend to be less focused on CD4+ Th1 cells and instead target other components of the MS immune system. For example, alemtuzumab appears to be able to influence production of regulatory T cells and may shift the unbalanced regulatory/effector T-cell ratio in MS back into homeostasis by removing lympho-mononuclear cells (CD4+ and CD8+) generally [31, 162].

Some of the clinical trials have helped to further understand the still evolving immuno-pathogenesis of MS. For example, a successful trial of rituximab targeting a marker on B cells indicated the involvement of B cells in acute MS lesions [162]. Further support for a role for B cells in disease pathogenesis also comes from a recent trial of ocrelizumab that depleted B cells and reduced RRMS lesion activity [172].

Recently, oral fingolimod has been proven effective for MS in clinical trials and is now available in Australia; this compound targets a protein critical for egress of T cells from lymph nodes and acts to down-regulate T cells moving from lymphoid tissues into the CNS [173]. Two clinical trials of fingolimod have shown halved relapse rates and reduced lesion activity in RRMS, and a third trial with oral cladribine has shown similar reductions (reviewed by Bermel and Cohen, 2011 [171]). Additional effective oral treatments for MS are now likely, such as teriflunomide, which has long been used in rheumatic diseases and has now proven efficacious in reducing MS relapses [174].

Immune suppressants such as methotrexate or mitoxantrone may also be given, especially for progressive forms of MS for which there are few clinical treatments.

However, recent trials of (a patient's own) mesenchymal stem cells show potential in neurorepair, or at least neuroprotection, in progressive MS [175, 176].

Although life expectancy in MS appears to be only moderately affected [177, 178], the overall prognosis means that irreversible limitation in ambulation occurs after a median time of eight years progression, a unilateral aid is required for walking after 20 years and patients become wheelchair-bound after a median time of 30 years [179].

In synopsis, recent observations of immuno-pathogenesis in MS patients have helped to reveal a remarkably complex and still evolving picture of the induction and progression of MS disease. The data suggest that many different T-cell populations can be involved in the induction, propagation and modulation of the disease and that the same cells, which are potentially encephalitogenic, can also provide a proper environment for re-myelination and repair [154]. B cells may also be involved. Cytokines such as TNF-α and IFN-γ, originally considered to be typically pro-inflammatory, are now known to also stimulate immunomodulatory processes as a second, and different, function [27, 163]. Thus, MS is a disorder in which many different components of the immune system interact to break self-tolerance and induce disease.

Importantly, there are many points in the immuno-pathogenetic sequence where both internal and external factors may possibly affect the initiation, precipitation and progression of MS. Relevant to this thesis, there may be a number of possible time points over the life course (discussed in Chapters 2 and 3) where environmental factors may influence disease induction through their effects on immuno-pathogenesis. By attempting to discern these critical points, we can begin to achieve the overall aim of understanding, and perhaps ultimately preventing, autoimmune disorders such as MS.

1.6 Conclusion

Organ-specific autoimmune disorders, such as MS, type 1 (insulin-dependent) diabetes and RA, constitute a significant public health problem in the Western world, yet their aetiology remains unknown. The concept of immunological self, and maintenance of immunologic tolerance to 'self' while retaining the ability to react to 'non-self', appears to be key to understanding the mechanisms of autoimmunity and disease. Several newly recognised, specialised groups of immune regulatory cells constitute a major

homeostatic mechanism whereby the tolerant state is maintained and autoimmune disease is avoided in the long term.

Loss, or breakdown, of tolerance to self-antigens appears to be a critical component in the pathogenesis of autoimmune disease. In genetically susceptible people, both intrinsic and extrinsic environmental factors may contribute to the initiation of autoimmune disease. However, still unknown is the nature of the trigger(s) initiating disease.

MS, a particularly disabling neurological autoimmune disorder, is heterogeneous in nature and its immuno-pathogenesis accordingly complex. Experimental animal models such as EAE, on which much prior work has been based, may represent only part of the spectrum of human MS disease. MS appears to be a largely autoimmune disease but a neurodegenerative component within the target tissue may also contribute to the initiation and propagation of disease.

No cure exists for MS or any of the major autoimmune diseases, and therapy—particularly immune therapy—is still somewhat developmental, or targeted towards the resulting organ damage rather than towards the underlying cellular mechanisms. A multifactorial aetiology is probable for MS as well as other organ-specific autoimmune disorders, a number of genetic, immunological and environmental factors possibly co-contributing, in a sequential way, to the final 'tipping of the balance' towards established disease.

The next chapter will review the epidemiological evidence for organ-specific autoimmune disorders, particularly MS, focusing on the nature of extrinsic environmental factors possibly involved in aetiology and their timing of action leading up to diagnosis of disease.

CHAPTER 2

DETERMINANTS AND THEIR TIMING IN MS AND OTHER ORGAN-SPECIFIC AUTOIMMUNE DISORDERS: EVIDENCE FOR GENETICS, ENVIRONMENT AND SUNLIGHT

2.0 Preface

A widely-used definition of the discipline of epidemiology is that of MacMahon and Pugh (1970), namely, 'the study of the distribution and determinants of disease frequency in human populations' [1], the ultimate goal being to identify and optimise ways of reducing the incidence and severity of disease in those populations. The purpose of this chapter is to summarise the current relevant epidemiological literature on specific organ-specific autoimmune disorders in terms of 'person, place and time', this conceptual framework allowing organisation of the possible determinants with which disease frequency can be associated [180]. Discussion will be restricted mainly to MS, because this is the main autoimmune disorder explored in this thesis. However, type 1 diabetes and RA will also be considered for comparison where appropriate. It will then be possible to identify where current knowledge is lacking, particularly for the southern hemisphere, in order to formulate the overarching research questions for this thesis, keeping in mind the primary aim of determination of possible causes of such autoimmune disorders.

This chapter will briefly consider first the genetic contribution to autoimmune disease, before focusing on the environmental factors and their likely immunological mechanisms that may contribute to the aetiology of MS, type 1 diabetes and RA. In this chapter, sunlight, in particular, will be discussed as a major candidate disease determinant for these autoimmune disorders. Chapter 3 will then continue the discussion of putative environmental factors, focusing on infections and other factors possibly contributing to such disease.

2.1 Genetic basis of autoimmune disease

Family pedigrees of disease phenotype provide one assessment of the contribution of genetic factors to disease susceptibility. In MS, for example, first-degree relatives (parents, children, siblings) of persons with MS have a 3 to 5% chance (i.e. absolute risk) of also developing the disease, this risk being 15 to 25 times the risk in the general background population (data from the Canadian Collaborative Project on Genetic Susceptibility to MS [CCPGSMS] [181]). Similar results were found also in the United Kingdom (UK) and Belgium [182, 183]. Since family micro-environmental factors can also produce familial disease, a study of adoptees was required to show that non-related family members had no increased risk over that of the general population [184]. A further study of half-siblings showed the same risk to each half-sibling whether they were raised together or separately [185], all indicating that the familial risk was more likely due to gene-sharing rather than to the family micro-environment. To further illustrate the disease complexity, though, the risk to a half-sibling (half-sibs sharing 25% of their DNA genetic material) was about one-third the risk to a full sibling (who share 50% of their DNA), rather than an expected risk of about one-half if only one gene was involved. This risk then decreased in more distant relationships, suggesting a multigene model of family inheritance [186]. More detailed genome-wide studies of MS over the past decade have supported the multigene concept [187, 188]; however, such studies have identified only about half of the genetic risk in terms of specific chromosome or gene associations [189, 190].

Autoimmune disorders are also more evident in some particular genetic populations [150, 191-194]. In MS, for example, risk is highest in those of Scottish ancestry and this appears to be due largely to genetic factors [195]. In general, MS is much more common (prevalence >50 per 100,000 population) in northern European populations (and in southern countries colonised by the same Europeans) and is least common (<5 per 100,000) in indigenous populations in the same areas and in Asia [196]. This variation in terms of 'person' is exemplified by another population with unusually high MS risk, the Sardinians, who have been shown to be one of the most intact and ancient human lineages; interestingly, the Sardinians also show high type 1 diabetes risk [4]. For RA, the highest frequency of this condition is found amongst Native American peoples rather than in Europeans [194], this being a further example of an autoimmune disorder linked to a particular genetic population.

The most important genes contributing generally to human autoimmune disease susceptibility are located in the HLA class II region of the MHC on the short arm of chromosome 6 [197, 198]. MHC class II molecules are central to adaptive immune responses against invading pathogens and to maintenance of tolerance to self-antigens (see Chapter 1). Because MHC class II molecules are 'self' molecules involved in (non-self) antigen presentation to CD4+ T cells (see Chapter 1), it is not surprising that this gene region holds the largest number of, and the longest recognised, associations with autoimmune disease of any similar-sized region across the genome [197]. Genes HLA-DR, HLA-DQ, or both (specific HLA-DR/DQ haplotypes involved in MS, type 1 diabetes and RA) encode key molecules in antigen presentation and T-cell repertoire determination [199], and thus genetic variations, and/or levels of gene expression, in this region have a comparatively large effect on susceptibility to such disease. Although the functional basis for the observed HLA class II associations in autoimmune disease remains incompletely understood, a breakdown in immunological tolerance to self-antigens through compromised regulatory processes (see Chapter 1) is a current view [197].

To estimate the effect of genetic factors on the cause of a disease, population-based twin studies can be used as a powerful tool to assess genetic and non-genetic factors in multifactorial immune-mediated diseases [200]. While the concordance rate for MS in fraternal twins in the large Canadian CCPGSMS study was approximately 4% (much the same as for regular siblings who shared the same amount, i.e. 50%, of their DNA), the pair-wise concordance rate in identical twins, sharing 100% of their DNA, was only 38% [201]. A later analysis of longitudinal twin data by the same research group showed that the proband-wise concordance among female monozygotic (i.e. identical) twins approached only 25.3% (SE $\pm$ 4.4), despite 100% sharing of genetic material [202]. In the UK, the monozygotic twin concordance rate was similarly only 25% [203]. Such twin concordance studies for MS in both Canada and the UK suggest the additional involvement of factors other than genetic.

Type 1 diabetes similarly shows a concordance rate in monozygotic twins of only 40 to 50% in several studies [204-206], leaving some 50 to 60% to be accounted for by non-genetic means; however, other studies found the monozygotic twin concordance rate for type 1 diabetes to be only around 20% [207, 208], leaving an even higher margin for environmental effects. RA also shows a particularly low monozygotic twin concordance rate of 12% [194, 209]. Thus, for all three of these organ-specific autoimmune diseases, concordance rates in dizygotic twins of only 2 to 4% that are less than half of those of

monozygotic twins, suggest that the disorders are affected by a number of genes, each with a moderate effect, rather than by a single dominant or higher-effect gene.

The evident multigene, non-Mendelian nature of these complex autoimmune diseases [41, 210] means that several genetic factors can occasionally combine in one individual to crucially affect their immunoregulatory tolerance mechanisms, predisposing that individual, in certain environmental contexts, to an autoimmune disease state. In MS, for example, a recent collaborative genome-wide association study of 9,772 MS cases from 15 countries revealed more than 20 previously suggested and another 29 novel risk loci for this disorder that, importantly, were all immune-related (though together having much less effect than the HLA genes [211]). These genetic loci particularly implicated T-helper-cell differentiation and thus immune dysregulation in the pathogenesis of MS [212].

Hafler (2004) gives an interesting evolutionary reasoning for the multigene concept in terms of microbes encountered by early humans coming out of Africa 30,000 to 50,000 years ago. Presuming that several different microbe-resistance genes were needed to provide maximal disease resistance at the population level, when these genes randomly (and rarely) came to be expressed together in an individual, a hyper-responsive immune system and subsequent autoimmune disease may have resulted [135]—this may have been the price that that rare individual paid for protection of the whole population [22, 135].

In synopsis, the curious, but well-established, identical-twin low disease concordance finding in many populations—that is, of two individuals having the same genetic background and same family environment, but both being concordant for MS or type 1 diabetes in less than half of these twin pairs—strongly suggests the importance of other non-genetic, non-shared environmental triggers, as well as 'chance' or other contributory random, stochastic events in the aetiology of such autoimmune disease [186]. Indeed, even in a recent, very 'deep' genomic analysis of monozygotic twins discordant for MS, no genetic differences in their T-helper immune cells (CD4+ T lymphocytes involved in pathophysiology of MS, see Chapter 1) that might definitively cause MS were found. Additionally, no epigenetic differences—that is, chemical modifications to DNA that effect gene expression but not nucleotide sequence—in the twins' T-helper cells were found, and nor were there any transcriptome differences (messenger RNA expression levels) in these cells [213]. Although only a few twin pairs

were analysed, such studies underline the major importance of the environment in autoimmune disease causation.

Further discussion of specific individual genes associated with increased disease risk is beyond the scope of this thesis; however, appropriate reference will be made particularly to gene-environment interactions that appear to be of primary importance in influencing the risk of developing autoimmune disease.

2.2 Interaction between genetics and environment

'Nature versus nurture' arguments about the pathogenesis of autoimmune disorders such as MS have now given way to evidence that both genetics and environment are important. Indeed, susceptibility is likely to be mediated by direct interactions between the environment and genes [189, 214]. How might such interaction occur? Rioux and Abbas (2005) [57] and others [215] suggest that the environment actively interacts with the multiple genetic factors, or genotypes, to orchestrate the developing immune system, producing the conditions required for subsequent autoimmune disease to be initiated or triggered by later environmental factors and/or random events (Figure 2.1). In part (b) of Figure 2.1, the environmental factors may be extrinsic, for example, exposure to pathogens or to a climatic variable(s), while intrinsic variables may be hormones, such as during pregnancy, or from the mother during the foetal period. Importantly, this concept expresses that an individual's functional immunological repertoire of specific antigen receptors (B-cell receptors and TCRs) is likely to be a *product* of the interaction between their genetic repertoire and the environment [57]. Indeed, differences in selection of TCRs after antigen stimulation between monozygotic twins discordant for MS have been shown [216].

A similar concept to explain (adult) phenotypic differences in genetically identical individuals is the 'early origins hypothesis' or the 'foetal basis of adult disease' promoted by D.J.P. Barker. This hypothesis postulates that nutrition and other environmental factors during foetal and early neonatal development alter susceptibility to several chronic disease conditions in later adult life [217-220]. A similar hypothesis of prenatal 'imprinting' was proposed particularly for vitamin D by J. McGrath in 2001 [221]. There is now much evidence for intra-uterine environmental imprinting, which occurs when developmentally plastic cellular pathways are affected during gestation,

enabling a single genotype to produce a broad range of adult phenotypes [222]. Recent work has shown that gene expression can be altered thus by environmental exposure at the epigenetic level, that is, literally 'above the genome', where heritable changes in gene expression occur in the absence of changes to the DNA sequence itself [223]. In MS, for example, the risk associated with the major MS HLA susceptibility locus has changed between successive generations, strongly implicating gene-environment interactions mediated by epigenetic changes [224].

Consistent with early life effects, a study by CCPGSMS showed that half-sibs with a maternal parent in common had an almost doubled half-sib risk of MS compared with half-sibs with a paternal parent in common; this suggests a significant maternal effect in MS occurrence [225]. These authors propose that environmental factors may have a substantial role in such an effect, and that 'parental imprinting remains a possible candidate'. Subsequent studies by the same group have confirmed this maternal parent-of-origin effect [226], and further identify gender-specific (i.e. female) epigenetic interactions as mediating this effect [227]. The possible importance of the maternal uterine environment is further suggested [228] by the excess concordance (two-fold) of MS for dizygotic (fraternal) twins compared to non-twin sibling pairs in the Canadian data, although these data did not reach statistical significance [202].

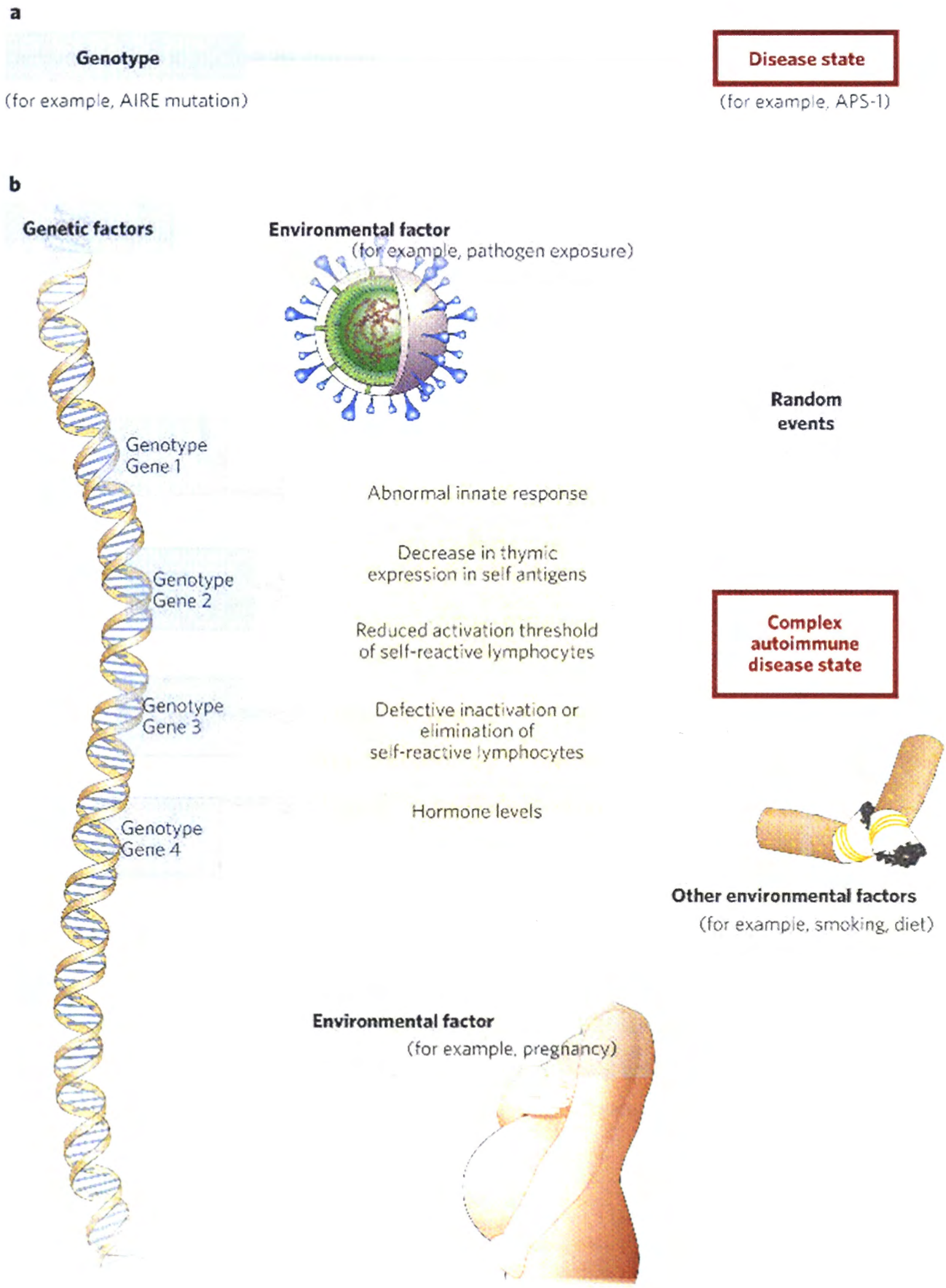


Figure 2.1: Comparison of (a) single gene disorders; for example, autoimmune polyendocrine syndrome (APS-1) caused by mutation of the autoimmune regulator (AIRE) gene (see Chapter 1, Section 1.4.1); and (b) complex traits where the disease state results from interactions between multiple genotypes and the environment. Individual genotypes can affect one or more components of the adaptive or innate immune systems; together these lead to an altered immune response to self-antigens. (Figure adapted from Rioux and Abbas, 2005 [57].)

Causation of autoimmune disorders such as MS and type 1 diabetes almost certainly involves a combination of genetic and environmental exposure factors. This may be considered in the context of Rothman and Greenland's 'component cause model' of disease causation [1], wherein 'sufficient cause' for a disease to occur is defined as a 'set of minimal conditions and events that finally produce disease'. Each 'sufficient cause' comprises a set, or series, of 'component causes', the latter being defined as 'an antecedent event, condition or characteristic that is *necessary* for occurrence of disease, but which *may not be sufficient by itself* to produce disease' (see Figure 2.2).

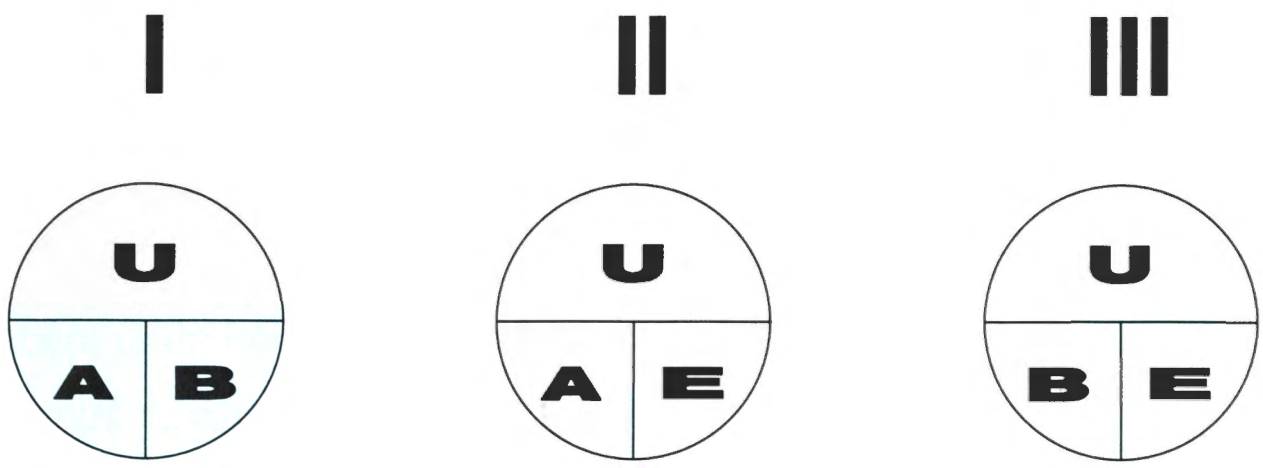
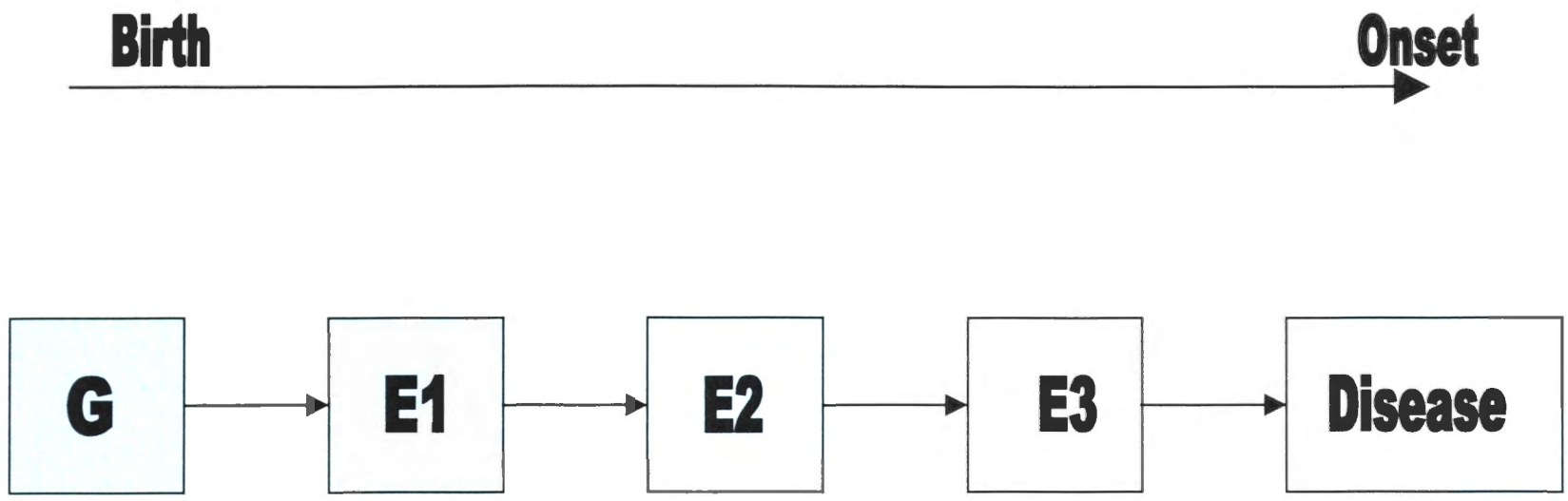


Figure 2.2: Component cause model of disease causation. Three sufficient causes of a disease characterised by component causes A, B, E and a ubiquitous component U. (Figure from Rothman and Greenland, 1998 [1].)

In Figure 2.2, the component causes, A, B and E, where only two of these three components are necessary for disease, act together with a ubiquitous component cause, U, to produce three possible sufficient causes. That is, these component causes acting together in the same sufficient cause may be thought of as interacting biologically to produce disease. In other words, 'biological interaction' may be defined as 'participation of two (or more) component causes in the same sufficient cause' [1].

Further, this interaction or 'joint action' need not be simultaneous: one component cause may act *many years before the other*, the condition being that the first component cause leaves some effect that interacts with the later component [1]. Figure 2.3 shows a possible 'life-course' sufficient-cause scenario for an example multifactorial disease with four successive component causes, three components being environmental and the first being genetic.



'Component causes do not need to act at the same time, and may interact and accumulate over the life course to result in sufficient cause.'

Figure 2.3: Life-course disease causation model for a multifactorial disorder with genetic (G) and environmental (E1 to E3) component causes.

Such a concept of an immune disorder such as MS or type 1 diabetes being 'a disease waiting to happen', wherein a combination of genetic and environmental factors must reach a sufficient level, or threshold, before the critical transition is made to clinical disease, is gaining support. For example, for MS, the critical factors, both genetic and environmental, may occur over 'a critical window of time', giving rise to persons with an 'MS trait' [228]. The 'MS trait' notion, which evolved from observations such as the low concordance rate in monozygotic twins, postulates that in some genetically susceptible individuals, a primary challenge, which may or may not be immunological, may cause damage to the BBB, but that this is insufficient to result in formation of demyelination plaques (see Chapter 1, Section 1.5). A second immunological challenge may then make the already-vulnerable BBB less resistant to subsequent events. There may also be other subtle changes within the CNS as part of the MS trait [228]. Such an individual then, over time, may or may not develop either subclinical or clinical MS depending on other, presumably environmental, factors. A similar sequential concept is Goodin's (2009) 'multi-hit causal cascade' hypothesis for MS [229], discussed in more detail in Chapter 3.

An analogous hypothesis for type 1 diabetes induction has been suggested, wherein genetic predisposition appears to determine almost entirely whether a person will develop immune reactivity against insulin-producing beta-cells in the pancreas, but environmental factors have a major effect on whether type 1 diabetes manifests itself clinically, this occurring after 80 to 90% of the beta-cells have been destroyed [230].

Such a possible life-course model for MS, or other autoimmune disorders such as type 1 diabetes or RA, is depicted in Figure 2.4, showing environmental factors acting at various 'critical windows of time'. The effect of environmental events in early life in particular will be discussed further at the end of this chapter in Section 2.6.

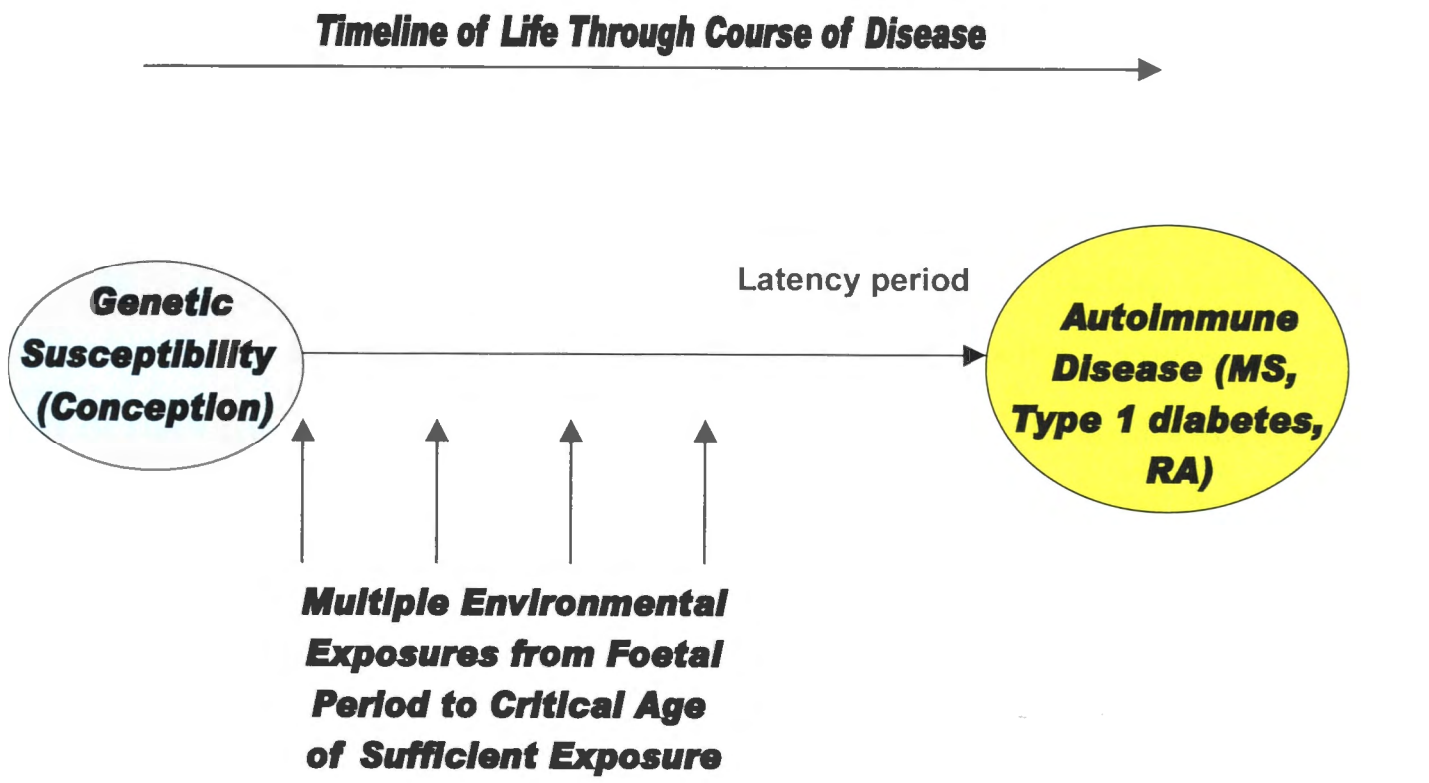


Figure 2.4: Aetiology of autoimmune disorders such as MS, type 1 diabetes and RA: a possible life-course model of timing of genetic and environmental component causes.

Summing up this section, a substantial contributory role for environmental factors in the aetiology of these organ-specific autoimmune disorders is suggested. Genetic ('person') factors are *necessary, but not sufficient*, to cause disease for most individuals. Environmental factors should be considered to both interact with genetic factors and may act alone at several possible time points over the life course leading to onset of disease (Figures 2.3 and 2.4). Such factors are of particular interest if they are modifiable, or even preventable, and if they also trigger disease exacerbations, such as seen in MS [231]. From a public health point of view, it is important to note that if there are a number of component causes involved in a sufficient cause for any disorder, intervention on only *one* of these (environmental) components may be all that is

required to prevent disease. The remainder of this chapter, and the next, will therefore focus on the environmental factors that may possibly be involved in aetiology of MS, type 1 diabetes and RA, and the 'place' and 'time' associations of these factors with disease occurrence.

2.3 Epidemiological evidence for environmental determinants

Ecological studies and candidate environmental determinants

Ecological studies, focused on group- or population-level disease determinants and disease occurrence, can inform about possible aetiological roles and assist in formulation of hypotheses to be tested at the individual level, or in screening of existing hypotheses [1]. Further, it is increasingly being recognised that such studies are valuable in their own right and that factors at the population level beyond those of individuals are often important to health [232]. For example, the well-known concept of herd immunity in infectious diseases implies that an individual's likelihood of contracting an infectious disease depends partly on the level of immunity in the person's population. Ecological studies allow the basic descriptive characteristics of a disease to be determined in relation to the parameters of 'person', 'place' and 'time'. That is, of basic interest in these studies is:

- Person: '*Who* is getting the disease?', including age, sex and race.
- Place: '*Where* are the rates of disease highest and lowest?', including country, geographic region within country (e.g. latitude gradient). For this parameter, studies of migrants provide hypotheses of possible relative roles of genetics and environment.
- Time: '*When* does the disease occur commonly or rarely?', 'Is the frequency of the disease different now from the past?' and 'Is there a cyclic or other pattern of the disease?' [233].

Despite ecological studies using data averaged over individuals, and sometimes proxy measures for exposure, together with often uncontrolled confounding factors, such studies can also be useful for detecting associations that signal the presence of effects that may be further investigated at the individual level [1]. An ecological approach is used in Chapter 4 of this thesis, for a number of immune disorders including type 1 diabetes.

From ecological studies, a range of factors have been variously associated with MS, including those in the geographic or physical category (e.g. latitude, altitude, temperature, UVR, toxins), biological factors (e.g. infections, diet, stress, hormones, smoking) and cultural factors such as socio-economic status (SES) [231, 234-236]. The least equivocal associations were found with low temperature, non-specific infections [235] and, more recently, low UVR and/or low vitamin D [6, 234, 237], EBV and smoking [238, 239]. For type 1 diabetes, the environmental factors most implicated include infections (e.g. enteroviruses), lack of breastfeeding, diet (e.g. cow's milk), low temperature, toxins [192, 240, 241] and, more recently, low vitamin D and/or UVR [242]. Environmental risk factors for RA lie mainly in the biological category and include infections, hormonal factors, adverse pregnancy outcomes, smoking, obesity, diet [194] and low vitamin D and/or UVR [25].

However, as shown by the component cause model of disease causation (see Figure 2.2), any of these factors as determined by ecological studies may be involved in any combination in the sequential determination, initiation (or induction) or final precipitation (triggering) of disease onset. They may contribute at different times to disease acquisition, as well as to the modification of its subsequent course [234].

Analytical epidemiology

A useful technique here is an analytical epidemiological approach, where each factor is carefully assessed at the individual level to determine whether the exposure factor is acting as an antecedent cause, a directly causal exposure, an intermediate, a confounder, an effect modifier, or as a consequence of disease.

Table 2.1: Framework for the interpretation of an epidemiological study (Hennekens and Buring, 1987 [233])

Is there a valid statistical association?
Is the association likely to be due to chance?
Is the association likely to be due to bias?
Is the association likely to be due to confounding?
Can this valid statistical association be judged as cause and effect?
Is there a strong association?
Is there biological credibility to the hypothesis?
Is there consistency with other studies?
Is the time sequence compatible?
Is there evidence of a dose-response relationship?

While a number of criteria (based on those first proposed by Hill, 1965) can be used to assess a valid statistical exposure-disease association for evidence of causality [233] (Table 2.1), Rothman and Greenland caution that only the criterion of temporality is strictly necessary to define causality (i.e. Is the time sequence compatible? In particular, does the exposure *precede* disease?), the other criteria assisting interpretation but not being exclusively required [1]. Additionally, whether an exposure factor is a true confounder of the association between exposure and disease, or is intermediate on the causal pathway between the exposure and disease (see Figure 2.5), must be distinguished. As shown in Figure 2.5, a confounder is a variable that is associated with the exposure and, independent of that exposure, is a risk factor for the disease. If, however, one mechanism of action of the exposure is to alter the level of the potential confounder, which in turn itself affects disease risk, then that factor is not a confounder but rather an intermediate step in the causal chain between the exposure and disease [233].

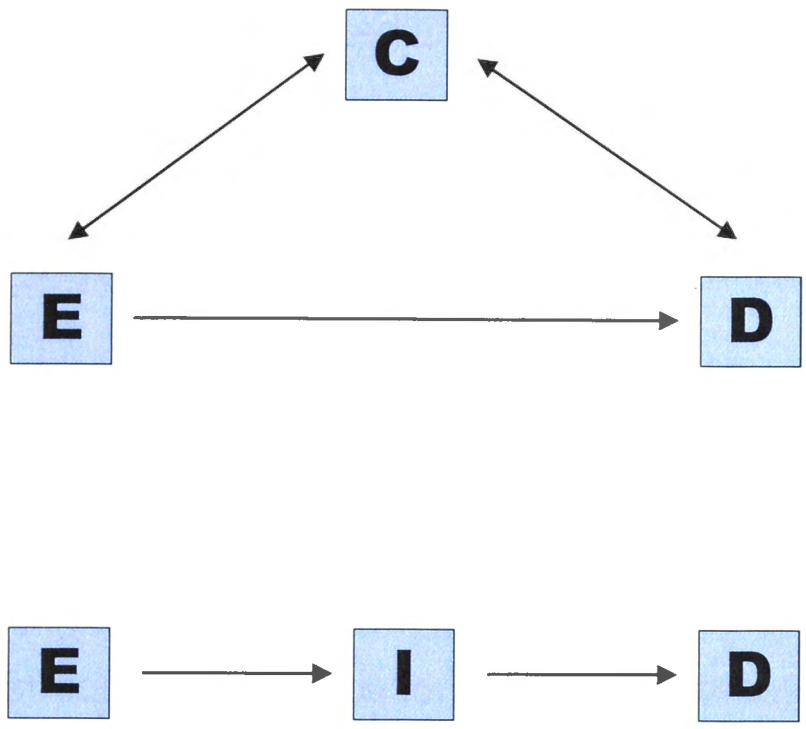


Figure 2.5: Interrelationship between an exposure (E), a confounding factor (C) and disease (D) (top of diagram), and an exposure (E), disease (D) and an intermediate factor (I), which is on the causal pathway and thus not a confounder (lower part of diagram).

Effect modification, conversely, may be present when the magnitude of the association between the exposure and disease varies by levels of a third factor [233]. To distinguish such relationships between exposure factor(s) and disease, individual-level epidemiological studies, such as randomised controlled trials (RCTs), cohort studies, or case-control studies, are required. While the effects of confounding can often be controlled in the design phase of analytical epidemiological studies (i.e. by randomisation, restriction or matching), the specific analytic techniques of data stratification and multivariate analysis can also be used to both control confounding and determine and describe effect modification (this is discussed further with respect to data analysis in Chapter 7). This thesis will use an individual-level observational epidemiological approach for Chapters 5 to 7 (timing of birth of MS cases) and Chapter 8 (order of birth of MS cases), aiming to determine the possible nature of environmental factors influencing MS risk from their life-course timing.

A review of analytical epidemiological studies of sunlight exposure (UVR and vitamin D) and effects on MS, type 1 diabetes and RA is given in Section 2.6 of this chapter. However, such individual-level analytical studies on specific, potentially causal factors

can be guided by information from ecological studies of disease occurrence in terms of person, place and time. Examples are longer-term changes in disease incidence (variation in time), geographic gradients in prevalence or incidence (variation in place) and seasonal or other cyclical changes in incidence (variation in time). These will now be considered in turn for MS and type 1 diabetes.

2.3.1 Increasing incidence of organ-specific autoimmune disorders

Variation in 'time'

One of the most convincing indications of the contribution of environmental factors to aetiology of autoimmune disorders is their increasing incidence over the last three decades in developed countries [3], such increases being faster than population-level genetic change would suggest or allow [4, 193, 240]. Although ascertainment of disease cases would undoubtedly have also improved over the same period, these increases in incidence do appear to be, mostly, real and thus suggest environmental causation [4, 240]. A substantial increase in MS incidence, for example, has been reported from high prevalence areas such as Norway, Finland and Sardinia in the northern hemisphere [243-245] and from Australia in the southern hemisphere [246, 247].

In Newcastle, New South Wales (NSW), Australia, where three different decades over the period 1950 to 1996 were compared using comparable diagnostic criteria and methods for each decade, average annual MS incidence rate doubled from 1.2 to 2.4 per 100,000 population over the 35-year study period. The largest increases were seen in females and in the older age groups. Prevalence also increased markedly from 19.6 to 59.1 per 100,000 of population over the same period. Although such trends in MS prevalence and incidence had been observed in the northern hemisphere, particularly in Europe (summarised in Table 2.2), this was the first such study to show a longitudinal increase in prevalence and incidence in the southern hemisphere over a period of this duration [246]. Consistent with these results, another (and longer) MS study in Tasmania (TAS), Australia showed a three-fold increase in age-standardised prevalence and a nearly two-fold increase in incidence over the period 1951 to 2009, part of the prevalence increase being attributed to increased longevity, decreased mortality and increased incidence [247].

Table 2.2: Increasing MS prevalence and crude incidence rates over time within areas in selected European countries

Country	Area	Prevalence (mean/10^5) (Year)	Reference	Incidence [95% CI] (per 10^5/yr) (Years)	Reference
Norway	North Norway	21 (1973)	[248]	2.6 [1.7, 3.7] (1974–78)	[249]‡
	(Troms/Finnmark)	73 (1993)	[249]	3.0 [2.1, 4.2] (1979–83)	‘ ‘
				3.5 [2.5, 4.8] (1984–88)	‘ ‘
				4.3 [3.0, 5.9] (1989–92)	‘ ‘
	Oslo	86* (1983)	[250]	3.6 [2.2, 6.0] (1972–76)	[251]
	(South Norway)	120 (1995)	[251]	4.4 [2.8, 6.9] (1977–81)	[251]‡
				4.9 [3.1, 7.6] (1982–86)	‘ ‘
				7.2 [5.0, 10.2] (1987–91)	‘ ‘
				8.7 [6.3, 11.9] (1992–96)	[251]
	SW Norway	?		1.8 (1953–57)	[252]
	(Hordaland Co.)	151 (2003)	[252]	4.1 [2.1, 6.1] (1973–77)	[253]
				4.7 [2.4, 7.0] (1978–82)	‘ ‘
				3.2 [0.6, 5.8] (1983–87)	‘ ‘
				6.0 (1993–97)	[252]

Country	Area	Prevalence (mean/10^5) (Year)	Reference	Incidence [95% CI] (per 10^5/yr) (Years)	Reference
	Central Norway	?		3.9 (1974–75)	[254]
	(Nord–Trondelag)	164 (2000)	[254]	5.6 (1998–99)	‘ ‘
Finland	Central Finland	59 (1993)	[255]	3.8 (1979–93)	[255]
		105 (2000)	‘ ‘	9.2 (1994–98)	‘ ‘
	Western Finland	119 (1983)	[256]	9.4 [7.7, 11.2] (1979–86)	[256]
	(Seinajoki)	202 (1993)	‘ ‘	11.6 [10.1, 13.8] (1987–93)	‘ ‘
Sweden	North Sweden	125 (1990)	[253]	2.6# [2.2, 3.0] (1974–88)	[253]
	(Vasterbotten Co.)	154 (1997)	‘ ‘	5.2 [4.4, 6.2] (1988–97)	‘ ‘
Denmark	Denmark	59 (1950)	[257]	4.1 [3.9, 4.3] (1970–79)	[258]
		112 (1990)	[258]	5.0 [4.8, 5.2] (1980–89)	‘ ‘
		154 (2005)	[257]		
UK	NE Scotland	127 (1970)	[259]	6.0 (1968–70)	[259]
		178 (1980)	‘ ‘	7.2 (1977–80)	‘ ‘

Country	Area	Prevalence (mean/10^5) (Year)	Reference	Incidence [95% CI] (per 10^5/yr) (Years)	Reference
	SE Scotland	203 (1995)	[195]	12.2 [10.8, 13.7] (1992–95)	[195]
	Nthn Ireland	51 (1951)	[260]+	2.7 (1937–51)	[260]+
		80 (1961)	‘ ‘	4.4 (1952–61)	‘ ‘
	NE Nthn Ireland	138 (1986)	[261]+	9.3 (1996–96)	[262]
		168 (1996)	[263]+		
		200 (2004)	[262]		
	Wales	101 (1985)	[264]	4.2 [2.6, 6.9] (1985–85)	[264]
		146 (2005)	‘ ‘	9.6 [7.1, 13.1] (2007–07)	‘ ‘
France	Lorraine	120 (2004)	[265]	5.5 [4.4, 6.6] (1990–02)	[265]
	France	95 (2004)	[266]	7.5 [7.3, 7.6] (2003–04)	[266]
Italy	North Sardinia (Sassari)	103 (1991)	[244]	1.2 [0.8, 1.8] (1965–69)	[267]
		144 (1997)	‘ ‘	4.9 [4.0, 6.0] (1980–84)	‘ ‘
				6.1 [5.1, 7.2] (1995–99)	‘ ‘

Country	Area	Prevalence (mean/10^5) (Year)	Reference	Incidence [95% CI] (per 10^5/yr) (Years)	Reference
	Central Sardinia	~100 (1985)	[268]	1.9 [1.3, 2.9] (1955–59)	[268]
	(Nuoro)	152 (1994)	‘ ‘	4.8 (1975–79)	‘ ‘
				6.4 [5.2, 7.8] (1990–96)	‘ ‘

*Vestfold County, Norway (near Oslo).
#Goteborg, southern Sweden.
‡Cited by Sundstrom et al., 2003 [253].
+Cited by Gray et al., 2008 [262].

In type 1 diabetes also, a marked increase in the incidence of this childhood immune disorder has been reported from Europe, North America, Norway, Sweden, Hungary, Finland, Austria, Israel, England and Switzerland in the northern hemisphere [269-285]. For example, in Switzerland, incidence increased from 4.5 to 10.5 per 100,000 of population over the 1965 to 2000 period, all of this increase being in children under five years of age, and most of the increase having occurred in the last 10 years [282]. Similarly in Finland, where incidence is highest worldwide, incidence increased from 31.4 to 64.2 per 100,000 per year from 1980 to 2005; again the highest rate of increase (4.7% per year) was in the 0 to 4 year-old age group [275]. For Europe (17 countries), the overall annual increase between 1989 and 2003 was 3.9% (95% CI 3.6, 4.2) [280].

In the southern hemisphere, a 3 to 6% increase in type 1 diabetes per year has been recorded in Australia [286, 287] and in New Zealand (NZ) [288] and Peru [277, 279], and the overall global increase in incidence, in 27 countries over three decades, has been estimated as 3% (95% CI 2.6, 3.3) per year [279].

2.3.2 Latitude gradients in prevalence or incidence

Variation in 'place'

Another of the more significant epidemiological features of organ-specific autoimmune disorders is a gradient of increasing prevalence with increase in geographical latitude. That is, there is a general, global increase in prevalence of both MS and type 1 diabetes away from the equator and towards Earth's poles, particularly in the northern hemisphere, where these diseases have been most studied (see Figure 2.6). Indeed, in MS, for example, while the effect of the environment is not necessarily mediated by a single factor, latitude demonstrates the strongest association [214]. Prevalence of MS in 2007/2008 varied globally from almost zero at the equator to greater than 100 (per 100,000) at some temperate latitudes [289]; these more recent estimates for MS are shown for Europe in Figure 2.7 (Multiple Sclerosis International Federation Atlas of MS database, http://www.atlasofms.org/index.aspx , accessed February 2011).

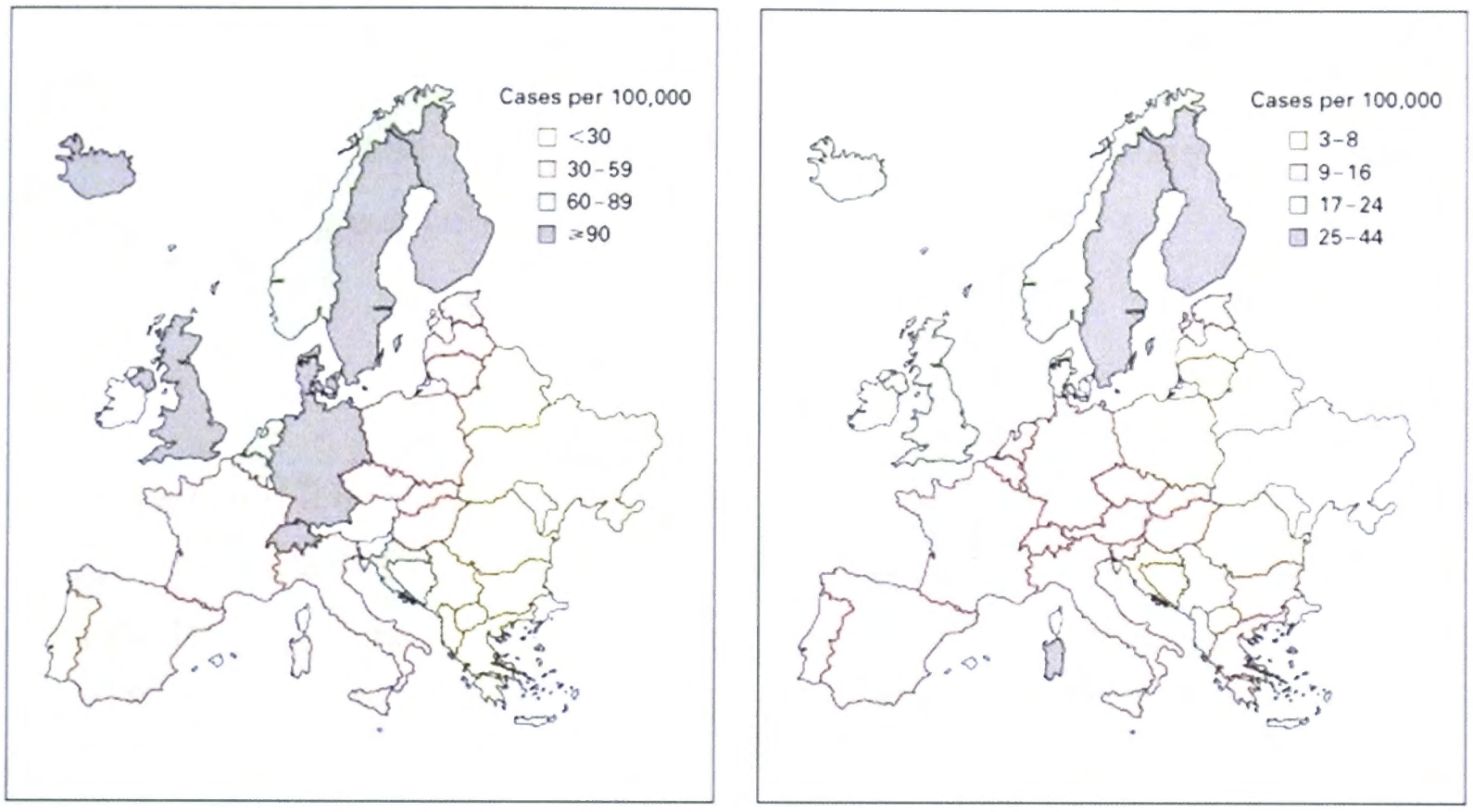


Figure 2.6: (a) The north-south gradient in prevalence of MS (based on data from Kurtzke, 2000 [290]) and (b) the gradient in incidence of type 1 diabetes (based on data from Green and Patterson, 2001 [291]) in Europe, as given by Bach (2002) [3].

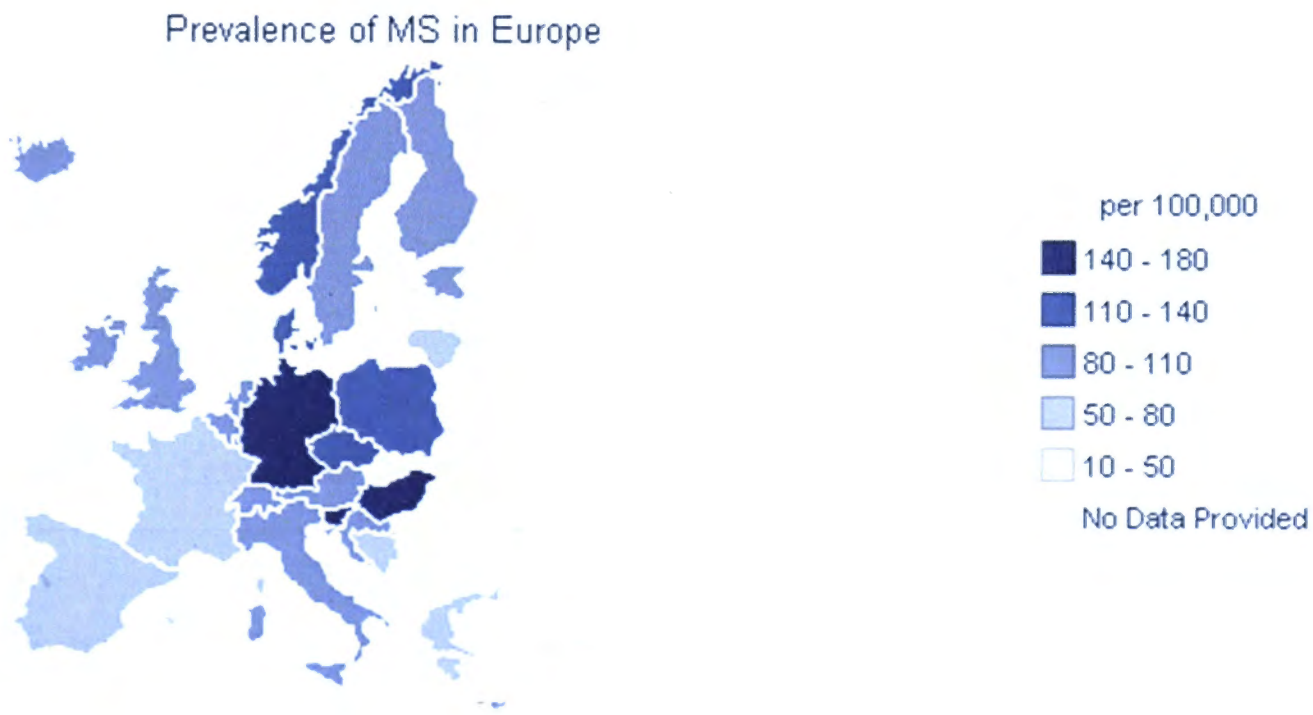


Figure 2.7: Prevalence of MS in Europe by country, 2008. (Data map from Atlas of MS database.)

RA also exhibits geographic variations suggestive of a similar north-south gradient in both prevalence and incidence in Europe and North America [194, 292].

Although there are some local exceptions to the general pattern (e.g. for MS: Sardinia [Italy], and Scotland and Northern Ireland [UK] are higher than expected, and Sami [northern ethnic group] areas in Norway lower than expected, see Figure 2.8), the general global distribution of these three autoimmune diseases suggests the action of environment shaping such disease occurrence.

MS

In Figures 2.8 and 2.9, I have collated data from recent reviews to show the prevalence and incidence of MS in Europe in more detail. In Figure 2.8, mean prevalence is plotted against the latitude of the capital city of each country, using prevalence data over 1986 to 2001 from Pugliatti et al. (2006) [293], supplemented by data for Scotland from Rothwell and Charlton (1998) [195].

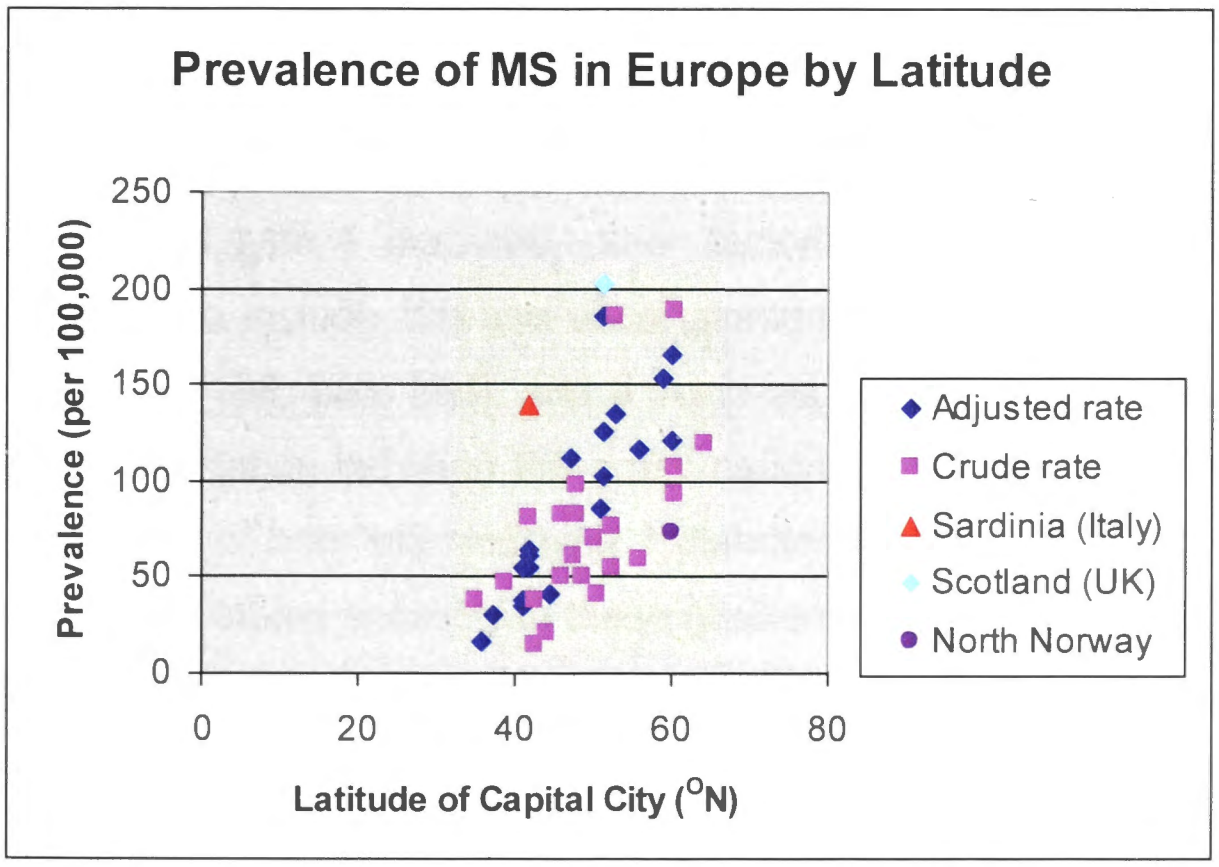


Figure 2.8: MS prevalence rates in Europe by latitude of country (rates shown adjusted for the standard European population, or crude where adjusted rates not available). Adjusted rates for Sardinia, Scotland and northern (Sami-populated) areas of Norway are shown separately. (Data from Pugliatti et al., 2006 [293] and Rothwell and Charlton, 1998 [195].)

MS incidence estimates over periods up to 15 years to 2001 show a similar pattern of increase with northern latitude of the capital cities (Figure 2.9).

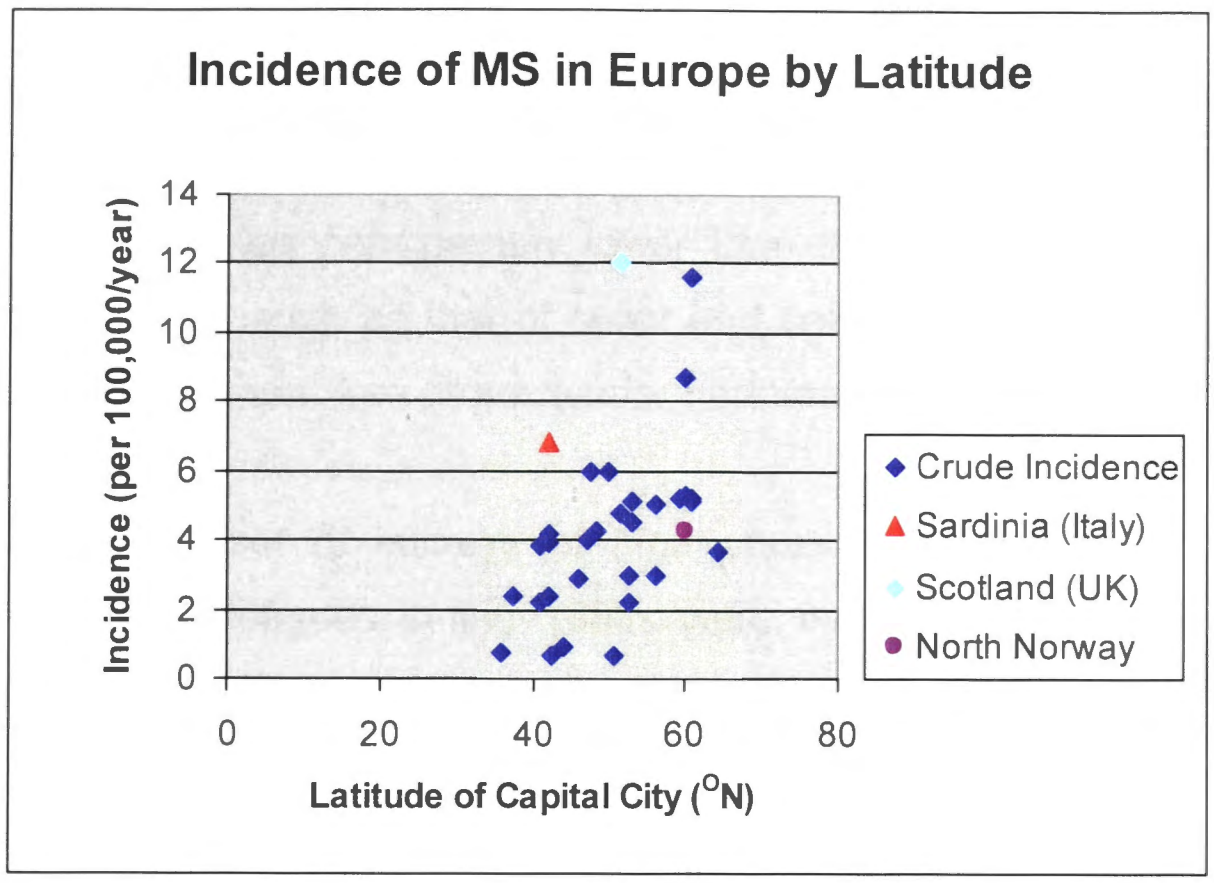


Figure 2.9: MS incidence rates in Europe by latitude of country. Crude rates for Sardinia, Scotland and northern (Sami-populated) areas of Norway are shown separately. (Data from Pugliatti et al., 2006 [293].)

For both MS and type 1 diabetes, other factors possibly confounding this latitude gradient in Europe include disease ascertainment differences between northern and southern Europe [196, 293, 294], and a socio-economic gradient. The latter is shown by a positive correlation between the gross national product in 12 European countries and the incidence of both MS and type 1 diabetes, as well as asthma [3]. In addition, the contribution of ethnic ancestry to these gradients has been more recently defined in both Europe and North America for MS [4, 193]. Nevertheless, as indicated by these gradients, environmental factors are generally agreed to be important in modifying genetic susceptibility and thus influencing disease distribution [193]. Within-country gradients are also marked in the UK [293] and France [189] for MS.

In North America, a large prospective study for MS (the Nurses' Health Study) compared women born in two different periods (1920 to 1946 and 1947 to 1964) in the US. A strong north-south latitude gradient was shown for the earlier cohort, though this appeared to have attenuated somewhat in the later cohort; a concomitant change in a causal environmental factor was, however, suggested [295]. Later work has shown the attenuation in the gradient to be mainly after 1980 [296] and particularly in the US, where decreased time outdoors because of a modern urban lifestyle, and thus decreased UVR exposure, was postulated to be one contributing factor [297].

In the southern hemisphere, the first epidemiological study of MS in Australia by Acheson in 1961 was based on mortality data, and found that MS mortality showed a close positive relationship with southern latitude [298]. This early study also showed that MS mortality was considerably lower than that in the UK [299]; this and later prevalence studies, such as that of Miller and colleagues (1990), suggested that the antipodean environment was protective for northern Europeans for MS [300].

A comprehensive set of surveys of nine areas of Australia and NZ by McLeod, Hammond and colleagues in the 1980s were the first prevalence studies in these climes. This research group showed a marked latitude gradient in MS prevalence for Australasia, a seven-fold increase in prevalence existing from subtropical North Queensland (latitude 19°S), to temperate Hobart, TAS (latitude 43°S), and NZ (latitude 34 to 47°S) [301, 302].

While the more appreciable MS gradient in NZ has more recently been attributed partially to both higher numbers of Scottish migrants in the south and higher proportions of genetically resistant indigenous Maori people in the north [4, 193], the still almost seven-fold gradient in Australia persists unexplained and cannot be easily attributed to differences in either SES or genetics [4, 301, 303]. Further, the latitude gradient in MS prevalence was also evident among British and Irish migrants to Australia [304].

A recent prospective study in Australia ('Ausimmune' study) on the incidence rate of 'first demyelinating events' (FDE, a highly predictive pre-diagnostic event for MS) has also established a four-fold latitudinal gradient from Brisbane (27°S) to TAS (43°S) [305], further suggesting environmental effects on MS-related disease. Australia is thus an ideal location for further MS study, as detailed in Chapters 5 to 7.

Age-related migration (variation in 'time' and 'place')

Studies of migrant populations moving between areas of differing prevalence rates have been used to help answer the 'genetics or environment' question, particularly for MS. In addition, by determining whether age at migration alters these prevalence patterns, the *timing* of important environmental exposures can be indicated. That is, even though such studies have many possible sources of error (reviewed by Gale and Martyn, 1995) [306, 307], migrants moving from a high-(MS) risk region to a low-risk region (e.g. UK/Europe to Asia or Africa) generally have a lower than expected risk of

disease, particularly if migration occurs before the age of 15 years. However, migration from a low-risk to a high-risk region tends to result in retention of the low risk of the home country and not show any clear age-at-migration effect [8, 306]. More recent studies are largely consistent with these findings [308-311], including a re-analysis of an earlier Australian migration study that had previously reported a 'critical' migration age well into adulthood [304] and that now supports the age-15 'limit' [312]. Thus, MS risk appears strongly associated with place of residence in early life.

Type 1 diabetes

In the northern hemisphere, the first globally-standardised incidence measures for type 1 diabetes over 15 countries were achieved in the early 1980s. These showed the average annual age-adjusted incidence for those under 15 years of age to decrease approximately seven-fold from Finland to France in western Europe, and eight-fold from Michigan in the northern US to California in the southern US [313]. An update on the worldwide incidence in 40 countries by the World Health Organization (WHO) DiaMond group in 1993 showed the highest incidences and the greatest intra-continental variation to be in Europe, where there was an eight-fold gradient from Finland to northern Greece [314]. In China, also, as part of the same WHO project, a strong north-south latitude gradient was shown, a doubled incidence being seen in the northern half of the country compared to the south, even though type 1 diabetes incidence overall was, globally, the lowest ever reported [315]. A further update by the WHO DiaMond group for 51 regions in 1990 to 1994 reported an overall 350-fold variation in annual incidence from 0.1/100,000 in China and Venezuela to 37/100,000 in Sardinia and Finland [316]. Subsequent analysis of these 1990 to 1994 data by latitude established an association between latitude and type 1 diabetes risk over all 51 regions [317]. Records over the next decade showed even higher incidences in northern Europe [291], and a north-south gradient as pictured in Figure 2.6b. Some of the incidence variation was shown to be due to ethnicity [313, 315], but a similar gradient was shown even within the genetically homogeneous country of Sweden [318].

Generally, southern hemisphere countries showed lower type 1 diabetes incidences than the northern hemisphere, and the lowest incidences overall were found in Asia (China) and South America [314, 316]. However, few latitude-gradient studies have been conducted in the southern hemisphere on this important childhood immune disorder, despite the average annual incidence being at least 10/100,000 in Australia and NZ [279, 288].

Migration studies also support the geographic differences for type 1 diabetes. For example, Japanese children in Hawaii show a four-fold risk compared to those in Japan, and Samoan children in NZ show similar differences. There is an overall trend for immigrant populations from low-incidence countries to quickly adopt the type 1 diabetes risk prevalent in the new environment [30].

Latitude as proxy environmental factor

It is important to note that geographic latitude is essentially a proxy for several putative environmental factors. Increasing northern, or southern, latitude is inversely associated with decrease in both of the climatic variables, temperature and solar radiation, the latter including UVR. For example, both temperature and UVR decrease towards Earth's poles, the rate of decrease in both of these physical variables being greater at certain times of the year (e.g. winter). In Australia, the decrease in annual averaged ambient UVR is 1 kJm^{-2} per 10° increase in latitude [319]. Geographic latitude may also be indirectly associated with diet and biological factors, such as infectious agents, through their more obvious association with climate, and again in particularly cold, or wet, seasons.

2.3.3 Seasonal variation

Cyclic variation in 'time'

Supporting the latitude-gradient evidence for the action of environmental factors on MS and type 1 diabetes, and perhaps even more strongly suggestive of this action, is evidence of seasonal changes in these disorders. Several of the aforementioned factors, including temperature, UVR and infections, vary seasonally as well as geographically, particularly at the higher latitudes. Possible associations of these with variations in disease parameters may give clues to the nature of these factors.

For MS, for example, clustering (in time rather than in space) can occur in the *time of birth* (within the year) of persons who later develop MS, in the *time of onset* of disease, in the *timing of* disease *relapses* (in RRMS), and in the *timing of CNS lesion activity*. Alternatively, or as well, a relative paucity of any of these MS parameters may occur at

any time of the year, since an environmental influence may be protective rather than adverse.

MS

Timing of MS births in the northern hemisphere has exhibited mainly a spring excess (relative to those in the general population), as seen in Denmark, Sweden and British Columbia (Canada) [7, 320-322]. An autumn excess in Sicily [323] has also been reported. More recently, a significant excess in MS births in May (northern spring) as well as a significant MS-births deficit in November (northern autumn) has been reported in a large, pooled Canadian, British, Danish and Swedish study [324]. A seasonal cycle in the month of birth of persons with MS may indicate a causal (or protective) environmental factor acting around the time of birth, which may be a critical time 'window' for such effects [7]. This is evaluated further in Chapters 5 to 7. Prior to this thesis, no published season- or month-of-birth data were available for MS in the southern hemisphere.

The time of onset of disease can also be influenced directly by seasonally varying environmental factors. MS onset, and onset of monosymptomatic optic neuritis (MON, closely related to MS), were studied in a meta-analysis of 21 northern hemisphere studies; both conditions showed highest frequencies in spring, and lowest in winter [325].

Relapses or progression of MS can also show analogous changes with season. In the northern hemisphere MS meta-analysis, MS exacerbations showed highest frequencies in spring and lowest frequencies in winter, similar to both MS and MON onset [325]. A more recent meta-analysis in both hemispheres showed similar relapse onset peaks in early spring but troughs in autumn, April being the peak month in the northern hemisphere and October the peak in the southern hemisphere [326]. In southern hemisphere TAS, Australia, lower MS relapse rates at the population level were observed in late summer (February) compared with the rest of the year; these were associated with both upper respiratory tract infections (positively) and erythemal UVR (inversely) 1.5 months prior [327].

Additionally, the *activity of MS lesions* detected by MRI in southern German MS patients was shown to exhibit a 'sinusoidal' seasonal pattern with an excess in spring and early summer and an autumn deficit [328]; in a further ecological extension of this

study, this pattern was shown to be inversely correlated with the circulating vitamin D (25(OH)D) levels two months prior, indicating a time-lagged relationship with this largely environmentally determined factor [329]. While another larger MRI study of MS lesional activity in patients from Europe and Canada did not find significant seasonal variation [330], a recent intensive MRI study in the US reported increased disease activity, based on new lesions, in spring and summer, consistent with the previous findings [331].

Type 1 diabetes

Seasonal variation in *month of birth* in type 1 diabetes has been reported in high-incidence areas, such as the UK, the Netherlands, Sweden, Sardinia and Slovenia, with excess births in spring or summer [332-336], but a seasonal pattern was not found in subsequent analyses of other European regions [337, 338]. More recently, however, Vaiserman and colleagues showed spring excesses in type 1 diabetes births in Ukraine that were more pronounced than elsewhere in high-incidence Europe [339]. Laron and colleagues also showed spring excesses in type 1 diabetes births in the high-incidence Jewish population in Israel, but not in the low-incidence Arab population [340]. In low-incidence Japan, no seasonal variation in either month of birth or month of onset was observed [341]. However, a recent, large study in the US has supported the spring excess in type 1 diabetes births in both sexes and in three racial groups, but only in the northern latitudes (Colorado, western Washington State and southern Ohio), no birth-month effect being seen in more southern locations [342]. In the southern hemisphere, a spring or summer peak in month of birth was seen in Canterbury, NZ, consistent with the northern hemisphere spring-summer findings [343, 344], whereas Laron and colleagues reported a lack of any season-of-birth pattern in a more heterogeneous Australian (Sydney) population [343].

The *onset of diabetes disease*, however, as measured by the time of diagnosis, has been more often studied in both hemispheres and does appear to show clear seasonal variation [240, 334, 345, 346]. Several studies from as early as the 1980s report higher occurrence during the cold autumn and winter months than during the warmer spring and summer months [347], thus mirroring the spring-summer seasonal variation in birth month [192].

In Melbourne, Australia, there were more new cases in late autumn and winter also [348] and in winter in NSW, Australia [349], and NZ [344]; however, a study in Victoria

(VIC), Australia, showed no significant variation [350]. A more recent national study in Australia showed significantly higher diagnosis rates in winter, but no apparent season-of-birth effect [351]. The autumn/winter pattern of onset has, in some studies, been associated with a preceding, perhaps precipitating, infection [352] or with both cold temperature and low sunshine hours [273]. Nevertheless, as indicated by Karvonen and colleagues, many physiological parameters, such as blood glucose, blood lipids, blood pressure and body weight, and health habits such as diet and physical activity, have a seasonal pattern. The seasonality in the diagnosis of type 1 diabetes may not be related only to causal factors that trigger the disease, but also with the expression of symptoms in people who are already at an advanced stage in developing the disease [347].

In synopsis, increasing incidence of autoimmune disease, latitude gradients in disease prevalence (or incidence), changes in risk in migrant populations and seasonal clustering in month of birth, month of onset or times of disease progression all suggest the action of environmental factors in causation of disorders such as MS and type 1 diabetes. Key factors may be involved either at the initiation stage of such diseases, or in subsequent precipitation of the clinical disorder, and even in the promotion of disease once precipitated. These factors vary spatially with geographic latitude and temporally throughout the year, as well as also possibly varying over a longer period—that is, from year to year. Therefore, the likely factors include: temperature, solar radiation including visible light and UVR and, less directly, climate-associated infections, and possibly diet.

2.4 Sunlight—a candidate determinant of disease

The more sunlight, the less autoimmune disease?

In 1960, an early ecological study reported that the North American gradient in MS prevalence by latitude of birth was 'explained' largely by the decrease in winter solar radiation [353]. This early study compared as exposures several climatic variables—midwinter solar radiation, latitude, average annual 'solar radiation' (measured by number of sunshine hours), a winter-severity index, a summer-temperature index and rainfall. Midwinter solar radiation, latitude, average annual sunshine hours and winter severity were, in decreasing order of magnitude, all significantly correlated with MS

prevalence by birthplace, whereas midsummer temperature and rainfall were non-significant. While the correlation coefficient for latitude (and winter severity) was positive, the coefficient for solar radiation was negative and of strong magnitude, indicating an inverse relationship with this climatic variable. Solar radiation itself, or some other factor indirectly associated with sunlight [353], appeared to be the most beneficial in preventing this autoimmune disease. Importantly, these and other early workers [354] showed that the protective influence of sunlight appeared to be independent of latitude [355]. However, the possibility that both latitude and solar radiation are markers for some other geographically variant factor still exists [6].

Immunosuppressant effects of sunlight have been suggested to explain such findings, these acting either through UVR [5, 6], and/or vitamin D [356-358] or through melatonin suppression by the visible-light portion of the sunlight spectrum [355]. Current knowledge of mechanisms of immunosuppression by UVR and vitamin D, in particular, giving plausibility to these hypotheses, are reviewed in Sections 2.4.1 and 2.4.2.

In Australia, an ecological analysis has shown that the well-accepted latitude gradient for MS prevalence [301] is inversely associated with regional ambient UVR ($r=-0.91$, $p=0.01$). This inverse association with MS was shown to be stronger than the positive association with melanoma skin cancer in Australia, and suggests a protective effect of higher UVR in northern Australia reducing MS prevalence in those regions [9]. A more recent geospatial analysis in North America also showed a strong inverse association between UVR and MS distribution, with a concomitant increase in MS risk in low ultraviolet (UV) areas [359]. Indeed, a global meta-analysis of MS prevalence in 54 studies over the previous 10 years has shown a lack of UVR to outweigh other MS risk factors by at least 20-fold in univariate regression analyses [360]. Moreover, in Australia, the recent prospective 'Ausimmune' study on FDE not only established an incidence gradient for this early precursor to MS [305], but showed that the incidence gradient was partly accounted for by decrease in past sun exposure and vitamin D status over these regions [361].

For type 1 diabetes, fewer population studies have been conducted, but an ecological analysis in the northern hemisphere has shown that the factors temperature and latitude, although collinear with each other, together accounted for approximately 40% of the variation in type 1 diabetes risk worldwide [313]. In the southern hemisphere, no corresponding studies on latitude and type 1 diabetes had been conducted prior to this thesis study (see Chapter 4). Subsequently, however, a further ecological study of

ambient UVR (of latitude of residence) and type 1 diabetes in Australia reported equivocal results: UVR was inversely associated with incidence at low population densities but was positively associated at high densities [351].

Thus, the ecological studies indicate that factors associated with latitude, and particularly UVR and/or vitamin D, are possible influences on the development of MS and other organ-specific autoimmune diseases such as type 1 diabetes, and that analytical studies are warranted. Although prior to this thesis few studies had been undertaken in the southern hemisphere, one advantage of Australia is that, unlike the large northern continents, its population is relatively genetically homogeneous and thereby suited to individual-level study to investigate the roles of 'place' and 'time'. That is, the 'person' (or ethnic group) factor can be held more constant than in some settings, and latitude or seasonal variation then constitutes variation in only 'place' or 'time', respectively.

Therefore, sun exposure, and thus UVR exposure, is a strong contender as an environmental factor influencing organ-specific diseases such as MS in a beneficial, protective way, even though the ecological studies do not discount other confounding, or directly causal, factors such as infections and even diet. As McMichael and Hall (2001) indicated, what are needed are individual-level studies on both kinds of factors, for example, UVR and infections, and then a focus on their mechanism(s) and at what age they might act [6].

The remainder of this section will focus on the immunosuppressive mechanisms of both UVR (see Section 2.4.1) and vitamin D (see Section 2.4.2) from sunlight for biological plausibility (i.e. how these factors might work). Then, the analytical epidemiological evidence for action of either of these factors in preventing, delaying or otherwise modulating MS, type 1 diabetes and RA is considered (Section 2.5).

2.4.1 Ultraviolet radiation—mechanisms of UVR immunosuppression

The effects of UVR from sunlight may be direct, as considered in this section, or indirect via production of vitamin D (considered in Section 2.4.2). Failure or breakdown of self-tolerance (i.e. failure of tolerance of one's own potentially antigenic substances) appears to be the fundamental cause of autoimmune disease (see Chapter 1). That is, autoimmune diseases develop when self-reactive lymphocytes are activated and

normal regulatory mechanisms are somehow compromised. How might an environmental factor like UVR in sunlight contribute to the maintenance of immunological tolerance?

While carcinogenic effects of solar radiation have been known for a century, it is only within the last 30 years that the profound immunological effect of exposing the skin to sunlight has been realised. This has given rise to the relatively recent science of photoimmunology, concerned with the effects of UVR on immunological processes. Exposure to UVR in sunlight is now known to induce a suppression of immune responses that plays a critical role in skin cancer induction [362]. Supporting this concept, data associating UV-induced immune suppression and skin cancer induction have come from both skin cancer patients and immunosuppressed renal transplant patients, as well as experimental animals. From both human and animal studies, UVR is found to induce changes that trigger a cascade of inflammatory events, leading ultimately to antigen-specific, systemic (as well as local), T-cell-mediated immunosuppression [363-366]. Key components of this cascade are epidermal cytokines, which modulate the immune response to antigens and divert the response toward a state of specific immunosuppression; this means that UVR can redirect the immune response from an effector to a suppressor pathway [366]. It is also important to note again that these effects are immediate and not mediated by vitamin D.

That UVR has any effect on the immune system at all is both unexpected and remarkable, given that UV does not penetrate deeply into the skin and that human life has evolved in an environment containing UVR. However, the essential role of UVR in the production of vitamin D, necessary for bone growth, has long been known. While systemic immunosuppression by UVR could have negative implications, for example, for responses to either tumours or invading pathogens, down-regulation of key immunological processes can also have important positive benefits in preventing over-expressed autoimmunity and consequent autoimmune disease.

UVR, in the non-visible portion of the solar radiation spectrum, covers the wavelength range from 200 to 400 nanometers (nm), visible light being in the 400 to 700nm range. UVC (200 to 290nm) is mostly absorbed by the atmosphere, leaving ~10% of the UVB (290 to 320nm) and most of the UVA (320 to 400nm) to reach the ground's surface. UVB, in particular, has been associated most often with risk of skin cancer, whereas photo-ageing has been attributed to the longer-wavelength and deeper penetrating

UVA [367]. Importantly, the UVB component of the spectrum also catalyses the vital endogenous production of vitamin D.

Most work on immunological effects of UVR has been conducted on UVB. The first step in the cascade of events leading ultimately to systemic immune suppression is the transformation of the electromagnetic energy of UVB into a biological signal in the skin. This is achieved by one of three possible chromophores, urocanic acid, DNA or lipid membranes, a major initiator identified being DNA damage in epidermal Langerhans cells (LCs) [362, 368-371]. The immune suppressive signal must next be transmitted to the immune system—this appears to be achieved by a number of immunoregulatory mechanisms, discussed next.

Immunomodulatory effects of UVR

Direct UVB effects on APCs

Direct UVB irradiation of the skin alters the antigen-presenting function of LCs (i.e. epidermal DCs), making them tolerogenic rather than immunogenic (see Chapter 1, Section 1.2.2). That is, UV-irradiated LCs do not efficiently stimulate effector T cells [369, 370], which are also rendered unresponsive (i.e. anergic or tolerant, see Chapter 1, Section 1.3.1) by means of IL-10 produced by UV-irradiated keratinocytes in the epidermis and/or mast cells in the dermis. The result is a shift away from effector T-cell immune responses, and tolerance induction [362, 368, 371], meaning a decrease in inflammatory autoimmune processes.

Of note, recent studies of patients with polymorphic light eruption (a dermatological condition with clinical onset in the teenage or adult years that manifests as an eczema-like eruption on sun-exposed areas of the body) have revealed a defect in UVB-induced immunosuppression. This appears to result from the ineffective migration of LCs from the epidermis to the draining lymph nodes where antigen presentation normally takes place [372].

UVB-induced immunomodulatory mediators

As well as local tolerance induction by LCs as part of the adaptive (antigen-specific) immune system, UVB irradiation of innate-system epidermal keratinocytes or dermal mast cells induces immunoregulatory cytokines (see Figure 2.10), including IL-10,

TNF-α and IL-4 (IL-4 ultimately inducing IL-10). IL-10, in particular, is crucial to the induction of *systemic* immune suppression and tolerance (see Chapter 1, Section 1.3.1) by down-regulating IL-12 secretion (by lymphoid DCs) at a distance from the skin. The decrease in IL-12 in turn reduces both inflammatory effector T-cell immune reactions [362, 371] and DNA repair [369].

IL-10 is also secreted in large amounts by infiltrating macrophages, particularly in humans [362, 368]. These non-LC macrophages have been shown to migrate to the epidermis 72 hours post-UV exposure and are involved in the activation of immunoregulatory (suppressor) T cells via IL-10. They also fail to secrete IL-12 upon activation, both these properties contributing to UV-induced immune suppression and tolerance induction [362, 368].

Additional innate-system mediators involved in UVB-induced immunosuppression include various neurogenic peptides and neuroendocrine hormones secreted by keratinocytes and by nerve cells on UV irradiation. Together with nitric oxide, these mediate suppression by down-regulating the antigen-presenting function of LCs and inducing IL-10 production [362].

UVB-induced regulatory T cells

Kripke's pioneering work in 1974 with murine UV-induced skin cancers indicated that UV exposure at a subcarcinogenic level resulted in immunosuppression that allowed subsequent growth of UV-induced skin cancers. Most importantly, this UV-induced immunosuppression could be transferred to normal, unirradiated mice using T cells from the spleen (this was called 'infectious tolerance'), confirming the existence of specific 'suppressor' cells of the T-cell type [362, 366, 371]. After a recent resurgence of interest in these 'suppressor' cells (now termed 'regulatory T cells', see Chapter 1, Section 1.3.1), there is now evidence from both murine and human studies that more than one type of these regulatory T cells with a role in tolerance induction can be induced by UVB irradiation. These include:

- ***Tr1-like CD4+ CD25+ regulatory T cells,*** which appear to be a unique and specialised subset of skin-specific Tr1-type regulatory cells (secreting high levels of IL-10) [370] but also displaying the classic nTreg (and iTreg) cell markers CD4+ CD25+ (see Chapter 1, Section 1.3.1) [362, 369, 370].

- ***NKT cells*** (see Chapter 1, Section 1.3.1), which rapidly produce high quantities of IL-10 and IL-4 [362, 368-371] and have been shown to mediate particularly systemic UVB-induced immunosuppression. Importantly, these were the cells responsible for (adoptive) transference of this immunosuppression to unirradiated recipient (murine) hosts [98].

The mechanisms of action of UVB-induced regulatory T cells have yet to be fully understood. Besides high levels of immunosuppressive IL-4 and IL-10 in the micro-environment, the necessary 'natural death' process of apoptosis of UV-irradiated, DNA-damaged (self) 'sunburn cells' [373] (that is, keratinocytes that could otherwise eventually progress to malignancy) and then engulfment of apoptotic bodies by LCs, seems to play an important role in tolerance induction and immunosuppression, as does subsequent apoptosis of DNA-damaged LCs themselves via decreased IL-12 [370] (Figure 2.10). As indicated by Saas and colleagues, any disruption in the orderly disposal of autoantigen-containing apoptotic cells can result in autoimmunity [111] and so UV-induced suppressor cells or cytokines that inhibit IL-12 and promote apoptosis are tolerogenic.

A summary of mechanisms of direct UV-induced immune suppression is shown in Figure 2.10.

UVA radiation

As the other component of solar UV radiation reaching humans, UVA has been less studied. However, recent work has focused attention on this higher-energy portion of the solar spectrum and is showing that UVA, too, can induce systemic immunosuppression, in both humans and animals, by apparently similar immunological mechanisms to those seen for UVB [362, 371, 374].

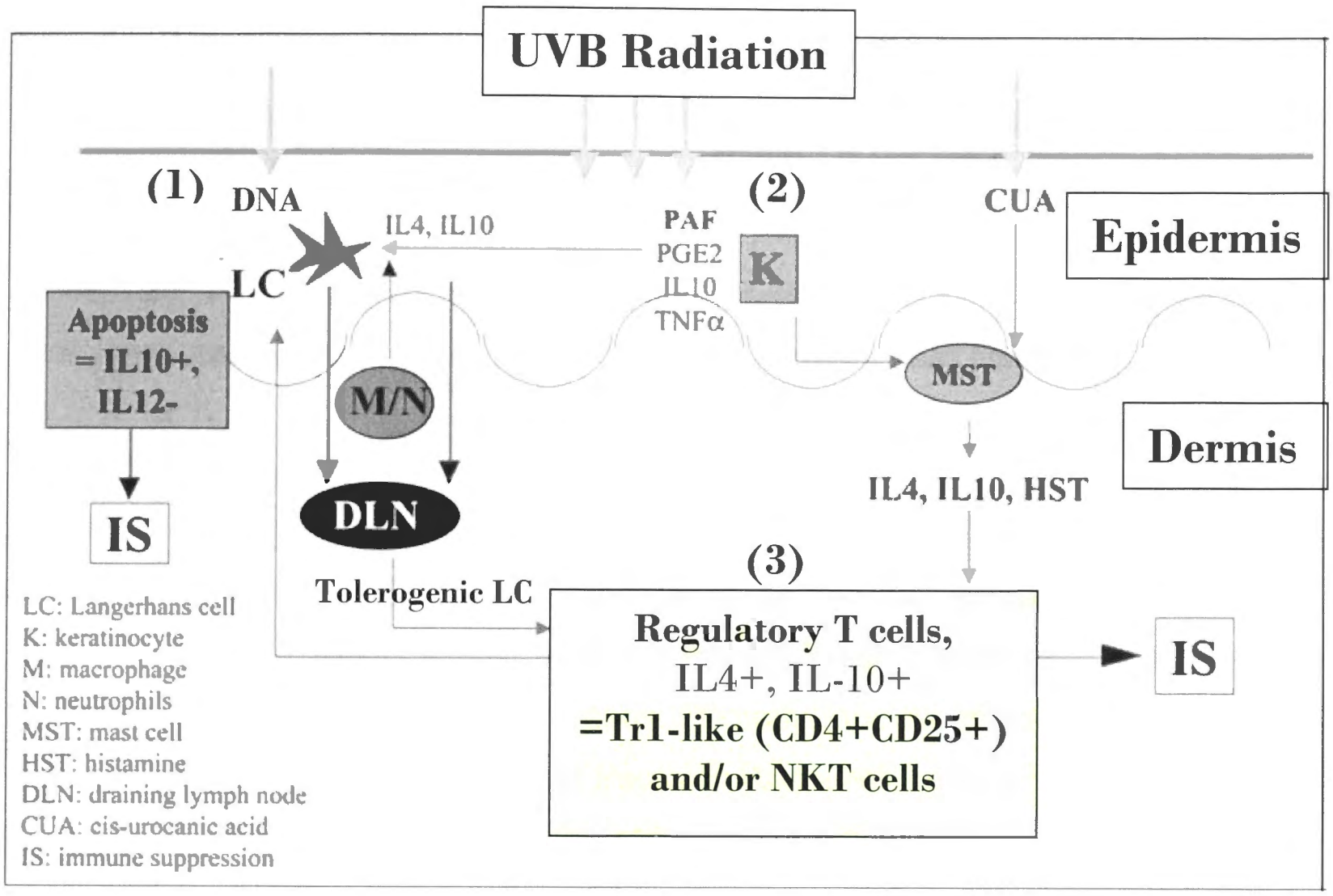


Figure 2.10: Mechanisms of UV-induced immunosuppression independent of vitamin D synthesis. (Adapted from Aubin, 2003 [368] and updated from Ullrich, 2005 [362], Schwarz, 2008 [369] and Timares et al., 2008 [370].)

1. **Epidermal LCs are the main targets of UV, which, via damage to DNA (conversion of DNA to pyrimidine dimers), inhibits the antigen-presenting activity of LCs and their capacity to stimulate allogeneic effector T cells, such as Th1 (via IL-12). Functionally UV-impaired LCs either migrate to the draining lymph node (DLN), where they fail to mature and become tolerogenic, or, if DNA damage is too great, undergo apoptosis. In human skin, the immune regulatory cytokines IL-10 and IL-4 are mainly produced by dermis macrophages and neutrophils (M/N) that infiltrate the epidermis after UV irradiation; UVB-induced infiltrating macrophages also fail to secrete IL-12, thus contributing to UV-induced immune suppression (IS) and tolerance induction, and to lack of DNA repair.**
2. **Keratinocytes (K) are also a target of UV light and they produce and release numerous soluble and immunosuppressive mediators, such as platelet-activating factor (PAF) and IL-12 inhibitory prostaglandin-E2 (PGE2). Mast cells (MST) are activated by PAF and directly by *cis*-urocanic acid (CUA, formed by UVB irradiation of epidermal *trans*-urocanic acid, a chromophore), and secrete immune regulatory cytokines; histamine (HST) can up-regulate PGE2 and IL-10 and potently suppress IL-12.**

3. **Regulatory T cells are induced by tolerogenic DNA-damaged LCs and mediate IS by releasing immunosuppressive IL-4 and IL-10; regulatory T cells also induce further apoptosis of APCs (LC) in the absence of IL-12.**

UVR in photomedicine—human autoimmune disease and UVR

Interestingly, several of the most significant findings about UV immunosuppressive mechanisms in the human field have resulted from a relatively recent subspecialty of medicine called 'photomedicine'. Beginning in the mid-1970s, this field uses different UV radiations therapeutically to treat cutaneous T-cell lymphoma, graft-versus-host disease and some autoimmune diseases. Phototherapy techniques employed include UVB irradiation, UVA irradiation, oral Psoralen+UVA irradiation (PUVA), photodynamic therapy and extracorporeal photopheresis (ECP) [373, 375]. ECP, or photochemotherapy, consists of exposing the patient's own (autologous) peripheral blood mononuclear cells, collected by apheresis, to 8-methoxypsoralen (8-MOP, a photosensitiser) photoactivated by UVA radiation, and then infusing these back into the patient.

This process can be effective in treating autoimmune diseases such as SLE, systemic sclerosis and RA, as well as various dermatological diseases and allograft (transplant) rejections where a specific decreased immune response is required. The beneficial effects of phototherapy are thought to act by inducing antigen-specific rather than general immunosuppression; UVB irradiation and ECP both can induce regulatory T cells as a possible main mechanism [369]. These two therapies also induce apoptosis of activated T cells, or of extracorporally treated mononuclear cells, respectively. It is thought that the processing and presentation of apoptotic T-cell antigens from clones of pathogenic T cells by macrophages and DCs might induce the generation of clone-specific regulatory T cells, and thus explain the resulting systemic, and specific, cell-mediated immunosuppression [370, 375].

Summary

In summary, there appear to be a number of different immunoregulatory mechanisms in humans and other mammals, for UVR at moderate, subcarcinogenic levels to effect, directly, significant systemic immunosuppression. As indicated by people with the uncommon disorder of polymorphic light eruption, this suppression appears to be a

natural protective mechanism to prevent autoimmunity to UV-induced auto-antigens from normal sun exposure. The evolutionary explanation of such UV-induced immunosuppression may be to prevent the UV-altered, DNA-damaged 'sunburn-cell' molecules being recognised as 'non-self' neo-antigens and then being over-reacted to during everyday exposure to solar radiation [362, 368, 370, 372]. That is, UVR can induce or re-establish immunological self-tolerance to skin-associated antigens and these mechanisms exist to protect the skin barrier and prevent or ameliorate autoimmune disease.

Thus, UVR in sunlight, even independent of vitamin D, is a contender, mechanistically, as an environmental factor possibly influencing the aetiology of MS and other such autoimmune diseases in a directly protective, beneficial way. *Lack* of UVR at critical stages in the life course may allow, or contribute to, the breakdown of immune self-tolerance and so may increase propensity towards autoimmune disease. Such UVR-induced immunosuppression may act through its direct immunoregulatory effects as discussed or, alternatively, indirectly through the production of vitamin D, considered next.

2.4.2 Vitamin D from sunlight—additional mechanisms of immunosuppression

There is increasing evidence that some of the beneficial effects of sunlight for the prevention of autoimmune disease are mediated through vitamin D.

Vitamin D terminology

'Vitamin D' is used here as a generic term, but the biologically active, hormonal form of vitamin D and its metabolism now need to be defined. As suggested by Vieth (2004) [376], the many terms and formulae conventions for the compounds in the vitamin D family are confusing and often misused. Following these recommendations, the terms 'active' and 'hormone' will be used only for the biologically active form, 1 alpha, 25-dihydroxyvitamin D3 ($1,25(OH)_2D_3$, or calcitriol), which is formed from the 'pre-hormone' form, 25-hydroxyvitamin D3 ($25(OH)D_3$, or calcidiol). The term 'vitamin D3' refers to the cholecalciferol molecule formed in the skin by the action of UVB on the precursor 7-dehydrocholesterol.

Vitamin D sources and production

Vitamin D3 (cholecalciferol), formed in the skin via UVB irradiation, is bioactivated in two enzyme-controlled steps: the first in the liver (and brain) to form the pre-hormone $25(OH)D_3$ and the second in the kidney (and neural tissue and macrophages) to form the active-vitamin D3 hormone $1,25(OH)_2D_3$ (here termed 'active-vitamin D').

While vitamin D3 can also be obtained from dietary sources, more than 90% is produced endogenously by conversion from 7-dehydrocholesterol in the skin, catalysed by the UVB received from sunlight [242]. Most importantly, vitamin D3 derived from sunlight exposure predominantly contributes to vitamin D3 stores (mainly in muscle and adipose tissue) and thus supplies vitamin D3 for future use; dietary sources (D3 and less active, plant-derived D2) provide a meagre, inconstant supply that is less likely to be stored [377]. These UVB-produced stores help determine the level of circulating $25(OH)D_3$ available at any time (and place) in the body, from which the active hormone is then derived.

New roles for vitamin D—effects beyond bone

Until just over two decades ago, the human vitamin D endocrine system was recognised only for its homeostatic regulation of calcium and phosphorus metabolism for bone formation and maintenance. From this classical viewpoint, a vitamin D deficit, through sunlight deprivation, was synonymous with the childhood condition of nutritional rickets that was endemic in the 19th and early 20th centuries in Europe and North America [378], this condition coinciding with the move from rural living to indoor work during the Western world 'industrial revolution' [379].

New molecular evidence in the 1980s then identified the nuclear vitamin D receptor (VDR) which was found to be present in most immune system cells, suggesting additional immunological functions for the light-sensitive vitamin D system. Molecular genetic techniques have now helped reveal that the biologically active (hormonal) form of vitamin D has widespread effects on cellular differentiation and proliferation, and can modulate not only immune responsiveness but also CNS development and function [380]. VDRs have now been found in many different organs and tissues in the body, including the brain and the beta-(islet) cells of the pancreas [381] and in rheumatoid tissues [382], suggesting vitamin D-related function in many areas of the body. The

main non-classical effects of vitamin D now appear to be anti-inflammatory, anti-infectious, immunomodulatory, anti-proliferative and as a neurotransmitter [379].

Importantly, the bioactivation of the hormonal form of vitamin D, which previously was thought to occur only in the kidney, is now known to occur in the brain and CNS [383], as well as within activated macrophages at inflammation sites [377] and possibly in several other organs [384]. Indeed, many human tissues are now known to possess the hydroxylase enzyme needed for the essential last step of bioactivation of the hormonal form of vitamin D. This extra-renal, paracrine production of vitamin D is particularly rapid and highly effective in controlling cell proliferation locally [384, 385], suggesting that the active form of vitamin D ('active-vitamin D') is important for rapid response in many different parts of the body.

Vitamin D as potent immunomodulator

Vitamin D is now being increasingly recognised as a potent immunomodulator based on seminal experimental work of research groups, such as that of Lemire and colleagues [386-394], as well as more recent work on human autoimmune disease reviewed by Arnson et al. (2007) [395], Smolders et al. (2008) [396] and Hewison (2010) [397].

Many different immunologically relevant cells have now been found to express the VDR, including monocytes and DCs, activated B and T lymphocytes, activated CD4+ (helper) and CD8+ (cytotoxic) T cells (see Chapter 1, Section 1.2.1), and T cells in RA, suggesting a number of important roles for vitamin D in the immune system. How might vitamin D, particularly the active hormonal form, 'active-vitamin D' ($1,25(OH)_2D_3$), act to suppress or regulate autoimmune processes?

Immunomodulatory effects of vitamin D

Effects on APCs and T cells

As discussed in Chapter 1 (see Section 1.2.2), DCs are highly specialised APCs critical for the initiation of CD4+ (helper) T-cell responses and are present, in different stages of maturation, in the circulation, in lymphoid organs and (as LCs) in non-lymphoid organs such as the skin (see Section 2.4.1). Human and mouse DCs, and myeloid DCs

in particular (see Chapter 1, Section 1.2.2, footnote), appear to be key targets of hormonally active $1,25(OH)_2D_3$, both *in vitro* and *in vivo* [53, 398, 399].

Earlier research suggested that active-vitamin D ($1,25(OH)_2D_3$) targeted DCs by inhibiting the production of IL-12, the DC-derived cytokine critical for Th1 cell development [393]. More recent work has shown that active-vitamin D inhibits the differentiation and maturation of DCs [399], leading to decreased production of both IL-12 (for Th1) and IL-23 (for Th17) [400, 401], and enhanced IL-10 production. These cytokine changes characterise the induction of tolerogenic DCs (see Chapter 1, Section 1.2.2), which in turn results in decreased effector T-cell activation, or T-cell anergy (hyporesponsiveness) [52, 53, 402]. Active-vitamin D ($1,25(OH)_2D_3$) also promotes myeloid DC apoptosis and thus affects all major stages of the DC life cycle: differentiation, maturation, activation and survival [399].

T lymphocytes have also been shown to be direct targets for the action of active-vitamin D ($1,25(OH)_2D_3$), which can inhibit, for example, the Th1-type cytokines such as IL-2 and IFN-γ [402] and IL-17 (Th17) (see Chapter 1), as well as affect development of Th2 cells [52]. In addition, VDR agonists (including active-vitamin D) profoundly affect the migration of effector T cells to the target organ, such as to the pancreatic islets in the type 1 diabetes mouse model. Further, in both humans and mice, active-vitamin D (together with dexamethasone) can act directly on naïve CD4+ T cells to induce differentiation *in vitro* into IL-10-producing regulatory T cells (discussed next), even in the absence of DCs [52].

Vitamin D-induced regulatory T cells

Foxp3+ (CD4+ CD25+) regulatory T cells: By acting on myeloid DCs and rendering them tolerogenic, active-vitamin D ($1,25(OH)_2D_3$) also indirectly promotes the differentiation of Foxp3+ (CD4+ CD25+) regulatory T cells (nTreg and iTreg, Chapter 1, Section 1.3.1) [52, 53]. Functionally, these vitamin D-induced regulatory T cells with suppressive properties can adoptively transfer tolerance to other naïve animals [403], as well as arrest the development of autoimmune disease *in vivo* [52]. In recent MS patient studies, the function of Foxp3+ (CD4+ CD25+) regulatory T cells in suppressing T-cell proliferation correlates with serum levels of $25(OH)D_3$ [401]; these regulatory T cells are also directly increased in numbers in MS patients by active-vitamin D [400]. Active-vitamin D can also enhance the recruitment of Foxp3+ (CD4+ CD25+)

regulatory T cells at inflammatory sites, by inducing the expression of homing receptors in these cells to aid their localisation in, for example, the epidermis [73].

Tr1-like regulatory T cells: As noted above, by direct effects on naïve T cells, Tr1-like regulatory T cells secreting anti-inflammatory IL-10 (similar, though not identical, to the Tr1 [see Chapter 1, Section 1.3.1] general type [52]) are able to be induced by active-vitamin D without APCs present, by using IL-10 as a positive autocrine factor. Upon *in vivo* transfer, these cells also protected against the development of EAE [105]. Thus, both direct and indirect (via DCs) actions of vitamin D result in the inhibition of the pro-inflammatory immune state, leading to an anti-inflammatory cytokine profile and tolerogenesis [396].

In summary, the main mechanisms by which active-vitamin D ($1,25(OH)_2D_3$) induces tolerance appear to be associated with profoundly reduced IL-12 production by myeloid DCs and by their consequent inability to induce full activation of CD4+ effector T cells, such as Th1 and Th17 (see Chapter 1), as well as by the ability of tolerogenic DCs to enhance and promote Foxp3+ regulatory T cells and Th2 cells under certain conditions [52]. As described in Chapter 1 (see Section 1.3.1), these nTreg and iTreg have a unique ability to suppress neighbouring T cells by direct cell contact, and amplify tolerance by inducing these same T cells to produce the immunosuppressive cytokines IL-10 and TGF-β. Similarly to UVB, active-vitamin D ($1,25(OH)_2D_3$) also promotes apoptotic activity, both of T lymphocytes which could otherwise be potentially dangerous autoimmune effector cells, and of DCs [399] (Figure 2.11).

In addition to being the major vitamin D targets, DCs themselves are able to synthesise $1,25(OH)_2D_3$ *in vitro*, as are macrophages and T cells; therefore, local production of $1,25(OH)_2D_3$ may contribute to regulatory T-cell induction or enhancement [52, 398].

Overall, vitamin D is able to alter the T-cell compartment into a more anti-inflammatory and regulated state, with the inhibition of Th1 and Th17 cells and promotion of regulatory T cells in particular. To summarise, the main mechanisms of vitamin D-induced immunosuppression are illustrated in Figure 2.11, in a similar way, for comparison to that for direct, non-vitamin D mediated, UV-induced mechanisms in Figure 2.10, Section 2.4.1. For this diagram, I have used information from Penna and Adorini (2000) [399], Correale et al. (2009) [400], Penna et al. (2007) [53], Smolders et al. (2008; 2009) [396, 401] and Adorini and Penna (2009) [52]; the initial production of active-vitamin D from the action of UVB from sunlight is also included in this diagram.

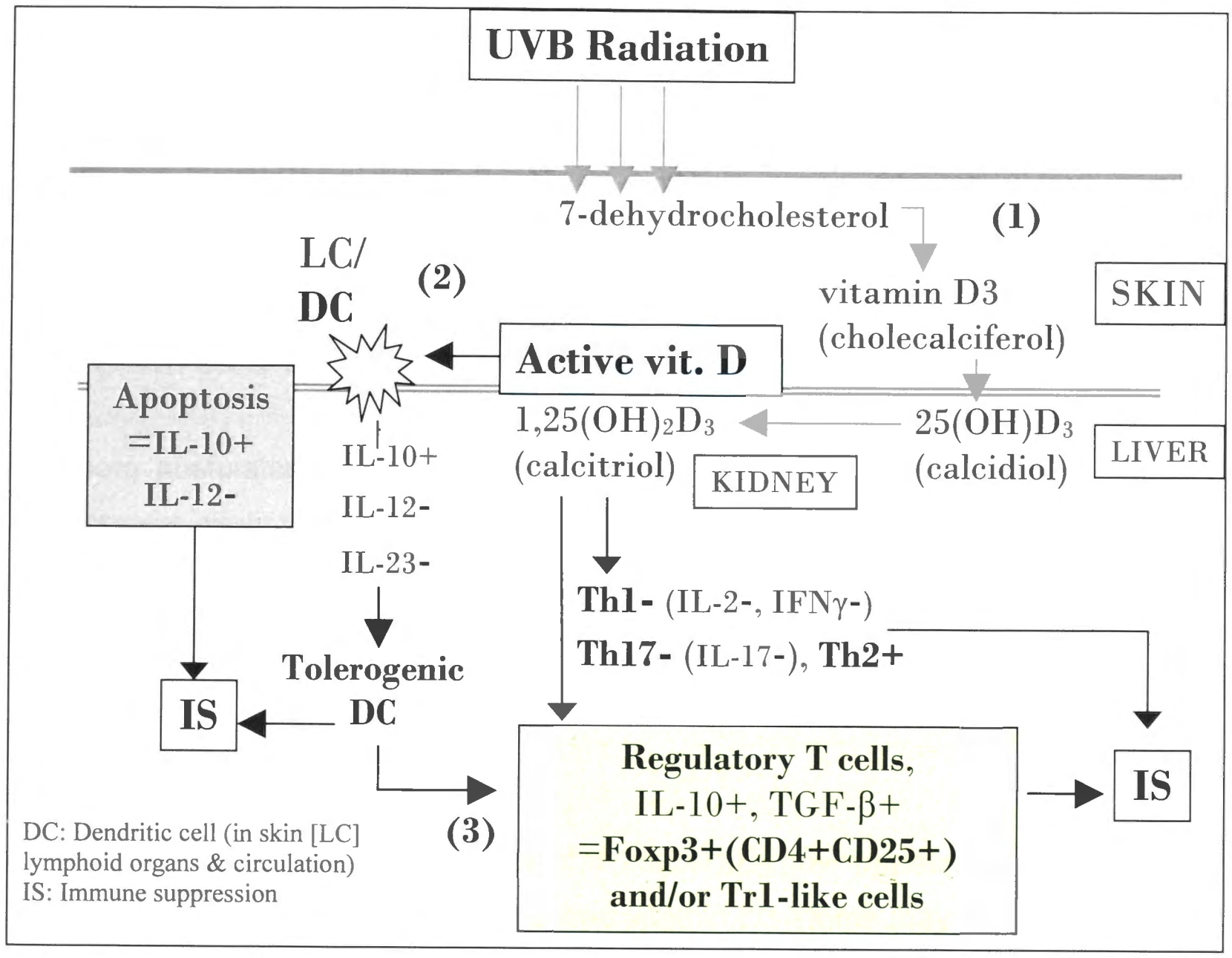


Figure 2.11: Production of active-vitamin D ($1,25(OH)_2D_3$) by UVB irradiation of the skin and mechanisms of vitamin D-induced immunosuppression. (Mechanisms summarised from Penna and Adorini, 2000 [399], Correale et al., 2009 [400], Penna et al., 2007 [53], Smolders et al., 2008; 2009 [396, 401] and Adorini and Penna, 2009 [52].)

1. **UVB from sunlight catalyses vitamin D3 (cholecalciferol) production from the precursor 7-dehydrocholesterol in the skin. Vitamin D3 bioactivation then occurs first in the liver (and brain) to form the pre-hormone $25(OH)D_3$, and second in the kidney (and other tissues) to form active-vitamin D, $1,25(OH)_2D_3$.**
2. **Antigen-presenting myeloid DCs in the circulation, lymphoid organs and skin (LC) are key targets of active-vitamin D, which inhibits differentiation and maturation of DCs leading to decreased IL-12 and IL-23, and enhanced IL-10 production. The tolerogenic DCs thus induced result in decreased T-cell activation and immunosuppression (IS); active-vitamin D also promotes DC apoptosis. Active-vitamin D further directly inhibits Th1 and Th17-type cytokines (IL-2 and IFN-γ, and IL-17), and may also affect development of Th2 cells.**

3. **Tolerogenic DCs also promote Foxp3+ (CD4+ CD25+) regulatory T cells, which suppress both by cell-to-cell contact and secretion of immunosuppressive IL-10 and TGF-beta, thus amplifying tolerance. Tr1-like regulatory cells secreting anti-inflammatory IL-10 are also directly induced by active-vitamin D from naïve T cells and further contribute to IS.**

Vitamin D status and autoimmune disease

Goldberg postulated in 1974 that insufficient UV light to support adequate vitamin D biosynthesis might be the environmental factor contributing to development of MS [356]. He also estimated that the daily amount of vitamin D that would be required to prevent MS, based on amounts of sunshine in low-prevalence areas and on the rate of formation in the skin, was 10-fold higher than the then-accepted anti-rachitic dose for dietary vitamin D3 [357].

Because highly localised production of $1,25(OH)_2D_3$ now appears to be required for immunological functions, the UV-produced vitamin D3 stores may be critical to vitamin D status [377]. The solar radiation intensity, varying with both latitude and season, determines the vitamin D3 synthesis rate. In Boston, US (latitude 42.2°N), there was found to be insufficient UVB photon energy in sunlight to initiate cutaneous production of vitamin D3, when human skin was exposed to the sun, during a four-month winter period from November to February. Further north, in Edmonton, Canada (latitude 52°N), the ineffective 'vitamin D winter' extended for six months from October to March [404, 405], and up to eight months of the year at even higher northern latitudes [406]. Recent data on the biologically effective UVB received at ground level at four widely-differing sites in the US (latitude range 18.3 to 63.7°N) support these findings and show that the *relative* production of vitamin D at higher latitudes in winter is markedly decreased. This is due to increased scattering and ozone absorption of the shorter UVB wavelengths that are the most effective in producing vitamin D [407].

Vitamin D deficiency or insufficiency, as measured by serum levels of the pre-hormone $25(OH)D_3$, has been documented in several organ-specific autoimmune disorders, including MS [122, 408, 409], type 1 diabetes [410] and RA [411, 412]. The fact that low vitamin D status is implicated in such disparate organ-specific autoimmune diseases suggests an effect of vitamin D on the common underlying autoimmune process itself, such as control of tolerance of self-antigens (see Chapter 1) [413].

For MS, for example, a recent genetic study has identified a vitamin D response element (a specific sequence of DNA in the promoter region of target genes, regulated by vitamin D) in the HLA-DRB1 (MHC class II) locus; this implies direct functional interaction between HLA-DRB1, the main susceptibility locus for MS, and vitamin D, which is thus a strong candidate for mediating the environmental effect [188]. While the exact role of this gene-environment interaction in MS disease aetiology is yet to be revealed, it is plausible that a lack of vitamin D in early childhood might reduce the expression of HLA-DRB1 in the thymus, resulting in loss of central tolerance (see Chapter 1) and perhaps increasing risk of autoimmune disease in later life [197, 214]. Thus, this interaction provides new insight into how vitamin D status may contribute to MS pathogenesis and pinpoints the MHC as the likely site of gene-environment interaction in this disorder [214].

In addition to vitamin D status data, genetic variations expressed in the vitamin D regulatory genes in these diseases add further weight to a role for vitamin D [214]—polymorphisms in the VDR genes have been correlated with increased susceptibility of MS, type 1 diabetes and RA in certain populations [413]. For example, variation in the VDR gene appears to be associated particularly with progressive forms of MS in an Australian population [414]. In type 1 diabetes, VDR gene variants linked to aetiology of disease may be influenced also by environmental UVR [415].

Another vitamin D-influencing gene is the 1α-hydroxylase enzyme gene controlling the rate-limiting step in making vitamin D active [214]; mutations in this gene cause a rare vitamin D-dependent type of rickets and this may be a risk factor also for MS [416]. In addition, a large Australian study reports an association between polymorphisms proximal to the 1α-hydroxylase gene and MS [187]. For type 1 diabetes, polymorphisms in the hydroxylase enzyme gene are associated with disease risk [417], all these studies suggesting that common inherited variation in vitamin D metabolism may affect susceptibility to type 1 diabetes and MS.

Vitamin D supplementation and autoimmune disease animal models

The disease systems in murine models of MS, type 1 diabetes and RA are all suppressed by $1,25(OH)_2D_3$ treatment *in vivo.* Both EAE and collagen-induced arthritis (CIA) development can be prevented by $1,25(OH)_2D_3$ supplementation [418-420], and injected $1,25(OH)_2D_3$ has been shown to slow the development of murine autoimmune

diabetes [421]. In these diabetic mice also, other autoimmune diseases such as EAE and CIA can be prevented by $1,25(OH)_2D_3$ [422].

Synopsis

Summing up, from these experimental and ecological studies and from the immunological mechanisms reviewed in Sections 2.4.1 to 2.4.2, UVR, both directly and indirectly through vitamin D, is a plausible candidate as a component determinant of autoimmune disease. The next section will consider analytical epidemiological evidence for possible protective effects of both UVR/sunlight and vitamin D on MS, type 1 diabetes and RA.

2.5 Analytical UVR studies and Vitamin D trials in human organ-specific autoimmune disorders

From the ecological studies reviewed, UVR directly, or mediated by vitamin D, could be important for the prevention of autoimmune disorders such as MS or type 1 diabetes at the population level. However, analytical studies at the individual level provide better epidemiological evidence for possible identification of the causal determinants of disease.

Multiple sclerosis

Table 2.3 summarises recent analytical studies of the effects of UVR/sunlight exposure, or trials of the effects or vitamin D supplementation, on the risk of MS; these are now discussed in turn.

UVR /sunlight exposure

Although perhaps better described as an ecological study, a 'case-control' study by Freedman and colleagues on the effects of residential and occupational solar radiation on MS mortality in the US, based on death certificates, showed that both residential and occupational sunlight exposure were inversely associated with MS mortality [423] (Table 2.3). The combined-effect odds ratio (OR) for outdoors- versus indoors-occupation for the high-sunlight residence group was 0.24 (95% CI 0.15, 0.38). A more

recent Swedish cohort study also found reduced risk of MS-related death with increasing occupational exposure to UV light, particularly in the highest exposure group (adjusted relative risk (RR) 0.48; 95% CI 0.28, 0.80) [424]. These results were consistent with those predicted by McMichael and Hall (1997) from their hypothesis that sunlight, and particularly UVR, were *protective* for MS [5]; these studies further suggest that MS risk may be affected by UVR exposure later than adolescence.

Another individual-level study of the inverse links between skin cancer and MS in the UK showed that skin cancer (both melanoma and non-melanoma), here being used as a proxy for sunlight exposure, was significantly less common in people with MS (rate ratio 0.49; 95% CI 0.24, 0.91). This adds to the prior analytical and ecological evidence that solar radiation may have a protective influence on the development of MS [425].

A further case-control study in TAS, Australia, linking the degree of exposure to sunlight with MS, is summarised in Table 2.3. This study by van der Mei and colleagues is the only such study on MS to date in Australia and in the southern hemisphere. The study found that higher exposure to sunlight during childhood and adolescence—that is, between the ages of six and 15 years—especially in winter, was associated with a decreased risk of MS, a dose-response relationship being evident. These authors also reported an independent dose-response relation between actinic skin damage (a measure of lifetime sun exposure not subject to recall bias) and MS [426].

A more recent North American case-control study of childhood sun exposure and MS using monozygotic twin pairs showed results consistent with these Australian data. Each of nine sun exposure-related activities conveyed protection against MS within the twin pairs, ORs ranging from 0.25 to 0.57 depending on the activity. For example, the risk of MS was substantially lower (OR 0.40; 95% CI 0.19, 0.83) for the twin who spent more time suntanning in comparison with the co-twin [427] (Table 2.3). Authors Islam and colleagues concluded that early sun avoidance seemed to precede the diagnosis of MS, and that this protective effect was independent of genetic susceptibility.

Kampman and co-workers further showed in a case-control study of Norwegians living above the Arctic Circle (where all winter, but not summer, vitamin D is provided by diet) that increased outdoor activity in summer in early life (particularly ages 16 to 20 years) reduced the odds of subsequent MS (OR 0.55; 95% CI 0.39, 0.78). This was also

achieved by cod-liver oil supplementation (in the low-activity group) and increased consumption of fish [428] (Table 2.3).

All these studies suggest that insufficient UVR, perhaps both early and later in life, may influence the development of MS. The latter study further suggests that dietary intake or supplementation of vitamin D-related foods may have similar effects and may be compensatory when sun exposure is less. A subsequent review by Kampman and Brustad (2008) suggests that latitudinal UVR and diet may interact to influence MS risk at a population level, vitamin D being the common mediator [429].

Vitamin D intake/supplementation

A large prospective cohort study of vitamin D intake has been completed by Munger and colleagues in the US, with more than 187,000 women being studied for either 10 years (Nurses' Health Study; 92,253 participants) or 20 years (Nurses' Health Study II; 95,310 participants) [430] (Table 2.3). Vitamin D intake at the study baseline was shown to be inversely correlated with subsequent MS incidence. The overall risk of MS (cohorts pooled) was 40% lower in women having taken vitamin D supplements of at least 400 IU/day in the form of multivitamins, compared with those not using supplements; sunlight was not accounted for. Dietary vitamin D levels were not associated with MS incidence. This study showed for the first time, prospectively, that sufficiently high levels of vitamin D can reduce the risk of developing MS, regardless of sunlight exposure. Interestingly, this study also suggested that high vitamin D intake was effective in lowering MS risk well into adult life.

However, because intake of vitamin D in Munger et al.'s (2004) study was largely from multivitamins, the possibility that the observed association was due to other micronutrients in the multivitamin could not be excluded [238]. An alternative approach was to use biomarkers of vitamin D status, specifically serum levels of 25(OH)D in many thousands of healthy young adults, made available from the (US military) Department of Defense Serum Repository, and determine how many people developed MS subsequent to the collection of the two serum samples. This nested case-control design enabled essentially a fully prospective study of adolescent vitamin D status prior to MS onset [238]. The results of this study (Munger et al., 2006; Table 2.3) supported the hypothesis of a protective effect of serum 25(OH)D for MS, at least in the majority white subpopulation. That is, among whites, MS risk declined with increasing levels of serum 25(OH)D—risk was 41% lower for every 50 nmol/L increase in serum 25(OH)D

(Table 2.3). Further, the reduction in risk of developing MS among individuals with high serum 25(OH)D levels was considerably stronger before the age of 20 years than at older ages [431][9]. While such 'prospective' analytical trials are not clinical trials, they do support the notion of the beneficial effects of sunlight for MS prevention. More definitive RCTs are thus needed, even though long follow-up periods would be required.

Vitamin D interventions have also been used for MS patients but few formal randomised trials of vitamin D supplementation on MS progression have yet been published. In one early patient study, 77% of MS patients exhibited vitamin D insufficiency (defined as serum 25(OH)D <50nmol/l) [408]. A small double-blind placebo-controlled intervention using 800mg calcium ± 1000 IU vitamin D daily supplementation for six months showed that vitamin D supplements increased serum levels of the anti-inflammatory cytokine TGF-β in MS patients [122]. An earlier patient study using calcium, magnesium and vitamin D as cod-liver oil (5000 IU/day) for one year showed vitamin D to more than halve the frequency of MS-related exacerbations, compared with that expected from the patients' case histories as controls [432]. Another small uncontrolled patient study suggested that vitamin D-containing fish oil together with other vitamin supplements and dietary/lifestyle advice reduced clinical severity of MS in 11 of 16 patients [433]. However, a recent randomised double-blind placebo-controlled trial in 23 MS patients in Australia failed to show any therapeutic benefit in terms of reduced MRI lesions of high-dose vitamin D supplementation (6,000 IU vitamin D2/day) compared with low-dose vitamin D (1,000 IU/day) [434]. Some eight to 10 such RCTs to determine efficacy and safety of vitamin D supplementation for MS treatment are now underway worldwide, some of these trials testing vitamin D by itself and others using vitamin D together with current IFN-β treatment for MS (see http://www.clinicaltrials.gov).

[9] Since this thesis was submitted, a nested case-control study in Sweden by Salzer et al. (Neurology 2012; 79: 2140-5) has been published showing that levels of 25(OH)D ≥ 75 nmol/L in prospectively collected blood samples were associated with a decreased risk of MS (OR 0.39; 95% CI 0.16, 0.98), supporting the protective effect of high 25(OH)D levels during the adult years preceding MS onset.

Table 2.3: Analytic observational studies on the association between UVR and/or vitamin D and MS. (Table adapted and extended from Ponsonby et al., 2005 [237] using additional data from Islam et al., 2007 [427], Kampman et al., 2007 [428] and Munger et al., 2006 [431])

Author, year	Location	Design	Population	Control or cohort n	Case n	Exposure assessment	Outcome assessment	Adjusted covariates	OR / RR	95% CI
UVR /Sunlight exposure										
Freedman et al., 2000 [423]	US (24 states)	Death certificate case-control	Deaths 1984-1995 with same residence at birth & death	Controls n=115,195	4282	1. Residential sunlight a) Low b) Med c) High	Cause of death MS v. other	Age, sex, race, SES	1.a) 1.00 b) 0.89 c) 0.53	(referent) 0.55, 0.63 0.48, 0.57
						2. Occupational sunlight a) Indoor worker b) Outdoor worker			2.a) 1.00 b) 0.75	(referent) 0.61, 0.80

Author, year	Location	Design	Population	Control or cohort n	Case n	Exposure assessment	Outcome assessment	Adjusted covariates	OR / RR	95% CI
van der Mei et al., 2003 [426]	TAS, Australia	Age- & sex-matched case-control	TAS residents with grandparent, both in TAS	Controls n=272	136	1. Higher sun exposures at ages 6-15 (average ≥2-3hrs/day, summer weekends & holidays)	MS	1. Melanin density, smoking history	1. 0.31	0.16, 0.59
						2. Higher actinic damage on the dorsum of left hand (grades 4-6 v. 3)		2. Melanin density, smoking history, sun exposure after diagnosis	2. 0.32	0.11, 0.88

Author, year	Location	Design	Population	Control or cohort n	Case n	Exposure assessment	Outcome assessment	Adjusted covariates	OR / RR	95% CI
Islam et al., 2007 [427]	North America	Monozygotic twin-paired case-control	Disease- and exposure-discordant monozygotic twin pairs in North American twin registry	Controls n=79	79	1. Outdoor activities in childhood: 9 sun exposures: a) 4 seasons b) day temper-ature (hot /cold) c) activity e.g. beach, suntanning, team sports	MS	1. Childhood infection, infectious mononucleosis, smoking, diet, age at menarche	a) 0.25 (spring) b) 0.40 (hot day) c) 0.40 (suntanning)	0.07, 0.89 0.18, 0.91 0.19, 0.83
						2. Overall sun exposure index		2. Sex, birth location, ancestry, age at diagnosis (i.e. stratification factors)	2. 0.75 (all pairs) 0.69 (females)	0.62, 0.90 0.53, 0.86

Author, year	Location	Design	Population	Control or cohort n	Case n	Exposure assessment	Outcome assessment	Adjusted covariates	OR / RR	95% CI
Kampman et al., 2007 [428]	Norway	Age-, sex- and place of birth-matched case-control	Norwegians born and living at latitudes 66-71°N (i.e. above Arctic Circle)	Controls n=402	152	1. Increased outdoor activities in summer in early life v. low activity	MS	1. Fish & cod-liver oil intake	1. 0.55 (particularly ages 16-20 yr)	0.39, 0.78
						2. Cod-liver oil supplementation		2. Fish consumption; Summer activity high v. low	2. 0.57 (in low-activity subgroup)	0.31, 1.05
						3. Consumption of fish ≥3times/week		3. Cod-liver oil intake	3. 0.55	0.33, 0.93

Author, year	Location	Design	Population	Control or cohort n	Case n	Exposure assessment	Outcome assessment	Adjusted covariates	OR / RR	95% CI
Vitamin D intake /supplementation										
Munger et al., 2004 [430]	US	Prospective cohort	Nurses' Health Study I and II	I) n=92,253 (followed 1980-2000) II) n=95,310 (1991-2001)	173	Total vitamin D intake at baseline a) Highest v. lowest quintile b) Vitamin D supplement use ≥400 IU/d v. nil	Incident MS	Age, smoker, latitude at birth	a) 0.67 b) 0.59	0.40, 1.12 (p*=0.03) 0.38, 0.91 (p*=0.006)

Author, year	Location	Design	Population	Control or cohort n	Case n	Exposure assessment	Outcome assessment	Adjusted covariates	OR / RR	95% CI
Munger et al., 2006 [431]	US	Prospective age-, sex-, race- & date of blood collection-matched nested case-control	US Military, Dept of Defense Serum Repository	Controls n=514 (from cohort followed 1992-2004)	257	Vitamin D status from two serum samples prior to first symptoms; 25(OH)D levels, within each racial/ethnic group: a) highest (>99.1 nmol/L) v. lowest (<63.3 nmol/L) quintile	MS	Latitude of residence at military entry	a) 0.38 (whites only, n=148)	0.19, 0.75 (p*=0.02)
						b) continuous, for 50 nmol/L increase in 25(OH)D			b) 0.59 (whites) (partic. serum <20 yrs) age	0.36, 0.97

p*=p for trend

Other autoimmune disorders—type 1 diabetes and RA

Tables 2.4 and 2.5 summarise recent analytical trials of the effects of vitamin D supplementation on risk of type 1 diabetes and RA, respectively.

Type 1 diabetes

Vitamin D has also been shown to be associated with a reduced risk of development of type 1 diabetes. A prospective birth cohort study of 10,366 children in Finland by Hypponen and colleagues, using vitamin D supplementation (2000 IU/day) during infancy, significantly reduced the development of type 1 diabetes evaluated 31 years later [435] (Table 2.4). Those children receiving regular doses of at least 2,000 IU/day during their first year of life had an adjusted RR of 0.22 (95% CI 0.05, 0.89) compared with those receiving less than this amount. Further, a subset of these children suspected of having rickets, and thus vitamin D deficiency, during the first year of life showed a three-fold risk (RR 3.0; 95% CI 1.0, 9.0) of developing type 1 diabetes compared with the remainder of children [435].

Two earlier case-control studies, again both in the northern hemisphere, show supporting results (Table 2.4). The EURODIAB substudy 2 was a large multicentre trial covering seven countries in Europe; the pooled estimate of diabetes risk over these countries gave an adjusted OR of 0.65 (95% CI 0.52, 0.83) for those children receiving vitamin D prophylaxis in early infancy compared with those not [436]. Interestingly, infants who received vitamin D supplements for one year or less showed a similar decrease in risk to those receiving vitamin D for more than one year, suggesting the early period after birth to be an important time for influencing diabetes risk.

The second case-control study, by Stene and co-workers in a county of Norway, surveying vitamin D-rich cod-liver oil, vitamin D supplements or multivitamins taken either during pregnancy or during the early infancy period, showed a reduced risk of type 1 diabetes in offspring when mothers took cod-liver oil during pregnancy (adjusted OR 0.36; 95% CI 0.14, 0.90). This suggests the importance of environmental risk factors *in utero* [437]. While cod-liver oil or vitamin D supplements during the first year of life were not found to be associated with diabetes risk, and nor were multivitamin supplements that were taken during pregnancy (Table 2.4), these unrecorded doses of

vitamin D may have been too low in comparison with those used in the Finnish cohort study [237].

A subsequent nationwide and much larger Norwegian case-control study by Stene and Joner [438] again investigated the effects of cod-liver oil and other vitamin D supplements both during pregnancy and during the first year of life. In this study, however, the beneficial effect of cod-liver oil taken at least five times per week during the first year of life was the main finding (adjusted OR for supplementation ≥five times per week 0.74; 95% CI 0.56, 0.99), the effects during pregnancy, both for cod-liver oil and other vitamin D supplements, being non-significant (Table 2.4).

A prospective birth cohort study by Fronczak and co-workers in Colorado, US, following 233 children identified as at risk for type 1 diabetes based on their heredity and tissue-typing screening at birth (Diabetes Autoimmunity Study in the Young [DAISY]), investigated the appearance of pancreatic islet autoimmunity as a pre-clinical type 1 diabetes outcome. This study compared maternal intake of vitamin D via food and vitamin D supplements, as well as fatty acid intake as exposure variables, all during the third trimester of pregnancy; the food frequency questionnaires were completed by the mothers shortly after delivery and prior to development of offspring disease. The study found that maternal intake of vitamin D via food during late pregnancy was significantly associated with a decreased risk of islet autoimmunity in offspring, independent of HLA genotype, family history of type 1 diabetes, presence of gestational diabetes mellitus and ethnicity (adjusted hazard ratio 0.37; 95% CI 0.17, 0.78). Conversely, vitamin D intake via supplements and fatty acid intake during pregnancy were not associated with the appearance of islet autoimmunity in offspring [439] (Table 2.4).

However, a recent Finnish prospective birth cohort study (Diabetes Prediction and Prevention Study [DIPP]) of children also at increased genetic (HLA genotype) risk of type 1 diabetes showed no association between maternal intake of vitamin D, either from food or from supplements, during pregnancy and the risk of advanced beta-cell autoimmunity or type 1 diabetes in the offspring, when adjusted for genetic risk and familial type 1 diabetes [440]. The maternal intake of vitamin D in this study was relatively low (44% receiving <5μg/day) and it is possible that only a few mothers received a potentially effective dose [440]. In contrast, an extensive Swedish birth cohort study (All Babies in southeast Sweden [ABIS]) reported that the maternal intake of vitamin D at a dose of at least 5μg/day from supplements during pregnancy

protected the infant from autoimmunity at one year of age but not at later ages [441]. The Swedish cohort was also derived from the general population rather than from at-risk children, raising the possibility that the effect of vitamin D may vary by HLA-related factors.

A further case-control study of type 1 diabetes by Tenconi and colleagues in North Italy (Table 2.4), investigating the 'administration of vitamin D' during lactation (but no doses given), found an inverse association with type 1 diabetes in the zero to 14 year age group (OR 0.31; 95% CI 0.11, 0.86). This similarly suggests the beneficial effects of vitamin D in early life [442]. In contrast, in a recent extension of the DAISY study of genetically at-risk children born in Colorado, US, there was no association of vitamin D intake during childhood (between two and 12 years of age) and subsequent islet autoimmunity or type 1 diabetes [443], suggesting a relatively early time 'window' for efficacy.

From a meta-analysis of the four case-control studies of type 1 diabetes discussed and summarised in Table 2.4 (i.e. not including the cohort studies of pre-clinical islet autoimmunity), Zipitis and Akobeng (2008) [444] concluded that children supplemented with vitamin D in infancy had a 29% reduction in the risk of developing type 1 diabetes compared with those not being supplemented. Further, although recall bias, subjective and non-quantitative measurement of vitamin D and confounding were still likely weaknesses in these case-control studies, this reduction in risk was also demonstrated in the stronger cohort study of Hypponen et al. (2001) (Table 2.4). The cohort study and one of the case-control studies [438] also showed evidence of a dose-response effect for oral vitamin D supplementation [444].

Table 2.4: Analytic observational studies on the association between vitamin D and/or related foods and type 1 diabetes. (Table adapted and extended from Ponsonby et al., 2005a [237] using additional data from Stene and Joner, 2003 [438], Fronczak et al., 2003 [439] and Tenconi et al., 2007 [442])

Author, year	Location	Design	Population	Control or cohort n	Case n	Exposure assessment	Outcome assessment	Adjusted covariates	OR / RR	95% CI
EURODIAB 1999 [436]	Seven European countries (Austria, Bulgaria, Latvia, Lithuania, Luxembourg, Romania, N. Ireland)	Multicentre case-control	Registry cases & population-based controls. Note control selection varied across centres, most commonly schools	n=2,335	820	Report of vitamin D supplementation in infancy with partial record validation	Type 1 diabetes by age 15 using validated registries	Age, sex, breastfeeding duration, maternal age, birth weight, study centre	0.65	0.52, 0.83

Author, year	Location	Design	Population	Control or cohort n	Case n	Exposure assessment	Outcome assessment	Adjusted covariates	OR / RR	95% CI
Stene et al., 2000 [437]	Vest-Agder county, Norway	Population-based case-control	Birth cohort of 1982–98, resident in county during 1998	Controls n=1,071	n=85	a) Maternal report, re cod-liver oil during pregnancy	Type 1 diabetes before age 15, on national register	Age, sex, breastfeeding duration, maternal education, other supplements	a) 0.36	0.14, 0.90
						b) Maternal report re multivitamin use during pregnancy			b)1.11	0.69, 1.77
						c) Maternal report re cod-liver oil during first year of life			c)0.82	0.47, 1.42
						d) Maternal report re vitamin D during first year of life			d)1.27	0.70, 2.31

Author, year	Location	Design	Population	Control or cohort n	Case n	Exposure assessment	Outcome assessment	Adjusted covariates	OR / RR	95% CI
Hyponnen et al., 2001 [435]	Northern Finland	Birth cohort	Live births due 1966	n=10,366 (91% of live births to one year, 86% to 1997)	81	Parental interview data at infant age 1 yr 1. Vitamin D supplementation dose a)Recommended (2000 IU) v. low (<2000 IU)	Type 1 diabetes by end of 1997 by registry, (type 2 diabetes checked for & excluded if age 20 or more at diagnosis)	Sex, gestational & maternal age, parity, maternal education, social status, standardised birth weight, infant growth rate	1a) 0.22	0.05, 0.89
						b)High (>2000 IU) v. low (<2000 IU)			b) 0.14	0.02, 1.01
						2. Vitamin D supplementation frequency				
						a) Irregular v. none			2a) 0.16	0.04, 0.74
						b) Regular v. none			b) 0.12	0.03, 0.51

Author, year	Location	Design	Population	Control or cohort n	Case n	Exposure assessment	Outcome assessment	Adjusted covariates	OR / RR	95% CI
Stene & Joner, 2003 [438]	Norway	Population-based case-control	Birth cohort of 1985–99, from national population register	Controls n=1,668	545	a) Maternal report, re cod-liver oil during pregnancy	Type 1 diabetes before age 15, on national register, diagnosed between 1997 & 2000 and born between Jan. 1985 & Dec. 1999; excluding those in pilot study (Stene et al., 2000 [437])	Age, sex, duration of breastfeeding, age of solid food introduction, maternal education, maternal age at delivery, family history of diabetes, maternal smoking, number of siblings	a) 1.00 (≥5 times per week)	0.74, 1.55
						b) Maternal report re multivitamin use during pregnancy			b) 0.98 (≥5 times per week)	0.73, 1.31
						c) Maternal report re cod-liver oil during first year of life			c) 0.74 (≥5 times per week)	0.56, 0.99
						d) Maternal report re vitamin D during first year of life			d) 0.97 (≥5 times per week)	0.73, 1.29

Author, year	Location	Design	Population	Control or cohort n	Case n	Exposure assessment	Outcome assessment	Adjusted covariates	OR / RR	95% CI
Fronczak et al., 2003 [439]	Denver, US	Prospective cohort	Birth cohort recruited from Denver metropolitan area in DAISY from Jan. 1996	n=233 children 'at risk' for type 1 diabetes on HLA screening at birth and/or having first-degree type 1 relatives, whose mothers completed FFQ within 3 months of delivery	16	Maternal report of vitamin D and polyunsaturated fatty acid intake via food and nutritional supplements during 3rd trimester, by food frequency questionnaire (FFQ) 1. Vit. D via food 2. Vit. D supplements (above or below 400 IU/day) 3. Fatty acid intake e.g. EPA & DHA from fish	Islet autoimmunity in offspring serum (=pre-clinical stage of type 1 diabetes) within average of 4-year (range 0.8-7.3 years) follow-up	HLA genotype, family history of type 1 diabetes, presence of gestational diabetes mellitus, ethnicity	HR (Hazard ratio): 1. 0.37 2. 3.09 (>400 IU) 3. 0.64 (EPA & DHA)	 0.17, 0.78 0.88, 10.83 0.24, 1.71

Author, year	Location	Design	Population	Control or cohort n	Case n	Exposure assessment	Outcome assessment	Adjusted covariates	OR / RR	95% CI
Tenconi et al., 2007 [442]	Pavia, North Italy	Age- & sex-matched case-control	Population recorded from 1988 to 2000 in registry of Pavia province	Controls (hospital) n=318	159	Maternal report of vitamin D administration during lactation	Type 1 diabetes (0-29 years of age) on validated registry between 1988 & 2000	Residency, family history of type 1 diabetes, drugs taken during pregnancy, type of delivery, bottle feeding, neonatal and common childhood viral diseases, genital surgical operations, scarlet fever, severe infections	0.31 (0–14 yr age group)	0.11, 0.86

Vit. D, Vitamin D; EPA, Eicosapentaenoic acid DHA, Docosahexaenoic acid

Table 2.5: Analytic observational studies on the association between vitamin D and RA. (Table adapted from Ponsonby et al., 2005 [237])

Author, year	Location	Design	Population	Control or cohort n	Case n	Exposure assessment	Outcome assessment	Adjusted covariates	OR / RR	95% CI
Merlino et al., 2004 [445]	Iowa, US	Prospective cohort	Women aged 55–69 yrs who did not have RA in 1986, followed up in 1992 or 1997 surveys	At-risk cohort n=29,368 (follow-up=11 years (314,181 person-years)	152	1. Baseline dietary vitamin D over past year by FFQ a)<169 IU/day b)169-289.9 IU/day c)≥290 IU/day 2. Baseline vitamin D supplements (questionnaire) a)Non-users b)<400 IU/day c)≥400 IU/day	Incident RA (validated)	Age, calorie intake, smoking, hormone replacement therapy, decaffeinated coffee intake, β—crypto-xanthin intake	1a) 1.00 b) 0.87 c) 0.72 (p*=0.16) 2a) 1.00 b) 0.65 c) 0.66 (p*=0.03)	(referent) 0.58, 1.29 0.46, 1.14 (referent) 0.36, 1.15 0.43, 1.00

p*=p for trend FFQ, food frequency questionnaire

RA

A prospective cohort study by Merlino and colleagues of 29,000 older women over a follow-up period of 11 years in Iowa, US, has shown that total vitamin D intake at baseline (sunlight not accounted for) was inversely associated with the risk of developing RA (adjusted RR 0.66; 95% CI 0.43, 1.00; for those women using vitamin D supplements of at least 400 IU/day compared with those using none) [445]. Vitamin D intake from diet alone showed no association with RA; however, these dosages were lower than in the supplements (Table 2.5). Interestingly, the beneficial effect of vitamin D supplementation appeared to be evident even in middle or older age, the cohort being made up of women aged 55 years and older at cohort entry [237].

An inverse association between serum 25(OH)D levels and disease activity was shown in RA patients compared with normal controls in Europe [446], and an inverse relationship between serum 25(OH)D levels and disease activity at baseline was seen in patients with early inflammatory polyarthritis (includes RA) in a cross-sectional UK study [447]. Both studies suggest an immunomodulatory role of vitamin D in RA. In an intervention study in RA patients, supplementation with a synthetic precursor to $1,25(OH)_2D_3$ decreased the severity of RA symptoms in a small case-control study [448]. Other patient intervention studies with variable results are reviewed by Zittermann (2003) [385] and, again, dosage of vitamin D appears to be important.

Summary

The large prospective cohort studies, in particular, summarised in Tables 2.3 to 2.5, support the hypothesis that high vitamin D intakes can reduce the risk of developing any of the three autoimmune diseases, MS, type 1 diabetes or RA [413]. These vitamin D studies in humans, together with the case-control studies of sun exposure and MS in Table 2.3, also now support mounting evidence from many animal-model studies (see Section 2.4) for UVR and vitamin D being protective for these disorders.

Additionally, it is now evident that both UVR and vitamin D are linked mechanistically with immunological self-tolerance, maintenance of which has been shown to be key to preventing autoimmune disease development. Both UVR and vitamin D are potent immunomodulators, acting to suppress inappropriate immunological responses against self-antigens. Both UVB and active-vitamin D ($1,25(OH)_2D_3$) down-regulate the action of APCs, and particularly DCs, preventing presentation of (self-)antigen to effector T

cells in both skin and lymph nodes, via cytokine mediators such as IL-10 (and reduced IL-12 and IL-23), thereby decreasing inflammatory autoimmune processes (see Sections 2.4.1 and 2.4.2). Both UVB and active-vitamin D also promote tolerogenic DCs, which then enhance the production of tolerance-crucial regulatory T cells. All of the main types of regulatory T cells shown to be involved in self-tolerance in Chapter 1 (see Section 1.3.1) are induced or enhanced by either UVR (UVB) or vitamin D. That is, UVB induces Tr1-like (CD4+ CD25+) regulatory T cells and NKT cells (producing immunosuppressive IL-10 and IL-4), both of which may up-regulate apoptosis necessary for normal disposal of auto-reactive cells (see Section 2.4.1). Immunosuppressive Foxp3+ (CD4+ CD25+) T cells and Tr1-like regulatory T cells (producing IL-10 and TGF-β) are induced by active-vitamin D, and again enhanced apoptosis appears to be an important vitamin D-induced regulatory mechanism, together with further potentiation of the highly anergic and immunosuppressive Foxp3+ (CD4+ CD25+) nTreg and iTreg cells (see Section 2.4.2).

Thus, UVR in sunlight, either directly or acting through vitamin D, appears to be a prime candidate as an extrinsic environmental factor possibly influencing the development of organ-specific autoimmune diseases such as MS. Interestingly, although vitamin D is increasingly considered to be the key environmental factor rather than UVR itself, recent work by Becklund and co-workers using the EAE murine model [449] showed that vitamin D dietary supplementation was not sufficient to clinically suppress this disorder, whereas continuous daily UVR could suppress EAE far more effectively. This suggests that UVR can suppress independently of vitamin D. This notion is also supported by the recent results for humans from the Australian MS 'Ausimmune' study, wherein both past sun exposure and vitamin D status were shown to be *independently* associated with a reduced risk of FDE of MS [361].

2.6 Timing of action of UVR/vitamin D

Several of the analytical trials summarised in Tables 2.3 to 2.5 have suggested the importance of either UVR or vitamin D acting early in life, or even *in utero*. For example, sun exposure (UVR) in childhood was associated with reduced MS risk in TAS, Australia [426], and in North America [427], but also in adolescence (16 to 20 years of age) in Norway, above the Arctic Circle [428] (Table 2.3). Serum vitamin D status prior to MS was similarly shown to be inversely associated with the risk of MS particularly in

adolescence (up to 20 years of age) but also, less strongly, after 20 years of age in US white people [431][10]. Age-related migration studies (see Section 2.3.2) further suggest major effects of environmental UVR acting prior to adulthood.

For type 1 diabetes, an autoimmune disorder with earlier onset (generally <30 years and often by 15 years of age) than MS, the beneficial effects of vitamin D intake as food, cod-liver oil or other supplementation appear to act at an early stage; that is, in infancy or the first year of life [435, 436, 438, 442] (Table 2.4) or during pregnancy [437, 439, 441]. While these individual type 1 diabetes studies varied slightly, some showing food or cod-liver oil intake to be superior to other vitamin D supplements, particularly in pregnancy [437, 439] (Table 2.4), the overall conclusion for both type 1 diabetes and MS is that UVR and/or vitamin D appear to be protective for these disorders when received, or are sufficient, relatively early in life.

These observations are consistent with increasing information on the effects of vitamin D deficiency, or insufficiency, as risk factor(s) for a variety of chronic diseases [450], early life effects in particular being important, as reviewed by Lucas and colleagues [451]. Ponsonby et al. (2010) further review this topic, concentrating on the prenatal period as possibly the most critical time for subsequent effects of low vitamin D status [452]. For example, maternal diabetes and obesity in pregnancy, known to contribute to vitamin D deficiency [214], are possibly associated with higher offspring MS risk [453]. A low vitamin D status is common in many pregnant women, even in sunny Australia, particularly in non-Caucasian populations and those not receiving supplementation (studies summarised in Table 2.6). Vitamin D deficiency during pregnancy has been similarly noted by other studies (for example, Datta et al., 2002 [454]) and in recent reviews (for example, Hollis and Wagner, 2006 [455]). In Australia, further, there is little fortification of food with vitamin D [456, 457] and several studies have recorded the high prevalence of low vitamin D status in both pregnant women and other Australian populations [457-461], even in subtropical regions [462]. Moreover, even in pregnant women previously considered to be 'not at risk' in Australia or NZ (i.e. Caucasians, not veiled or dark-skinned and in abundant sunshine areas), low vitamin D status has been recently found to be widely prevalent [463-465].

---

[10] Supporting the early-life findings, a longitudinal study by Mirzaei et al. (2011) showed that gestational vitamin D, as measured by maternal milk intake, maternal dietary vitamin D intake and predicted maternal serum 25(OH)D in mothers of nurses in the US Nurses' Health Study II, was inversely associated with subsequent MS risk in their nurse-daughter offspring [675].

Table 2.6: Prevalence of low vitamin D status in pregnancy. (Table adapted and extended from Ponsonby et al., 2010 [452] using additional data from Maghbooli et al., 2007 [469])

Year	Country	Population (all pregnant women)	Stage of pregnancy	25(OH)D definition	Prevalence % (n/N)	Reference
1997	Iran	Iranians (age 16-40 yr) attending largest Tehran hospital	Delivery at term	<25 nmol/L	80.0% (40/50)	Bassir et al., 2001 [470]
1997-2001	US	Pittsburgh (40°N) residents (>90% on prenatal vitamins)—African-American	4-21 weeks gestation	<37.5 nmol/L	44.9% (89/194)	Bodnar et al., 2007 [471]
			37-42 weeks		29.2% (54/185)	
		—White	4-21 weeks gestation	<37.5 nmol/L	2.0% (4/199)	
			37-42 weeks		5.0% (10/199)	
1999	Northern Ireland	Caucasians attending Belfast hospital (54-55°N)—no supplements	12 weeks gestation	<25 nmol/L	44.2% (34/77)	Holmes et al., 2009 [472]
			20 weeks		50.6% (39/77)	
			35 weeks		20.8% (16/77)	
		—on supplements	12 weeks gestation	<25 nmol/L	4.5% (1/22)	
			20 weeks		22.7% (2/22)	
			35 weeks		0% (0/22)	

Year	Country	Population (all pregnant women)	Stage of pregnancy	25(OH)D definition	Prevalence % (n/N)	Reference
1999–2000	Australia	Veiled and/or dark-skinned attending antenatal clinic, Melbourne	First antenatal visit	<22.5 nmol/L	80.5% (66/82)	Grover & Morley, 2001 [459]
1999–2000	UAE	Attending two hospitals in Kuwait	Delivery	<25 nmol/L	40.0% (86/214)	Molla et al., 2005 [473]
2002	Iran	Attending Tehran university hospitals	Delivery	<35 nmol/L	66.8% (380/552)	Maghbooli et al., 2007 [469]
2002	India (North)	Attending Lucknow (26.8°N) hospital	Delivery	<22.5 nmol/L	84.0% (174/207)	Sachan et al., 2005 [474]
2002–2004	The Netherlands	The Hague residents—Turkish (midwife clinic) —Moroccan —other non-Western —Western	12 weeks gestation	<25 nmol/L	83.5% (66/79) 81.2% (56/69) 59.0% (62/105) 8.0% (8/105)	van der Meer et al., 2006 [475]
2005–2006	India (North)	Barabanki district residents (26.8°N)	Second trimester	<50 nmol/L	74.1%(103/139)	Sahu et al., 2009 [476]

As discussed in Section 2.2, foetal life experience may well influence the onset of chronic disease in later life (i.e. Barker's 'foetal origins hypothesis'). Low vitamin D status during the prenatal period may thus be a major 'imprinting' influence, as hypothesised by McGrath in 2001 [221]. The mechanisms through which vitamin D might act include modulation of the immune system early in life. For example, in a German study of winter-born newborns, low vitamin D status was associated with low cord blood levels of the immunosuppressive cytokine IL-10 important in regulating immunological tolerance (see Sections 2.4.1 and 2.4.2) and preventing allergy [466].

In another study, vitamin D acted to inhibit further differentiation of naïve cord blood T cells into either Th1- or Th2-type effector T cells [467], underlining the regulatory importance of vitamin D in the early developing immune system. The importance of vitamin D availability *in utero* is further suggested by the recently shown epigenetic control of a 24-hydroxylase enzyme regulating vitamin D availability and activity specifically in the placenta [468].

Other plausible mechanisms of early life vitamin D action, particularly for MS and other CNS disorders, include neurotrophic and neuroprotective effects on CNS development [477], because this period coincides with the rapid development of the nervous system [214]. For example, vitamin D has been shown to have a potent effect on the induction of human nerve growth factor synthesis [478, 479].

The incidence of the human CNS disorder, schizophrenia, may also be influenced protectively by early life vitamin D. In a notable Finnish birth cohort study by McGrath and colleagues, regular vitamin D supplementation during the first year of life (assessed for offspring by maternal self-report) was associated with a reduced risk of schizophrenia, particularly in males (RR 0.12, 95% CI 0.02, 0.90) [480]. Many animal studies by McGrath and other collaborators further link vitamin D and the developing brain and spinal cord [481-484]; for example, maternal vitamin D depletion alters neurogenesis in the developing rat brain [485]. The specific VDR is also found throughout the CNS of the foetal rat [486], all suggesting an early role for vitamin D in the development of the mammalian CNS.

2.7 Conclusion

Ecological studies in terms of 'person', 'place' and 'time', the immunological mechanisms involved, and the analytical epidemiological studies reviewed herein all point to a role for sunlight as a possible environmental determinant of MS, and of other autoimmune disorders such as type 1 diabetes and RA. Sunlight exposure, acting through UVR directly and/or enhanced by vitamin D generation early in life, may thus be one component cause that, together with genetics, can (inversely) affect the risk of developing such organ-specific autoimmune disorders later in life.

The next chapter, Chapter 3, will consider other possible environmental factors that may additionally, or alternatively, be involved as possible determinants of autoimmune disorders, and at what stages in the life course these may act.

CHAPTER 3

INFECTIOUS AND OTHER ENVIRONMENTAL DETERMINANTS AND THEIR TIMING IN MS AND OTHER ORGAN-SPECIFIC AUTOIMMUNE DISORDERS

This chapter forms a continuation of the review of possible '*causes that can explain patterns of disease occurrence*' [1] (see Chapter 2) and now focuses on environmental factors other than sunlight for autoimmune disease. As in Chapter 2, discussion will be restricted mainly to MS, with type 1 diabetes and RA considered where appropriate for comparison. This chapter will consider possible infectious and other contributory causes of these autoimmune disorders, with a special focus on the putative timing of environmental factors during the period before disease onset.

3.1 Infections as autoimmune disease determinants—causal and/or protective?

3.1.1 Possible infectious causes of autoimmune disorders

As discussed in Chapter 2, the 'person', 'place' and 'time' characteristics of autoimmune disease epidemiology suggest the action of factors such as infectious agents—among other environmental factors such as sunlight—as possible determinants of such disease. In particular, the geographical gradient of increasing prevalence (or incidence) with increasing latitude north and south of the equator, together with corresponding differences in populations migrating from high- to low-prevalence regions (or *vice versa*) (i.e. variation in 'place') seen in both MS and type 1 diabetes, suggest environmental factors similarly varying with distance from the equator. These factors include infections—perhaps dependent on climatic factors such as temperature—and 'person' or cultural/agricultural factors such as diet. The infectious hypothesis is also strongly supported by the different temporal patterns of the disease (variation in 'time') in different geographic areas; incidence rates have changed over time in many regions but have remained stable in other areas [487]. In addition, a seasonal cycle in disease onset or other disease parameters further suggests an environmental factor, such as infections, that typically cycles from cold to warmer parts of the year, particularly in temperate regions.

The notion that autoimmune disease might be infection-determined is not new. For example, apart from the geographic and temporal patterns discussed, MS in the North Atlantic (Danish) Faroe Islands appeared to be an infectious, point-source epidemic occurring in a previously-unexposed population after occupation by British troops there in 1940 [488]. Kurtzke postulated a widespread transmissible agent that causes an asymptomatic, persistent, primary infection. That is, the 'prevalence hypothesis', wherein the disease is most common where the causal agent is most widespread [489]; then, years after the primary infection and only rarely, this agent could cause neurological symptoms of MS [8].

With regard to possible mechanisms overall, there have been two parallel hypotheses for the pathogenesis of organ-specific autoimmune disorders such as MS:

1. autoimmune, with immune injury to an essentially normal target organ (the CNS in the case of MS)
2. infectious, with chronic or re-activated latent infection provoking immune injury.

More recently, these hypotheses have fused into a third:

3. a common pathogen drives or initiates autoimmunity, whether or not the pathogen resides in the target organ [490].

Infectious agents—mechanisms of breaking self-tolerance

Section 1.3.2 in Chapter 1 introduced the concept of temporarily 'breaking' tolerance of self-antigens in order to satisfactorily respond to infection, as, for example, when infectious microbes such as viruses are held within host (self) cells. Indeed, the immune system has evolved primarily to protect ourselves from invasion by foreign organisms. So, we might ask, what could make the immune system turn on itself? What makes it break self-tolerance to the extent that self-tissues are attacked and destroyed? A popular answer, as noted by Benoist and Mathis (2001), is that autoimmune disease is a by-product of the immune response to microbial infection [491].

In support of this view, several associations between infectious agents and autoimmune disorders have been observed; for example, between β-haemolytic streptococci and rheumatic fever; B3 Coxsackie viruses and myocarditis; herpesviruses

and MS; enteroviruses or rubella and type 1 diabetes; the spirochaete *Borellia* and Lyme arthritis, to name a few [491].

How might such microbial agents incite autoimmune disease? As discussed in Chapter 1, a necessary feature of the immune system is that all individuals have potentially self-reacting T-cell lymphocytes, but that these cells remain innocuous unless activated. Multiple mechanisms to explain how viruses may trigger autoimmune disease have been investigated, including non-specific virus-induced general activation of the immune system, as well as antigen-specific immune responses to viral antigens that cross-react with self-antigens and ultimately cause auto-reactive immuno-pathologies [492, 493].

The initial *innate* immune system response to invading microbes, involving (microbe)-pattern recognition by the host's cells (see Chapter 1), triggers signalling pathways leading to cellular activation, expression of co-stimulatory molecules on APCs, activation of APC antigen-presenting capacity and their production of type 1 interferons, pro-inflammatory cytokines and chemokines, which initiate and direct the immune response against the pathogen. Thus, pathogens act as adjuvants for the immune response, while also providing the antigen source for *adaptive* T-cell and B-cell activation and effector function (Chapter 1). It is then relatively easy to imagine how auto-reactive cells normally present might trigger an aberrant destructive immune response in this highly inflammatory environment [492].

Molecular mimicry

Figure 3.1 indicates possible mechanisms of infection-induced autoimmune disease. Most evidence in animal disease models supports the concept that such disease is caused by adaptive, cross-reactive immune responses due to similarities between microbial and self-antigens, known as molecular mimicry [492, 494] (Figure 3.1a).

Molecular (or epitope) mimicry is a mechanism by which infectious agents (or other exogenous substances) may trigger an immune response against auto-antigens because their antigens are immunologically similar to the host antigens but differ sufficiently to induce an immune response when presented to T cells. Such mimicry is thought to result from the necessary flexibility of the TCR recognition system, wherein T cells can be activated by the majority of different peptides likely to be encountered from pathogens, these peptides being bound to one or more MHC (i.e. self) molecules.

That is, a side effect of this 'degeneracy' of the immune recognition system is the potential induction of reaction-to-self by the immune response to microbial antigens [492]. This 'breakdown of tolerance' to self-antigens means that the pathogen-specific immune response generated then cross-reacts with host structures to cause tissue damage and disease [493, 494].

For example, MS may be initiated by an infection by a virus sharing antigenic structures with human CNS tissue. The self-antigens involved, such as MBP or PLP, cross-react with the host's anti-viral immune response, leading to demyelination. Subsequent viral infections may then cause exacerbations, or relapses, of the disease by reactivating the immune response against viral antigens and auto-antigens. In type 1 diabetes, as another example, mimicry may be based on sequence homology between glutamate decarboxylase, an enzyme concentrated in pancreatic beta-cells, and an enzyme involved in replication of the enterovirus Coxsackie virus B [494].

As evidence for the possibility of such a mechanism in humans, molecular mimicry has been shown to trigger autoimmune disease in several experimental animal models. These include Theiler's murine encephalomyelitis virus (TMEV), engineered to express mimics of encephalitogenic myelin epitopes, a model of human MS; autoimmune demyelinating disease associated with Semliki Forest virus (SFV); and various models of type 1 diabetes (reviewed by Ercolini and Miller, 2008 [495], Munz et al., 2009 [493] and Getts and Miller, 2010 [492]) (Table 3.1).

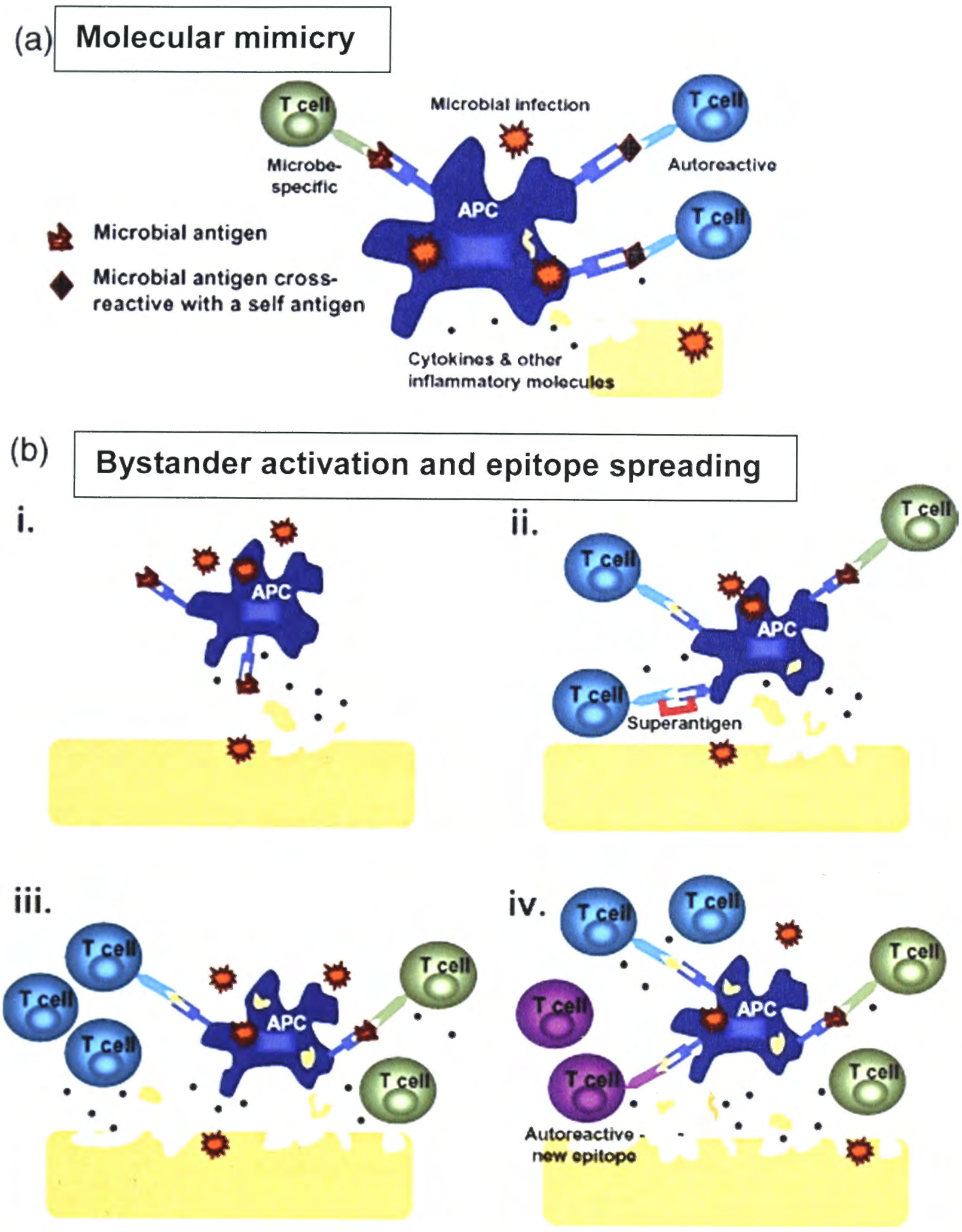


Figure 3.1: Possible mechanisms of infection-induced autoimmune disease.

(a) Auto-reactive T cells can be activated via molecular mimicry by cross-reactive recognition of a viral antigen that has similarity to self-antigen.

(b)(i) Microbial infection stimulates TLRs and other pattern-recognition receptors on APCs, leading to the production of pro-inflammatory mediators, which can lead in turn to tissue damage.

(ii) Self-antigen that is released from damaged tissue can be taken up by activated APCs, processed and presented to auto-reactive T cells (concomitant with presentation of virus antigen to virus-specific T cells) in a process known as bystander activation. Alternatively, an infection can lead to microbial superantigen-induced activation of a subset of T cells, some of which could be specific for self-antigen.

(iii) Further tissue destruction by activated T cells and inflammatory mediators causes the release of more self-antigen from tissues.
(iv) The response can then spread to involve T cells (or antibodies) specific for other self-antigens in a process known as epitope spreading. (Figure adapted from Getts and Miller, 2010 [492].)

Table 3.1: Selected murine models of infection-induced autoimmune diseases. (Table adapted from Getts and Miller, 2010 [492])

Disease modelled	Infectious agent	Mechanism(s) of disease initiation/exacerbation*
MS	TMEV	Bystander activation and epitope spreading
	TMEV expressing PLP139	Molecular identity
	TMEV expressing PLP 139 mimics	Molecular mimicry
	Lymphocytic choriomeningitis virus (LCMV) (in mice expressing LCMV protein in CNS)	Molecular identity
	SFV	Molecular mimicry
	EAE + bacterial superantigen staphylococcal enterotoxin B*	Superantigen*
Type 1 diabetes	Coxsackie B4	Bystander activation
	LCMV (in mice expressing LCMV protein in pancreas)	Molecular identity
	Pichinde virus (in mice expressing LCMV protein in pancreas)*	Molecular mimicry*
RA	CIA + murine arthritogenic mycoplasma superantigen*	Superantigen*

*In these examples, the indicated infectious agent does not cause disease but results in exacerbation of disease established by other means.

In addition, there are many experimental models of molecular *identity* (rather than mimicry) (Table 3.1), in which an exact microbial protein or epitope is expressed

transgenically in a particular tissue. Under these conditions, animals develop autoimmune disease in that tissue after infection with the protein-expressing microbe, rather than spontaneously. These approaches, although artificial and not proving causation, do indicate that T cells specific for a 'self'-antigen can become activated by infection with a microorganism containing an identical antigen, resulting in autoimmune disease [492].

There are further indications of possible mimicry cross-reactions in humans with autoimmune disorders such as MS, RA and SLE, whose self-reactive T and B cells have been shown to be at higher frequencies and activation states, these evident clonal expansions being also persistent and long term. For example, patients with MS have predominant clonal expansions of CD4+ T cells specific for the EBV nuclear antigen 1 (EBNA1), and EBNA1-specific T cells recognise myelin antigens more frequently than other auto-antigens that are not associated with MS [492, 496]. (MS and EBV are further discussed in detail in Section 3.1.3.)

Bystander activation and epitope spreading

As shown in Figure 3.1b, APCs activated during the inflammatory response to microbes can also stimulate the activation and proliferation of self-reactive T or B cells by 'bystander activation' (Figure 3.1b, i to ii); in this case, self-antigen, perhaps exposed by the initial immune response, is presented subsequent to tissue destruction. Further tissue destruction (Figure 3.1b, iii) can then compound this, and may lead to an immune response by a broader set of T cells to other epitopes of the same self-protein or even to other self-proteins, a phenomenon known as 'epitope spreading' (Figure 3.1b, iv). Epitope spreading has been shown in EAE ('mouse MS'), as well as in TMEV-induced demyelinating disease in mice (Table 3.1), and in the spontaneously arising NOD mouse model of type 1 diabetes [492]. By such antigen-non-specific mechanisms for initiating autoimmune disease, no particular microbe is implicated; however, a combination of antigen-specific (i.e. mimicry) and non-specific mechanisms can also occur, resulting in overall activation of auto-reactive T cells that expand, differentiate and ultimately become pathogenic.

Other mechanisms of breaking self-tolerance

Recent reviews have suggested that viral infection can also cause changes in normal immunoregulatory mechanisms, such as those involving regulatory T cells [497]. For

example, susceptibility to TMEV-induced demyelinating disease in mice is mediated (in a particular susceptible strain) by virus-induced activation of regulatory T cells, which interfere with virus clearance leading to persistent CNS infection and later initiation of autoimmune disease via epitope spreading. Conversely, in a mouse strain that is disease resistant, infection fails to activate these regulatory T cells and the virus is rapidly cleared [492].

Table 3.2 summarises specific viral pathogens that have been implicated in selected human autoimmune disorders and evidence for, and/or proposed mechanisms (in italics) for, their action. (MS and type 1 diabetes will be also discussed more fully in subsequent Sections 3.1.3 and 3.1.4.)

Table 3.2: Viral pathogens implicated in selected human autoimmune disorders. (Table adapted from Munz et al., 2009 [493])

Virus	Autoimmune disorder	Evidence/*proposed mechanism(s)*	Selected references
EBV	MS	• Increased risk to develop MS after primary symptomatic infection • Increased antibody responses in healthy individuals who will develop MS • Increased seroprevalence • Altered T-cell and humoral immune responses • Localisation in diseased tissue • *Molecular mimicry/bystander activation*	[498] [499] [500] [501] [502] [503] [504] [496] [505]
EBV	RA	• Higher viral loads in circulating blood cells • Altered immune responses • Localisation in diseased tissue	[506] [510] [507] [511] [508] [509]
HHV6 (Human herpes virus 6)	MS	• Localisation in diseased tissue • Clonally expanded CSF-infiltrating T cells recognise virus-encoded antigen	[512] [513]
Measles virus	MS	• Infection can result in demyelination • Higher titres of virus-specific IgG and increased frequencies of virus-specific T cells in CSF	[514] [515] [516]

Virus	Autoimmune disorder	Evidence/*proposed mechanism(s)*	Selected references
Coxsackie virus	Type 1 diabetes	• Altered immune responses • Experimental infection causes type 1 diabetes • Enterovirus positive beta-cells detected in pancreata from type 1 diabetes subjects • *Bystander activation, molecular mimicry*	[517] [137] [518] [519]
Rubella virus	Type 1 diabetes	• Tropism for pancreatic beta-cells • *Molecular mimicry*	[520] [521]
Parvovirus B19	RA	• Detection of viral DNA in synovial tissue • Phenotype of acute infection can mimic early RA	[522] [523]

3.1.2 Infections as protection from autoimmune disease, and their timing

As well as possibly being causal determinants of autoimmune disease, infections may also be protective and thus *prevent* development of these, and various allergic, disorders [524]. For example, Leibowitz and colleagues suggested as early as 1966 that the risk of MS was increased among persons who spent their childhood in a home with a high level of sanitation [525]. This concept, based on what is now known as the 'hygiene hypothesis', seeks to explain some of the more obvious epidemiological observations of autoimmune (and allergic) disorders, particularly their rising incidence over the last few decades (i.e. variation in 'time').

Hygiene hypothesis—evidence and mechanisms

The 'hygiene hypothesis' was proposed originally for allergic disorders by Strachan in 1989, who observed an inverse correlation between hay fever and the number of siblings, particularly older siblings, in the family when following a cohort of more than 17,000 British children for 23 years from their birth in 1958 [526]. Strachan reasoned that birth order and the size of the family reflected the degree of exposure to common infections within households in early childhood, and that these infections have a role in preventing allergic disease [3]. The resulting hypothesis was then extended to include autoimmune disorders such as type 1 diabetes and MS in 2002 by Bach, who proposed that the rise in allergic and autoimmune diseases in developed countries was related to the decrease in infectious diseases over the same period as a result of increased sanitation and better socio-economic conditions (Figure 3.2) [3].

Figure 3.2 illustrates this inverse relation at the population level between the incidence of various prototypical infectious diseases (panel A) and the incidence of immune disorders (panel B) over the 50-year period from 1950 to 2000. While some of the immune disorder increase may be attributable to better diagnosis or improved access to medical facilities in economically developed countries, the increase in those countries over a short period of time is less easily dismissed for readily diagnosed autoimmune diseases such as type 1 diabetes, and perhaps also for MS that can be definitively, if not easily, diagnosed [527].

The concomitant decline in infectious diseases over the same period is particularly clear for hepatitis A, childhood diarrhoea and especially parasitic infections such as helminthiasis, all of which are chronic disorders in countries with substandard health systems. In Western countries, conversely, the spread of infections was limited after the industrial revolution by public health measures such as decontamination of water, pasteurisation of milk, antibiotics and vaccination against common childhood infections [3, 527]. Thus, the hygiene hypothesis accounts for differences between developing and developed countries, where both intestinal and parasitic infections in particular have decreased, particularly among children [3].

The timing of infections and of disease onset is of fundamental importance to this hypothesis (timing discussed in more detail subsequently); for example, intestinal colonisation with gram-negative bacteria occurs later in developed, compared with

developing, countries [3], and type 1 diabetes now occurs earlier in life than in the past [527]. Indeed, type 1 diabetes has now become a serious public health problem in some European countries such as Finland, and in central and eastern Europe, where an increasing number of cases in the zero to four year age group has been reported [275, 280].

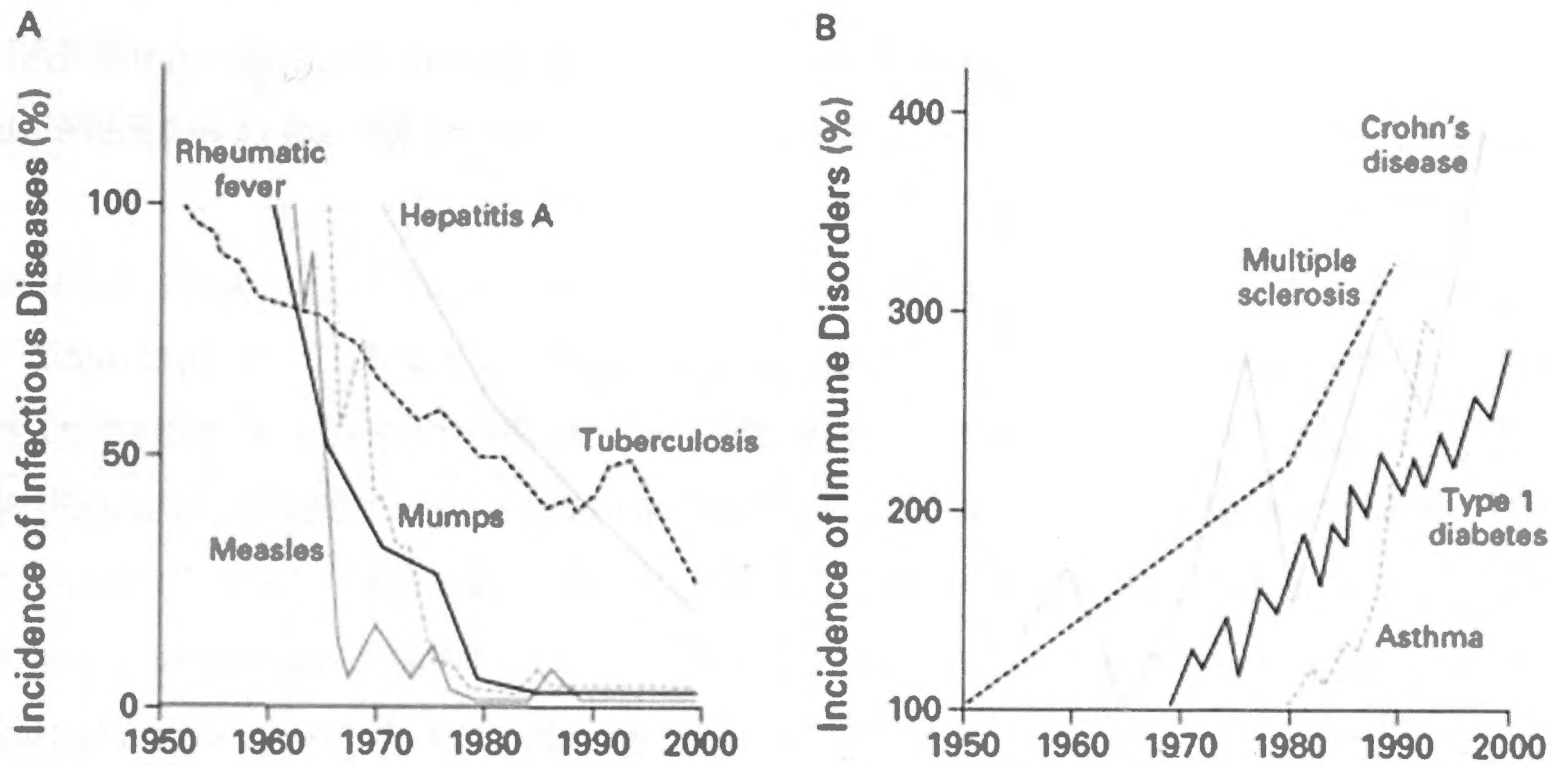


Figure 3.2: Inverse relationship between the incidence of prototypical infectious diseases (panel A) and the incidence of immune disorders (panel B) from 1950 to 2000. (Figure from Bach, 2002 [3]; data sources given in Bach, 2002 [3].)

Which epidemiological features of autoimmune disease can the hygiene hypothesis explain?

As well as variation in 'time', the global distribution (variation in 'place') of autoimmune diseases appears to mirror that of various infectious diseases, including hepatitis A, gastrointestinal infections and parasitic infections. As discussed in Chapter 2 (see Section 2.3.2), there is an overall north-south latitude gradient for autoimmune disorders in the northern hemisphere, and a south-north gradient in the southern hemisphere. There is also a west-east gradient in autoimmune disease in Europe— the incidence of type 1 diabetes in Bulgaria or Romania being lower compared with western Europe but also rapidly increasing [291]—and a west-east gradient is also evident in the US for MS [189]. Genetic differences, and/or differences in sun exposure, do not fully explain such gradients; for example, in Europe the incidence of type 1 diabetes is six-fold higher in Finland compared with the adjacent Karelian republic of

Russia, even though the genetic background is the same [528] and sun exposure also similar [527]. Sanitary conditions, and related infections, may be the discriminating factor here.

Migration studies (see Chapter 2, Section 2.3.2) have shown that offspring of immigrants acquire the same risk as the host country by the first generation for MS [304, 529] and type 1 diabetes [530]. Asian immigrants to the US show increasing risk of MS when migrating from their low-incidence countries [531], and immigrant families from Pakistan to the UK similarly show increasing frequency of type 1 diabetes [532].

Several factors may explain these differences in 'place', including confounding by SES. As discussed in Chapter 2, there is a positive correlation between gross national product and the incidence of asthma, MS and type 1 diabetes in Europe [3]. This is also found in smaller, within-country regions such as Northern Ireland, where the low incidence of type 1 diabetes is correlated with low average SES [533]. In Australia, frequency of MS was significantly higher in those who left school at an older age and who achieved a higher educational level [534]. However, SES itself may not be the factor directly responsible but may be a marker for sanitary conditions linked to early life infections. Several epidemiological studies have indicated a positive correlation between sanitary conditions and MS [525] or type 1 diabetes [533], suggesting a possible beneficial role of infections consistent with the hygiene hypothesis.

Protective infections in animal models

The most direct evidence of the protective influence of infections on autoimmune diseases comes from animal models, such as the NOD mouse model that typically develops spontaneous type 1 diabetes. These mice develop this autoimmune disorder spontaneously only if bred under sanitary, pathogen-free conditions; in 'conventional' facilities, the incidence of type 1 diabetes is very low to zero. Conversely, sterile-bred mice can be completely protected from developing type 1 diabetes by infecting them with a variety of bacteria, viruses and parasites [3]. Similarly, administration of mycobacteria (as complete Freund's adjuvant) can prevent murine EAE [535]; even bacterial or parasitic extracts, rather than living microbes or parasites, are sufficient to achieve such protection for type 1 diabetes [536, 537].

Human patient studies

A small recent MS patient cohort study in Argentina showed an inverse association between intestinal helminth parasitic infection and the rates of MS progression [538], indicating the beneficial effect of parasitic infection on this autoimmune disorder, consistent with the hygiene hypothesis. To further test this effect, RCTs of helminth immunomodulation on the clinical course of MS are now in progress or imminent, including one using a low dose of live hookworm larvae applied to the skin as 'controlled parasite exposure' and measuring immunoregulation as increase in various types of regulatory T cells (see www.clinicaltrials.gov). Similar studies on other autoimmune disorders such as Crohn's disease or ulcerative colitis have achieved encouraging results; for example, patient symptoms were markedly improved in both these conditions by deliberate administration of ova from the swine parasite *Trichuris suis* [527], supporting the hygiene hypothesis as a mechanistic concept.

Mechanisms of protection by infections

A number of possible mechanisms by which infectious agents might result in a protective effect against autoimmune disease have been proposed. Antigenic competition has been known for some time, whereby two immune responses elicited by distinct antigens at the same time tend to inhibit each other; for example, strong immune responses against antigens from infectious agents may inhibit responses to weaker antigens such as auto-antigens [539]. Recent attention has been given to competition by lymphocytes for cytokines, growth factors and for recognition for MHC/self-peptide complexes—all of which are necessary for the proliferation and differentiation of T and B lymphocytes and for the maintenance of lymphocyte homeostasis—to explain such antigenic competition [527].

Regulatory T cells, shown in Chapters 1 and 2 to be critically involved in many different areas and functions of the immune system, form another likely mechanism of maintaining immune homeostasis by means of bystander suppression [539, 540]. That is, immunosuppression can occur against antigens other than the specific antigen of the infectious agent and thus result in a lack of response by effector T cells to otherwise-antigenic self-tissues. Indeed, there are several known instances of infectious agents, such as gut-associated parasites, 'hijacking' the regulatory T-cell machinery of the human immune system and effecting general immunosuppression,

precisely in order to prevent efficient clearance of the pathogen to allow their long-term persistence in the host [541].

Which regulatory T cells might be involved in dampening the inflammatory reactions characteristic of autoimmune disease? Transfer experiments in a murine parasite model suggest the role of Foxp3+ (CD4+ CD25+) regulatory T cells [542]. Consistent with this, parasite-infected MS patients followed prospectively in an extension to the aforementioned Argentinian study [538] appeared to develop three different regulatory T-cell populations—Foxp3+ (CD4+ CD25+) regulatory T cells, IL-10 -secreting cells (Tr1) and TGF-β -secreting cells (Th3) (see Chapter 1)—as shown by decreases in these cell types three months after anti-parasite treatment [543].

Non-antigen-specific mechanisms may also be important, though they are as yet ill-defined [539]. For example, although TLRs of the innate immune system can trigger or exacerbate inflammatory autoimmune responses (as a by-product of pathogen recognition) (see Section 3.1.1), surprisingly, it has been observed that TLR stimulation can also *prevent* the onset of spontaneous autoimmune diseases such as type 1 diabetes in NOD mice [527, 544]. That is, in the gut, for example, the TLR system is required for recognition not only for pro-inflammatory effector T-cell responses against pathogens, but also for tolerance of commensal microbes leading to necessary anti-inflammatory or dampening action by regulatory T cells. In addition, TLR signalling is required for induction of oral tolerance to dietary antigens encountered in the gut. Thus, TLRs and regulatory T cells together result in normal intestinal homeostasis and also mucosal immunity, and these, both, in turn can affect long-term systemic responses including autoimmunity [540, 541].

That is, pathogens may have evolved to exploit, and even imitate, our symbiotic relationship with commensal gut flora, which we 'tolerate' by means of immunosuppressive regulatory T cells that generally prevent over-exuberant immune responses. Probiotic microorganisms, found to be beneficial in the treatment of inflammatory bowel diseases, similarly induce regulatory T-cell populations, and it now appears that parasitic helminths can also act this way. Thus, the presence of symbiotic and pathogenic microbes in the gut or other peripheral tissues could result in a supply of activated regulatory T cells (both natural and inducible, see Chapter 1) that would maintain host immune homeostasis over time, perhaps at a relatively unresponsive (towards the chronic pathogen) level. The microbes or helminths would thereby

decrease the chance of aberrant immune responses to self-antigens such as are otherwise seen in autoimmune (or allergic) disorders [541].

In summary, although underlying mechanisms of the hygiene hypothesis are, as yet, not fully understood, the key concept is that infectious agents interact with the innate and adaptive arms of the immune system as it develops and while it is susceptible to modulation, thus helping to shape the type and degree of immune responses through later life.

The importance of timing of infections

J.F. Bach noted in 2002 that when infections are associated inversely (i.e. protectively) with disease, they often occur early in childhood [3]. This concept was also noted by Poskanzer some 40 years earlier for poliomyelitis paralysis, wherein infection with the poliomyelitis virus at a late age was believed to be responsible for the development of paralysis. The 'polio' or 'late-exposure' hypothesis for MS, for example, thus proposed that MS could be caused by an infectious agent that is harmless and confers protective immunity when acquired in early childhood, but pathogenic when acquired later in life [545]. An example of this hypothesis applied to MS was the attention paid to measles virus for nearly two decades [189], for which it was observed, in case-control studies, that age at onset of measles was later in MS patients than in controls.

The 'polio/late-exposure' hypothesis evolved into the more general hygiene hypotheses of Leibowitz (1966) [525], Strachan (1989) [526] and Bach (2002) [3], where exposure to infections early in life was protective, as in the poliomyelitis model. However, there was no specific microbe identified—rather, an autoimmune (or allergic) reaction could be triggered by multiple microbes in genetically susceptible people, and risk of clinical disease increased with age at infection [490, 546]. Consistent with this notion, children who experience more infections in the first year of life have a reduced risk of type 1 diabetes [145].

A proxy measure that has been used for early exposure to childhood infections is the degree of social mixing early in life. For example, a case-control study in Yorkshire, UK, showed a correlation between frequency of day-care centre attendance during the first six months of life and type 1 diabetes [547].

Another proxy measure of early exposure to childhood infections, as noted by Strachan (1989) [526], is given by the number of siblings, and particularly older siblings, in the family, these siblings being postulated to bring common childhood infections into the home. That is, the birth-order, and number of older siblings, of the child may influence their time and degree of exposure to such infections (i.e. more older siblings—later birth order—would result in higher exposure at an earlier age, and greater beneficial stimulation of the immune system).

For example, higher (later) birth order (comparing third or later born with first born) was associated with a significant decrease in the risk of type 1 diabetes in a large cohort study in Northern Ireland, UK (adjusted RR 0.75; 95% CI 0.62, 0.90) [548]. Correspondingly, the frequency of type 1 diabetes was found to be higher in first-born children and decreased progressively (15% risk reduction per child born) with increasing birth order in a prospective family study in the UK; this could be explained by a lower exposure of first borns than siblings to infections [549]. In another population-based cohort study in Austria, there was borderline significance for lower risks of type 1 diabetes in second and later-born siblings [550]. However, in a further Northern Ireland, UK, case-control study, there was significant reduction in type 1 diabetes risk in children having three or more siblings at home per se (rather than older siblings in particular) (OR 0.58; 95% CI 0.39, 0.85) [551]. A recent meta-analysis of six cohort and 25 case-control studies found some overall evidence of a lower risk of type 1 diabetes with increasing birth order (second or later born compared with first born), but particularly in the subgroup of children aged under five years (adjusted OR 0.84; 95% CI 0.75, 0.93; n=25 studies) [552].

Variable results have also been seen for MS (discussed in more detail in Chapter 8). In an Australian case-control study in TAS, exposure to *younger* infant siblings during the first six years of life, independent of birth order, was found to be associated with reduced MS risk; nevertheless, this marker may still be indicative of early exposure to common childhood infections [553]. For MS in Canada, no general pattern was found over all family sizes in a large population-based cohort, although higher (i.e. later) birth order was associated with greater MS risk in larger families [554]. However, a large Swedish case-control study found both older and younger siblings to be protective against MS [555], while a nationwide cohort study in Denmark failed to find any association between any sibship characteristics and MS risk [556].

More work is needed on sibship structure and its relation to the hygiene hypothesis for MS. Birth order, sibling number and the hygiene/late-exposure hypothesis will be explored for MS in Australia in Chapter 8 of this thesis, where analysis of national MS survey data at the individual level will be presented.

In summary, infections as determinants of autoimmune disease can be *either* causal or protective [524], and in many cases the timing of these infections (late versus early exposure) may determine the particular outcome. Moreover, the hygiene hypothesis is not necessarily in conflict with directly causal mechanisms, because infectious or other microbial exposure during the critical stage of development of the immune system can provide non-specific protection that does not preclude specific viruses, or other infectious agents, acting to precipitate the immune-mediated destruction of the particular target organ (e.g. brain and spinal cord myelin for MS, or pancreatic islet cells for type 1 diabetes) [539].

Evidence for specific infectious agents and their associated mechanisms will now be considered in more detail for MS, before other autoimmune disorders are considered briefly in Section 3.1.4.

3.1.3 MS and infections

As discussed in Chapter 1, MS is characterised by a loss of the myelin sheath surrounding axons in the CNS. Demyelination is associated with elevated levels of CD4+ T cells specific for major myelin proteins, and the disease is generally thought to be autoimmune (see Chapter 1, Section 1.5). Although it is not known what initiates, or eventually triggers, the development of MS, it is well established that relapses or disease flares in RRMS patients are often associated with exogenous infections, particularly upper respiratory infections [495]. For example, in TAS, Australia, a population-based cohort of MS patients followed prospectively for two to three years showed a seasonal cycle in relapse rates that was positively correlated with upper respiratory tract infections ($r=0.39$, $p=0.014$) [327].

Active MS plaques are characterised by inflammatory cell infiltrates that have features consistent with an active infection: T and B lymphocytes, plasma cells and macrophages or microglia [557]. Further evidence supporting infection as a cause of MS is that the brain and CSF of more than 90% of MS patients have high

concentrations of IgG antibody, seen as oligoclonal bands. This is also evident in other CNS diseases including subacute sclerosing panencephalitis (SSPE) caused by measles virus, chronic progressive rubella panencephalitis caused by rubella virus, mumps virus meningitis and cryptococcal meningitis caused by fungal *Cryptococcus.* However, while the causative organisms that the IgG antibodies are directed against are known in these latter CNS diseases, the specific organism in MS has been difficult to identify [557], although more recent work suggests that these antibodies may be directed against nuclear antigens of EBV (discussed in more detail subsequently).

Indeed, although more than 24 viral infectious agents have been linked to MS, and several experimental rodent infectious models of demyelination exist that closely resemble different aspects of clinical MS (animal demyelination models reviewed in detail by Ercolini and Miller, 2008 [495]), no single infectious organism has been consistently found in this complex autoimmune disorder in humans. Some studies have found evidence of specific pathogens in MS plaques or CSF; for example, MS plaques were found to contain HHV6 antigens that were not found in tissues from other neurological disorders [512] (see Table 3.2), and CSF from MS patients was shown to have higher levels of (bacterial) *Chlamydia pneumoniae* DNA and antibody compared with that from patients with other neurological diseases [558]. Evidence of EBV infection has been found in brain-infiltrating B cells and plasma cells in MS cases and not in other inflammatory neurological diseases [504], although this finding has not been confirmed in subsequent studies. These three main infectious organisms linked with MS will now be discussed with regard to MS epidemiology.

Chlamydia pneumoniae

C. pneumoniae is an obligate intracellular parasite of macrophages that establishes a banal respiratory infection in a large number of normal individuals; 70 to 80% of healthy individuals are seropositive by the seventh decade, and 10% of community-acquired pneumonia is caused by this species [559]. Although evidence of this pathogen was originally found in CSF of MS patients [558] and in brain tissue [560], other studies have been inconsistent or have been unable to relate the finding to any clinical disease parameter (such as course, duration, MRI activity, disability) [8, 490, 557]. However, a nested case-control study within the American Nurses' Health Study found an association between *C. pneumoniae* infection, as measured by *C. pneumoniae*-specific IgG antibodies, and an increased risk of development of MS (OR 1.7; 95% CI 1.1, 2.7). There was a stronger association between seropositivity and progressive disease (OR

7.3; 95% CI 1.4, 37.2), perhaps suggesting a co-factor role for this infection in disease progression [561]. Thus, the role of *C. pneumoniae* in MS is unclear.

HHV6

This ubiquitous neurotropic and lymphotropic herpesvirus causes a common childhood febrile illness, infantile exanthema subitum (roseola), and rarely a mononucleosis syndrome [490]. Viral antigens occur in the brain tissue of MS patients [512], and IgM (but not IgG) antibody titres are elevated in serum [513] (see Table 3.2). However, these findings have not been confirmed by other research groups, and more recent studies with mixed findings mean that a causal association with MS has not been substantiated [490, 557], even though the evidence for HHV6 appears stronger than for *C. pneumoniae*. As HHV6 is a virus known to be latent in T cells, the detection of DNA and antibody of HHV6 in brain in MS, as well as in other neurological diseases, might reflect reactivation of the virus from latency in blood T cells trafficking through the brains of patients with inflammatory CNS disease [557]. That is, the observed changes may be a consequence of such disease, rather than its cause [8].

Moreover, HHV6 contributes little to explain aspects of MS epidemiology, particularly the age-at-migration data (see Chapter 2, Section 2.3.2) [8]. As HHV6 infects virtually all children by the age of two years, it is not only difficult to compare MS risk for infected versus non-infected individuals, but this early age of seroconversion does not explain why the critical age for migration differences, particularly from high- to low-incidence regions, is generally agreed to be up to the age of 15 years [306, 312].

Epstein-Barr virus

EBV is a ubiquitous lymphotropic herpesvirus that maintains a nonlytic, latent infection of lymphoid B cells, and causes infectious mononucleosis (IM) [490], particularly when infection occurs in adolescence or adulthood [8]. EBV has been a leading candidate trigger for several autoimmune diseases since the initial finding of raised EBV-specific antibody titres in SLE patients in 1971 [562] (Table 3.2). Like HHV6, EBV fulfils the prediction that agents inducing or exacerbating autoimmune disease are most likely to be ubiquitous pathogens of high prevalence in the population [21, 199]. EBV is further biologically plausible because it establishes a lifelong dormant infection (with continuous virus production due to reactivation) and, importantly, modulates the immune system (by rescuing infected B cells via latent antigen expression and

assisting their differentiation into memory B cells in which the virus persists). In addition, EBV continuously stimulates strong T-cell responses via chronic antigen presence (this immune control being crucial to prevent EBV-associated malignancies, such as Burkitt's lymphoma and naso-pharyngeal carcinoma) [562].

EBV sero-epidemiology

In developing countries, almost all children are infected with EBV in the first years of life, and prevalence of seropositivity is typically higher than 90% at the age of four years. In most developed countries, EBV infection is delayed in many children until adolescence, when the infection is more likely to result in IM; prevalence of EBV seropositivity here also follows a latitude gradient parallel to that of MS [238].

Large prospective sero-epidemiological studies consistently showed that MS patients were almost universally seropositive for EBV compared with healthy age-matched controls [499, 500, 563], suggesting that infection with EBV is required for the development of MS. Such longitudinal studies of serum samples collected before the onset of disease in healthy adult populations further showed that MS risk increased significantly with elevated EBV-antibody titres more than 10 years *before* the onset of symptoms [500, 563, 564], particularly for EBNA1[11]-specific IgG, which also exhibits a dose-dependent relationship with MS risk. Moreover, these EBNA1-specific antibodies have been found to interact with the major MS-susceptibility gene locus HLA-DRB1*15 (see Chapter 2) in determining MS risk [565].

Importantly, a recent study by Levin and colleagues prospectively followed a subsample (10 cases) of individuals who were initially seronegative for EBV (i.e. uninfected) and who later became infected with EBV (i.e. seropositive) and still later developed MS. These cases were selected from more than eight million military personnel whose serum had been stored in the US Department of Defense Serum Repository, a total of 305 of whom developed MS, including some individuals who were seropositive at the outset. *All* of the 10 EBV-negative MS cases became EBV-positive before MS onset, whereas *none* of those individuals among the matched controls who remained seronegative throughout the study developed MS; in addition, only 35.7% of the controls seroconverted (p=0.0008 for difference in seroconversion rates between

[11] EBNA1 is the dominant latency-associated antigen of EBV that is consistently expressed in healthy virus carriers and is crucial for viral persistence. It represents a key target antigen for CD4+ T-cell mediated immune control mechanisms of EBV infection in healthy individuals [562].

cases and controls) [566]. The mean time interval between primary EBV infection and MS onset was estimated to be 5.6 years (range 2.3 to 9.4 years). This study shows that MS risk is extremely low among individuals not infected with EBV, but that it increases markedly in the same individuals following EBV infection—that is, to a similar rate as that manifested by individuals who were already EBV-positive at the study baseline [566]. Thus, primary infection with EBV was proposed to significantly increase MS risk.

IM

Recent epidemiological studies have shown an even stronger positive association between a history of IM and occurrence of MS, further suggesting a causal link with EBV. For example, a large Scandinavian cohort study of IM patients followed for occurrence of MS showed a more than two-fold risk of developing MS compared to subjects who acquired the virus asymptomatically (standardised incidence ratio 2.27; 95% CI 1.87, 2.75); this increased risk also persisted for at least 30 years after the infection [503]. A similar increase in MS risk in those with an IM history was shown in a case-control study in TAS, Australia (RR 2.01; 95% CI 1.11, 3.62) [553], and in a meta-analysis of 14 cohort and case-control studies, including the latter Australian study and others from Europe and the US (RR 2.3; 95% CI 1.7-3.0; $p<10^{-8}$) [501]. Remarkably consistent with these findings, the most recent meta-analysis of 18 individual-level case-control and cohort studies (including a recent large Canadian MS cohort with an OR of 2.06 [95% CI 1.71, 2.48] for history of IM [567]) showed an overall RR of MS of 2.17 (95% CI 1.97, 2.39), firmly establishing that IM is a significant risk factor for MS [568].

IM may also modify the MS risk associated with the MS-susceptibility gene locus HLA-DRB1*15, a 2.4-fold (95% CI 2.0, 3.0) increased MS risk associated with this gene locus in IM-negative individuals becoming a seven-fold (95% CI 3.3, 15.4) increased risk in IM-positive people [569].

Possible mechanisms for EBV infection in MS pathogenesis

Figure 3.3 illustrates some potential mechanisms that may underlie the association of EBV infection with MS, none of which depend on direct infection of the CNS. The first hypothesis is that the immune response to EBV infection in genetically susceptible individuals cross-reacts with myelin antigens, and this response could be by T cells and/or antibodies [8, 562]. That is, there is some evidence that EBV could break

immune tolerance to myelin antigens through molecular mimicry [570] (Figure 3.3 (1)). For example, CD4+ T cells specific to EBNA1 are increased in frequency and recognise a broader range of epitopes in individuals with MS than in healthy controls; two EBV peptides, one of which was from EBNA1, have also been recognised as targets of the immune response in the CSF of MS patients. Further, EBNA1-specific cell lines cross-reacting with myelin antigens have been isolated from MS patients [238, 571].

Alternatively, EBV infection may assist in the maintenance and increased survival of auto-reactive B cells (which would otherwise be neutralised or controlled during establishment of peripheral tolerance, see Chapter 1, Section 1.2) as shown in Figure 3.3 (2). Otherwise, EBV may transactivate the expression of human endogenous retroviruses (HERV), which are cytotoxic for oligodendrocytes (cells responsible for myelinating the CNS), as in Figure 3.3 (3). Another possibility is that the altered immune responses to EBV, seen as a persistently dysregulated EBV infection [570], are an 'epiphenomenon' resulting from host factors that predispose for autoimmune disease but are not directly involved in MS pathogenesis [562], as in Figure 3.3 (4).

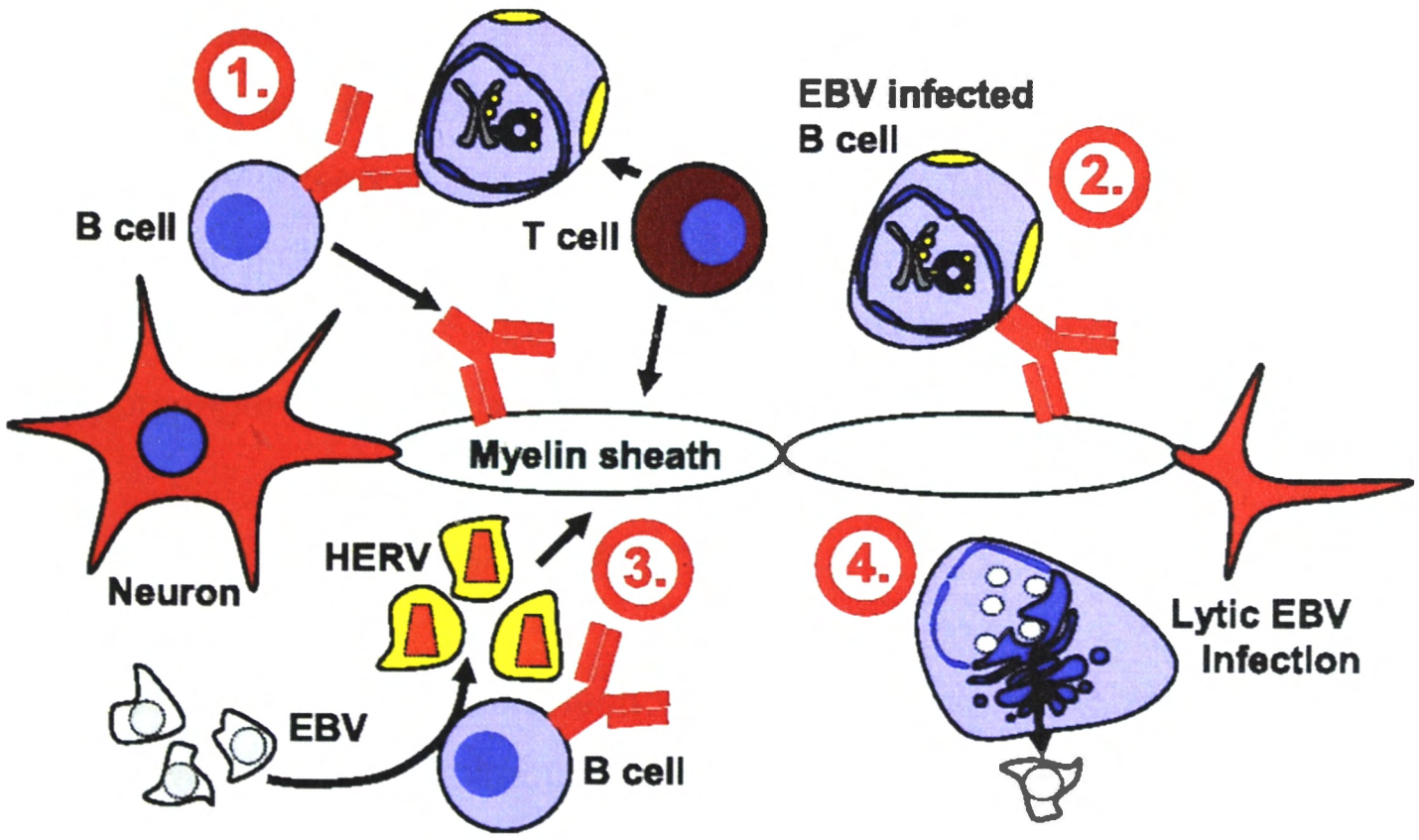


Figure 3.3: Potential mechanisms responsible for the association of EBV infection with MS.

(1) EBV-specific T cells or antibodies could cross-react with auto-antigens expressed in the CNS and attack the myelin sheath of axons.

(2) Latent EBV antigens could sustain the survival of auto-reactive B cells.

(3) EBV infection transactivates retroviral elements such as HERV, which in turn mediate cell death of oligodendrocytes.
(4) Auto-reactive B-cell activation could initiate EBV replication and in turn augment EBV-specific T- and B-cell responses. (Figure from Lunemann et al., 2007 [562].)

Other proposed mechanisms for which there is some evidence include the activation of superantigens (see Section 3.1.1), or an increased expression of alpha-B-crystallin heat-shock protein in lymphoid cells post-infection [8] that provokes a CD4+ T-cell response attacking alpha-B-crystallin in oligodendrocytes in a 'mistaken self' scenario resulting in demyelination [572]. The predominant mechanisms may even vary across different patients, and may reflect the observed heterogeneity of MS disease in the affected population (see Chapter 1, Section 1.5).

In synopsis, although many associations between EBV infection and MS have been observed (level of evidence reviewed and summarised by Lucas et al., 2011 [572]) and possible mechanisms for these associations elucidated, a direct causal role of EBV in the pathogenesis of MS is still debated [8, 562]. However, EBV may act as a 'primer' or initiator of the pathological process (i.e. *necessary but not sufficient*, see Chapter 2, Section 2.2), and the observation that MS tends to occur several years after IM [573] or primary EBV infection [566] suggests that other factors may be needed to eventually trigger clinical disease [8]. Further, as will now be discussed in the next subsection, the *timing* of primary EBV infection may be the most important key to action of this infectious agent for this complex autoimmune disorder, MS.

EBV, timing and the hygiene hypothesis

As discussed earlier in this section, the hygiene/late-exposure hypothesis posits that MS, as with other autoimmune disorders, is an autoimmune reaction that can be triggered by multiple microorganisms in genetically susceptible individuals, and risk of clinical disease increases with age at infection. As reviewed by Ascherio and Munger (2010), the hypothesis in this general form could explain several features of MS epidemiology (see Chapter 2, Section 2.3), including:

- the latitude gradient
- the protection from MS of individuals migrating from low-risk to high-risk areas
- the higher MS rates among individuals of higher SES

- (possibly) the attenuation of the latitude gradient within the US, if this can be explained by improved conditions in hygiene in the south and thus a lower incidence of childhood infections [571].

Also supporting this hypothesis is:

- a trend towards later age at infection with childhood viruses in MS cases compared with controls [574]
- lower MS risk among individuals exposed to older and/or younger siblings in early childhood [553, 555].

However, if EBV is the causative infectious organism for MS, the hygiene hypothesis is limited, in this general form, in explaining the substantial differences in MS risk between EBV-negative individuals, EBV-seropositive (but asymptomatic) individuals, and those who become EBV-positive via clinical IM.

As discussed earlier in this section, one of the most striking and consistent observations is that MS is extremely rare among EBV-negative individuals, and this finding is confirmed and emphasised in recent studies of paediatric MS [575, 576]. However, if EBV infection occurs *early* in childhood, as happens in most populations but particularly in developing countries, the risk of MS is about 10-fold higher compared with EBV-negative individuals [238]. Yet, if EBV infection is delayed (hypothetically by higher levels of hygiene in early childhood) and infection then occurs during adolescence and young adulthood, the infection is more likely to manifest as IM. The risk of subsequent MS is then at least 20-fold higher than for EBV-negative individuals (who ostensibly share the same high-hygiene early life environment and should have *high*, rather than negligible, MS risk). That is, there is a two- to three-fold difference in MS risk between the two classes of EBV-positive individuals (compared with EBV-negative people) depending on whether EBV infection occurs *early* in childhood (generally asymptomatically) or *later* in adolescence or adulthood as clinical IM [571] (Figure 3.4).

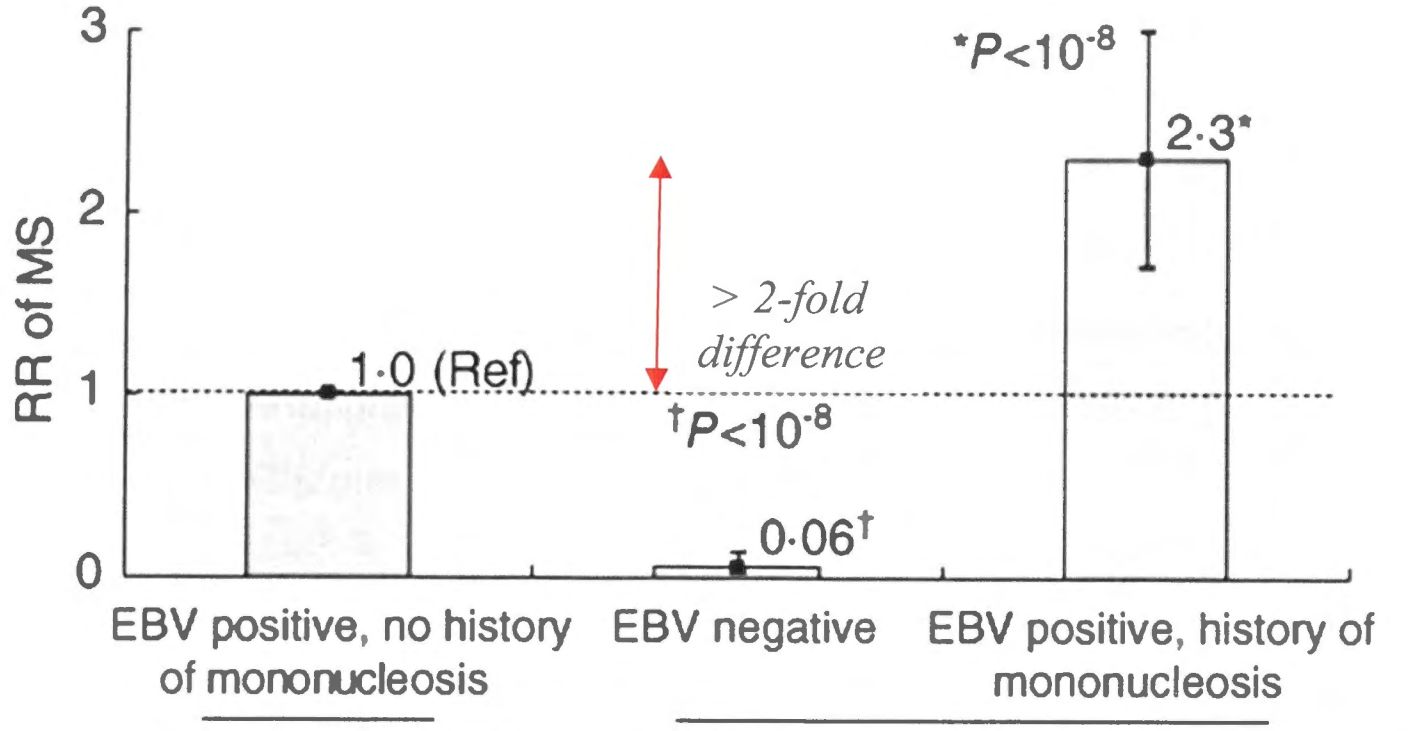


Figure 3.4: The relative risk of developing MS according to EBV infection and history of IM. Bars represent the 95% confidence intervals of the RR estimates. (Figure adapted from Ascherio and Munger, 2010 [571].)

According to the hygiene hypothesis in its strictest sense, individuals *not* infected with EBV should have *high* MS risk, because EBV infection in early childhood is strongly associated with *non*-hygienic conditions. That is, most children in developing countries are EBV-seropositive at four to six years of age, whereas in most developed countries more than half may be still uninfected at these ages [577]. Thus, if EBV is to be a determining infectious agent of MS, the *timing* of primary EBV infection appears to be key to the MS-risk patterns in Figure 3.4 and needs to be incorporated into any model of the hygiene hypothesis for MS [307].

Which features of MS epidemiology are consistent with the 'EBV variant of the hygiene hypothesis' [8]?

As reviewers Ascherio and Munger indicate, a role of EBV in MS causation could explain the following remarkable similarities between the epidemiology of MS and IM:

- age of peak incidence
- latitude gradient
- rarity in populations in which EBV infections occur early in life (including Japan and most of Asia)
- earlier age at peak onset in women than men

- lower incidence in blacks, Asians and Eskimos than in whites, and positive correlation with SES [8, 571].

In addition:

- early age at EBV infection could explain the retention of low MS risk among people migrating from areas of low MS prevalence to areas of high prevalence (see Chapter 2, Section 2.3.2).

However, migration in the reverse direction—from high to low MS prevalence areas—and the associated *reduction* in risk, cannot be similarly explained by EBV infection; nor can the possible occurrence of an MS 'epidemic' in the Faroe Islands [238]. Other factors may therefore be involved, including latitude-related factors such as vitamin D (see Chapter 2), or infection with other microbes, such as other herpesviruses, that may modify the effect of EBV infection. Or, there may be different EBV strains involved in different regions, some more likely to increase MS risk than others [571].

Summing up, there is some supportive, but not conclusive, evidence for the role of a generally hygienic environment (i.e. no particular infectious agents) in early life and the development of MS [307]. There is also compelling evidence for EBV infection being a strong risk factor for MS [238]. Although the mechanisms by which EBV infection increases MS risk are still unclear, some combination of the hygiene hypothesis and an adverse EBV infection may be important, wherein insufficient early life infections are associated with an adverse characteristic of EBV infection, such as later age of infection, which may lead to a dysregulated host immune response to EBV [307]. Therefore, more work on the effect of early childhood infections on MS is needed—particularly in the southern hemisphere where such work is sparse—to gain further evidence of the possible protective role of these infections in development of MS.

Birth order and sibship structure of MS cases at the individual level in Australia will be considered in detail in Chapter 8.

Other MS determinants

To complete this review of the pathogenesis of MS, the influence of other candidate environmental factors, and their possible interactive role with factors already detailed, will be briefly considered here.

Smoking

There is strong and growing evidence that cigarette smoking is an important risk factor for MS [236, 238, 578-580], even though the smoking-MS link is independent of latitude and ancestry and does not contribute directly to explaining the geographical variation in MS incidence [189]. A recent meta-analysis of 14 studies shows that smokers have about a 50% higher risk of developing MS than non-smokers (RR 1.5; 95% CI 1.3, 1.7) [581], and evidence suggests that smoking may adversely affect MS disease progression particularly [238, 582, 583].

Several mechanisms have been proposed for these effects, including a systemic effect—indicated by an increased risk for smokers also for other autoimmune diseases such as RA, SLE and Grave's disease—and direct effects on the BBB and the CNS, perhaps by nitric oxide. Among the plethora of possible mechanisms (reviewed in detail by Shirani and Tremlett, 2010 [580]), the established fact that smoking also increases the frequency and persistence of respiratory infections is notable because of the link between such infections and the increased frequency and damaging effects of MS relapses [580]. Further evidence of possible interaction between MS-linked infectious agents and smoking includes higher levels of specific antibodies to *Chlamydia pneumoniae* [584] and EBV [503] in smokers than in non-smokers.

Interestingly, with regard to the timing of action of this factor, passive smoking was shown to be associated with a significantly higher risk of childhood MS [585], but maternal smoking during pregnancy did not appear to be associated with higher MS risk in the offspring [586].

In summary, the main candidate determinants of MS—UVR/vitamin D, EBV infection and smoking—are listed in Table 3.3 with other currently considered candidate factors, together with their summarised epidemiological evidence and putative mechanisms [289].

Table 3.3: Summary of potential environmental risk factors for MS. (Table adapted from Handel et al., 2010 [289])

Environmental factor	Source of evidence[+]	Putative mechanism[+]
UVR/Vitamin D	Epidemiological case-control/cohort studies assessing the association of MS risk with disease geography, sun exposure, outdoor occupation or vitamin D levels **(Chapter 2, Table 2.3)**	Levels of vitamin D experienced *in utero* can have long-lasting effects on the development of numerous organ systems, including the CNS; during life, vitamin D has clear immunomodulatory functions **(Chapter 2)**
Infection: EBV	Epidemiological case-control studies assessing the association of MS risk with IM or antibody titres **(Chapter 3)**	Clonal expansion of B lymphocytes in the CNS or EBV infection triggers autoimmunity via molecular mimicry **(Chapter 3)**
Smoking	Cohort and case-control studies **(Chapter 3)**	Nitric oxide-mediated demyelination, axonal loss and epigenetic effects **(Chapter 3)**
Sex hormones and/or oral contraceptive pill	Cohort and case-control studies	Altered antigen reactivity, tolerance and epigenetic effects
Stressful life events*	Cohort and case-control studies	Dysregulation of the hypothalamic-pituitary-adrenal axis
Respiratory tract infections	Epidemiological studies	Immunological trigger for inflammatory demyelination
Organic solvents	Epidemiological studies	Damage to BBB
Diet	Epidemiological studies	Vitamin D supplementation from oily fish

* The evidence is conflicting.
[+] References for factors additional to UVR/vitamin D, EBV and smoking are given in Handel et al. (2010) [289].

3.1.4 Other organ-specific autoimmune disorders—evidence for infections and other factors

Type 1 diabetes

Type 1 (insulin-dependent) diabetes mellitus is considered to be an autoimmune disease in which T lymphocytes infiltrate the pancreatic islets and destroy the insulin-

producing beta-cell population (see Chapter 1). As markers of the autoimmune process, auto-antibodies to insulin, glutamate decarboxylase and tyrosine phosphatase appear successively often years before clinical symptoms of type 1 diabetes, but auto-antibodies alone do not cause the disease, which is hypothesised to be largely T-cell mediated [198, 587].

Apart from dietary vitamin D (considered in Chapter 2), two types of environmental factors have particularly been associated with type 1 diabetes in epidemiological and immunological studies—exposure to enteroviral infections and other dietary factors such as cow's milk proteins early in childhood—both of which appear to act by affecting the immune system in the gut [587, 588].

Viral infections

While congenital rubella [520], mumps, measles, cytomegalovirus and retroviruses [589], as well as rotaviruses [590], have all been linked with type 1 diabetes, the current main candidates as initial determinants of type 1 diabetes are enteroviruses, and particularly the Coxsackie virus group [136] (see Table 3.2). Some four decades ago, antibodies against Coxsackievirus B serotypes were found to be more frequent in patients with newly diagnosed diabetes than in control subjects [591] and this has since been confirmed using more recent techniques [587, 592].

Enteroviruses are frequent among children and adolescents but usually are subclinical or manifest with mild respiratory symptoms; the group includes the well-known polioviruses. The primary replication of the virus occurs in the lymphoid tissues of the pharynx and small intestine, and can then spread to various organs including the pancreas. Enterovirus has been isolated from the pancreas of patients newly diagnosed with type 1 diabetes (Table 3.2); enterovirus RNA has also been detected in the blood of significant numbers of similarly newly diagnosed patients [198].

The first epidemiological reports linking enterovirus infections to type 1 diabetes were also more than 40 years ago; for example, the seasonal variation of the onset of diabetes was observed to parallel that of enterovirus infections [593] (Table 3.4). This has since been confirmed by other, prospective studies following the first appearance of diabetes-associated auto-antibodies (reviewed by Akerblom et al., 2002 [587] and Knip et al., 2005 [198]). Prospective studies (included in Table 3.4) have been particularly informative in determining not only the higher frequency of enteroviral

infections in those who progressed to clinical diabetes, but also the temporal relationship between enteroviral infections and the appearance of auto-antibodies, these considered to reflect the initiation of β-cell damage. Several of these studies have been conducted in Finland, where the incidence of type 1 diabetes is particularly high.

Table 3.4: Evidence supporting a causal relationship between enterovirus infections and type 1 diabetes. (Table adapted from Akerblom et al., 2002 [587] and Knip et al., 2005 [198])

Seasonal variation in the appearance of auto-antibodies and the incidence of diabetes resembles that of enterovirus infections
Epidemiological association observed in several studies in different countries using various methods (virus antibodies, viral RNA, cellular immunity)
Association observed in prospective studies (DiMe, DIPP)* in addition to age, sex and HLA-matched case-control studies
Clustering of enterovirus infections in prospective studies to period immediately preceding appearance of auto-antibodies (DiMe, DIPP)*
Association specific for enteroviruses and not observed in other virus infections
Inverse correlation between frequency of enterovirus infections and risk of type 1 diabetes in different populations and periods in ecological studies (analogy with poliomyelitis)
No evidence of abnormal regulation of enterovirus-specific immune responses after standardised enterovirus exposure (e.g. poliovirus vaccinations)
β-cell damage in patients with severe enterovirus infections (consistency with animal and *in vitro* studies showing that enteroviruses can infect and damage β-cells)

*DiMe, Childhood Diabetes in Finland study (non-diabetic siblings of type 1 diabetes patients followed) [594-596]. DIPP, Finnish Diabetes Prediction and Prevention trial (infants with diabetes-associated HLA-DQ alleles followed from birth) [597, 598].

However, it should be noted that other prospective studies in Germany and the US failed to find such associations between enterovirus infections and β-cell autoimmunity [599, 600], possibly because of limited statistical power, limited sampling times and virus detection strategies [198]. The role of enterovirus infections is thus still debated, although recent studies have again strengthened the view that enteroviruses contribute significantly to the pathogenesis of type 1 diabetes [601]. Current large international prospective trials such as The Environmental Determinants of Diabetes in the Young

(TEDDY) should help establish the role of this and other putative environmental factors more clearly [588].

Interestingly, ecological studies have suggested that the frequency of enterovirus infections is actually very low in Finland, even though the incidence of type 1 diabetes is the highest in the world [602]. In addition, the frequency of enterovirus infections decreased rapidly during the three-decade period before 2000, while the incidence of type 1 diabetes increased over the same period [603], as seen also for other infections in Europe as a whole (see Section 3.1.2, hygiene hypothesis [3]). Paralytic poliomyelitis, a well-known enterovirus disease, has also followed the same epidemiological pattern and, as discussed in Section 3.1.3 for MS, was found to depend on the age at exposure to the virus (i.e. the late-exposure, or 'polio', version of the hygiene hypothesis). It is possible that, analogous with paralytic polio, type 1 diabetes may be similarly caused by late exposure to an enterovirus, consistent with the observation of higher disease incidence where the frequency of enterovirus infections is low [587] (Table 3.4). This means that enteroviruses such as Coxsackievirus B may indeed have an ambiguous (i.e. causal or possibly protective depending on conditions) role in the context of autoimmune type 1 diabetes [497, 592].

Possible ways by which enterovirus could cause β-cell damage include two main mechanisms: infection and direct destruction of the β-cells (Table 3.4) and induction of an autoimmune response against β-cells. Evidence of both types exists from animal models (reviewed by Akerblom et al., 2002 [587]), the latter type including molecular mimicry and bystander activation of auto-reactive clones (see Section 3.1.1), together with epitope spreading [198]. Recent studies suggest that certain diabetogenic enterovirus variants establish *persistent* infection by induction of IL-10 [497], particularly in the intestinal mucosa and the pancreatic islets where local inflammation then leads to breakdown of tolerance in young, genetically susceptible individuals [592, 601].

Importantly, with regard to timing, several analytical studies suggest that the autoimmune process could be initiated before birth by maternal enterovirus infections, indicating the possible role of *in utero* exposures [595, 604]. Certainly, the appearance of auto-antibodies as early as three to six months after birth [605] suggests early exposure to causal factors. For example, gestational infection could induce immunological tolerance to the virus, and this recognition as 'self' by the offspring immune system could have consequences for viral persistence (i.e. latency) and later re-infection [588].

Other factors—dietary

Early exposure to dietary cow's milk in infancy may be another important environmental factor affecting risk of subsequent type 1 diabetes, the primary antigen involved possibly being bovine insulin [198]. While the vast majority of infants are exposed to cow's milk proteins in the first months of life without any harmful effects, it seems that some with a genetic risk for diabetes are vulnerable to the diabetogenic effect of such proteins. That is, the intestinal immune system is thought to play a key role as a modulator of (non-self, bovine) insulin-specific immunity and in establishing, and preventing the breakdown of, immunological tolerance (see Chapter 1) to these and other dietary antigens. The pathogenic mechanisms involved in type 1 diabetes, by which self-insulin proteins become antigenic and induce (human-) insulin-specific auto-antibodies, may thus be related to the aberrant regulation of oral tolerance (to bovine insulin) in the gut [198]. However, most of the supporting evidence for cow's milk as a determinant factor comes from the Finnish population and its role is still debated [592]. Alternatively, the increased immunity to cow's milk proteins, and to other proteins such as wheat proteins in dietary cereals (seen in both type 1 diabetes and coeliac disease), may instead reflect a general impairment in mucosal immunity in the intestine [606]. To answer the question of whether cow's milk is a possible determinant, the current Finnish prospective Trial to Reduce IDDM in the Genetically at Risk (TRIGR) will attempt to prevent type 1 diabetes by eliminating cow's milk in infant nutrition before the onset of islet autoimmunity (see www.clinicaltrials.gov).

In addition, other environmental factors, such as gastrointestinal infections, including bacterial infections, may interact with the dietary antigen-induced disease process by modifying the normal gut flora or causing changes in the intestinal cytokine environment [587, 592]. Still other candidate environmental determinants of type 1 diabetes are reviewed fully by Akerblom et al. (2002) [587] and van Belle et al. (2011) [592]. Some of these are considered to belong to the group of likely *initiators* of β-cell damage, such as enteroviral infections, dietary proteins and antenatal and perinatal factors, while others, such as high growth rate and stressful life events, may belong to the *promoters* or *precipitators* of the disease process [240]. The large prospective cohort TEDDY trial being conducted over six centres in the US and Europe will investigate several different infectious, dietary, psychosocial and other environmental and genetic factors proposed to initiate or protect against type 1 diabetes [588]. As in other organ-specific autoimmune disorders such as MS, it is likely that different

combinations of genetic and non-genetic risk factors act together, or in sequence [198], to produce disease in different individuals [240].

Rheumatoid arthritis

RA is a chronic inflammatory joint disease resulting in bone and joint destruction [194, 292]. The disease is characterised by persistent synovitis, systemic inflammation and auto-antibodies (particularly to rheumatoid factor and citrullinated peptide). RA affects 0.5 to 1.0% of adults in industrialised countries and is most typical in women after age 50 [607, 608].

Like MS and type 1 diabetes, RA appears to be a multifactorial disease resulting from interactions between genetic (mainly HLA-DRB1 and, in European populations, the tyrosine-phosphatase gene on chromosome 1) and environmental factors, all of which may act many years before disease onset. Among environmental factors implicated in the development of RA, smoking shows the strongest association with susceptibility and is also linked to worse clinical outcomes [292]. Other candidates include hormonal factors, adverse pregnancies, obesity and diet, as well as infections [194, 607]; vitamin D intake (see Chapter 2, Table 2.5) and dietary omega-3 fatty acids appear protective [292].

Infections

There has long been a belief that RA is caused by an infection, although no single organism has been identified. However, RA has been observed to begin within a few weeks of an infection in a substantial proportion of cases, and vaccination can trigger RA in some people [194]. Moreover, an infectious aetiology is strongly suggested by examples of infection-induced, chronic erosive arthritis in humans, and the known ability of infectious agents, especially viruses, to induce immuno-inflammatory dysregulation [609].

Infections linked with RA include mainly EBV and human parvovirus B19 (see Table 3.2), as well as bacteria such as *Proteus,* and mycoplasma [292, 607]. For example, sera from RA patients show high titres of EBV antigens and of antibodies to latent and replicative EBV antigens; in addition, EBV RNA has been identified in B cells in synovial tissue from RA patients [610]. Evidence supporting a role for human

parvovirus B19 includes the presence of viral DNA in the synovial fluid, synovial cells and/or tissue of RA patients [609, 611].

As found for both MS and type 1 diabetes, environmental factors acting early in life may be particularly important in determining subsequent RA. For example, high birth weight was found to be associated with increased RA risk in the American Nurses' Health Study [612]. Early life studies also suggest that a history of infections in infancy may protect against RA in adulthood [292], suggesting a hygiene hypothesis mechanistic effect for this disorder in company with other autoimmune conditions.

Summary

Summing up, the evidence for the effects of environmental factors on pathogenesis of the three organ-specific autoimmune disorders—MS, type 1 diabetes and RA—shows similarity in overall principles and underlying mechanisms, even though some specific factors may vary. Lack of UVR and/or vitamin D appears to be a plausible candidate determinant common to all three disorders (see Chapter 2), while infections appear to be causal and/or protective for the same disorders, depending on conditions and specific timing (see Section 3.1). To tease out the evident importance of timing in such disorders, and to construct a framework for the subsequent analyses in this thesis, the final section of this review will focus particularly on MS and will consider the overall timing and integration of the major environmental factors considered in Chapters 2 and 3.

3.2 Overall timing and integration of environmental factors for MS

Chapters 2 and 3 have reviewed the accumulating evidence for the major candidate environmental factors implicated in MS, type 1 diabetes and RA. Sunlight as UVR/vitamin D and infections—particularly EBV for MS (and possibly RA) and enteroviruses for type 1 diabetes—have been shown to be important environmental determinants based on current epidemiological and immunological evidence for these autoimmune disorders. Smoking has also been shown to be important for MS and RA.

The different life periods to which the evidence pertains for the major environmental factors for MS—UVR/vitamin D, EBV infection and smoking—will now be considered,

before integrating these together in a possible causal-pathway model in the final subsection of this chapter.

Prenatal and perinatal exposure

From northern hemisphere studies, the period around the time of birth, including before birth and shortly after, appears to be important for its subsequent influence on MS risk. As discussed in Chapter 2, a seasonal pattern in month of birth found in Canadian MS cases [324] implicates causal factors acting in early development; the observation that these factors seem to be latitude related adds further weight to this finding [189]. In addition, maternal parent-of-origin effects, and higher MS concordance rates among dizygotic (non-identical) twins than among siblings of the same family [202, 214] (see Chapter 2), further support a critical role for the prenatal environment in determining subsequent MS risk.

Either sunlight/vitamin D or EBV could play a role in determining future MS risk in this important period (maternal smoking appears unrelated). However, while maternal transmission of EBV during pregnancy has been observed [613], there is little further evidence that viral infection *in utero* or as a neonate determines the subsequent risk of MS [289]. Conversely, there is substantial and accumulating evidence for action of UVR or vitamin D in early life, including *in utero*, where vitamin D appears to have significant effects (see Chapter 2, Section 2.6). This factor remains potentially the most important modulator of subsequent MS risk acting during this early period.

Childhood and adolescence

Studies of migration with respect to age implicate childhood and early adolescence as a risk period for the subsequent development of MS, even though these studies are often small and prone to error [189, 307]. Overall, these studies suggest that people younger than 15 years at the time of migration tend to adopt the MS risk of the host country, whereas those older than 15 years have a risk of MS similar to their home country [306, 312] (see Chapter 2). Consistent with this, childhood sun exposure is associated with decreased risk of MS in Australia, as well as in the northern hemisphere, and greatest reduction of MS risk with vitamin D supplementation was seen in individuals less than 20 years old in the US (see Chapter 2, Table 2.3).

A role for the environment in modifying MS risk during childhood and adolescence is further suggested by space-time cluster analysis; a Sardinian study found MS patients were more likely to live near each other than expected between age one to three years [614], and an earlier Norwegian study found such clustering between 13 and 20 years [615]. These studies together suggest that environmental factors that are active during adolescence or early adulthood alter MS risk [289].

EBV infection, in particular, is linked primarily with the childhood and adolescent period. Importantly, people who develop adult-onset MS have significantly elevated EBV-antibody levels early in adult life [500, 564], suggesting that there is a period in adolescence and early adulthood during which EBV infection predisposes to subsequent development of MS [289]. Related to this, the risk of MS is particularly high in people who, because of a hygienic childhood environment, fail to become infected with EBV in early childhood and instead contract EBV-induced IM later in childhood or adolescence (see Section 3.1.3), particularly if this occurs after the age of 15 years [616].

Also related to this period, and the hygiene hypothesis, it appears likely that infection with other microbes, or helminths, during this time may influence subsequent risk of developing MS by protective modulation of the immune system (see Section 3.1.2).

Adulthood

The age at MS onset is variable, implying that environmental factors might continue to alter MS susceptibility even in adulthood. In addition, in the most common form of MS, RRMS, relapse rates appear linked to seasonal changes (see Chapter 2), suggesting a role for environmental factors also as pre-onset 'triggers'.[12]

Outdoor occupation and/or higher vitamin D levels in adulthood may reduce MS risk (see Chapter 2, Table 2.3); however, fully prospective trials of vitamin D supplementation in adulthood still need to be carried out, and the main effects of UVR/vitamin D with respect to MS seem to be prior to adulthood. EBV, also, appears to have its major effects prior to adulthood, most of the adult population being already seropositive (see Section 3.1.3). Other viruses could act during adulthood as a final 'trigger' for demyelination; for example, viral respiratory tract infections are associated

[12] The term 'trigger' has been used variously in the literature, to mean either an 'initiator' of autoimmune disease, or a (later) 'promoter' or 'precipitator' of such disease. The latter meaning is used here.

with new-onset MS during adulthood [617] and, as discussed earlier, these infections may also be linked to relapses.

Smoking, in particular, later in life seems to be related to MS susceptibility (Section 3.1.3) and may also affect the clinical course of disease. Patients with RRMS have an increased risk of developing secondary progressive MS if they have ever smoked, and the risk of presenting initially with primary progressive MS is significantly higher in smokers than in non-smokers [582].

Summing up the epidemiological and other evidence, and assuming a multifactorial causation sequence for clinical MS (see Chapter 1), the most influential environmental risk factor appears to act early in life and determine geographical prevalence gradients in ethnically homogeneous populations. As latitude is the most strongly associated environmental factor, an obvious candidate factor in this period is sunlight/vitamin D exposure, with other factors such as diet possibly accounting for geographical exceptions. The next factor in the life course also seems to be geographically distributed and may be infections; these may be further influenced, or primed, by early or late exposure, as in EBV-induced IM [6, 189].

Integration of main potential factors in possible causal model

Goodin (2009) has postulated a possible life-course mathematical model for the timing and the role of environmental factors in the aetiology of MS. The model comprises a 'causal cascade' of a series of factors, these acting sequentially and at 'appropriately timed' periods before disease onset [229]. This idea was also proposed earlier by McMichael and Hall (2001) for the sequence of UVR, infections and other possible factors for MS [6] and, further, is similar to the 'two-hit hypothesis' for schizophrenia described by Torrey and colleagues (1997) [642].

In Goodin's model for MS, the two most influential periods for exposure to such factors are postulated to be near birth and during childhood or adolescence, with a third period in adulthood closer to onset [229]. While the two leading environmental candidate factors are proposed by Goodin (and several other reviewers, based on current evidence, Table 3.3) to be vitamin D deficiency and EBV infection, the mathematical model also allows for the eventuality that these factors are replaced by other factors, should future evidence indicate this [229]. The model further proposes a third factor

acting long after the first two, the identity of this factor, or factors, being presently unknown (Figure 3.5).

Possible Causal Cascade to MS Pathogenesis

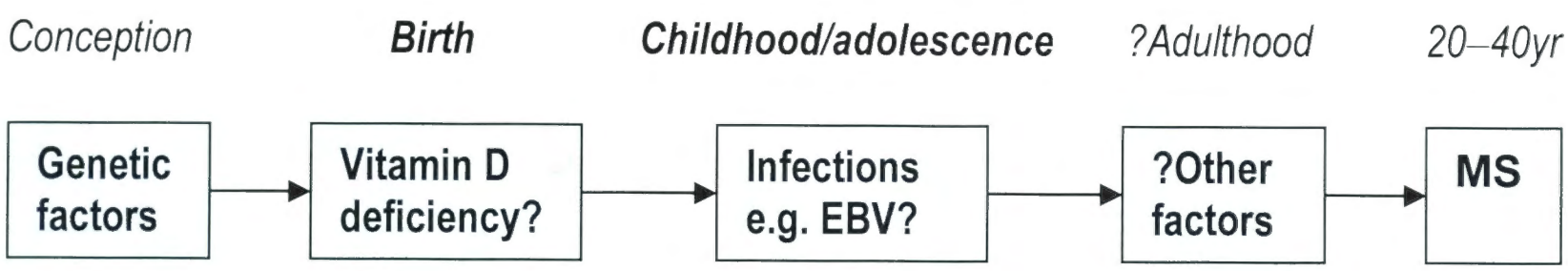
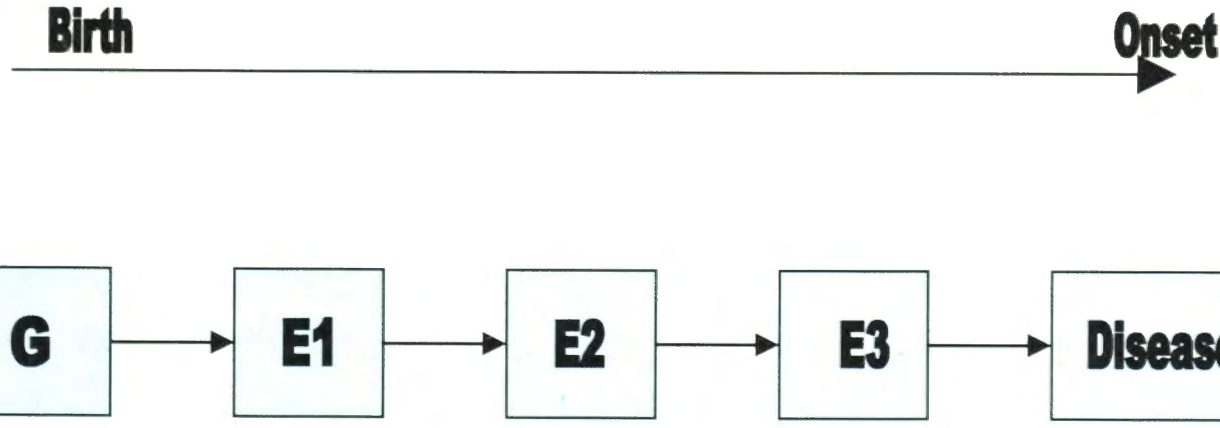


Figure 3.5: 'Most probable' causal pathway leading to MS, showing timing of environmental factors. (Figure based on possible pathways according to 'causal cascade' model of Goodin, 2009[13] [229].)

Goodin proposes a possible causal pathway for MS, with the first environmental factor being 'necessary but not sufficient' but, importantly, acting near the time of birth. The second factor is again 'necessary but not sufficient', and is further postulated to be acting in childhood or adolescence. Together with the third, still necessary, factor acting post-childhood/adolescence (Figure 3.5), this causal-pathway model is essentially based on the 'life-course component cause model' of disease causation [1] pictured in Chapter 2 (Figure 2.3) and repeated here as Figure 3.6, but with the *timing* for each factor now specified in the causal hypothesis.



'Component causes do not need to act at the same time, and may interact and accumulate over the life course to result in sufficient cause.'

Figure 3.6: Life-course disease causation model for a multifactorial disorder with genetic (G) and environmental (E1 to E3) component causes.

[13] It should be noted that Goodin's causal pathway does not specifically include the possibility of *protective* infections in childhood that may act to modify later causal effects of infections such as EBV in adolescence (see Section 3.1.3).

Importantly, Goodin's life-course model gives a basis on which to frame the research questions for MS for this thesis. That is, if the *timing* of effects of environmental factors can be determined for MS, it may then become possible to discern the likely *nature* of the environmental factors involved. For example, if a month-of-birth pattern is found to be evident in MS cases (but not the general population) in Australia, an environmental factor acting around the time of birth (the first of Goodin's 'critical' periods), would be indicated. Additionally, an association with latitude or with seasonal perinatal UVR might further increase knowledge of the factor's nature (see Chapters 5 to 7). Similarly, an investigation of Goodin's second 'critical' period, childhood, by means of a hygiene-hypothesis proxy marker for childhood infections—birth order—may give information on both the importance of this period and the possible nature of the factor(s) acting in this period (see Chapter 8).

Thus, this thesis will approach candidate environmental determinants of MS, by means of the first two possibly critical life-course periods indicated in Goodin's (2009) model [229]. That is, the period around the time of birth and the childhood period will be investigated separately using individual-level data from the 1981 Australian MS Survey (see Chapters 5 to 8), in order to attempt to determine the nature of the possible determinant factors acting in each of these periods.

3.3 Conclusion

Epidemiological evidence in terms of 'person', 'place' and 'time' and the immunological mechanisms reviewed herein point to a role for infections as a possible environmental determinant of MS and other autoimmune disorders such as type 1 diabetes and RA. Such infections, acting early in life, may be a component cause that, together with genetics and other interacting factors such as UVR/vitamin D (see Chapter 2), may affect the risk of developing such organ-specific autoimmune disorders later in life.

3.4 Postscript

This chapter has reviewed the evidence for infections as a possible determinant environmental factor for MS and other organ-specific autoimmune disorders—type 1 diabetes and RA—and considered how the major factors identified in Chapters 2 and 3

for MS may be related to each other according to an overall timing-based life-course model.

Using such a model as a conceptual framework, the main analysis chapters of this thesis will investigate whether there is, first, a timing-of-birth pattern for MS cases in Australia—that might indicate particular determinants acting near birth—and whether such a pattern is related to southern latitude and/or perinatal UVR (see Chapters 5 to 7). Second, Chapter 8 will investigate a further life-course stage—childhood—and attempt to discern the nature of environmental factors acting then, both studies using available individual-level data from a national MS survey.

Prior to this, in addition to the main MS 'timing' analyses in Chapters 5 to 8, available population-level data for type 1 diabetes, RA and other immune disorders will be explored for possible southern hemisphere associations with latitude and UVR in Chapter 4, as corroborative evidence for corresponding ecological associations between latitude/UVR and MS in Australia that were already known [9, 301]. These disorders other than MS had not previously been investigated in this way in Australia, even at the ecological level, prior to commencement of this thesis.

CHAPTER 4

DISTRIBUTION OF AUTOIMMUNE DISEASE IN AUSTRALIA: ECOLOGICAL ANALYSIS OF LATITUDE, REGIONAL UVR AND IMMUNE-RELATED DISEASE PREVALENCE

4.1 Introduction

As reviewed in Chapter 1, autoimmune diseases such as MS, type 1 diabetes mellitus and RA are organ-specific immune system disorders that share common features of self-reactive T cells and the presence of auto-antibodies. As a group, these disorders affect some 5% of the global population, particularly in Western countries [2]; at least 0.1% of children and adolescents (up to 20 years of age) are affected by type 1 diabetes [26, 28] and incidence worldwide is rapidly increasing, particularly in younger age groups (see Chapter 2). At least two-thirds of MS and RA patients are female, but a relatively equal risk between males and females is evident in childhood onset type 1 (insulin-dependent) diabetes.

While their precise aetiologies are unknown, these autoimmune disorders are generally agreed to reflect interactions of polygenic traits with various ill-defined environmental factors (see Chapters 2 and 3) that result in altered homeostatic balance between self-reactive and regulatory T cells to cause ongoing disease (see Chapter 1). Ecological study at the population level may help elucidate the role of environmental factors in the aetiology of the organ-specific autoimmune diseases MS, type 1 diabetes and RA, as well as other immune-related disturbances such as asthma and eczema/dermatitis.

MS, type 1 diabetes and, to a lesser extent, RA display a latitudinal gradient in disease frequency in the northern hemisphere, particularly in Western Europe and North America, with the prevalence of these disorders increasing at higher latitudes (see Chapter 2). MS exhibits a similar prevalence gradient in the southern hemisphere, in Australia and NZ [302, 304]. In Australia, however, where the opportunity exists to study gradients in prevalence rates across a large-area population (Figure 4.1) that is less ethnically and genetically diverse than across Europe, analyses for other immune-related disorders had not been conducted prior to this thesis study.

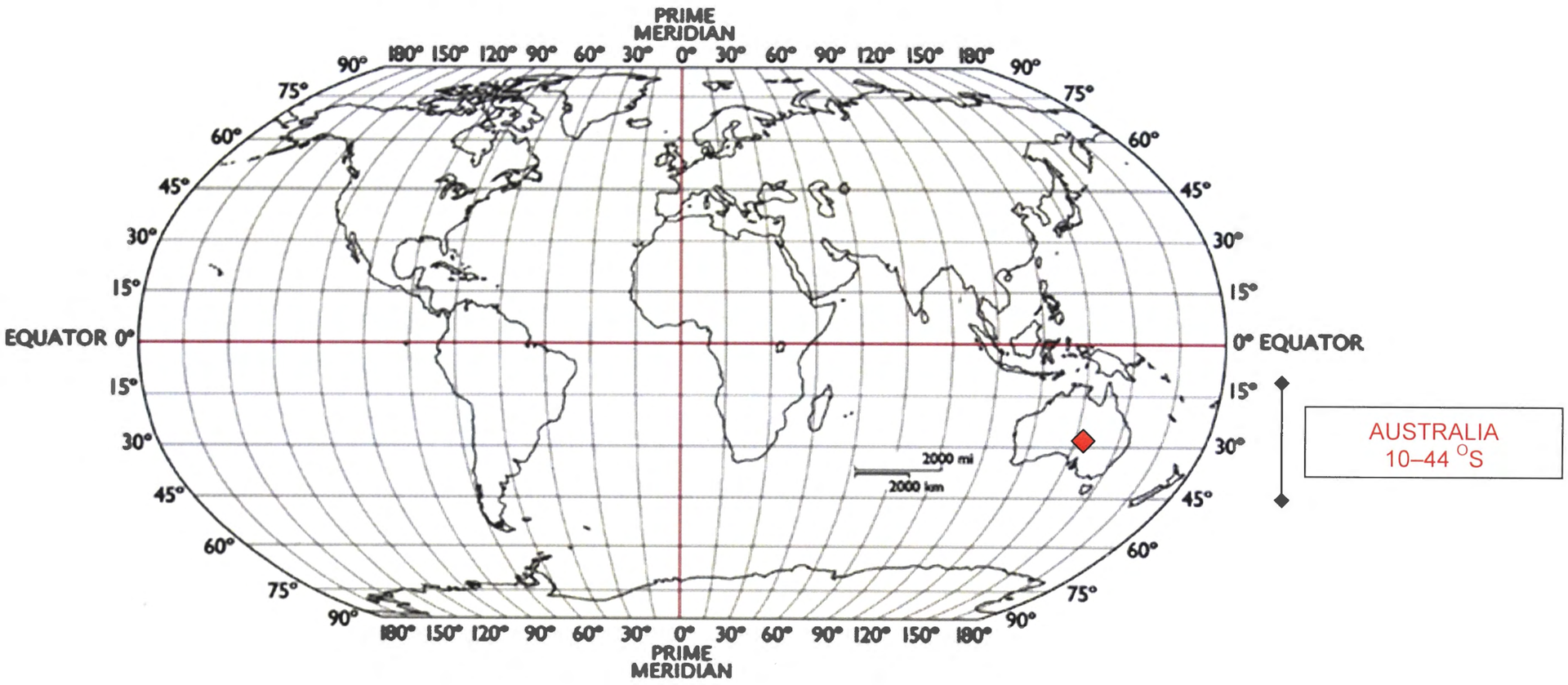


Figure 4.1: World map showing position and latitude range of Australia compared with other continents and particularly those in the northern hemisphere. (Figure from GraphicMaps.com; accessed at http://www.worldatlas.com/aatlas/printpage/latlog.htm.)

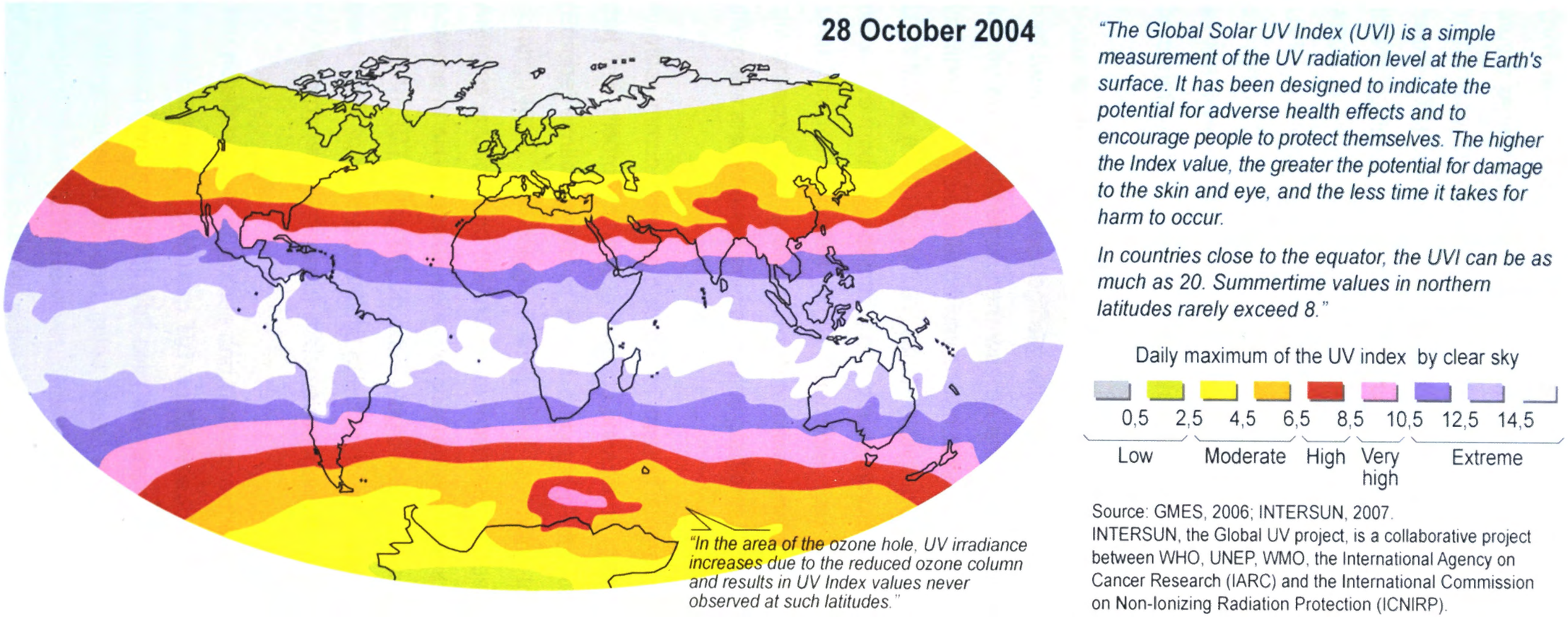


Figure 4.2: World map showing variation in solar UVR reaching the Earth's surface with latitude. (Figure from http://maps.grida.no/go/graphic/the-global-solar-uv-index).

UVR reaching the Earth's surface varies inversely with latitude (Figure 4.2); UVR is thus a prominent latitude-related environmental factor (see Chapter 2). Recent photo-immunological work shows that UVR suppresses cellular immunity, attenuating T-helper (Th)-cell-mediated immune responses [362, 363]. The mechanisms by which this is achieved have been described in Chapter 2 and include tolerogenic effects on APCs and induction of regulatory T cells. Vitamin D may also be a major mediator of UVR (see Chapter 2). Breakdown of these mechanisms of immune self-tolerance is thought to be significantly involved in the development of autoimmune disorders such as MS, type 1 diabetes and RA (see Chapter 1). UVR might therefore be expected to be beneficial for the prevention of these disorders by preventing tolerance breakdown.

However, few organ-specific autoimmune diseases or other immune-related disorders had been assessed ecologically with respect to UVR prior to this thesis. An exception for the southern hemisphere was an analysis of MS in Australia, where the regional variation in MS prevalence was shown to be strongly inversely associated with ambient UVR levels ($r=-0.91$; $p=0.01$) [9]. This finding supported the possibility of UVR being a protective modulator of immune and autoimmune processes involved in the aetiology of such immune disorders [5, 6]. Therefore, type 1 diabetes and RA were chosen for similar ecological analysis, to determine whether these organ-specific autoimmune disorders also showed environmental gradients as observed for MS in Australia.

Another immune-related disorder of interest is atopic eczema. There is some evidence that this disorder shows a latitude-prevalence gradient in the northern hemisphere (Europe) similar to that of type 1 diabetes (A.J. McMichael, unpublished data, pers. commn) [618]. An ecological analysis of worldwide International Study of Asthma and Allergies in Childhood (ISAAC) data from 146 centres further showed that symptoms of eczema in six- to seven-year-old children correlated positively with increasing latitude [619]. An RCT in the UK on adult atopic eczema has shown UVR, particularly narrow-band UVB, to have a beneficial effect on this immune-related disorder [620]. Moreover, UV therapy (see Chapter 2) is also a recognised treatment for atopic dermatitis.

The possible effect of latitude and UVR on asthma is of further interest. UVR has been thought to effect a shift from Th1- to Th2-mediated processes by down-regulating Th1-mediated immunity [363]. Th2 cells are responsible for immediate-type hypersensitivity to allergens such as dust mites; thus, UVR may have the potential to exacerbate allergic disease [621]. However, recent work has cast doubt on the mutual antagonism of Th1 and Th2 cytokine expression, particularly in humans [2, 26, 622, 623] and

understanding of immune disorders is now considerably more complex, involving several types of regulatory T cells (see Chapter 1). For example, UVB can directly inhibit both Th1- and Th17-type immune responses, particularly through IL-10 (see Chapter 1), but also suppress Th2 and allergic responses [114]. Conversely, vitamin D inhibits both Th1 and Th17 cells while inducing Foxp3+ regulatory T cells (Chapter 1) and, under certain conditions, also Th2 cells [52] (see Chapter 2, Figure 2.11). Importantly, asthma can coexist with disorders such as type 1 diabetes, RA and coeliac disease in children, suggesting a common environmental influence [624]. Therefore, eczema/dermatitis and asthma were chosen to be similarly analysed for regional prevalence gradients, to compare with any associations evident for type 1 diabetes and RA.

This chapter examines the possible association between latitude and prevalence of the immune-related disorders, type 1 diabetes mellitis, RA, eczema/dermatitis and asthma in Australia, a country with a relatively genetically homogeneous population[14]. Regional differences in ambient UVR were further considered in order to examine possible associations between regional and seasonal UVR levels and the prevalence of these four immune disorders in Australia.

4.2 Methods

4.2.1 Outcome measures

Prevalence data source—1995 NHS

The Australian NHS is a nationwide cross-sectional survey carried out every three to five years by the ABS, responsible also for the five-yearly Australian Census of Population and Housing. The 1995 NHS was conducted during the 12-month period from January 1995 to January 1996, but on a quarterly basis to take account of possible seasonal effects. Approximately 54,000 people from all states and territories of Australia and across all age groups provided information about their own health status in response to a series of questionnaires. Residents of a stratified, multi-stage

[14] The Australian population is largely of European origin, the indigenous population comprising just 1.5% of the total population in 1986 and 2.0% in 1996 (ABS publication 4705.0: Population Distribution, Indigenous Australians, 1996; accessed at http://www.abs.gov.au/AUSSTATS/abs@nsf/DetailsPage/4705.01996?OpenDocument)

area, random sample of 23,800 private households (houses, apartments) and households in non-private dwellings (such as hotels, hostels and caravan parks) were included; the base sample size corresponded to approximately one-third of 1% of the Australian population. Residents of hospitals and other institutions (such as nursing homes, prisons and military institutions) were excluded, and the Northern Territory (NT) sample was also predominantly urban (because of vast, sparsely populated areas in that territory) [13].

Residents of the selected households were interviewed in person by trained ABS interviewers, and all persons within each state/territory had a known, and in the main, equal, chance of selection. Responses were received from 91.5% of sampled households (i.e. total unweighted response rate) and 97% of people from these households completed the questionnaires fully. Independent quarterly population estimates were used by ABS for population standardisation, and prevalence rate estimates were adjusted for household size. The estimation procedure used information on the patterns of response to counter known biases in target variables resulting from partial response [13].

The 'Personal Interview Questionnaire' used to obtain health information was designed by the ABS to be administered using standard ABS procedures for conducting population interview surveys, and was fully field tested to ensure particular aims of the survey were addressed. The survey interviewers recorded information on recent illness and long-term conditions, as reported by respondents. Specific questioning about conditions, including diabetes, arthritis and asthma, was followed by 'actions-based' questions on recent visits to hospitals, clinics etc. or health professionals, and about days away from work or school. Information on medications usage, including insulin and specific medications for asthma, arthritis and allergies, was further elicited [13, 625]. Disease conditions were classified by categories of the International Classification of Diseases, 9th Revision. However, information recorded in the survey was self-reported and not medically verified.

Prevalence estimates—summary data

For this thesis study, summary-data prevalence estimates from the published survey results [626] for four different immune-related disorders were used. The age- and sex-standardised prevalence rates, per 1000 population, of type 1 diabetes, RA,

eczema/dermatitis and asthma were compared over the eight major state and territory regions of Australia, as listed in Table 4.1.

4.2.2 Exposure measures

Latitude

The latitude of the capital city of each state or territory (Figure 4.3), expressed in decimal degrees south, was used for each region (Table 4.1).

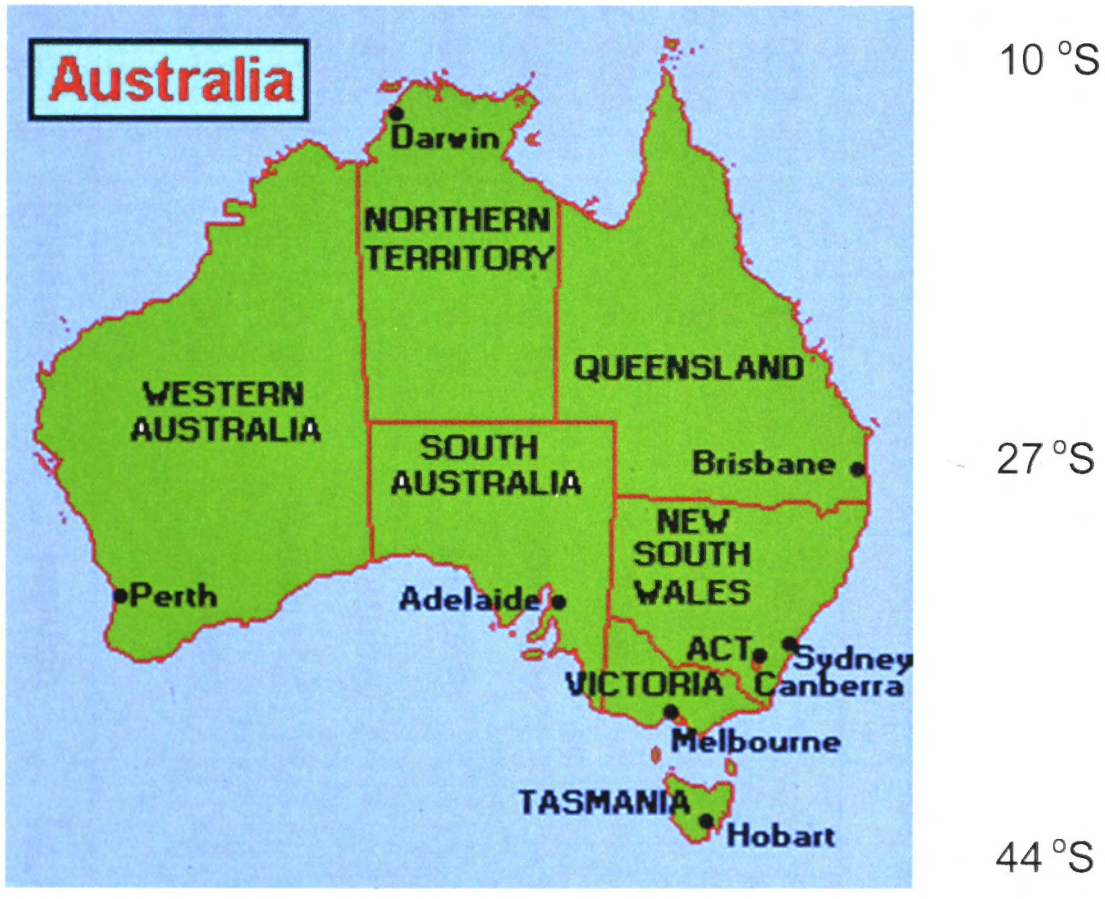


Figure 4.3: Australian states and territories and their capital cities, shown together with the overall latitude range. (Figure from http://www.cse.unsw.edu.au.)

Table 4.1: Australian state and territory regions and approximate latitude ranges; regional capitals and their latitudes[+] (decimal degrees south) shown together with midsummer (January) and midwinter (June) solar-noon UVR[#] (mW/m^2) for each regional capital

State/Territory region		Latitude range for region (°S)	Regional capital	Latitude of capital	Midsummer UVR (mW/m^2)	Midwinter UVR (mW/m^2)
NT	Northern Territory	11–26	Darwin	12.4	339.5	205.9
QLD	Queensland	10–29	Brisbane	27.5	332.4	103.9
WA	Western Australia	14–35	Perth	31.9	326.1	82.9
NSW	New South Wales	28–37	Sydney	33.9	306.1	66.5
SA	South Australia	26–38	Adelaide	34.9	303.4	60.7
ACT	Australian Capital Territory	35–35.5	Canberra	35.3	302.7	55.7
VIC	Victoria	34–39	Melbourne	37.8	287.6	48.6
TAS	Tasmania	41–43.5	Hobart	42.9	256.7	31.4

[+] Latitude values obtained from Geoscience Australia mapping agency (http://www.ga.gov.au).
[#] UVR data are monthly averages from daily (at local solar noon) erythemally-weighted clear-sky UV Index values derived from satellite ozone data over the period 1979 to1993 (supplied by Dr Lilia Lemus-Deschamps, BMRC, Australia) and expressed as mW/m^2.

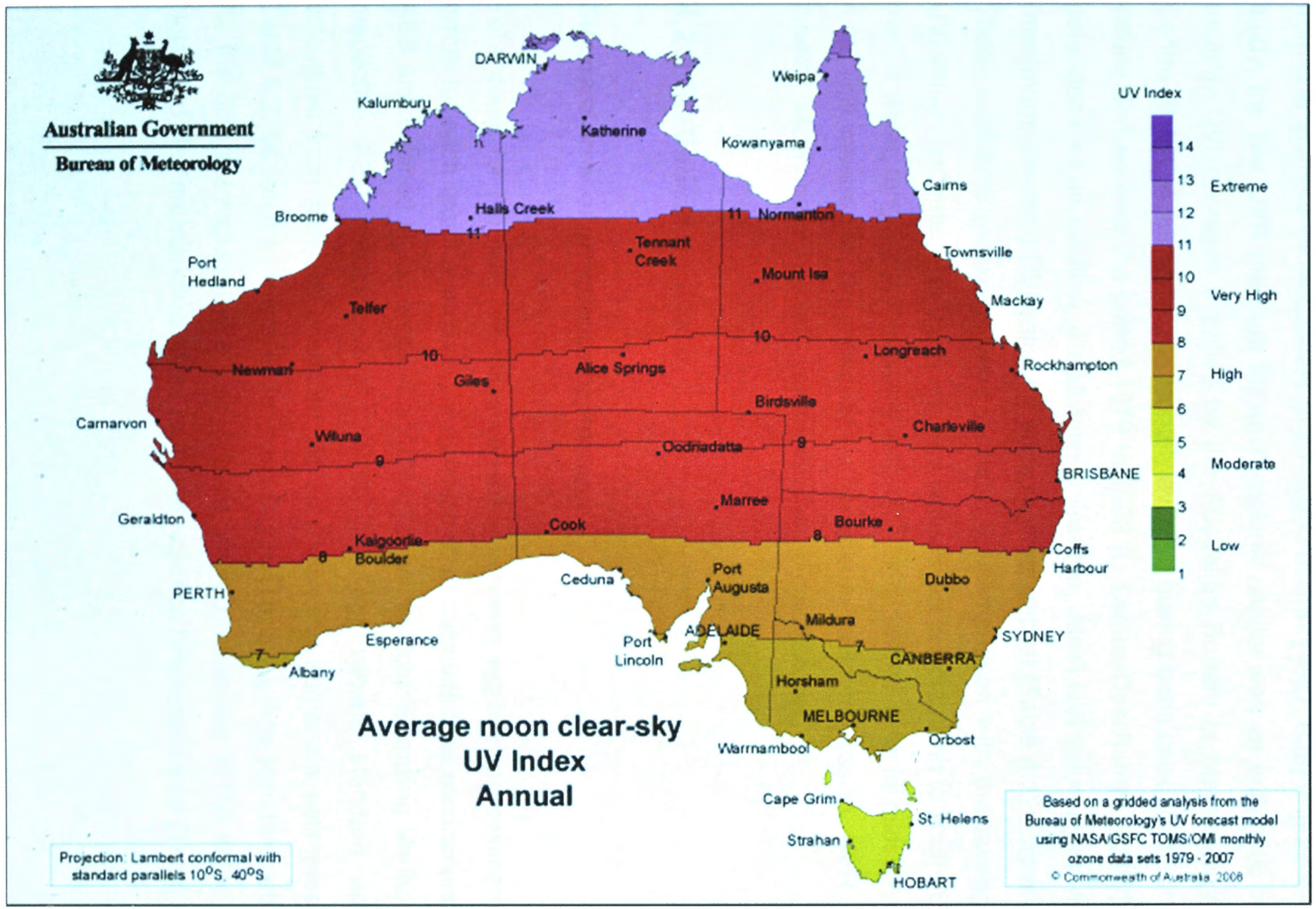


Figure 4.4: Average annual UV Index values for Australia, 1979 to 2007 (Figure from Bureau of Meteorology, http://www.bom.gov.au/jsp/ncc/climate_averages/uv-index/index.jsp.)

UV radiation

Three ambient UVR measures were examined, each relating to the WHO's Global Solar Ultraviolet Index (UVI) measuring UVR intensity (see Figure 4.2), and expressed in milliWatts per square metre (mW/m^2) (where 1 UVI unit=$25mW/m^2$). The UVI is a daily (at local solar noon) erythemally-weighted clear-sky UV-intensity value derived from total ozone mapping spectrometer (TOMS) satellite data and in Australia ranges from one to about 14 (equivalent to 25 to 350 mW/m^2) [627, 628] (Figure 4.4). For this study, the first UVR measure for each regional capital was an arithmetic mean of 12 monthly UVI averages supplied by the Australian Bureau of Meteorology Research Centre (BMRC Australia), each monthly average having been calculated from daily UVI values derived over the period 1979 to 1993 (L. Lemus-Deschamps, BMRC Australia, pers. commn). In addition, a midwinter (minimum, June) UVI value and a midsummer (maximum, January) UVI value for each regional capital (Table 4.1) were investigated. These additional measures were included for comparison with the average monthly UVI values because over the Australian latitude range from north to south, UVR varies over a wider range (approximately two-fold) in midwinter than in midsummer (Table 4.1). This means that the independent variable in linear regression, UVR, may have greater discriminatory power in midwinter than in midsummer.

4.2.3 Statistical analysis

Standard error of prevalence estimates

To account for the NHS sampling variability between regions, approximate standard errors for each regional prevalence estimate were calculated, as recommended by the ABS and outlined in Appendix H and the 'Technical Note: Sampling Variability' in the respective ABS publications [13, 626]. The approximate standard errors were calculated from standard errors for the corresponding numerator estimates shown in Table 4.2, as listed by the ABS for each region [13, 626]. Age-standardisation factors for the different regions were also applied, as recommended in the same Technical Note, and the final standard errors for the prevalence rates calculated (Table 4.2).

Table 4.2: Regional prevalence rates[+] of type 1 diabetes, RA, eczema/dermatitis and asthma (in bold with SE[#] in parentheses) together with numerator data on which SE estimates were based

State/ Territory		Type 1 diabetes	RA	Eczema/ dermatitis	Asthma	Regional population
NT	No. cases	200	1,500	2,900	17,200	145,300
	Prevalence	**2.9 (1.80)**	**18.2 (5.56)**	**19.2 (4.43)**	**127.2 (13.11)**	
QLD	No. cases	10,500	82,100	96,900	438,000	3,277,800
	Prevalence	**3.2 (0.62)**	**25.7 (1.34)**	**29.5 (1.40)**	**132.6 (2.51)**	
WA	No. cases	6,900	49,000	82,400	201,500	1,732,400
	Prevalence	**4.2 (0.79)**	**29.8 (1.66)**	**47.1 (1.90)**	**115.2 (2.76)**	
NSW	No. cases	26,300	170,700	181,900	633,700	6,120,500
	Prevalence	**4.2 (0.56)**	**27.3 (1.24)**	**29.8 (1.30)**	**103.9 (2.10)**	
SA	No. cases	8,300	41,900	73,000	163,500	1,474,800
	Prevalence	**5.4 (0.63)**	**26.8 (1.15)**	**50.3 (1.55)**	**112.4 (2.16)**	
ACT	No. cases	1,100	5,600	14,400	35,500	304,900
	Prevalence	**4.6 (0.99)**	**21.4 (1.65)**	**45.5 (3.98)**	**111.9 (2.65)**	
VIC	No. cases	23,800	106,400	171,500	501,500	4,503,100
	Prevalence	**5.2 (0.50)**	**23.2 (0.88)**	**38.2 (1.07)**	**111.8 (1.63)**	
TAS	No. cases	2,100	19,100	20,700	48,800	473,600
	Prevalence	**4.5 (1.14)**	**39.5 (2.38)**	**44.0 (2.53)**	**102.1 (3.48)**	

[+] Rate per 1000, age- and sex-standardised to the 1995/96 Australian population (data from ABS [626]).
[#] Approximate SE calculated from ABS-provided SE for numerator [13, 626].

Weighting of prevalence estimates

The reciprocal of the variance of each prevalence rate was then used to weight the estimates to take account of differing sample sizes in the regions. For example, the region with the lowest population, NT, and the largest SE (and variance) (Table 4.2), contributed less weight to the overall regression analysis of the eight regions.

Linear regression

Associations between each of the environmental variables and the immune disorder prevalence rates were examined by variance-weighted least squares regression, using the statistical software program Stata 7.0 (Stata Corporation, College Station, Texas). The eight regional prevalence rate estimates were regressed, first, versus the latitude values of the regional capital cities and, second, versus each of the three UVR measures for the regional capitals. The magnitude of change in disease prevalence rates over the north-south range was compared by substituting latitude values for northernmost Darwin and southernmost Hobart into the regression equations.

Sensitivity analysis to check use of capital-city latitudes

Since the use of capital-city latitudes was convenient but may not have been representative of the regional population distribution, particularly for larger regions with the capital at either a northern or southern extreme of the region rather than medially placed (e.g. QLD and NT), a sensitivity analysis was carried out. This tested the effect of using an alternative latitude value midway between the two main population centres in these regions. Results using the substitute latitude values were then compared with those obtained from the regional capitals.

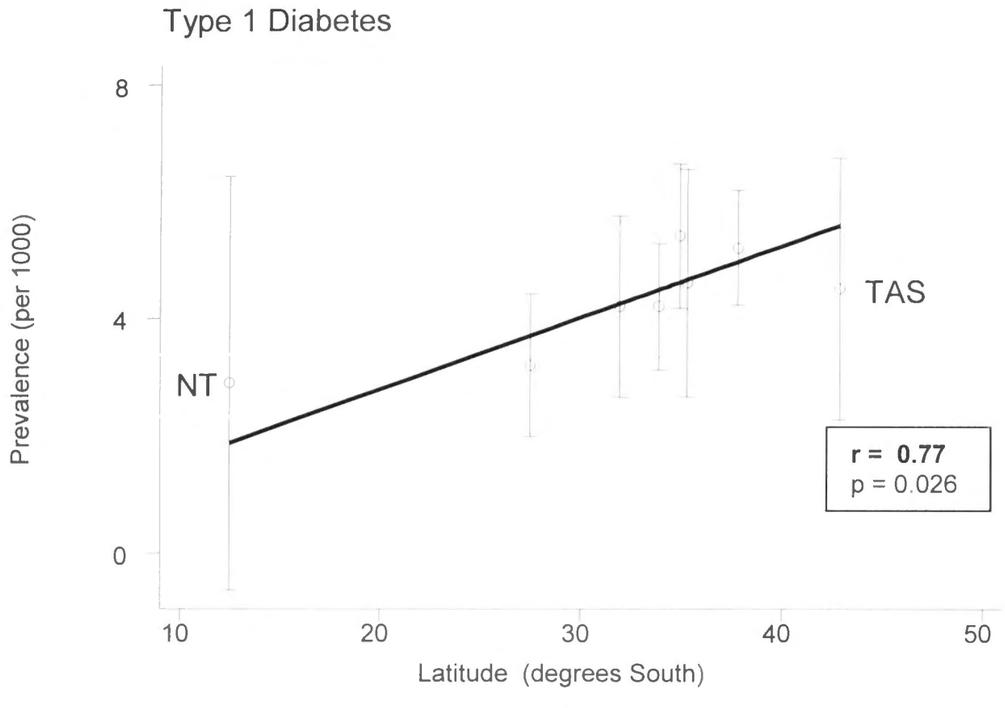
4.3 Results

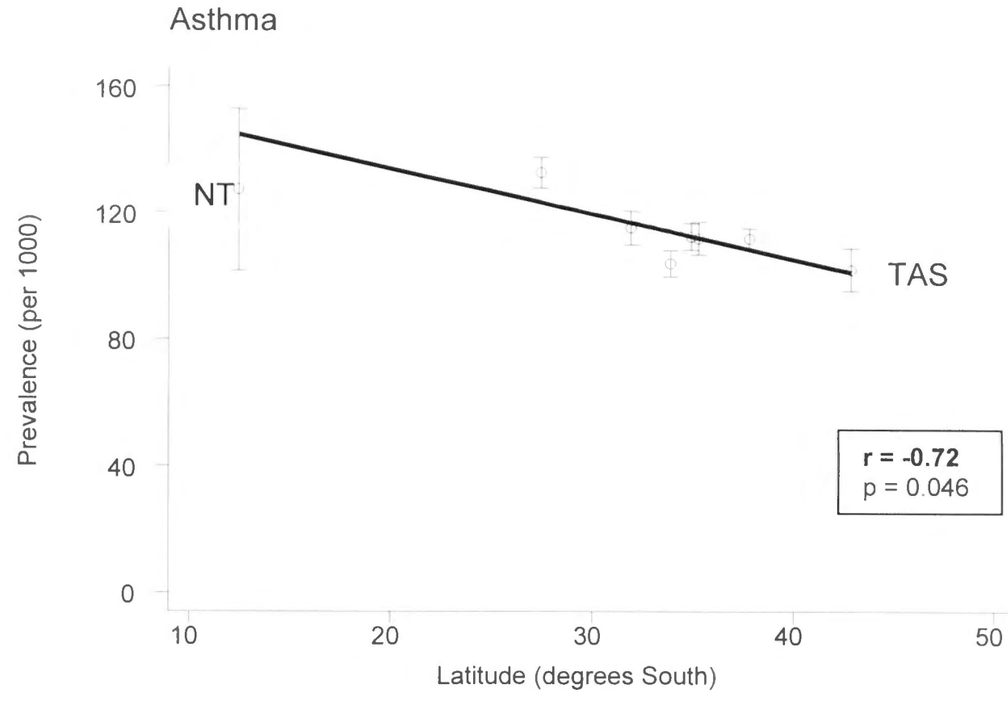
4.3.1 Latitude and immune disorders

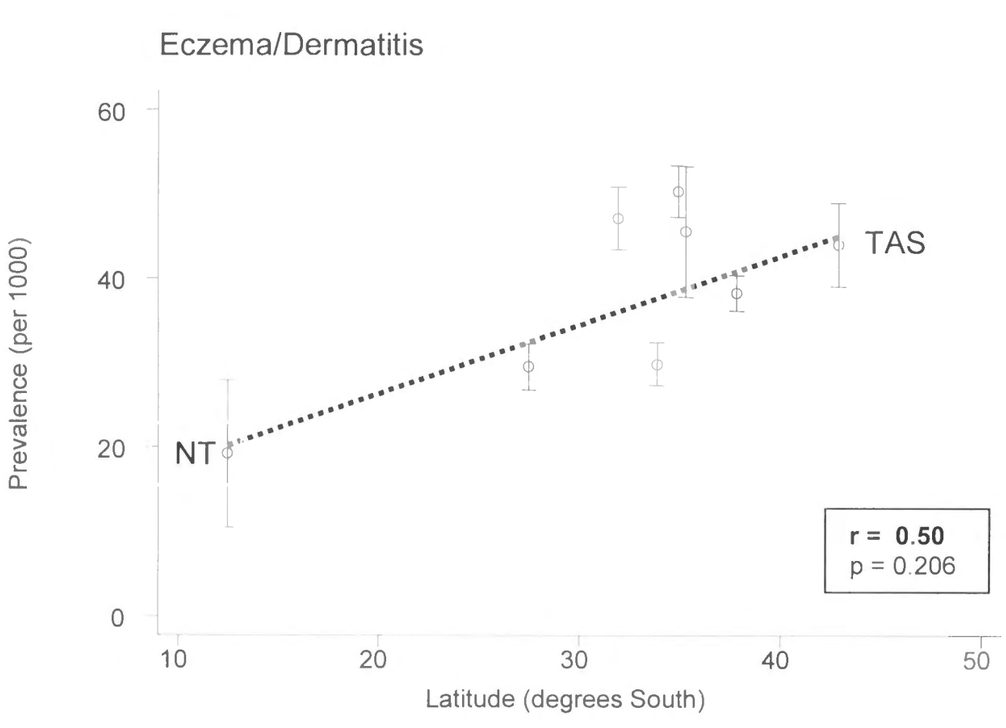
The relationships between latitude of the regional capitals and immune disorder prevalence are shown in Figure 4.5 as regression lines fitted to the eight regional prevalence estimates for each immune disorder; 95% confidence intervals are also

shown and the width of these indicate, reciprocally, the relative weighting applied to each estimate in the regression analysis.

Prevalence of type 1 diabetes was positively correlated with latitude (Pearson r=0.77; p=0.026), the prevalence increasing 2.97-fold over the north-south latitude gradient. Conversely, asthma prevalence was inversely correlated with latitude (Pearson r=-0.72; p=0.046), the prevalence rate decreasing 0.7-fold—that is, by approximately one-third—over the same latitude range. Although both eczema/dermatitis and, to a much lesser extent, RA showed trends of increasing prevalence with increasing latitude, these were not statistically significant (Figure 4.5).


Type 1 Diabetes
Prevalence (per 1000)
8
4
0
NT
TAS
r = 0.77
p = 0.026
10
20
30
40
50
Latitude (degrees South)


Asthma
Prevalence (per 1000)
160
120
80
40
0
NT
TAS
r = -0.72
p = 0.046
10
20
30
40
50
Latitude (degrees South)


Eczema/Dermatitis
Prevalence (per 1000)
60
40
20
0
NT
TAS
r = 0.50
p = 0.206
10
20
30
40
50
Latitude (degrees South)

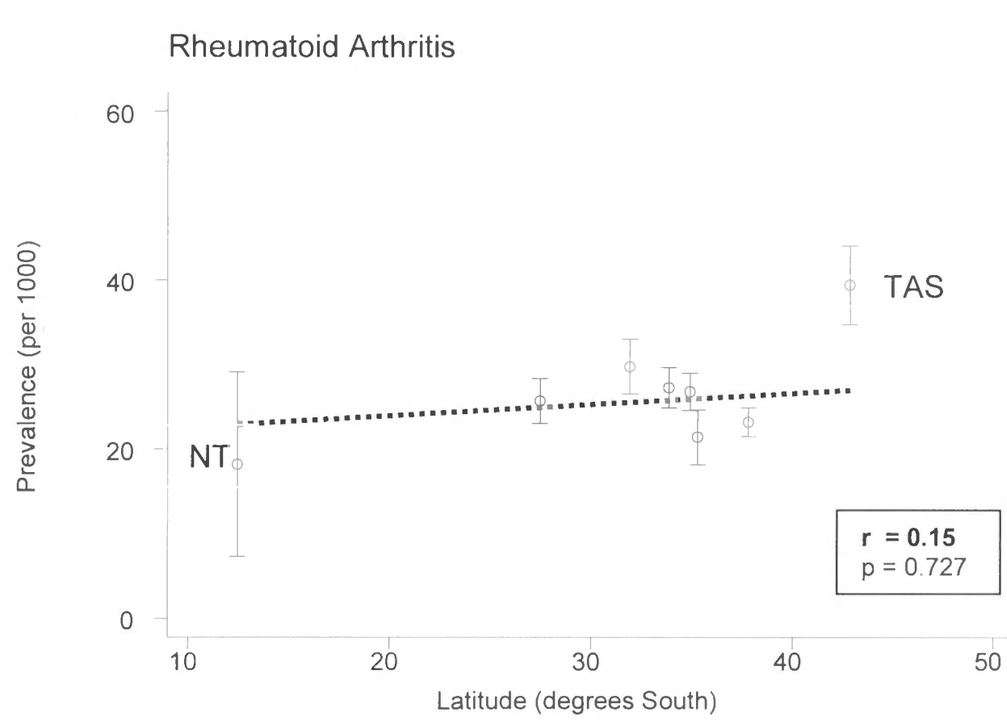

Rheumatoid Arthritis
Prevalence (per 1000)
60
40
20
0
NT
TAS
r = 0.15
p = 0.727
10
20
30
40
50
Latitude (degrees South)

Figure 4.5: Associations between latitude and prevalence per 1000 of type 1 diabetes, asthma, eczema/dermatitis and RA. (95% confidence intervals of prevalence estimates shown as error bars; solid lines denote statistically significant association, $p < 0.05$; dotted lines denote non-significant trend.)

Sensitivity analysis

Since the aim of this study was to relate regional disease prevalence to the latitude where most of the population resided within each region, the proportion of the population living in the regional capitals was checked using both the 1991 and 1996 Australian censuses. In four of the eight regions, WA, SA, VIC and the ACT, at least 70% of the state or territory populations resided in the capital metropolitan area (ABS Census data for 1991/96, not shown). Among the remaining regions, where the population was more dispersed outside of the capital, only the NT and QLD covered a wide latitudinal range and had their capital cities located non-centrally in terms of the regional latitude range, thus potentially biasing the associations.

Alternative latitude values for both the NT and QLD were therefore investigated in this sensitivity analysis: For the NT, a latitude value of 18.1 degrees south, midway between the two main population centres, Darwin and Alice Springs, was substituted. For QLD, a latitude value of 23.4 degrees south, midway between Brisbane and Townsville, was used. However, the effect on the initial associations of using these alternative latitudes was to strengthen the statistical significance of the associations for type 1 diabetes, asthma and eczema/dermatitis when one or both midway latitudes were substituted. For example, for type 1 diabetes and asthma, the maximum increase in statistical significance of the associations (from p=0.026 to p=0.010 for diabetes (r=0.77 to 0.84), and from p=0.046 to p=0.013 (r=-0.72 to -0.82) for asthma) resulted when midway latitudes were used for both regions. Less deviation resulted when only one latitude value was substituted, but this was still in the more significant direction, that is, away from the 'null', or no-association, situation.

Both approaches unavoidably entail some exposure misclassification for dispersed regional populations. Whereas the latter (midway-latitude) approach may lead to better estimation of the prevalence-latitude correlation, by using the capital latitude values for type 1 diabetes and asthma as first conducted (see Table 4.1), the estimates were somewhat biased towards the null hypothesis of 'no association' and thus towards under-estimation of the Pearson correlation coefficient, 'r'. Similarly for eczema/dermatitis the maximum rise in statistical significance (from p=0.206 to p=0.174 [r=0.50 to 0.54]) was obtained when the QLD latitude, only, was altered but this was still non-significant at the 5% level. For RA, conversely, using both midway latitudes gave maximal change but the statistical significance of the observed trend was lowered (from p=0.727 to p=0.795 [r=0.15 to 0.11]).

In summary, the use of the regional-capital latitude values, rather than latitude values of the midway points, was more conservative and did not alter the overall conclusions of whether or not the observed associations were statistically significant at the 5% level.

4.3.2 Regional UVR and immune disorders

Because the use of the regional capital-city latitudes was able to be justified by the sensitivity analysis, capital-city solar-noon UVR values were also used. Midday, or noon-time, exposure to the maximal UVR level in sunlight has been shown to suppress the systemic immune response in humans in Australia [629]. The amount of skin-damaging UV radiation expected to reach the Earth's surface when the sun is highest in the sky (i.e. UVI) is thus an appropriate UVR measure and is now an internationally recognised standard for expressing solar UV radiation levels [359]. Type 1 diabetes prevalence was inversely correlated with regional average monthly (solar-noon) UVR (Pearson r=-0.80; p=0.018), whereas asthma prevalence was positively correlated with the same UVR measure (Pearson r=0.73; p=0.040) (Figure 4.6). Eczema/dermatitis and RA prevalences showed inverse trends with all of the measures of UVR (average monthly, midwinter [June] and midsummer [January]), but none of these was significant (Table 4.3).

Comparing the three different ambient UVR measures, the prevalence of asthma was positively correlated with either average monthly or midwinter UVR, but not midsummer UVR (at the 5% level). While type 1 diabetes prevalence was inversely correlated with all three ambient UVR measures, it was more closely correlated with average monthly or midwinter UVR than with midsummer UVR (Table 4.3).

Table 4.3: Correlations between immune disorder prevalence in the 1995 NHS and regional solar-noon UVR levels for 1979 to 1993. (Pearson correlation coefficients[+] [p-values[#] in parentheses])

Immune disorder	Average UVR	Midwinter UVR	Midsummer UVR
Type 1 diabetes	-0.80 (0.018)*	-0.77 (0.024)*	-0.72 (0.045)*
Eczema/dermatitis	-0.47 (0.243)	-0.49 (0.215)	-0.34 (0.412)
RA	-0.08 (0.845)	-0.06 (0.880)	-0.12 (0.774)
Asthma	0.73 (0.040)*	0.72 (0.044)*	0.68 (0.060)

[+]Pearson correlation coefficients (correlations based on the eight Australian state/territory regions) between immune disorder prevalence and solar-noon UVR, for each type of immune disorder and three different measures of UVR—monthly average over year, midwinter (minimum) and midsummer (maximum).
[#]Strength of association examined by variance-weighted least squares regression.
*Statistically significant association, $p < 0.05$.

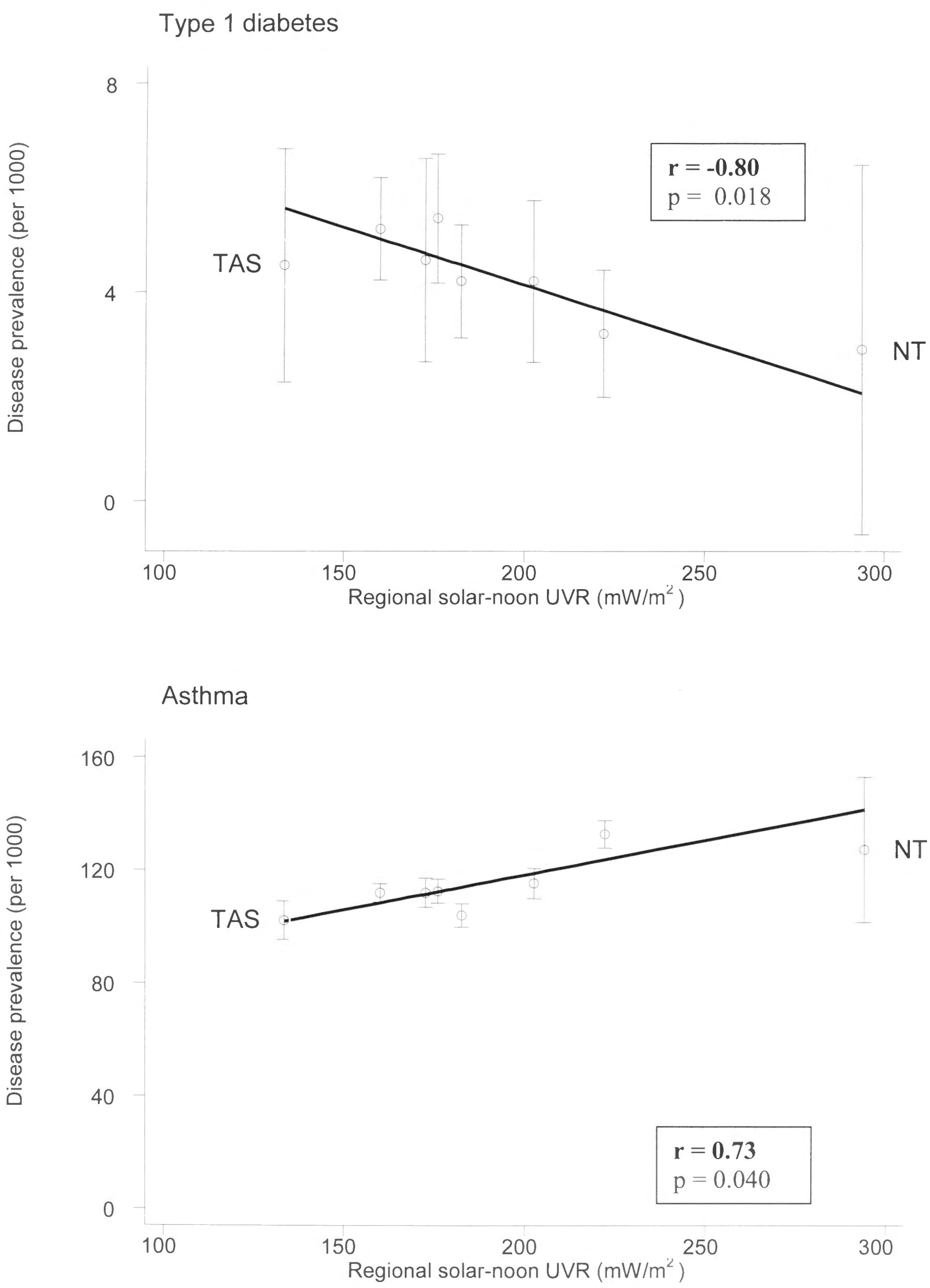


Figure 4.6: Associations between average monthly ambient UVR and prevalence per 1000 of type 1 diabetes and asthma. (95% confidence intervals of prevalence estimates shown as error bars.)

4.4 Discussion

Findings

Using summary age- and sex-standardised data on self-reported prevalence of immune disorders from the ABS 1995 NHS, strong gradients in type 1 diabetes prevalence with latitude and, inversely, with all UVR measures were found. The magnitude of change for type 1 diabetes in Australia was an approximately three-fold increase in prevalence from the lower-latitude northernmost region, the NT, to the higher-latitude southernmost region, TAS. These ecological data provide support for a previously proposed beneficial effect of UVR on autoimmune disorders such as type 1 diabetes [5]. In contrast to type 1 diabetes, asthma correlated negatively with latitude and positively with regional annual or midwinter UVR, but the magnitude of change for asthma prevalence from the north to south of Australia was only 0.7-fold. For both RA and eczema/dermatitis there were possible trends but no statistically significant associations between latitude or UVR and disease prevalence.

Strengths of study

A strength of this ecological analysis derives from the relative genetic homogeneity of the Australian population and relative uniformity also in the standard of national health care [301]. A further advantage is the wide latitude range and the resulting wide range of ambient UVR levels over the Australian continent. A high total response rate to the survey was achieved by the ABS over all regions.

Sources of error

Outcome classification: One potential source of (probably random) error lies in the self-reported, non-verified nature of the data, which could have resulted in some misclassification of disease conditions. This is particularly so for eczema/dermatitis, the classification of which appears from the questionnaire to have been based largely on medications usage; there were no direct disease-specific questions on eczema [625]. The classification 'eczema' is often used as an 'umbrella' term, encompassing various dry, itchy skin conditions [630]; 'dermatitis' similarly can include contact dermatitis. The 'eczema/dermatitis' category in the NHS further included heat eczema and sunburn [13], and may have led to possible regional differences in prevalence.

RA, also, is difficult to differentiate from unspecified polyarthritis [292]. Therefore, misclassification between rheumatoid, osteoarthritis and general arthritic or 'rheumatism' conditions could have occurred even though 'arthritis' was allocated four specific questions. Likewise, outcome misclassification of self-reported 'asthma' may have occurred, this condition being sometimes difficult to differentiate from general 'wheezing' disorders, in spite of specific questions on medications usage [625].

Type 1 diabetes, conversely, was classified on the basis of 24 direct questions for diabetes and insulin usage, including duration of use, both past and future expected use and age at first use [625]. Therefore, the relative validity of classification of type 1 diabetes, a serious disease requiring specific treatment, was probably more regionally consistent.

Exposure classification: Some exposure misclassification at the ecological (i.e. state or territory) level compared to the individual level could have occurred because actual personal UVR exposure depends on behaviour in relation to the sun (i.e. whether sun-avoidant or taking sun protection measures) as well as on regional ambient UVR. For example, regional UVR would not have been a good measure of personal sun exposure among sun-avoidant people, and if the proportion of such people in each local population varied markedly, then population-level exposure assessment would be biased. In fact, the recorded proportion of 'sun-avoidant' people was small and did not vary substantially by region. That is, the percentage of people who reported in the survey 'don't go out in sun' was 3.4% for the NT, 2.0% for QLD and 2.3% for TAS [626], which is perhaps surprising given that the outside temperatures in northern Australia are much higher than in southern Australia in summer or winter. Further, for most individuals, personal UVR exposure varies only between 5 and 15% of daily total ambient UVR [631]. Thus, ambient UVR levels may provide a reasonable measure of average personal sun exposure at the population level.

Another possible source of measurement error—using capital-city latitudes and UVR values as representative of each region—was shown by sensitivity analysis to be not significantly biasing the results, even for the two regions having greater non-metropolitan distribution together with the most non-central capitals and the widest latitude ranges, namely, QLD and the NT. Further, mean state latitude calculated from population distribution, as used in Chapter 5, is subsequently shown to be minorly different from the capital-city latitudes used here (see Chapter 5, Section 5.1, Figure

5.2). However, only current residence was considered in this study and the contribution of prior residence areas could not be taken into account, nor the timing of the critical UVR exposure, which may have occurred earlier in life, nor even possible migration after disease initiation, because prevalence rather than incidence data were used. It was also not possible to control for latitude when examining the relationships between the immune disorders and UVR, because of the high degree of collinearity between UVR and latitude.

Possible confounders

Potential ecological confounders include other possible causal factors for immune-related disorders that vary with latitude or UVR; for example, regional infection patterns that may be associated with climatic differences. Infectious agents have been linked to the aetiology of immune disorders, including type 1 diabetes [136, 198, 240, 601, 632], RA [194, 292, 494, 607, 633] (see Chapter 3) and asthma [307, 634]. As discussed in Chapter 3 (see Section 3.1.2), a *lack* of infections in early life may also adversely affect immune development—that is, the 'hygiene hypothesis' [3]. Infectious agents may be either causal or even protective for such immune disorders [292, 497, 587, 592], depending on conditions and particularly their timing of action in the life course (Chapter 3).

Other confounding factors could include environmental temperature and dietary differences [198, 240, 592] (Chapter 3). These potential confounders warrant more detailed consideration in future research. There may also be interaction between some of these environmental exposures; for example, infection and nutritional factors [241]. However, as indicated by Ponsonby and colleagues, it is not possible to determine such interactions in ecological analyses, since there is a lack of data on joint environmental exposures at the individual level [358].

In this study, UVR has been shown to have either an inverse, or no, association with the immune-related disorders analysed, except for asthma, which showed a positive association between UVR and prevalence. The significance of this finding for asthma is uncertain, particularly as the prevalence increase over the UVR range was only low. UVR-induced Th2 up-regulation mediated by vitamin D [52] would explain this result, even though other workers have noted suppression rather than stimulation of Th2 by UVR [365], particularly through IL-10 [114]. However, much asthma at the population level may even be non-allergic [635] with Th2 mechanisms not involved. In addition,

the contribution of regional allergen levels to asthma prevalence could be important. For example, an 11-fold higher level of mean house dust mite allergen concentration in homes in Sydney, NSW (low latitude, warm and more humid), compared to TAS (higher latitude, cool and dry) [636] may explain the inverse latitude gradient for asthma.

An inverse latitude gradient for asthma was also found in a subsequent Australian study but this association was accounted for by regional (average daily) temperature [637]. This reinforces the likelihood that the findings here for asthma may be explained by factors related to regional temperature, such as house dust mite levels, an examination beyond the scope of this investigation.

Latitude, UVR and type 1 diabetes

The association found between latitude and type 1 diabetes prevalence in Australia is consistent with similar incidence gradients found in Western Europe and North America [30, 313, 314] and within China [315], and has now been shown for the first time in the southern hemisphere. The corresponding inverse association between regional ambient UVR levels in Australia and type 1 diabetes prevalence extends this latitude-gradient finding for type 1 diabetes and is also now consistent with the strong inverse UVR association found for MS in another Australian ecological study [9]. Subsequent to this chapter's findings (published 2003, see Appendix I), ecological analyses of both MS and type 1 diabetes largely in the northern hemisphere have further confirmed inverse associations between UVR levels and disease prevalence or incidence rates. For example, northern hemisphere findings for MS confirm a strong ecological association between UVR and MS prevalence in North America by geospatial analysis [359], and in an ethnically homogeneous farming population in France [189, 289]. For type 1 diabetes, worldwide incidence rates for 1990 to 1994 (WHO multinational 'DiaMond' project) [316] have been subsequently analysed versus regional UVB radiation levels taking per capita health expenditure into account, an inverse relationship between UVR and risk of type 1 diabetes being reported over 51 regions [317]. In Australia, however, a recent study of regional residential UVR and type 1 diabetes incidence (also subsequent to this thesis study) was able to support the present findings for type 1 diabetes only in low-density (i.e. rural) populations, the association being reversed in high-density (urban) populations [351].

The findings in this chapter for type 1 diabetes are also consistent with photo-immunological work showing that UVB irradiation has systemic as well as local

immunosuppressive effects in humans and animals [362-365, 368, 371], as does UVA irradiation [362, 371, 374] (see Chapter 2). UVR exposure may be protective against disorders such as type 1 diabetes, by down-regulating effector T-cell (Th1/Th17) autoimmune responses by several different immunoregulatory mechanisms, including direct effects on antigen presentation, and promotion and induction of regulatory T cells [362, 368-371, 373] (see Chapter 2).

As discussed in Chapter 2 (Section 2.4.2), another of these possible mechanisms for immunosuppression involves UVR-induced vitamin D, which, like direct UVR, can inhibit Th1 and Th17 cell responses (via effects on DCs) and promote and induce regulatory T cells, but also enhance Th2 cells [52]. That is, the proposed protective role of UVR for both MS and type 1 diabetes may be mediated at least in part through its important role in vitamin D synthesis in the skin. For example, a recent multicentre incidence study of FDE, a common precursor to MS, in Australia showed that a four-fold increase in FDE incidence from Brisbane, QLD, to TAS regions [305] was partly accounted for by both past sun exposure and vitamin D status [361]. Some 90% of plasma vitamin D in humans is produced endogenously via skin exposure to UVR in sunshine [242], and particularly so in Australia where foods are not generally fortified with vitamin D [456, 457, 638]. There are also recent reports of vitamin D deficiency in some Australian populations [458, 459, 461], particularly in winter [457, 462].

For type 1 diabetes, this possible vitamin D mechanism is consistent with reports of decreased risk of this disorder in offspring of mothers supplemented with cod-liver oil or other vitamin D foods in pregnancy [437, 439], and decreased incidence of type 1 diabetes in children supplemented with vitamin D in infancy [435, 436, 438, 442] (see Chapter 2, Section 2.5). The Finnish birth cohort study, for example, reported an incidence rate ratio (IRR) of 0.12 (95% CI 0.03, 0.51) for diagnosis of type 1 diabetes by the age of one year, comparing regular versus no vitamin D supplementation in the first year of life [435]. A recent meta-analysis of four case-control studies further concluded that children supplemented with vitamin D in infancy had a 29% lower risk of developing type 1 diabetes than those not supplemented [444] (see Chapter 2). These studies suggest that vitamin D supplementation may prevent initiation of type 1 diabetes, and thus accord with the vitamin D-protective hypothesis proposed for this and other organ-specific autoimmune disorders, such as MS [25, 357].

This finding for ambient UVR and type 1 diabetes in Australia supports the specific prediction that autoimmune diseases other than MS, such as type 1 diabetes, should

show latitude and/or UVR gradients analogous to those seen for MS if these immune disorders are similarly influenced by UVR exposure [5].

Type 1 diabetes and MS—underlying similarities and common mechanisms?

Although type 1 diabetes and MS are organ-specific autoimmune disorders with disparate target organs and clinical manifestations, there is evidence of underlying similarities in epidemiology, as well as genetics and pathogenesis, and possibly similar causal influences for these two autoimmune conditions in particular. As reviewed by Handel and colleagues, type 1 diabetes and MS may be more similar than any other pair of autoimmune conditions. Importantly, most of this similarity appears to be attributable to environmental rather than genetic factors, and study of one of these autoimmune disorders may therefore illuminate pathogenesis and possible prevention of either [211].

The present UVR-associated ecological findings for type 1 diabetes that are consistent with those for MS in Australia, and with findings for both disorders in the northern hemisphere, support the notion of possibly similar aetiological mechanisms for these two autoimmune disorders. Individual-level study of such disorders is now required and, in particular, the *timing* of putative environmental exposures, such as UVR, in the life course before disease onset [6]. Investigation of possible seasonal and other early life influences (seasonal variation being a property of UVR and a feature also of other candidate factors such as infections) is now needed, particularly for these disorders in the southern hemisphere; therefore, the remainder of this thesis will focus on MS and utilise an existing Australian MS dataset for these purposes.

4.5 Conclusion

In conclusion, ecological analysis of existing prevalence data from the 1995 NHS has demonstrated a regional gradient of type 1 diabetes prevalence within Australia that is inversely associated with regional ambient UVR levels. The inverse association with UVR is consistent with type 1 diabetes latitudinal gradients seen in the northern hemisphere, and consistent with that found for another organ-specific autoimmune disease, MS, in Australia and North America. The finding is consistent also with photo-immunological evidence of UVR-induced immunosuppression and suggests a beneficial effect of UVR in preventing both these immune-related conditions. Analytical

epidemiology studies, investigating risks of type 1 diabetes and MS in relation to other facets of UVR exposure in humans, such as seasonal (monthly) variation and critical timing of exposure, are now required—these will be explored for MS in the following chapters of this thesis.

CHAPTER 5

TIMING OF BIRTH AND MS RISK IN AUSTRALIA: THE 1981 AUSTRALIAN MS SURVEY—VERIFICATION OF NUMERATORS FOR TIMING-OF-BIRTH ANALYSES

5.0 Preface

The ecological study in Chapter 4 adds to previous findings and suggests that at the population level, a gradient of increasing latitude and decreasing UVR is associated with an increase in prevalence of organ-specific autoimmune disorders such as MS and type 1 diabetes. UVR varies seasonally as well as latitudinally and it would now be informative to investigate this latitude-UVR-disease relationship in more detail using data at the individual level.

Additionally, the *timing* of action of putative environmental determinants, such as UVR, on such organ-specific disorders may indicate the *nature* of these factors (see Chapter 2). For example, a causal (or protective) factor that varies seasonally and operates around the time of birth may contribute to a seasonal or other temporal birth 'pattern' among these immune disease patients that is not seen in the general population. That is, such patients may be born in higher (or lower) frequencies in some months relative to others, if the environmental factor is indeed causal or protective and if the critical period for action of this factor is close to the time of birth. Such a 'peak', or excess, in births of people that subsequently develop the autoimmune disease may be a result of seasonally varying environmental factors acting around the time of birth, including UVR, temperature and infections. That is, the risk of developing an organ-specific autoimmune disease, such as MS or type 1 diabetes, at some time after birth may be influenced by the time of the year in which gestation and birth occurs.

Disease registers of incident cases have frequently been used to examine seasonal birth patterns—for example, in psychotic disorders such as schizophrenia in several northern and southern hemisphere countries [639-642]. Neurological autoimmune disorders such as MS have also been studied thus, but to a lesser extent than schizophrenia and largely in the northern hemisphere [7, 324, 643].

In these case-register studies, consecutive patients are typically assessed and diagnosed at a disease clinic over a specified period of time (i.e. longitudinal data). An example for both schizophrenia and MS is the Danish study by Templer and co-workers (1992) where 'all schizophrenia patients with onset between 1970 and 1987 (~9000 cases)' and 'all MS patients with onset between 1950 and 1984 (~6000 cases)' were registered. They were then compared with the general population born between 1901 and 1960 (~4.5M persons) [321]—that is, those born around the same time as the cases taking into account time to disease onset (an historical or retrospective cohort study). Timing of birth of the disease cases relative to the population controls can then be examined and any periodicity existing across the years determined [642].

Type 1 diabetes has been similarly studied for temporal birth patterns using incident case registers, but nearly all of these studies have been based in the northern hemisphere (see Chapter 2). A recent exception is the study by Elliott and colleagues on timing of both birth and disease onset in Australia using a national type 1 diabetes register, which found no season-of-birth effect [351].

In the case of MS, a large population-based longitudinal study by Willer and colleagues has been completed in the northern hemisphere, where a timing-of-birth pattern has been shown using register data from 19 MS clinics in major cities across Canada [324]. These 19 registers formed the CCPGSMS, totalling 17,874 Canadian patients, and were studied together with additional incident case data from a population study in Scotland plus cases derived from death records from the UK (another 11,502 British patients). Pooled with further Scandinavian data, this large register study showed a deficit in MS births in November together with an increase in MS births in May [324]. Another register study in British Columbia, Canada confirmed the same May to November peak-and-trough pattern in MS-case births there [644]. Willer and colleagues (2005) further reported a greater timing-of-birth effect size in the country with highest MS prevalence in their study, namely Scotland [324].

In Australia, such national registers for MS are not available. However, a comprehensive epidemiological survey of known MS cases in Australia was conducted by Professor J. McLeod, University of Sydney, NSW, on 30 June 1981, this date chosen to coincide with a national census. These data provided much detailed epidemiological and clinical information for MS in Australia and the southern hemisphere, and resulted in several key publications (Table 5.1) that put antipodean MS 'on the map'. Although not designed for the present purpose, this national MS

survey dataset provides an opportunity to examine the timing-of-birth hypothesis for MS in the southern hemisphere. Thus, MS has been chosen for detailed individual study for the remainder of this thesis.

How do we best use an existing cross-sectional dataset to analyse timing of birth in MS?

In order to examine a timing-of-birth effect on MS risk, a dataset of MS incidence rates by birth month and by birth year is needed, to allow a longitudinal type of study to be conducted. It is possible to create such a dataset from McLeod's 1981 MS prevalence survey, by linking with available reference-population data, including the Australian 1981 Census of Population and Housing. That is, the McLeod dataset provides the MS cases by birth month and year (the *numerators*), and the Australian census and other available register data will provide the number of births by month and by year for the reference population (the *denominators*), as shown diagrammatically in Figure 5.1.

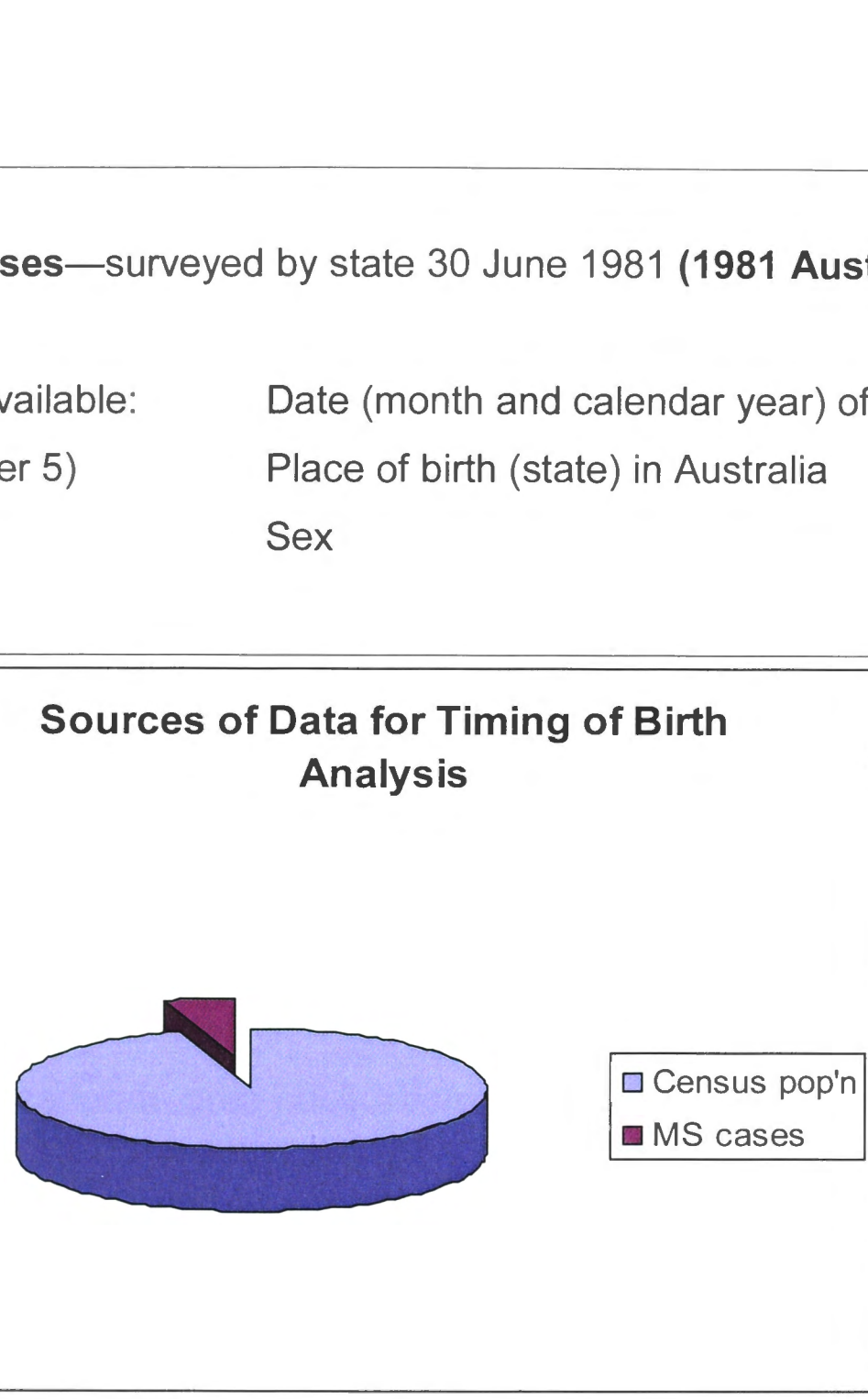
The present chapter describes the available unit-record, cases-only, cross-sectional 1981 MS Survey dataset, focusing on the numerators required for analysis of seasonal or other timing-of-birth patterns. For initial exploratory and data-checking purposes, the 1981 Australian census data from the same day in 1981 will also be used to compare MS prevalence rates by survey area, sex and age with those described in the relevant McLeod research-group publications given in Table 5.1. Early-, mid- and late-born cohorts will be further investigated with the aim of determining the optimal sample for timing-of-birth analysis.

The next chapter, Chapter 6, will then focus on determining appropriate reference-population denominators for the surveyed MS cases, specifically for timing-of-birth analysis. The 1981 Australian Census of Population and Housing will be used, together with other Australian births-registration data necessary to supplement the available census data (Figure 5.1).

MS Cases—surveyed by state 30 June 1981 **(1981 Australian MS Survey)**

Data available: Date (month and calendar year) of birth (1920 to 1950)
(Chapter 5) Place of birth (state) in Australia
Sex

Sources of Data for Timing of Birth Analysis



Reference (census) population—census date 30 June 1981

Data available: Age in completed years on 30.06.81
(→ census year of birth)
Country of birth
Sex

Unavailable: Month and place of birth (state)
(→ to be estimated using supplementary Australian births registrations (by state) covering same years of birth
(Chapter 6)

Figure 5.1: Sources of data for analysis of timing of birth in MS cases.

5.1 Introduction—the 1981 Australian MS Survey

The nationwide MS survey, coordinated by Professor J. McLeod, Discipline of Medicine, University of Sydney, was conducted in 1981 throughout Australia, with the prevalence day being taken as 30 June 1981, the same day as a national census conducted by the ABS. Ascertained MS cases were confirmed by the study neurologists in each survey area, and standardisation of procedures was ensured by regular meetings of survey coordinators from each area [299, 645].

A primary aim of this original 1981 survey was to determine the relationship between MS frequency and latitude within Australia [301]. Such a relationship between southern latitude and MS prevalence had been suggested by earlier, separate, regional and metropolitan surveys in 1961 and was confirmed by the 1981 survey in the McLeod group's separate-area publications [301, 645-647] (Table 5.1).

The Australian MS survey further aimed to compare MS prevalence in 1981 with that recorded 20 years earlier in 1961 [301, 646-648] (Table 5.1). Additional clinical findings resulted, this survey being the largest population-based clinical study of MS then undertaken [649, 650]. For example, clinical prognoses and survival estimates confirmed those found in the northern hemisphere [177]; for example, a worse prognosis for later age of onset [650].

Table 5.1: Selected analyses and findings based on the 1981 MS Survey and published by the McLeod research group, 1987 to 2011

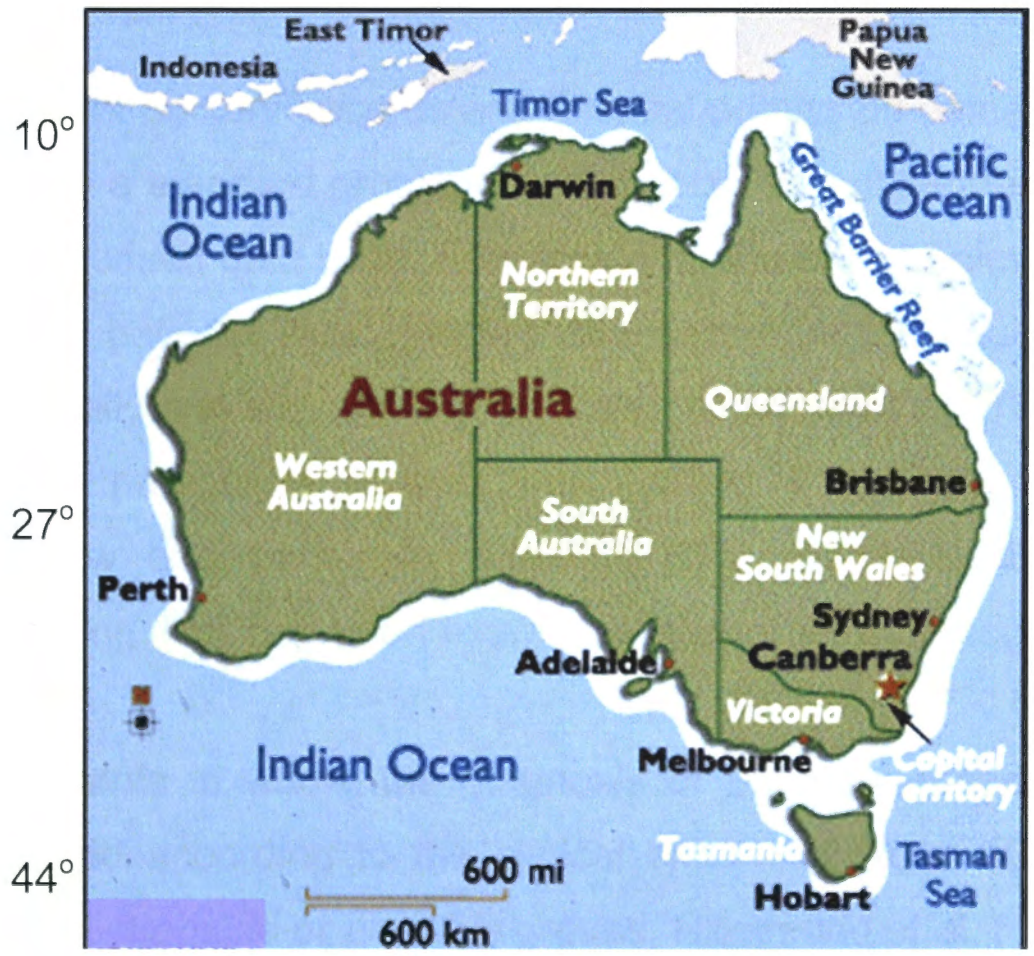
Survey area analysed	Main subject/findings	Publication reference
QLD (tropical/subtropical zones)	1981/1961 prevalence; 1981 prevalence versus latitude difference within QLD	Hammond et al., 1987 [646]
Perth SD (WA), Newcastle SDist (NSW), Hobart SD (TAS)	1981/1961 prevalence and incidence; 1981/1961 prevalence versus latitude relationship	Hammond et al., 1988a [647]
QLD, Perth SD (WA), Newcastle SDist (NSW), Hobart SD (TAS)	Clinical differences between medium- (QLD) and high-frequency (3 cities) prevalence zones	Hammond et al., 1988b [649]
WA	1981/1961 prevalence	Hammond et al., 1988c [648]
QLD, WA, SA, Newcastle SDist (NSW), Hobart SD (TAS)	Mortality-prevalence-latitude relationship (mortality—all states and UK)	Hammond et al., 1989a [299]
QLD (tropical/subtropical), WA, Perth SD (WA), SA, Adelaide SD (SA), Newcastle SDist (NSW), Hobart SD (TAS)	1981 prevalence versus latitude gradient	Hammond et al., 1989b [645]
Nine regions, Australia & NZ	Prevalence-latitude gradient - tropical QLD/southern NZ	Miller et al., 1990 [302]
NSW (whole state), SA	1981/1961 prevalence; 1981 prevalence versus latitude relationship (including all states and survey areas in Hammond et al., 1989b)	McLeod et al., 1994 [301]
QLD, WA, NSW, SA, Hobart SD (TAS)	Socio-economic factors	Hammond et al., 1996 [534]

Survey area analysed	Main subject/findings	Publication reference
QLD, WA, NSW, SA, Hobart SD (TAS)	Migration study: risk with respect to age of UK/Ireland migrants to Australia (latitude-gradient differences, prevalence, age range)	Hammond et al., 2000a [304]; McLeod et al., 2011 [312]*
QLD, NSW, WA, SA, Hobart SD (TAS), VIC	1981 Clinical prognostic factors for disability	Hammond et al., 2000b [650]
Newcastle SDist (NSW)	1996/1981/1961 prevalence and incidence	Barnett et al., 2003 [246]
Newcastle SDist (NSW)	Long-term prognostic survival and disability	McLeod et al., 2007 [177]

* Data in Hammond et al. (2000a) [304] re-analysed in McLeod et al. (2011) [312].
SD, statistical division; SDist, statistical district.

Survey case ascertainment and diagnostic classification

All MS cases were recorded and verified by the McLeod research group for the states of QLD [646]; WA [648]; NSW (including ACT) and SA [301]; and the Hobart (TAS state capital) statistical division (SD) [647] on 30 June 1981. These states and areas are shown together with their mean south latitude values (mean latitude based on the census-population distribution in 1981), and their capital-city latitude values, in Figure 5.2. The NT and the TAS rural region outside of Hobart were not able to be surveyed, because of a lack of neurologists in those regions (J. McLeod, pers. commn). Case ascertainment from the state of VIC unfortunately was also incomplete (J. McLeod, pers. commn) [534] and so VIC MS cases in the supplied dataset could not be included in this analysis. The required adjustments to the dataset to overcome these data limitations will be discussed further in Chapter 6.



Latitude (°S) of survey areas	Mean	Capital
QLD; Brisbane	25.1	27.5
WA; Perth	31.3	31.9
NSW; Sydney	33.4	33.9
SA; Adelaide	34.9	34.9
SD; Hobart (TAS)	42.8	42.9

Figure 5.2: Australian states and territories and their capital cities, shown together with mean south latitude and capital-city latitude of 1981-surveyed areas (mean south latitude as given by Hammond et al., 2000a [304] and based on 1981-census state population distribution; NSW includes ACT). (Figure from GraphicMaps.com, accessed 23/03/10 WorldAtlas.com.)

In this 1981 Australian survey, the MS cases were sourced from the following:

- Diagnostic indices of major hospitals.
- Practising doctors: a circular letter was sent to all neurologists, internal medicine specialists, neurosurgeons and general practitioners requesting the name, most recent address, approximate date of diagnosis and hospital where notes might be found, of any patient known to them, either currently or in the past, in whom the likely diagnosis was MS. Follow-up procedures included a second mailing of this letter if no reply was received to the first, and personal telephone calls in some instances.
- MS Society (state/territory).
- Department of Veterans' Affairs (State) Diagnostic Index.
- Chronic care hospitals.
- Commonwealth statistician: a request was made to notify the survey of any deaths from MS subsequent to prevalence day (30 June 1981) [301, 646-648].

Details of doctors' records and hospital admission notes were transferred by the survey team to a standard protocol form suitable for subsequent entry of its information into a computerised data base. Permission for the main surveyors (neurologists) to contact all notified patients was obtained and subsequently these patients were all interviewed personally and examined in all states except NSW [301, 646-648]. In NSW, where the highest number of patients was notified, it was only possible for the surveyors to interview and personally examine 57% of the total. However, almost all of the remaining 43% had been examined by another neurologist [301].

All patients in whom the diagnosis of MS was considered to be correct were then classified according to the clinical criteria of Rose et al. (1976) [159] into clinically *definite*, *probable* or *possible* groups. Hammond et al. (1987) [646] and the subsequent McLeod group's publications stress that no laboratory results (e.g. CSF analysis, evoked-potential studies or brain scans) were considered in the allocation of individual patients to particular diagnostic categories. The disability status of the patients on prevalence day was assessed according to the Kurtzke disability status scale (DSS) [651].

For case validation, a 10% sample of the protocol forms was selected on an alphabetical basis and submitted to an independent neurologist arbiter for assessment of the correctness of the diagnosis and of the diagnostic classification into which each

patient had been placed [301, 646-648]. In NSW, this 10% sample was taken from the 57% in whom a personal interview and examination had been possible [301]. In all cases, the veracity of diagnosis of MS was not disputed but the category of the disease was reclassified in 6 to 30% of cases [301].

Survey data variables

The resulting unit-record 1981 MS-case dataset (as received for this thesis by this candidate) comprised the following survey information:

- Survey ID number (five-digit number, the first digit indicating the survey area, for example, 20,001 ... =NSW; 40,001 ... =QLD).
- Name; place of residence (Australian postcode or overseas code); day, month and year of birth; place of birth (postcode or overseas code); sex; age left school; highest level of education.
- Residence history from birth to MS onset.
- Family history: parental birthplaces; race; whether twin; total number siblings and place in family; parental ages at (case's) birth; whether cat or dog in house.
- Onset of illness: month and year of onset (first symptoms); residence at onset; whether first symptoms multiple or single (and which first symptom).
- Source of case information.
- Postcode on each of two intended prevalence dates, P1 (1976) and P2 (1981) [301].
- Status and diagnosis on P1 and P2: Two sets of clinical variables for the two intended prevalence dates were included in the dataset. Variables included: diagnostic category; clinical course (e.g. relapsing-remitting and secondary progressive v. primary progressive); clinical type (further clinical subcategories, seven types); grade of stair disability (zero to four); symptoms (Kurtzke grade zero to five or six) for each of pyramidal, cerebellar, brain stem, sensory, bowel and bladder, visual, mental or other disability; Kurtzke DSS (zero to nine).
- Date and postcode of death (if patient had died prior to P1 or P2).

Table 5.2 lists the variable names in the dataset as received, and distinguishes those variables that were primary (i.e. from the original patient interview) and those that had been derived by subsequent McLeod group investigators for various other analyses.

Table 5.2: 1981 MS Survey dataset: original and derived variables and their description (variables used in this thesis, for either case verification or analysis, shown in bold)

Data category	Variable names	Variable description
ORIGINAL (primary) data:		
ID survey no.	**svn1**	Survey (case) number identifier
Name, residence	surname1, surname2, frstnam1, **pc**	Surname name, first name; postcode of residence (in Australia)
Date & place of birth, sex	dd, **mm, yy, pcbirth, sex**	Day, month and year of birth; postcode of birth; sex
Education	agelsch, hied	Age left school; highest education
Residence history (<onset)	pcres1– pcres15	Postcode of residence 1–15 years prior to onset
Parents' birthplace	mpc, fpc	Postcode of birth of mother, father
Race	race	Race (Aboriginal or not)
Family information	**twin, sibs, fampl**, matage, patage	Whether twin; total number of siblings; place in family; age of mother, father, when born
Pets	catdog, cat, dog	Whether cat and/or dog in household
Onset details	mmonset, **yyonset**, pconset	Month, year and postcode of residence at onset
First symptoms	fstsymm, fstsyms	Whether multiple first symptoms; which single first symptom
Information source	srcin	Source of information
Prevalence day P1 (1976)	pcp1; symp1	Postcode on prevalence day P1; whether symptoms on prevalence day P1
P1 diagnosis	diagctp1, clintyp1, clincrp1	Diagnostic category, clinical type and clinical course on P1

Data category	Variable names	Variable description
P1 clinical info.	stdisp1, pyram1, cerbp1, bstemp1, sensp1, bowelp1, visp1, mentp1, otherp1; kurp1	Symptoms on P1 (and Kurtzke grades): stair disability, pyramidal, cerebellar, brainstem, sensory, bowel and bladder, visual, mental, other; Kurtzke DSS on P1
P1 death info.	**ddiep1, mdiep1, pcdiep1**	Day, month and postcode of death (if dead on P1)
Prevalence day P2 (1981)	**pcp2**; marp2	Postcode on prevalence day P2; marital status on prevalence day P2
P2 diagnosis	**diagctp2, clintyp2, clincrp2**	Diagnostic category, clinical type and clinical course on P2
P2 clinical info.	stdisp2, pyram2, cerbp2, bstemp2, sensp2, bowelp2, visp2, mentp2, otherp2; kurp2	Symptoms on P2 (and Kurtzke grades): stair disability, pyramidal, cerebellar, brainstem, sensory, bowel and bladder, visual, mental, other; Kurtzke DSS on P2
P2 death info.	**ddiep2, mdiep2, pcdiep2**	Day, month and postcode of death (if dead on P2)
DERIVED (by subsequent investigators) data:		
Residence history (<onset)	yrrs1– yrrs15	Residence history pre-onset in number of years
Age on P1, P2 dates	age76, age81	Age (years) in 1976 and 1981
MS on P1, P2 dates	ms81, ms76	Whether MS had been diagnosed by 1981 and/or 1976
Age of onset	ageonset	Age at onset
Duration of MS	dur76, dur81	Duration of MS in 1976 and/or 1981
Residence on P1, P2 dates	statep1, statep2	State/territory of residence on P1 and P2
Ascertainment information	instap1, instap2	Whether in state of ascertainment on P1 and P2
New ID	newid	New? ID number

Data category	Variable names	Variable description
Birth in Australia	**ausborn**	Whether born in Australia
Residence on P2	placep2	Place of residence in 1981, P2

5.2 Methods

5.2.1 Initial treatment of national case data

The 1981 MS dataset, as received, comprised identified case data (n=3,277) from the states of QLD, WA, NSW (including ACT), SA and VIC, and the Hobart SD in the state of TAS. Data records were first de-identified by removing names, and the unique survey ID numbers used for individual identification thereafter.

Identification of dead-in-1981 cases

Because the dataset was found to contain some records of deaths prior to the 1981 prevalence day (such cases having been part of the dataset for earlier prevalence-survey dates, Table 5.2), these cases needed to be identified and excluded from the current analyses.[15] The 143 cases dead in 1976 and a further 27 cases dead by 1981 were identified and a new data variable 'alive81' generated for the cases who were alive on 30 June 1981 (n=3,107) (Table 5.3).

[15] Prevalence in 1981 is defined as the ratio of persons with an acceptable diagnosis of MS that are living in the defined area on 30 June 1981 to the total number of persons in the population of the same area on the same day [646-648]. Therefore, inclusion of previously surveyed and now dead (in 1981) MS cases would have over-estimated the 1981 prevalence.

Table 5.3: 1981 MS Survey dataset. Further variables (and their coding) derived specifically for the purposes of this thesis

Variable description	New generated variables (coding)
Confirmation of alive in 1981	**alive81** (1,0)
Confirmation of MS diagnosis in 1981	**diagMS81** (1,0)
Residence in 1981 (state)	**state81** (2-7)*
Birthplace in Australia (state)	**statborn** (2-7, 9)#

* 'state81' coded as 2, 3, 4, 5, 6 or 7, these codes being the first digit of the 4-digit state postcodes and thus denoting the states of NSW/ACT, VIC, QLD, SA, WA and the area of Hobart SD in the state of TAS, respectively.
# 'statborn' coded as for 'state81', with the additional '9' denoting NT born.

Verification of 1981 MS diagnosis

The diagnostic category variable 'diagctp2' (Table 5.2) comprised five main categories (coded one to five), of which only the first three—namely 'definite', 'probable' and 'possible'—were considered by Hammond, McLeod and colleagues to constitute 'diagnosed MS' [301, 646-649]. A new variable 'diagMS81' was generated (Table 5.3), comprising just the diagnostic categories one to three, the first two of these categories ('definite' and 'probable') constituting 88% of the diagnosed cases in the final dataset used. The few cases with missing 'diagctp2' data were considered individually (one case from SA and three cases from VIC) and were retained only if all three other main clinical variables (namely, clinical course, clinical type and Kurtzke DSS) were recorded for 1981. At the same time, it was also ensured that the final total sample numbers by survey area agreed with those published by the McLeod group [301, 646-648]. A total of 2,912 living cases in the supplied dataset were thus verified as having been diagnosed with MS in 1981.

Survey area—verification of place in 1981

Although this thesis study will be primarily concerned with place of *birth* of each MS case rather than place of ascertainment (in order to examine any latitudinal effect on timing of birth of MS—discussed further in Chapters 6 and 7), it was necessary to examine also where each MS case was residing and thus surveyed in 1981, because

some of the states or areas were insufficiently ascertained or not surveyed at all (i.e. VIC, NT and non-Hobart TAS) and adjustments would be required to allow for this. A 'state81' variable was generated (Table 5.3) to describe the place (i.e. survey area, usually state) in Australia where each MS case was resident in 1981, using data from the relevant residence postcodes together with the survey ID number (whose first digit indicated original survey area). Where there were inconsistent or missing residence data, the McLeod group's published total sample numbers by survey area were used as the definitive guide [301, 646-648].

Exclusion of VIC cases

As outlined in Section 5.1, cases surveyed in the state of VIC (i.e. 'state81'=3) were then necessarily removed from the dataset.

Adjustment for incomplete ascertainment of VIC, NT and non-Hobart TAS

Because VIC, NT and non-Hobart TAS were not able to be surveyed in 1981 (J. McLeod, pers. commn), adjustments to the dataset were required to overcome this data limitation for timing-of-birth analysis. This was achieved by adjusting the reference-population denominators in particular and is described fully in Chapter 6 (see Section 6.2.3).

Region of birth (state)

A 'statborn' (denoting state born) data variable was generated (Table 5.3) from the original 'postcode of birth' variable to enable examination of the effect of region of birth (and its mean latitude) on the timing of birth of Australia-born cases (these cases denoted 'ausborn' in the dataset).

The 'ausborn' variable in the dataset, which had been generated by previous investigators to denote whether cases were born in Australia or elsewhere (see Table 5.2), was also used to help determine inconsistencies in coding of the 'postcode of birth' variable. For example, this variable indicated that double-digit postcodes, as opposed to the usual Australian four-digit postcodes, denoted an overseas country of birth. Also, single-digit postcodes one to six occasionally apparent in the dataset, but only in Australia-born cases were determined as being state codes based on an early ABS

system of denoting the states NSW to TAS one to six in order of their population sizes (J. Wall, ABS, pers. commn).

5.2.2 Completion of NSW data—addition of missing records

When the supplied dataset was found to contain only the 57% of the NSW sample that had been followed up and fully verified by the survey neurologists [301], the remaining 43% of NSW data needed to be separately obtained after assurance by the surveyors that all of these remaining cases had indeed been diagnosed with MS in 1981 (J. McLeod, pers. commn). Many (but not all) of the more recent McLeod group's publications—for example McLeod et al. (1994) [301] and Hammond et al. (2000a, b) [304, 650] (Table 5.1)—had examined and described the complete NSW sample, including this 'missing' 43%.

However, unlike the rest of the supplied dataset, the supplementary (43%) NSW sample lacked some data variables required, particularly for the timing-of-birth analyses; for example, *month of birth* and *place of birth*. To gain these data, the original, hard-copy patient files of the 43% NSW sample (831 cases, including 11 from ACT) were accessed by this candidate at the University of Sydney's Discipline of Medicine department; these clinical case notes then provided the required information for the majority (~75%) of the supplementary cases. For the remainder for which birth-date and/or birthplace details were missing in the clinical notes (~25%), a request was then made to the NSW Registry of Births, Deaths and Marriages for this information, using full names and sex as the main search parameters.

The NSW/ACT supplementary data in Microsoft Excel 2000 (n=831 cases) were then incorporated into the overall numerator dataset in the statistical analysis program Stata 8.0 (release 8.0, 2003; Statacorp, College Station, Texas). To aid later identification, these records were also given a unique prefix to their (new) ID survey numbers.

5.2.3 Prevalence comparisons with previous 1981 MS survey publications

After cleaning and completing the unit-record case dataset, the data were summarised using available 1981 census totals in order to obtain basic prevalence estimates by

survey area, sex, age group and year-of-birth cohort. Where applicable, these prevalence estimates were then compared with those published by the McLeod research group, these previous results generally being for just a few, or even a single, survey area per publication (Table 5.1). All MS cases surveyed, whether Australia- or overseas-born, were thus included in these earlier, and the present, estimates.

5.2.4 Consideration of dataset numerators for main analysis

Because the 1981 MS dataset was originally cross-sectional, having been sampled at just one point in time, both numerator case data and the (reference) census-population data were finally considered in detail by decade of birth, in order to investigate any inherent biases and any possible need for restriction of the data for timing-of-birth analysis.

5.3 Results

5.3.1 Numerator dataset cleaning—summary

Including the 831 supplementary NSW/ACT cases, all of whom were confirmed as alive in 1981 and diagnosed with MS (J. McLeod, pers. commn), 3,938 of the total 4,086 cases of the dataset were verified as alive on the prevalence day 30 June 1981 and coded as 'alive81'=1. Living cases numbering 3,743 were then confirmed as having been diagnosed with MS in 1981 ('diagMS81'=1). After exclusion of seven cases because their place of residence in 1981 ('state81') could not be confirmed, a total of 3,736 MS cases remained (Table 5.4).

Table 5.4: Number of MS cases in the initial (cleaned) 1981 MS Survey dataset by survey area. (Corresponding sample numbers used by the McLeod research group, and their relevant publications, given in parentheses)

Survey area	Number of MS cases	
QLD	420	(420)[1]
WA	318	(318)[2]
NSW/ACT	1907	(1907)[3]
SA	378	(378)[3]
VIC	588	
Hobart SD, TAS	125	(125)[4]
Total	3736	
TOTAL excluding VIC	3148	

[1] Hammond et al. (1987) [646].
[2] Hammond et al. (1988c) [648].
[3] McLeod et al. (1994) [301].
[4] Hammond et al. (1988a) [647].

An initial numerator dataset of 3,148 MS cases resulted after exclusion of the 588 cases identified as from the VIC survey area in 1981 ('state81'=3). The dataset comprised the same sample numbers as given by the McLeod group for the total MS cases in the separate survey areas [301, 646-648] (Table 5.4) and at this stage included MS cases born in Australia and elsewhere.

5.3.2 Basic statistics of MS cases

Total number of cases and prevalence by sex

The total 3,148 MS cases comprised 2,208 females and 940 males (70.14 and 29.86%, respectively) (Table 5.5), giving a female to male case ratio of 2.35. Crude prevalence, defined as the ratio of persons with an acceptable diagnosis of MS living in the defined area on 30 June 1981 to the total number of persons in the population of the same area on the same day [646-648], was calculated by sex using survey area population numbers from the 1981 Australian Census (Table 5.5).

Table 5.5: Total survey population and 1981 MS prevalence by sex

Sex	Number MS cases	Survey population[#] (1981 Census)	Prevalence (per 100,000)
Males	940	5,174,328	18.2
Females	2208	5,195,641	42.5
Total	3148	10,369,969	30.4 (30.4*)

* Adjusted for sex using 1981 Australian census population.
[#] Note: Survey population totals are for the surveyed areas, QLD, WA, NSW/ACT, SA and Hobart SD (TAS), only.

Year of birth and age range

Year of birth ranged over seven decades, from 1901 to 1968 for males (Figure 5.3) and 1897 to 1969 for females (Figure 5.4). Equivalent age in June 1981, expressed in completed years on the census date, ranged from 11 to 84 years for females and 12 to 80 years for males.

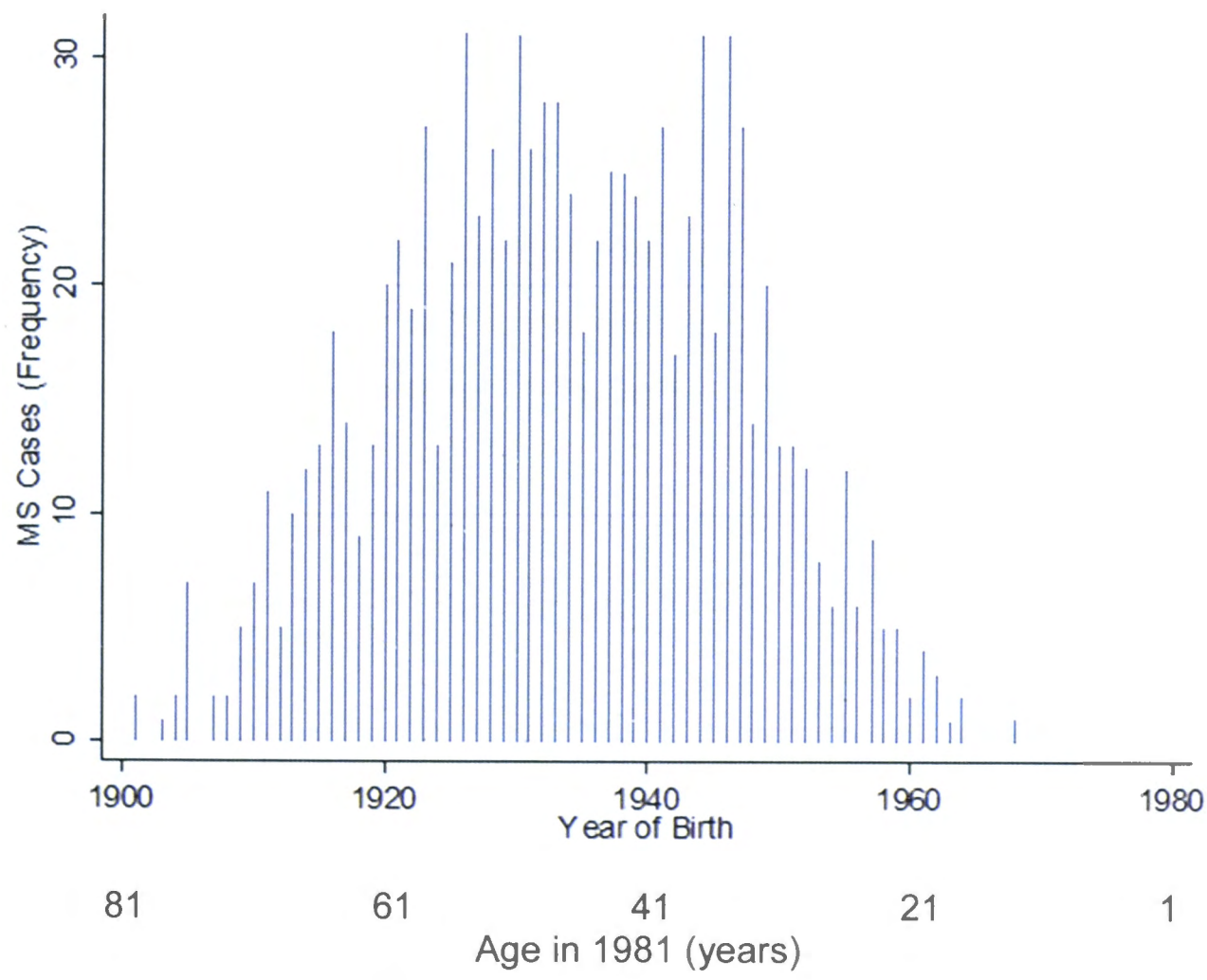


Figure 5.3: Year-of-birth distribution for males in 1981 MS Survey (n=940 cases).

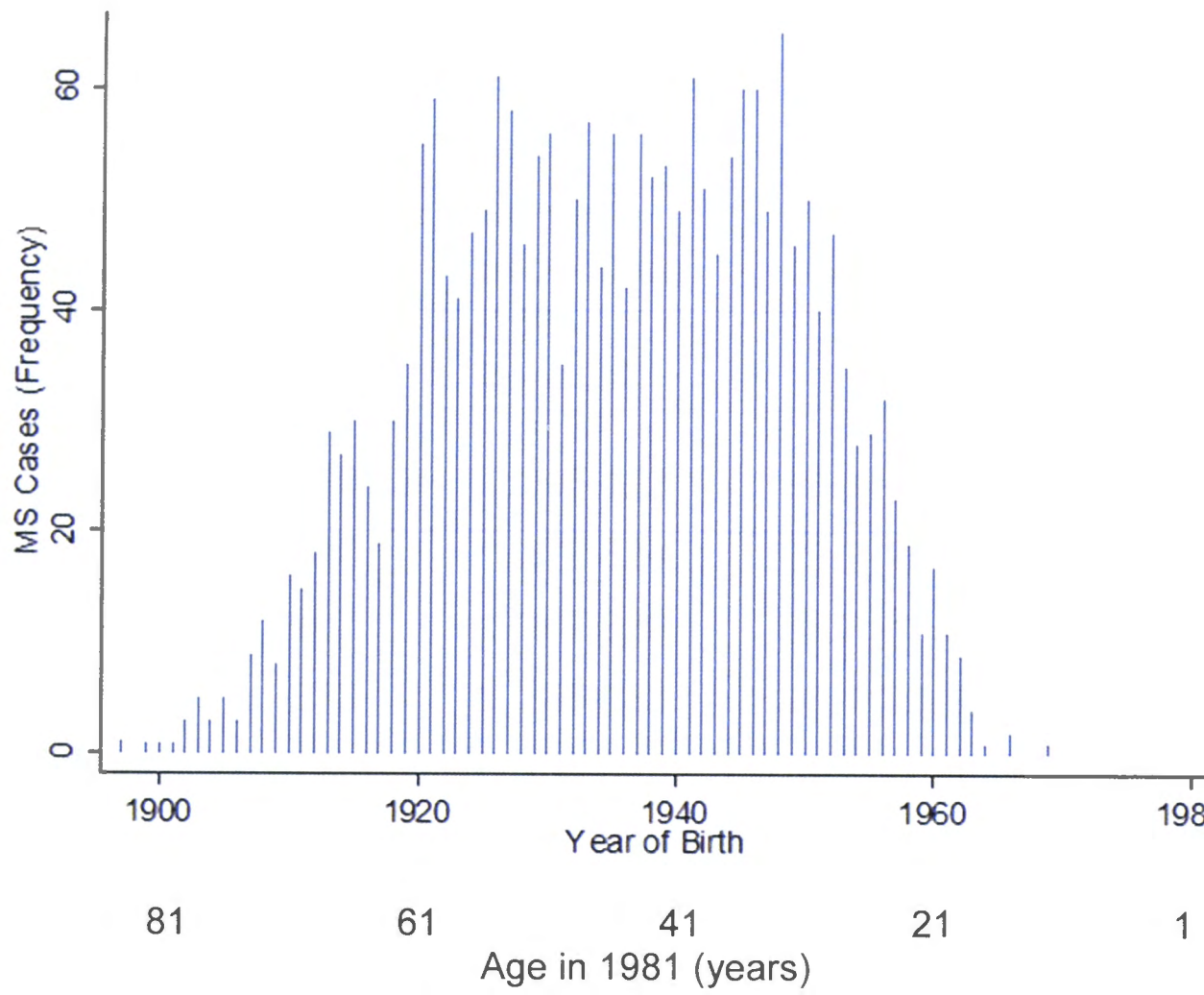


Figure 5.4: Year-of-birth distribution for females in 1981 MS Survey (n=2,208 cases).

MS prevalence by survey area

Prevalence (both crude and age-standardised to the 1981 Australian population) is shown by survey area, together with corresponding prevalence estimates given by the relevant McLeod group's publications, in Table 5.6.

Table 5.6: MS prevalence (age-standardised to 1981 Australian population, in parentheses) by survey area

Survey area	Mean[+] latitude °S	Number MS cases	Survey population (1981 Census)	Prevalence per 100,000	Prevalence per 100,000 by McLeod group
QLD	25.1	420	2,295,123	18.3 (18.6)	18.3 (18.6)[1]
WA	31.3	318	1,273,624	25.0 (25.9)	25.0 (25.9)[2]
NSW/ACT	33.4	1907	5,347,826	35.7 (35.2)	37.2 (36.6)[3 #]
SA	34.9	378	1,285,033	29.4 (28.8)	29.4 (28.8)[3]
Hobart SD (TAS)	42.8	125	168,363	74.2 (75.6)	74.2 (75.6)[4]
TOTAL		3148	10,369,969*	30.4 (30.3)	

[+]Mean latitude of survey area in decimal degrees as given by Hammond et al. (2000a) [304], calculated from the population distribution by SD in the 1981 Census.
[1] Hammond et al. (1987) [646].
[2] Hammond et al. (1988c) [648].
[3] McLeod et al. (1994) [301].
[4] Hammond et al. (1988a) [647].
[#] NSW/ACT prevalence (both crude and age-standardised) given by McLeod et al. (1994) [301] for 'NSW' is consistent with the ACT population total not having been included in the denominator.
* 'Total' survey population is total of areas surveyed.

Age-standardised MS prevalence increased with increasing mean south latitude from survey areas QLD to Hobart SD (TAS) (Figure 5.5). This trend in prevalence from the north to the south of Australia, here estimated over all four states, QLD, WA, NSW/ACT, SA and the fifth state city SD, Hobart (TAS), is consistent with similar trends shown separately by Hammond, McLeod and colleagues for: (i) within the state of QLD [646], (ii) between three Australian state cities, Perth (WA), Newcastle (NSW) and Hobart (TAS) [647] and (iii) overall among three states (QLD, WA and SA) and two other state cities (Newcastle and Hobart) [299, 645]. McLeod et al. (1994) later reported MS prevalence for all of the survey areas, including NSW, and reported age-standardised prevalence for the southernmost Hobart SD (TAS) survey area (75.6/100,000) as four times higher than that of the northernmost state surveyed, QLD (18.6/100,000) [301]. Consistent with this, Hammond et al. (2000a) reported a prevalence ratio of 3.84 (95% CI 3.05, 4.83) for TAS-resident compared with QLD-resident persons for just the Australia-born portion of the MS survey population [304].

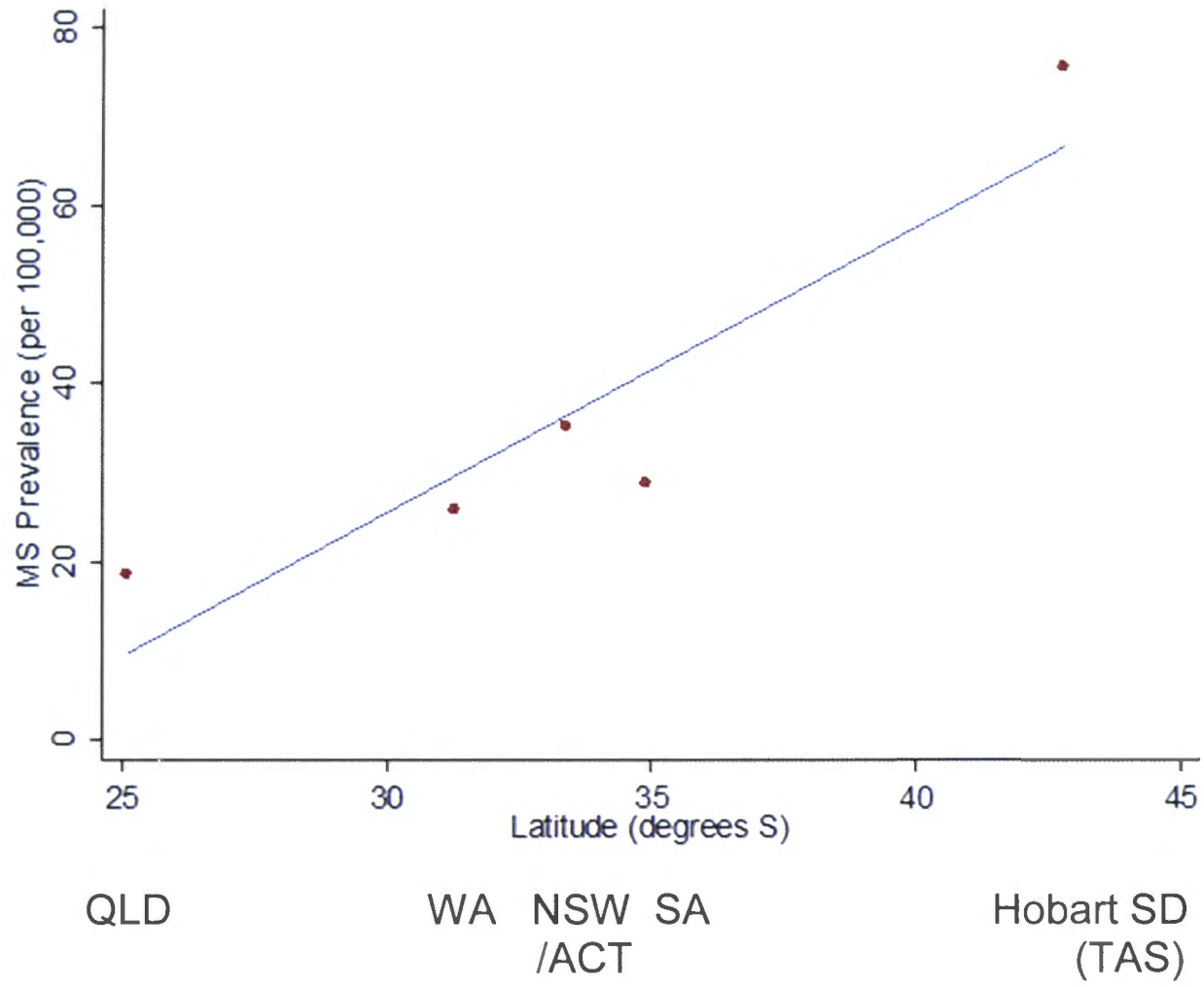


Figure 5.5: Mean latitude of survey area and age-standardised prevalence of MS (p=0.029 for linear regression; n=3,148 cases).

Prevalence by survey area and sex

MS prevalence by both survey area and sex is shown in Table 5.7, and Figure 5.6.

Table 5.7: MS prevalence (age-standardised to 1981 Australian population, in parentheses) by survey area and sex

Survey area*	MS prevalence (per 100,000)	
	Males (n=940 cases)	Females (n=2,208)
QLD	11.3 (11.6)	25.4 (25.8)
WA	12.0 (12.5)	38.2 (39.5)
NSW/ACT	21.6 (21.3)	49.6 (48.9)
SA	18.1 (17.7)	40.5 (39.6)
Hobart SD (TAS)	53.2 (54.5)	94.6 (95.8)

*Survey areas listed in order of increasing mean south latitude (see Table 5.6)

Figure 5.7 indicates a linear relationship between mean south latitude-and age-standardised MS prevalence in both sexes.

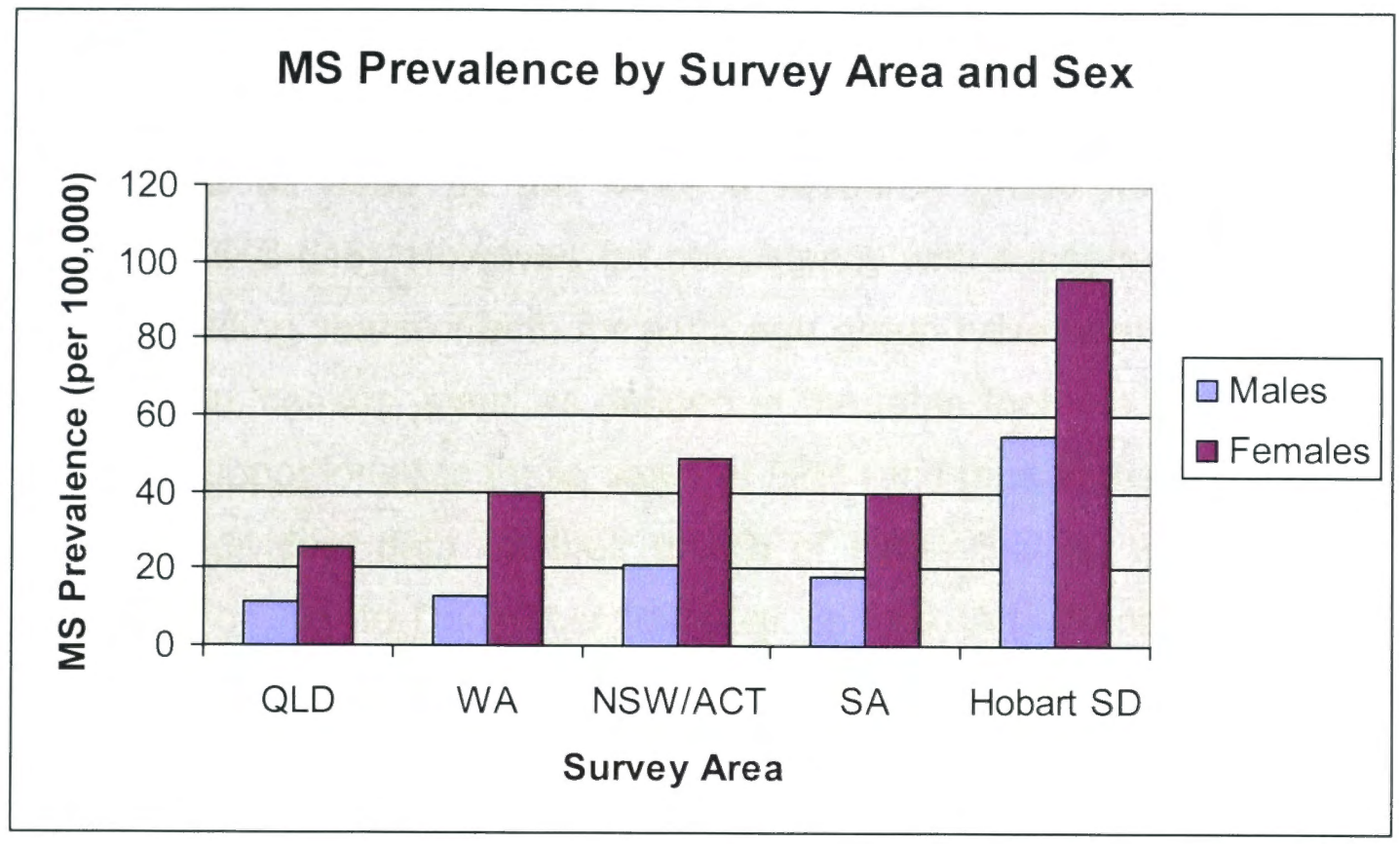


Figure 5.6: Age-standardised MS prevalence by survey area and sex (n=3,148 cases).

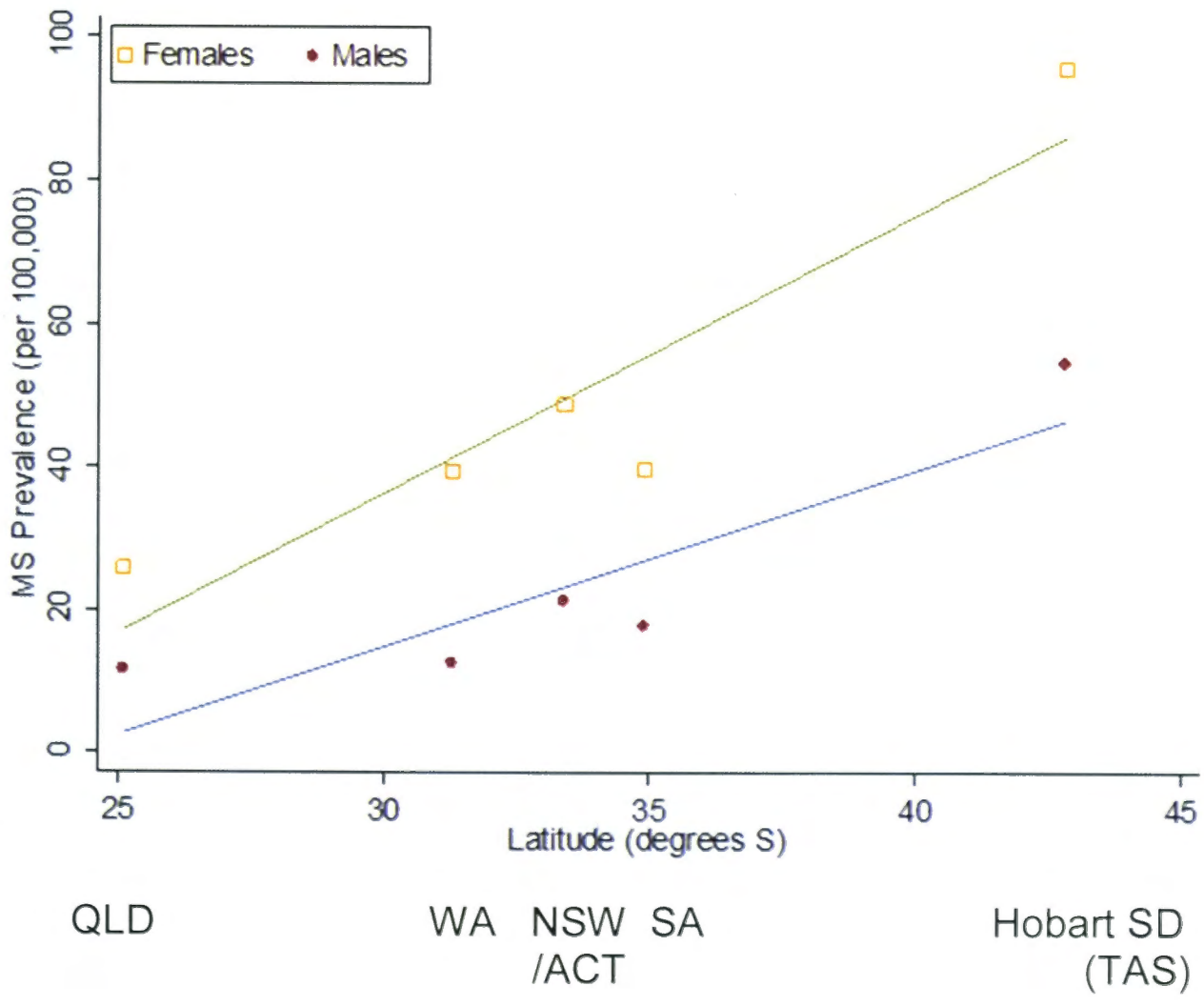


Figure 5.7: Age-standardised MS prevalence by mean latitude of survey area and sex (p=0.024 and 0.043 for linear regressions for females [n=2,205] and males [n=940], respectively).

Age-specific prevalence—MS case data by (10-year) age group

Age-specific prevalence of MS for the total survey area is given in Table 5.8, using ten-year age groups as used by the McLeod research group for the survey areas separately [301, 646-648]. However, for consistency with subsequent analyses in this thesis, corresponding years of birth for each age group have been calculated and are here expressed in 'census years' as defined in the table footnote. The number of MS cases has been apportioned to these years of birth (and thus to the corresponding age groups) using individual data on their month of birth—that is, whether born in the January to June or July to December half-year periods. Age-standardised prevalence for the overall survey population, adjusted to the age structure of the total Australian census population in 1981, is also shown.

Table 5.8: Age-specific MS prevalence in 1981 (n=3,148)

Age group (years[+])	Year-of-birth group[#]	Number of MS cases	Survey population (data from 1981 Census[†])	Prevalence (per 100,000)
0–9	July 1971–June 1981	0	1,680,309	0
10–19	July 1961–June 1971	29	1,798,196	1.6
20–29	July 1951–June 1961	346	1,728,980	20.0
30–39	July 1941–June 1951	740	1,545,420	47.9
40–49	July 1931–June 1941	752	1,097,364	68.5
50–59	July 1921–June 1931	742	1,057,326	70.2
60–69	July 1911–June 1921	435	816,029	53.3
70+	Before July 1911	104	646,345	16.1
Total		3148	10,369,969	30.4 (30.3*)

[+]Age in completed years on census date 30 June 1981.
[#]Year of birth calculated in 'census years' spanning 1 July to 30 June, because the Australian census records only 'age in completed years on 30 June 1981' rather than actual year of birth. For example, 'age 10 (completed) years' on 30 June 1981 equates to a date of birth between July 1970 and June 1971 (inclusive) census year. Similarly, 'age 19' on 30 June equates to a date of birth between July 1961 and June 1962 census year. Therefore, the years of birth for the 10 to 19 years age group span the 10 census-year period July 1961 to June 1971 (inclusive). (Similarly, the years of birth for the age 20 to 29 years age group span the 10 census-year period July 1951 to June 1961 [inclusive], and so on.)
[†]Census total for Hobart SD as given by Hammond et al. (1988a) [647].
*Prevalence age-standardised to the 1981 Australian census population.

For the total survey area, highest MS prevalence was seen in the 40 to 49 and 50 to 59 age groups, consistent with the McLeod group's results for the survey areas separately [301, 646-648]. The overall age-standardised prevalence was 30.3/100,000.

Age-specific prevalence by sex

Age-specific prevalence for MS cases is shown by sex in Table 5.9, together with overall prevalence, both crude and adjusted for age structure of the 1981 total Australian population. Figure 5.8 illustrates this generally found clear difference in prevalence by sex, particularly in the mid-age groups [301, 646-648, 652].

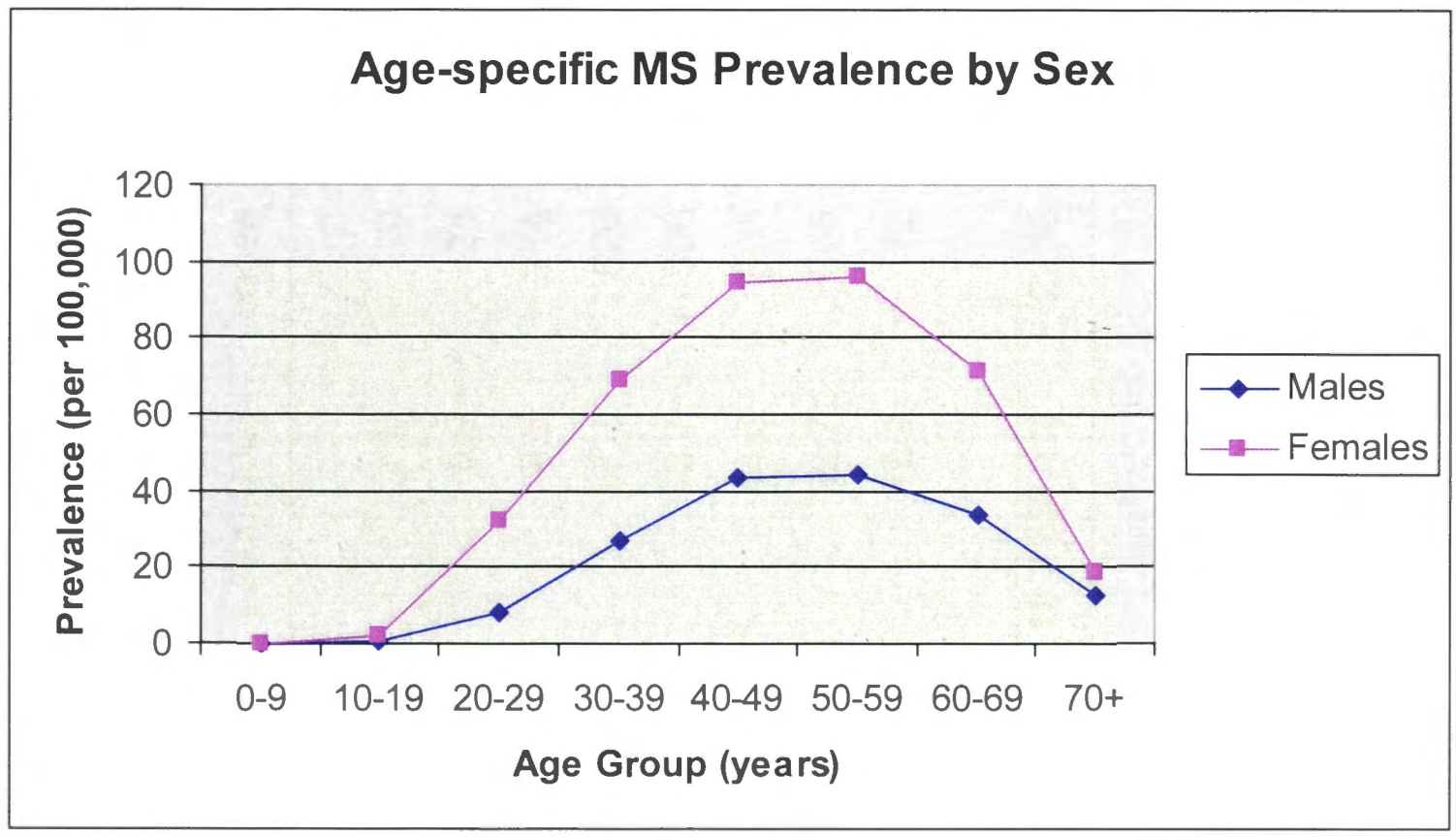


Figure 5.8: Age-specific MS prevalence by sex (data from Table 5.9; n=2,205 and 940 for females and males, respectively).

Table 5.9: Age-specific MS prevalence by sex

Age group (years[+])	Year-of-birth group[#]	Males			Females		
		MS cases	Survey population	Prevalence (/100,000)	MS cases	Survey population	Prevalence (/100,000)
0–9	July 1971–June 1981	0	860,560	0	0	819,749	0
10–19	July 1961–June 1971	8	919,983	0.9	21	878,213	2.4
20–29	July 1951–June 1961	72	869,880	8.3	274	859,100	31.9
30–39	July 1941–June 1951	214	784,161	27.3	526	761,259	69.1
40–49	July 1931–June 1941	246	562,508	43.7	506	534,856	94.6
50–59	July 1921–June 1931	238	533,913	44.6	504	523,413	96.3
60–69	July 1911–June 1921	130	386,024	33.7	305	430,005	70.9
70+	Before July 1911	32	257,299	12.4	72	389,046	18.5
Total		940	5,174,328	18.2 (18.1*)	2208	5,195,641	42.5 (42.4*)

[+] Age in completed years on census date 30 June 1981.
[#] Census years, as defined in Table 5.8.
*Prevalence age-standardised to 1981 Australian census population.

Prevalence by survey area and birth cohort

To consider the effects of both survey area and age structure together, the age range of MS cases (on 30 June 1981) was considered as three cohorts, namely *early*-born (age 60+ completed years [on census date], i.e. born before July 1921), *mid*-born (age 30-59 completed years, i.e. born between July 1921 and June 1951 [inclusive]) and *late*-born (age up to 29 completed years, i.e. born after June 1951). Crude and age-standardised MS prevalence by both survey area and birth cohort are given in Table 5.10.

Table 5.10: MS prevalence (age-standardised to 1981 Australian population, in parentheses) by survey area and birth cohort (n=3,148)

Survey area*	MS prevalence (per 100,000)		
	Early-born cohort (born <July 1921)	Mid-born cohort (born July 1921–June 1951)	Late-born cohort (born > June 1951)
QLD	21.2 (21.1)	38.6 (38.7)	3.7 (3.7)
WA	34.8 (35.0)	52.9 (53.4)	3.9 (3.9)
NSW/ACT	41.4 (41.1)	68.7 (68.6)	9.9 (9.8)
SA	39.7 (39.8)	59.0 (58.4)	4.9 (4.8)
Hobart SD (TAS)	105.4 (105.7)	148.4 (148.2)	16.2 (15.9)

*Survey areas listed in order of increasing mean south latitude (see Table 5.6).

MS prevalence in general was highest in the mid-born cohort in all survey areas (Figure 5.9). As shown in Figure 5.10, age-standardised MS prevalence increased linearly with increasing mean south latitude of the survey areas in both early- and mid-born cohorts (p=0.024, 0.029, respectively, for linear regression). Although the late-born cohort shows a slight increase in prevalence with mean south latitude in Figure 5.10, this was not statistically significant (p=0.072).

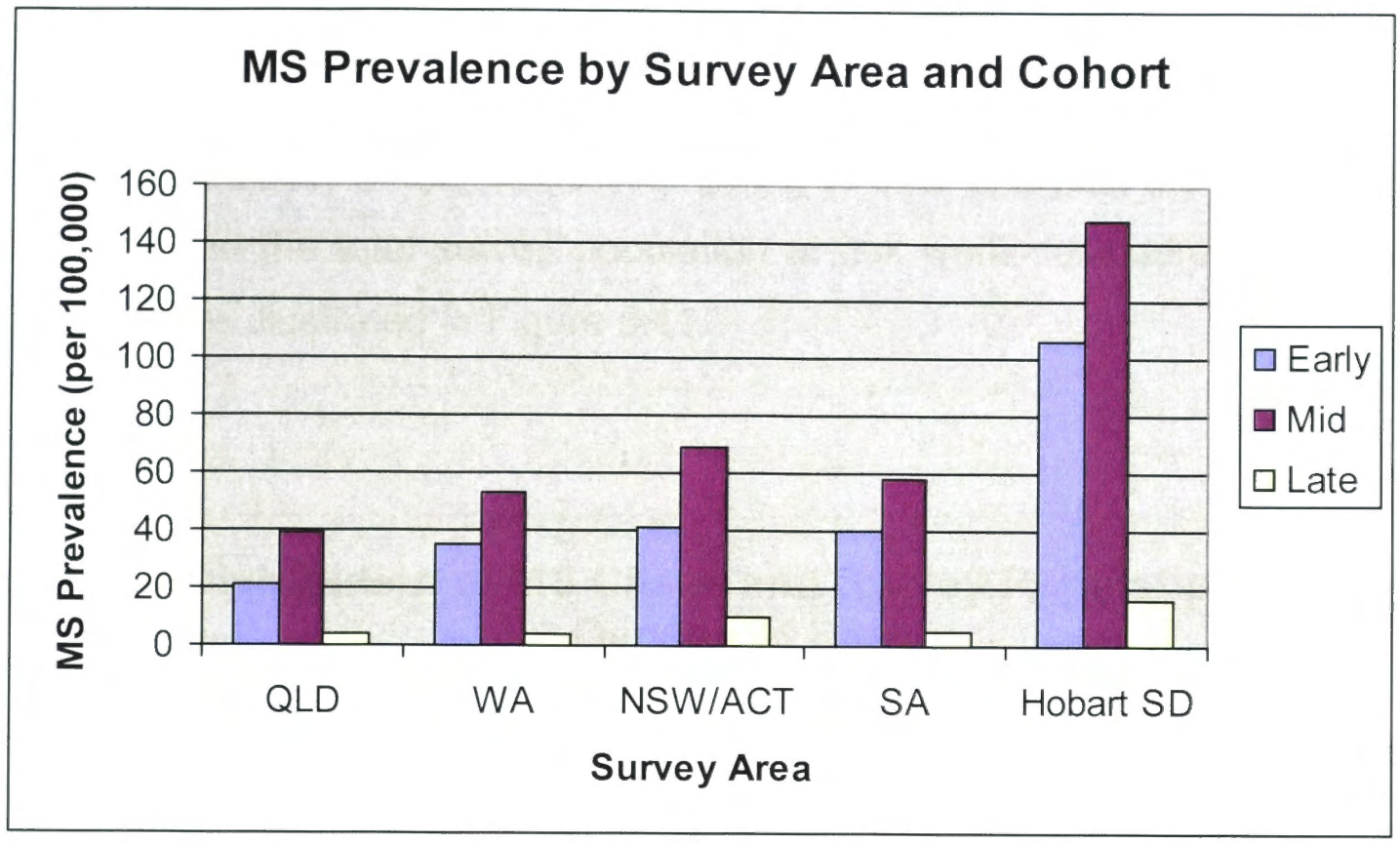


Figure 5.9: Age-standardised MS prevalence by survey area and birth cohort (data from Table 5.10; n=3,148).

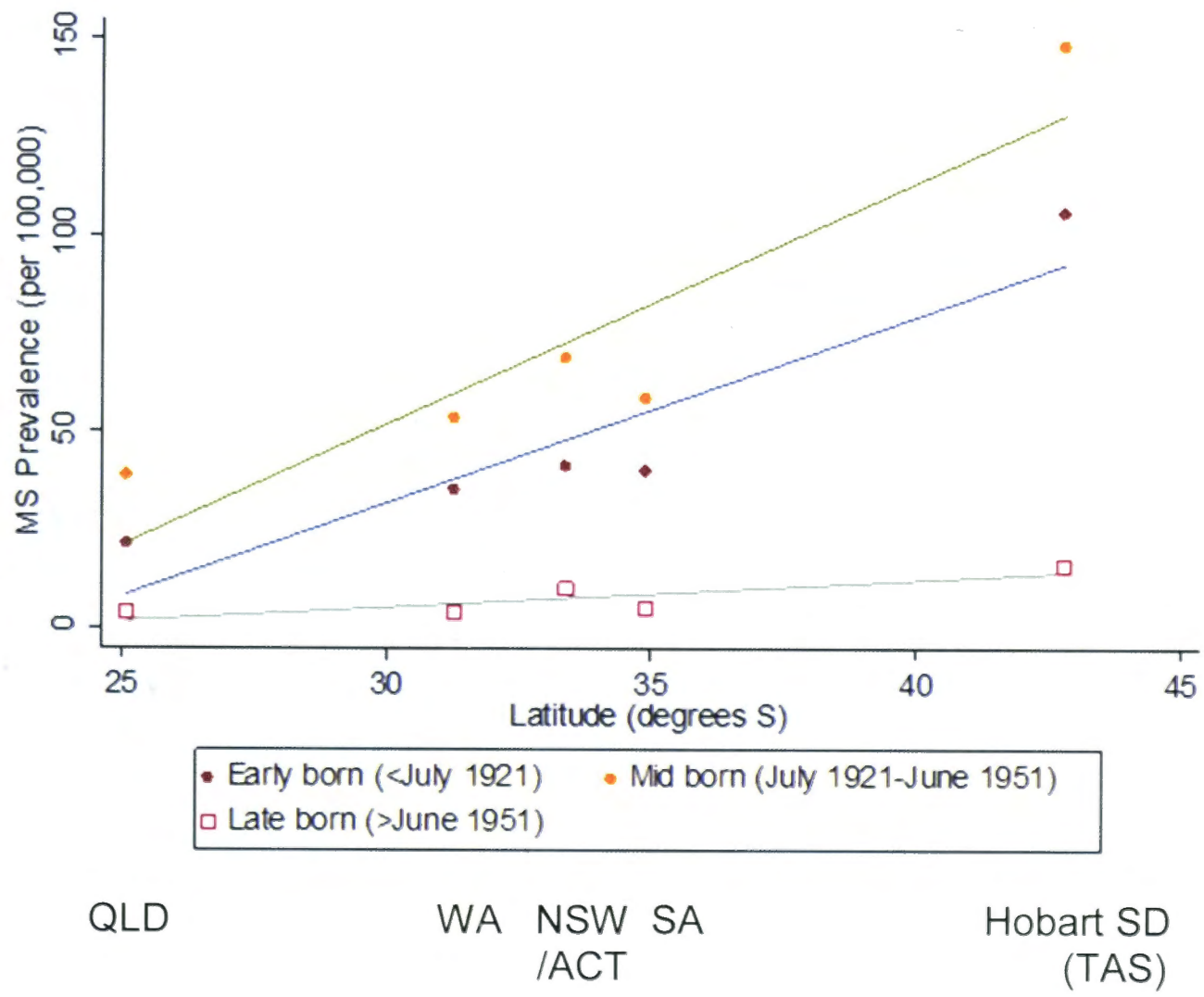


Figure 5.10: Age-standardised MS prevalence by birth cohort and mean south latitude of survey area (prevalence data from Table 5.10; n=3,148).

5.3.3 Numerators for timing-of-birth analysis

To assess the suitability of this dataset for timing-of-birth analysis, the numbers of both the MS cases and the total survey population at risk were considered separately, by decade of birth, as illustrated in Figure 5.11.

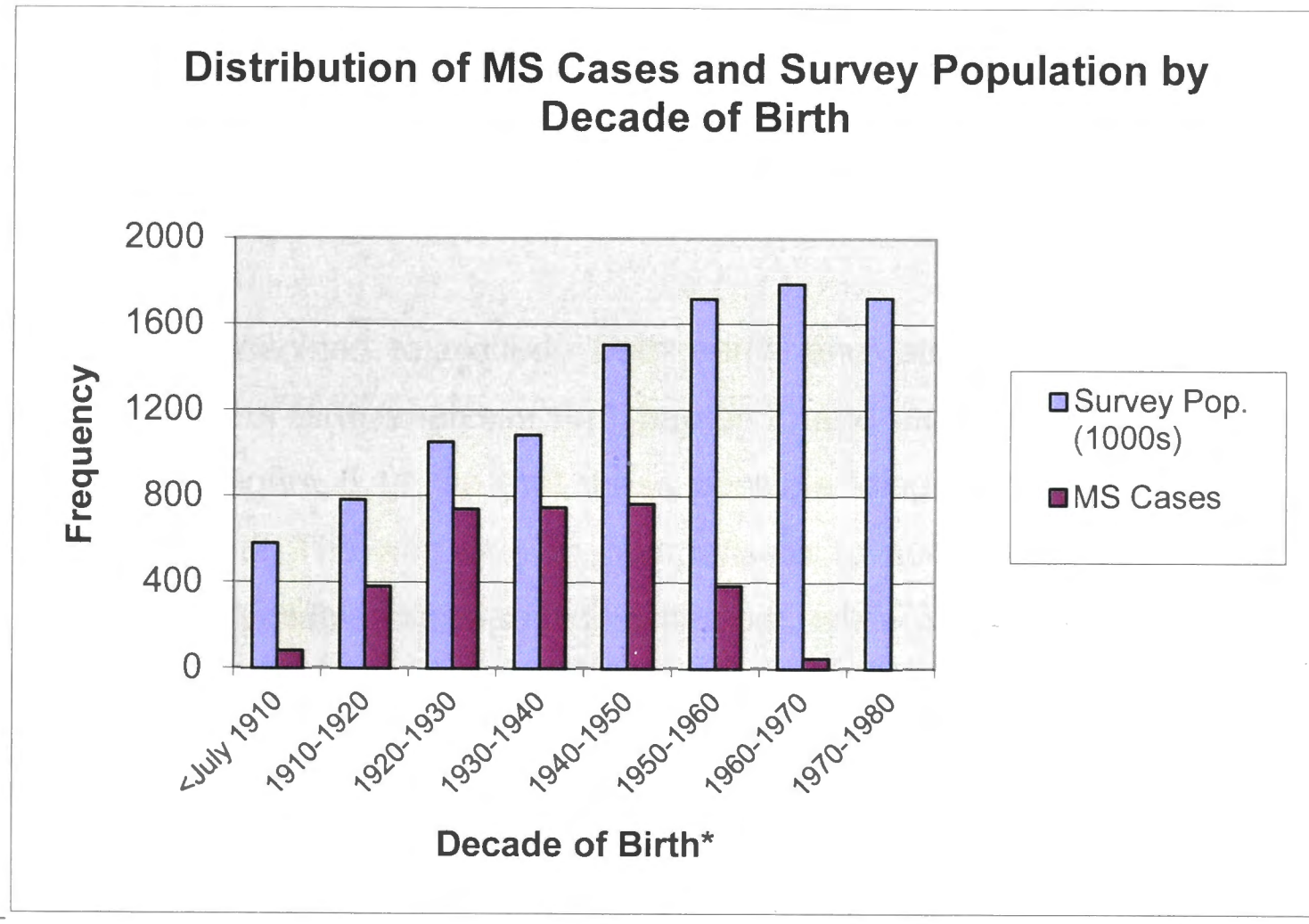


Figure 5.11: Distributions of MS cases and total survey (census) population by decade of birth. (*Decades of birth each span July to June in accordance with census years, for example, July 1910 to June 1920 but, for convenience for subsequent analysis, now differ by a shift of one year from the year-of-birth periods in Tables 5.8, 5.9 and 5.10; decades thus correspond with age [in 1981] of one to 10, 11 to 20, 21 to 30 ... 61 to 70 and 71+ completed years.)

While the total survey (census) population distribution follows what might be expected as a typical 'survival curve', the MS cases in Figure 5.11 suggest a decrease in survival of MS cases compared with the total at-risk survey population in the early birth decades (in the left-hand part of the figure), particularly in those born prior to mid-1910 (71+ age group). That is, while the MS cases to the left of the figure can be assumed to have all reached MS onset if that was going to occur, there appears to be a slight survival difference in these earliest birth decades (i.e. possible survival bias).

However, in the right-hand part of the figure the large differences between the distributions of the MS cases and the total survey population are likely to be related more to disease onset than to survival, given that onset of MS can occur up until the mid-years of life (age of onset ranges from eight to 60 years of age in this current dataset) and that MS overall does not appear to cause substantial mortality [177, 178, 653]. That is, there is a very clear difference in frequency of MS cases compared with the total survey population in the later birth decades in Figure 5.11—that is, fewer MS cases than expected—particularly in those born after mid-June 1950 (i.e. those who are one to 30 completed years of age in 1981). Some at-risk survey-population persons born in these decades are thus likely to be 'future' MS cases that have not yet been diagnosed by the date of the survey (i.e. diagnosis bias).

Therefore, it was decided to exclude both early- and late-born MS cases from the subsequent timing of birth analyses (in Chapter 7) and include only the three central birth decades in Figure 5.11, so that these possible 'diagnosis' and 'survival' biases could be minimised. The aim was to compensate to some extent for the inherent limitations of the available cross-sectional dataset, while still retaining sufficient case numbers for analysis.

Restriction to years of birth 1920 to 1950 and to Australia born

Restriction to years of birth 1920 to 1950

For the purpose of timing-of-birth analysis in Chapter 7, MS cases born in the three central birth decades in Figure 5.11 (from July 1920 to June 1950 census years) were retained, together with the remaining half census year at each end of this period in order to maximise the available sample. That is, the MS-case sample used comprised those born during the complete calendar years 1920 to 1950. Thus, excluding early-born (before January 1920) and late-born (after December 1950) MS cases, the number of MS cases with their year of birth within the restricted 1920 to 1950 period was 2,322 (73.8 % of the available total sample of 3,148).

Restriction to Australia born

For subsequent analysis, and particularly for the effect of *birthplace* on timing of birth, only the MS cases born in Australia were able to have relevant population month-of-birth denominators derived (see Chapter 6) for the purpose of answering the specific

research questions posed for this thesis (see Chapter 7). MS cases not born in Australia (n=680 [29.3% of those born 1920 to 1950]) or of unknown birthplace (n=10 [0.43%]) were therefore omitted from the cleaned and year-of-birth-restricted 1981 dataset, reducing the number of MS cases for timing-of-birth analysis to 1,632. The resulting number of Australia-born MS cases with their year of birth within the restricted 1920 to 1950 period is shown by sex and survey area in Table 5.11.

Table 5.11: Number of MS cases born in Australia between January 1920 and December 1950 (inclusive), by sex and survey area

Survey area	Number of MS cases born 1920 to 1950 in Australia		
	Males	Females	Total
QLD	80	161	241
WA	28	110	138
NSW/ACT	292	686	978
SA	61	147	208
Hobart SD (TAS)	26	41	67
Total	487	1145	1632*

*Month of birth was missing for one MS case from 1920 to 1950 Australia-born cohort, leaving 1,631 MS cases available for timing-of-birth analysis.

Summing up, the January 1920 to December 1950 year-of-birth period, including ages 30 to 61 completed years in 1981 (and approximating—to within one year—the 'mid-born' cohort of total MS cases shown in Table 5.10), has been chosen for analysis, to maximise the Australia-born sample while minimising probable diagnosis and survival biases evident in Figures 5.11 (for total sample) and 5.12 (for Australia-born cases).

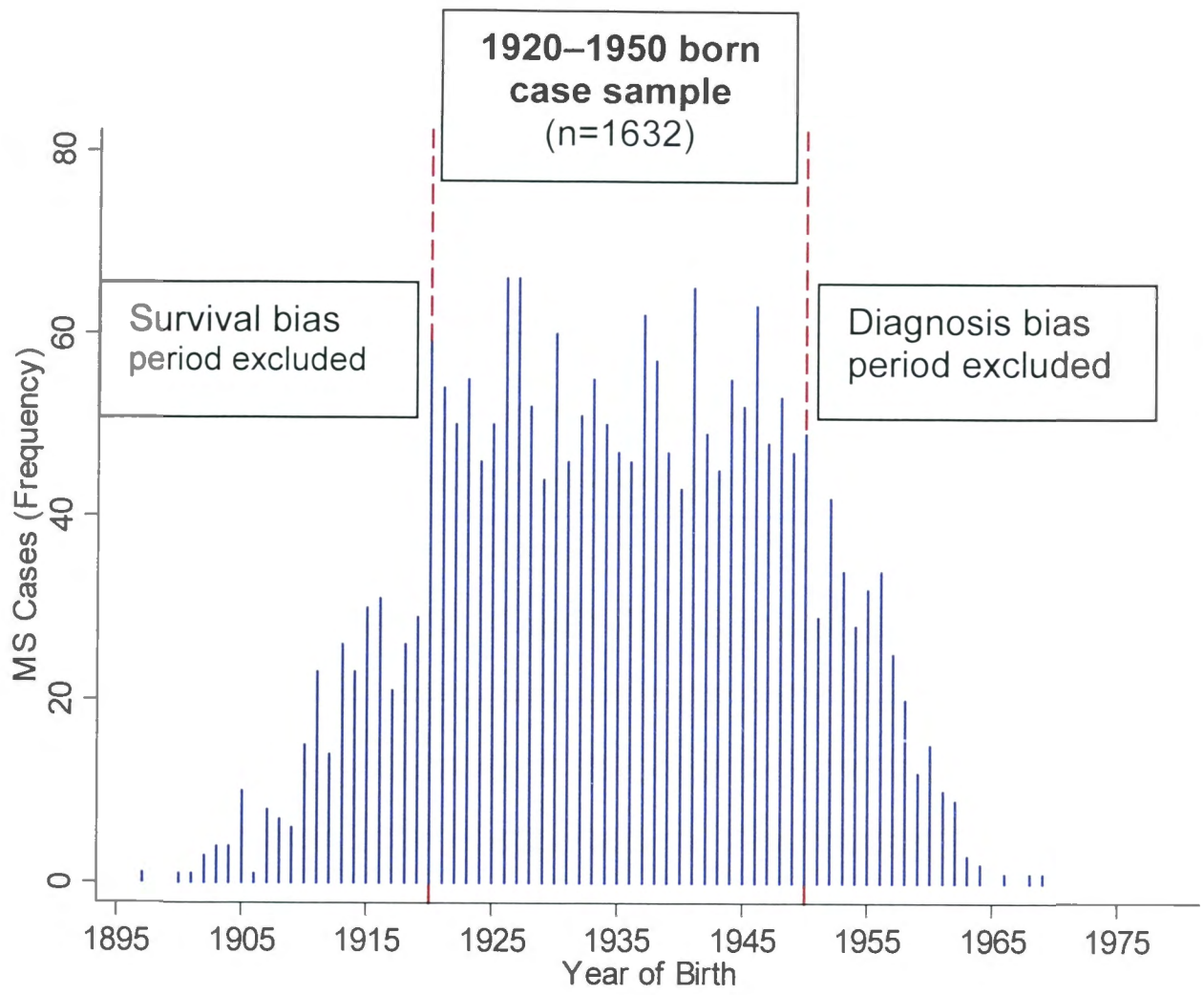


Figure 5.12: Year-of-birth distribution of Australia-born MS cases in the 1981 dataset (n=2,214), indicating final sample of 1920- to 1950-born cases used in timing-of-birth analysis (shown bounded by dashed lines; n=1,632).

The resulting 1920- to 1950-born cohort with known month of birth (n=1,631), required for timing-of-birth analysis, comprised 73.7% of the total 2,214 Australia-born cases in the cleaned numerator dataset (Figure 5.12). The mean age of onset of MS, calculated from year-of-birth and year-of-onset data for this 1920- to 1950-born cohort (aged 30 to 61 completed years in June 1981), was 32.9 $\pm$ 9.0 (s.d.) years. Of these MS cases, 85% were classified as relapsing-remitting (RRMS) from onset (i.e. 15% were primary progressive).

5.4 Discussion

The 1981 MS Survey by the McLeod group of researchers contributed much epidemiological and clinical information to the study of MS both in Australia and internationally (Table 5.1). As summarised by McLeod et al. (1994), MS in Australia is clinically remarkably similar to MS in predominantly white populations in the northern hemisphere [301]. Females are affected about twice as often as males and the age of onset is mostly between 20 and 50 years. About 80% of patients have an exacerbating-(i.e. relapsing-)remitting clinical course in the early stage of their disease, the remaining

20% having a progressive course from onset [301, 649]. The mean duration of disease from onset to death is more than 25 years, about 45% of patients having only mild to moderate disability after a period of 15 years. The clinical profile appears to be very similar throughout Australia, although male patients in the hotter climate of QLD show a greater tendency to develop progressive disease, and hence more disability, than those in more southern states such as TAS [649].

By using and extending (in Chapters 6 and 7) this existing, well-coordinated Australian survey, more information on possible environmental factors causing MS is now possible. However, the survey dataset in its original form required initial checking and verification of the numerator data (described in this chapter) as well as subsequent derivation of suitable denominators (described in Chapter 6) to maximise the usefulness of the dataset for this current thesis study. The main purpose of this chapter was to describe how a comprehensive national survey of MS patients in 1981 was prepared in order to construct (in Chapter 6) a longitudinal dataset suitable for analysis (see Chapter 7) of timing of birth in these MS cases.

The verification and organisation of the MS cases in the unit-record dataset, as received, has been detailed in Sections 5.2.1 and 5.2.2 of this chapter. Cases present on the earlier prevalence day, P1, in 1976 and who had died by P2, 1981, were removed. Diagnostic categories as defined by the McLeod research group were retained and those cases in other categories were removed from the dataset. Where some diagnostic data were missing, it was possible to make a judgement about inclusion or exclusion of these few cases by using other clinical data variables. In addition, several NSW cases found to be missing from the dataset (831 cases, 43% of the total NSW/ACT sample) were fortunately able to be accessed as original, hand-written clinical files at the University of Sydney. Where this supplementary information required for analysis of timing of birth was still incomplete, particularly for month and place of birth (approximately 25% of the 831 cases), the missing parameters were obtained from the NSW Registry of Births, Deaths and Marriages. The final case numbers were then consistent with those reported for each survey area by the McLeod group in their relevant publications (Table 5.6).

The summary prevalence data in this chapter have been given for comparison with previous McLeod group findings for individual survey areas, and 1981 census totals have been used as denominators here. It should be noted that these prevalence data include all persons in Australia on the census date, whether born in Australia or

overseas, as given in most of the McLeod group publications listed in Table 5.1. The new estimates by survey area given in this thesis (Table 5.6), for example, are thus comparable to those in the publications for the survey areas separately; however, the discrepancy for prevalence in the NSW/ACT survey area appears to be a result of the ACT population not having been included in the denominator by McLeod et al. (1994) [301]. Most importantly, this chapter reinforces the existence of an overall association between latitude of residence and prevalence of MS in Australia reported in the McLeod group's summary publications [299, 304], and corroborates the statement made by McLeod and colleagues (1994) that MS prevalence in residents of Hobart SD, TAS is four times higher than that in QLD residents [301]. Consistent with this, a four-fold difference in incidence of FDE (a precursor to MS) from Brisbane (QLD) to TAS has been found in the 'Ausimmune' study in Australia [305].

The remaining results on age-specific MS prevalence for the overall Australian population have also been included for comparison with previous McLeod group findings for separate areas; similar 10-year age groups have been used. In the present study, 'census age' (i.e. age in completed years on the census date) has been used to derive 'census years' of birth, as these will be required in the subsequent analyses for timing of birth in Chapter 7. Therefore, while the distribution of MS prevalence by 'age' for the total survey population was slightly different from any of those given by McLeod and colleagues for the separate survey areas, the overall prevalence age-standardised to the 1981 Australian population (30.3 per 100,000), was found to be the same as calculated for the McLeod group's survey areas when considered together as a single sample (data not shown). The highest prevalence rates in the 40 to 49 and 50 to 59 year age groups, and the differing age-specific prevalence rates for males and females, were also consistent with previously reported findings [301, 646-648].

Bringing the results for 'latitude' and 'age' together, the prevalence rates by survey area and birth cohort showed highest overall prevalence in the mid-born cohort (Table 5.10); a linear latitude-prevalence relationship was also evident for both early- and mid-born cohorts, but not for the later-born cohort (Figure 5.10). This information, together with the separate decade-of-birth distributions for MS cases and the total (census) survey population (Figure 5.11), suggested possible biases present in both the early-born and late-born cohorts, and particularly in the later born due to under-representation of (not yet) diagnosed cases. This conclusion underpinned the decision to restrict the final sample of MS cases for timing-of-birth analysis to essentially the 'mid-born' cohort only, in order to limit any effects of either late onset of MS or survival

differences between cases and controls. The restricted sample was chosen, for convenience, to be those MS cases born between calendar years January 1920 and December 1950, inclusive; this sample should therefore avoid major error resulting from use of an initially cross-sectional dataset, while maximising sample size as much as possible.

With regard to regional latitude as an exposure factor for MS acting together with timing of birth (both to be discussed further in Chapter 7), the latitude of *birthplace* is arguably considerably more relevant than the latitude of later-life residence, if the time around birth is critical to the initiation of MS. That is, potentially causal ('causal' including either detrimental or beneficial here) environmental factors acting around the *time* of birth may be linked to *place* of birth, or to its (southern) latitude; if so, testing for such causal factors for MS should be more sensitive if place of *birth* rather than later place of residence, or survey area, is used. Hence, the following chapters (Chapters 6 and 7) will investigate time and place of birth within Australia, as factors potentially affecting MS outcome, rather than using the survey area as an exposure factor, as discussed so far. Accordingly, only those MS cases born in Australia will now be relevant to the present thesis study (see research questions in Chapter 7).

A case sample of 1,631 Australia-born MS cases with known month of birth results. The focus on individual birthplace rather than place of residence within Australia in the present study further ensures that the timing-of-birth data will not be diluted by inclusion of overseas-born migrant cases, particularly any northern hemisphere-born who may show a reciprocal seasonal birth timing. In the 1981 survey, the number of overseas-born cases represented 29% of the total 1920- to 1950-born persons, the Australian migrants in 1981 originating mainly from the UK and Ireland [304, 647]. In comparison with the present thesis strategy of restriction to Australia-born persons only, register timing-of-birth studies, such as that of Willer et al. (2005) [324] in the northern hemisphere, have generally included all cases ascertained in a particular area, regardless of where born and regardless of when any overseas-born persons migrated to the area.

Chapter 6 will detail the construction of a longitudinal type of dataset from the basic cross-sectional, cases-only 1981 MS Survey described here. Numerator data (cases) and suitable denominator data (population controls) will be combined into the same dataset, month-by-month and year-by-year, to give a series of 'MS rates' relative to the 1981 Australian reference population, by birth month and year, and by place of birth.

CHAPTER 6

TIMING OF BIRTH AND MS: DERIVATION OF DENOMINATORS FOR TIMING-OF-BIRTH ANALYSES

6.0 Preface

How do we provide suitable population controls for the MS cases in the existing MS survey dataset?

To answer this question, this chapter will focus on determining appropriate population denominators for MS rates for the 1981 surveyed cases and detail the construction of a longitudinal by-year and by-month MS-rates dataset from the cross-sectional 1981 MS Survey discussed in Chapter 5. Suitable denominators will be calculated for the available MS-case numerators, taking into account region of birth within Australia and sex. The Australian census, taken on the same day as the survey in 1981, will be used together with other necessary Australian births-registration data. The denominators will also be adjusted to account for the non-surveyed state of VIC and the NT, in particular. This will enable timing-of-birth analysis for MS cases in Chapter 7; that is, MS risk expressed as MS-births incidence rate by month and place of birth.

As outlined in the previous chapter, only Australia-born persons will be included in the final dataset, and only those Australians born between January 1920 and December 1950 will now be considered.

6.1 Introduction

As discussed in Chapter 5, the original MS dataset was a cross-sectional prevalence survey. For timing of birth (by month), year of birth will also be taken into account rather than simply combining all years of birth together for analysis, enabling MS incidence rates to be obtained. Reconstruction of such a longitudinal dataset from the original point-prevalence cross-sectional survey, and incorporation of the appropriate reference-population denominators, will be detailed in Sections 6.2.1 to 6.2.4.

Comparison with register studies

As outlined in Chapter 5 (see Section 5.0), most longitudinal timing-of-birth studies use registers of incident cases, wherein disease cases are registered over a period of time as they are diagnosed or otherwise notified. Registration may cover, for example, a period of 20 years, say 1960 to 1980, and these cases may be found to have been born between 1920 and 1950, depending on the average age of onset of disease. For such a disease register, the same range of year of birth from a total-population births register—that is, 1920 to 1950—would generally be used to obtain month-of-birth denominators for the reference population, enabling a (retrospective) case-cohort study. These reference-population births would ideally be from the same ascertainment area as the disease cases, for example, Australia (as in Table 6.1, left-hand column).

In the present study, an existing survey dataset is the available data source, and MS cases have been ascertained at one point in time (30 June 1981). The most appropriate reference population here is the total population surveyed at the same point in time, namely the 30 June 1981 census population (Table 6.1, right-hand column). The main difference compared with the register study is that both cases and the reference population are *survivors* of the original birth cohorts in the register population at the time of survey/census. That is, both cases and the reference population were alive and *in Australia* on the survey/census date. Further, in the present study, both the cases and the reference population may have been born overseas or in Australia (however, this study will choose to restrict to just Australia born for both cases and the reference population for timing-of-birth analysis), whereas the total-population births registrations, at least, will represent those who have been born only in the relevant registration area. Table 6.1 summarises these main differences (bolded italics) between the available data for the present thesis study and usual register studies.

Table 6.1: Similarities (italics) and differences (bolded italics) in numerator (MS cases) and denominator (reference) populations between an example register study and the present thesis study

Example **Register Study** (longitudinal)	**Thesis Study** (cross-sectional → longitudinal)
MS cases registered in ascertainment area (e.g. Australia) between e.g. 1960 - 1980, and *born (in Australia or overseas)* between e.g. 1920 to 1950	**MS cases** surveyed in 1981, therefore ***alive*** and ***in Australia when surveyed***, and *born (in Australia or overseas)* between 1920 to 1950 → ***Australia born only***
Reference population (denominators): Total population registered as born in ascertainment /register area (e.g. Australia) between 1920 to 1950 (i.e. *Australia-born* register)	**Reference population (denominators):** Total population ***alive and in Australia on census date in 1981*** (i.e. census population), and ***born (in Australia or overseas)*** between 1920 to 1950 → *Australia born only*

Place (region) of birth

As discussed in Chapter 5 (see Section 5.4), the present study will also investigate timing of birth by place of *birth* rather than by place of ascertainment. Birthplace—that is, *where* a person is at this early stage in the life course of this disorder—is likely to be more relevant and have more effect on subsequent environmentally initiated disease than residence in 1981, particularly if the influential environmental factor is, as hypothesised, acting early in life. In comparison with register studies using place of ascertainment, which may also include immigrants to the ascertainment area (for example, Willer and colleagues' Canadian MS study [324]), the present study should have fewer dilution effects. As indicated by Torrey and colleagues for timing-of-birth studies in disorders such as schizophrenia, failure to account for individuals born elsewhere and later immigrating to the study area may influence results for countries such as Canada and Australia with high immigration rates [642].

However, while birthplace (i.e. state) data are readily available for the MS cases (i.e. the numerators) in the 1981 survey, such data are more difficult to obtain for the denominators when a register is not the starting point. That is, the 1981 census did not record region of birth (state) within Australia for the reference population, nor did it record month of birth; only age in completed years on the census date was available, from which year of birth in 'census years' could be derived, as summarised in Table 6.2. This table again gives comparison with a register type of study, and indicates how additional national births-registration data could be used to gain the required denominator data for analysis. That is, the shaded box in the lower right-hand column of Table 6.2 shows the required reference-population data that were absent from the 1981 census; namely, month of birth and birthplace (state) within Australia. These parameters could be derived from available Australian births-registration data (Table 6.2, shaded box in lower left-hand column) to address this issue; these derivations are detailed subsequently in Section 6.2.2.

Table 6.2: Availability of data for timing-of-birth analysis, in an example register study and the present thesis study, for both cases and the reference populations. (Directly available data are shown bolded; data not directly available are shown in italics; shaded boxes and arrow emphasise the main areas of difference and indicate which data had to be sourced for this thesis study and where it could be sourced from)

Example **Register Study** (longitudinal)		**Thesis Study** (cross-sectional → longitudinal)	
Cases: ***(Disease Register)*** (Not necessarily surviving or in ascertainment /register area [e.g. Australia] post-registration)	• **Register area** • **Year of birth** • **Month of birth** • *Birthplace (state) within register area?*	**Cases:** ***(1981 Survey)*** (All alive and in Australia when surveyed 30 June 1981)	• **Birthplace area** (Australia or not) • **Year of birth** • **Month of birth** • **Birthplace (state) within Australia**
Reference Population: ***(Births Register)*** (Similarly to cases, not necessarily surviving or in ascertainment /register area [e.g. Australia] post-registration)	• **Birthplace area** (=register area, e.g.Australia) • **Year of birth** • **Month of birth** • **Birthplace (state) within register area** (e.g. Australia)	**Reference Population:** ***(1981 Census)*** (Similarly to cases, all alive and in Australia on census date 30 June 1981)	• **Birthplace area** (Australia or not)[+] • **Year of birth** (in census years)[#] • *Month of birth*** • *Birthplace*** *(state) within Australia*

[+] Derived from proportion Australia born by age in 1981 (census table).
[#] From age in completed years on census date 1981.
** To be derived from supplementary Australian births-registration tables (left-hand column) for 1920 to 1950 born.

Adjustment for migration between time of birth and 1981

Finally, because the Australian MS survey was not able to ascertain cases in the areas encompassing the state of VIC, the NT and the area of the state of TAS outside of the Hobart SD (see Chapter 5), further adjustment to the dataset to compensate for this lack of ascertainment in 1981 was required before analysis. As *birthplace* was the required exposure factor rather than place of residence or ascertainment, an estimate of net migration between the time of birth and time of survey/census was necessary in order to adjust the data. Adjustments were made either to the numerators (for the unsurveyed state/territory areas of VIC and NT) or to the denominators (for all of the other surveyed state/territory areas), as detailed in Section 6.2.3. Although relatively small, these adjustments were considered necessary to account for ascertainment losses, which would have occurred by cases migrating to (or from) the unsurveyed areas during the period from their birth to 1981.

The main purpose of this chapter is to derive the appropriate at-birth denominators for the MS cases (i.e. by month and place of birth) using 1981 census and other available supplementary data (as shown in Table 6.2), and then to adjust either the numerators or these denominators to take account of net migration to or from the unsurveyed areas. The aim of this chapter is to describe the construction of an MS-rates dataset, using the MS-case numerators and the derived denominator estimates in a longitudinal format, and to examine necessary assumptions and limitations of these data.

6.2 Methods

6.2.1 Required MS-rates dataset

Format of dataset

The format of the required longitudinal dataset is shown in Table 6.3. The six main variables describing birth (columns 2 to 7) are *year, month, place* and *sex,* as well as the *number of MS births* (often zero) and the *number of reference-population births.* Also shown is *census year* (column 1), necessary when deriving the reference-population births and defined in Chapter 5 (see Section 5.3.2, Table 5.8). The relationship between census years (July to June) and calendar years (January to

December) is indicated at the right-hand side of Table 6.3. Thus, the dataset will include the case dataset of MS cases, expressed in frequency form by year and month of birth, together with the corresponding population denominators also by (same) year and month of birth. Likewise, to examine a regional (latitude) effect on MS (and henceforth, latitude of *birth* rather than ascertainment area) these data will be separated by place of birth (i.e. state) for both numerators and denominators (column 5, Table 6.3).

Table 6.3: Format of MS-rates dataset: MS and reference-population births by month (January to December denoted as 1 to 12) for the first three calendar years, January 1920 to December 1922, for males (sex=1) born in NSW/ACT (place=2). (Calendar years and census years indicated to right of table. July 1920/June 1921 census-year example given in text is shown in bold)

	1	2	3	4	5	6	7
	CensYear*	Year#	Month	Sex	Place	MS births	Pop. births
1.	20	20	1	1	2	0	1426.993
2.	20	20	2	1	2	2	1317.430
3.	20	20	3	1	2	1	1408.955
4.	20	20	4	1	2	0	1372.212
5.	20	20	5	1	2	1	1457.056
6.	20	20	6	1	2	0	1672.174
7.	**21**	**20**	**7**	**1**	**2**	**2**	**1828.180**
8.	**21**	**20**	**8**	**1**	**2**	**0**	**1761.388**
9.	**21**	**20**	**9**	**1**	**2**	**1**	**1669.807**
10.	**21**	**20**	**10**	**1**	**2**	**0**	**1640.198**
11.	**21**	**20**	**11**	**1**	**2**	**0**	**1601.637**
12.	**21**	**20**	**12**	**1**	**2**	**0**	**1651.215**
13.	**21**	**21**	**1**	**1**	**2**	**2**	**1602.326**
14.	**21**	**21**	**2**	**1**	**2**	**0**	**1476.316**
15.	**21**	**21**	**3**	**1**	**2**	**0**	**1528.648**
16.	**21**	**21**	**4**	**1**	**2**	**1**	**1572.717**
17.	**21**	**21**	**5**	**1**	**2**	**0**	**1577.537**
18.	**21**	**21**	**6**	**1**	**2**	**0**	**1582.357**
19.	22	21	7	1	2	0	1622.595
20.	22	21	8	1	2	2	1695.139
21.	22	21	9	1	2	0	1588.652
22.	22	21	10	1	2	0	1568.686
23.	22	21	11	1	2	0	1572.013
24.	22	21	12	1	2	1	1574.010
25.	22	22	1	1	2	1	1617.936
26.	22	22	2	1	2	1	1365.695
27.	22	22	3	1	2	1	1647.220
28.	22	22	4	1	2	0	1455.543
29.	22	22	5	1	2	0	1589.983
30.	22	22	6	1	2	2	1508.787
31.	23	22	7	1	2	0	1694.207
32.	23	22	8	1	2	1	1739.096
33.	23	22	9	1	2	0	1596.948
34.	23	22	10	1	2	0	1656.120
35.	23	22	11	1	2	1	1639.116
36.	23	22	12	1	2	0	1491.528

1919/20 census-year

1920 calendar year
Jan. – Dec.

1920/21 census-year
Jul. – Jun.

1921 calendar year

1921/22 census-year

1922 calendar year

1922/23 census-year

*CensYear '21' denotes 1920/1921 census year, i.e. July 1920 to June 1921.
#Year '20' denotes 1920 calendar year, i.e. January to December 1920.

For timing-of-birth analysis, the required outcome measure describing MS risk is *MS-births rate,* where 'rate' is the number of MS births each month relative to the number of births in the reference Australian population:

$$\text{MS-births rate (by month and year)} = \frac{\text{No. MS births per month per year}}{\text{No. reference-population births per (same) month per (same) year}}$$

This MS-births rate is required also by *place of birth* (in Australia) and sex.

The 1981 (McLeod) MS Survey can be used to create such an 'MS-rates' dataset based on comparative birth rates, thus providing an opportunity to examine timing-of-birth effects on MS risk while also taking place of birth into account. The structure of this MS-rates dataset would be:

- Number of records=5,208 (=31 years x 12 months x 7 states/territories x 2 sexes):
  - birth year: 1920 to 1950 (n=31)
  - (birth) month: January to December (n=12)
  - (birth) place (state/territory): NT to TAS (n=7)
  - sex (n=2).

- For each record:
  - the 1981 MS Survey, described in Chapter 5, gives the number of MS cases born (per month, year, place and sex)
  - the 1981 census data, together with other Australian births-registration data (Table 6.2), will be used to derive the reference-population births denominators.

Thus, this MS-rates dataset comprises six variables pertaining to births—the four variables, year, month, place and sex (columns 1 to 5 in Table 6.3)—as well as:

- The total number of births in the Australian reference population in each month and year, separately for each birthplace and each sex, of persons who were still alive and in Australia on 30 June 1981 (column 7, Table 6.3), this

denominator to be derived using data from the 1981 census and national registration data covering births over the relevant period.

- The number of these births of persons who, by 30 June 1981, had been diagnosed as an MS case and who were still alive and in Australia on that date (column 6, Table 6.3) (case ascertainment using data from the 1981 MS Survey in Chapter 5).

Numerator data

For column 6 in Table 6.3, the numbers of MS births per month for every year over the 31-year period from 1920 to 1950, by birthplace and sex, were directly obtained by contracting unit-record data in the 1981 MS Survey dataset to frequencies of MS births per month per year, using Stata 8.0 software (release 8.0, 2003; Statacorp, College Station, Texas).

An example of these contracted frequency data for just the 1920/1921 census year (i.e. July 1920 to June 1921) is given in tabular form for males in Table 6.4, and for females in Table 6.5. For example, from Table 6.4, the MS survey dataset comprised two males born in NSW/ACT in July 1920; this is then seen in Table 6.3 in column 6 as the seventh record.

Table 6.4: Number (frequency) of Australian male MS cases born July 1920 to June 1921, by birth month and birth state (data contracted from unit records in 1981 MS Survey)

Year/month	Birthplace (state/territory)*								
	NSW/ ACT	VIC	QLD	SA	WA	TAS	NT	Not known	Total
1920 July	**2**	0	0	0	0	0	0	0	2
Aug.	0	0	0	0	0	0	0	0	0
Sept.	**1**	0	0	0	0	0	0	0	1
Oct.	0	0	0	**2**	0	0	0	**1**	3
Nov.	0	0	0	0	0	**1**	0	0	1
Dec.	0	0	0	0	0	0	0	0	0
1921 Jan.	**2**	0	0	0	0	0	0	0	2
Feb.	0	0	**2**	0	0	0	0	0	2
Mar.	0	0	0	0	0	0	0	0	0
Apr.	**1**	0	0	0	0	0	0	0	1
May	0	0	0	0	0	0	0	0	0
June	0	0	**1**	0	0	0	0	0	1
Total	6	0	3	2	0	1	0	1	**13**

* Birthplace states or territories listed in order of decreasing population size.

Table 6.5: Number (frequency) of Australian female MS cases born July 1920 to June 1921, by birth month and birth state (data contracted from unit records in 1981 MS Survey)

Year/month	Birthplace (state/territory)*								
	NSW/ ACT	VIC	QLD	SA	WA	TAS	NT	Not known	Total
1920 July	**1**	0	**1**	**1**	**3**	0	0	0	6
Aug.	**2**	0	0	0	0	**1**	0	0	3
Sept.	**1**	**2**	0	**1**	0	0	0	0	4
Oct.	**2**	0	0	**1**	0	0	0	0	3
Nov.	**1**	0	**1**	**1**	**1**	0	0	0	4
Dec.	**1**	0	0	0	0	**1**	0	0	2
1921 Jan.	**1**	0	0	0	**1**	0	0	0	2
Feb.	**1**	0	0	**1**	**1**	0	0	0	3
Mar.	0	0	0	0	0	0	0	0	0
Apr.	**6**	**1**	0	0	0	0	0	0	7
May	**2**	0	0	**1**	0	0	0	0	3
June	**4**	0	0	0	0	0	0	0	4
Total	22	3	2	6	6	2	0	0	**41**

* Birthplace states or territories listed in order of decreasing population size.

In order to calculate the MS risk for males born in NSW/ACT in July 1920, for example, the total number of males born in NSW/ACT in July 1920 (i.e. the denominator) also needs to be ascertained. Further, as shown in Tables 6.1 and 6.2, because the numerator data in Table 6.4 represent *surviving* males in 1981, the corresponding denominator must also take this *survival to 1981* into account.

6.2.2 Derivation of population denominators for MS rates

This section describes how the values for column 7 in Table 6.1 were derived; that is, the number of the reference Australian population *born* by month, year, birthplace and sex in Australia over the 1920 to 1950 period *that were still alive on 30 June 1981* (and still in Australia).

To derive the reference-population denominators, the 1981 Australian census taken on the same day as the MS survey provides a suitable starting point, being a similarly cross-sectional sample and including only those persons alive on 30 June 1981 and those residing in Australia on that date (Tables 6.1 and 6.2). However, the 1981 census did not contain the month-of-birth and place-of-birth information required for timing-of-birth analysis (Table 6.2). Fortunately, summary births-registration data provided by the ABS recorded the number of Australian births registered by sex and state (and by SD—the latter required for TAS because Hobart SD was surveyed rather than the whole state) for each month and year. It was therefore possible to estimate the number of persons born in each month and year from 1920 to 1950 *and still alive* on 30 June 1981, in each state/area and for each sex, by combining the ABS births-registrations data with the 1981 census data, as indicated in Table 6.2 and described in the following subsections.

Census data

Census year of birth

The 1981 census totals for each sex were given by 'census age'; that is, 'age in completed years on the census date, 30 June 1981', rather than directly by year of birth. For the purpose of denominator derivation, this 1981 'census age' could be equated to year of birth expressed in 'census years'; that is, in years spanning July of one year to June of the following year, as defined in Chapter 5 (footnote to Table 5.8). 'Census

year', as defined, was shifted six months from the calendar years, as illustrated in Table 6.3. (Later, once the denominators per month were derived and the MS-rates dataset reconstructed, timing of birth could be analysed simply by calendar year.)

Thus, the 30 June 1981 Census gave the numbers of males and females by census age, from which the numbers born in a given census year and surviving to 1981 could be deduced.

Restriction to Australia born: estimated numbers of Australia-born males and females for each census year of birth

The 1981 census totals by year included all those in Australia on census day, 30 June 1981, whereas only Australia-born persons were required for timing-of-birth analysis. For example, from the census, 68,328 men and 73,560 women born during the July 1920 to June 1921 census year were alive and in Australia on 30 June 1981, but these were not necessarily born in Australia. An additional 1981 census table provided the proportion of Australia born by census age group and sex, these age groups covering periods of five or 10 years and being for the whole region of Australia (Table 6.6). Assuming that the proportion Australia born did not differ markedly within these age groups, nor between regions within Australia, an estimate of the number of Australia born per census year of birth could be obtained for each sex.

Table 6.6: Proportion of persons Australia born, by age- and year-of-birth group and sex (data from the 1981 Census)

Year-of-birth group	Age group (years*)	Proportion Australia born	
		Males	Females
July 1916–June 1921	60-64	**0.71841**	0.75844
July 1921–June 1926	55-59	0.67489	0.72443
July 1926–June 1936	45-54	0.64837	0.70598
July 1936–June 1946	35-44	0.64400	0.68079
July 1946–June 1956	25-34	0.72455	0.73184

*Census age in completed years on census date 30 June 1981.

For example, of the total 68,328 males born during the 1920/1921 census year and alive on 30 June 1981, the proportion of these that were Australia born was 0.71841

(Table 6.6), giving an estimate of 49,087.5 for the number of males *born in Australia* between July 1920 and June 1921 and still alive in 1981 (Table 6.7, column 4). The resulting Australia-born estimates for both males and females for each year of birth over the 1920 to 1950 (calendar year) study period are shown in Table 6.7 in columns 4 and 6, respectively.

The Australia-born estimated totals (columns 4 and 6 of Table 6.7) for each census year of birth were then ready to be apportioned into month-of-birth and place-of-birth subtotals using the supplementary Australian births-registration data, these subtotals being the basis for the eventual denominators for analysis of timing of birth.

Table 6.7: Number of Australia-born persons in 1981 by sex and census year of birth over the 1920 to 1950 calendar-year period (columns 4 and 6). This is calculated from the total numbers of males and females (columns 3 and 5) and the proportions of Australia-born persons by census-age group and sex (Table 6.6). (All data from the 1981 Census; numeric example given in text shown in bold)

Year of birth (census year July–June[#])	Census age (years)	Males		Females	
		Total in 1981	Australia born 1981	Total in 1981	Australia born 1981
1919 /1920	61	60596	43532.77	63505	48164.73
1920 /1921	60	**68328**	**49087.50**	73560	55790.85
1921 /1922	59	69569	46951.42	71346	51685.18
1922 /1923	58	70852	47817.31	71748	51976.40
1923 /1924	57	72164	48702.76	71339	51680.11
1924 /1925	56	74607	50351.52	74285	53814.28
1925 /1926	55	76150	51392.87	74425	53915.71
1926 /1927	54	75548	48983.06	73116	51618.43
1927 /1928	53	78316	50777.74	74288	52445.84
1928 /1929	52	78142	50664.93	74198	52382.30
1929 /1930	51	75594	49012.88	71410	50414.03
1930 /1931	50	79569	51590.15	77363	54616.73
1931 /1932	49	73866	47892.50	69325	48942.06
1932 /1933	48	72876	47250.61	69242	48883.47
1933 /1934	47	72603	47073.61	68311	48226.20
1934 /1935	46	73546	47685.02	70541	49800.53
1935 /1936	45	78115	50647.42	74714	52746.59
1936 /1937	44	78370	50470.28	74408	50656.22
1937 /1938	43	81044	52192.34	77471	52741.48
1938 /1939	42	85337	54957.03	81868	55734.92
1939 /1940	41	85072	54786.37	79652	54226.28
1940 /1941	40	91776	59103.74	87703	59707.32
1941 /1942	39	92782	59751.61	88440	60209.07
1942 /1943	38	91988	59240.27	88435	60205.66
1943 /1944	37	100513	64730.37	96998	66035.27
1944 /1945	36	105153	67718.53	102351	69679.54
1945 /1946	35	107240	69062.56	103333	70348.07
1946 /1947	34	126914	91955.54	123426	90328.08
1947 /1948	33	119398	86509.82	116407	85191.30

Year of birth (census year July–June[#])	Census age (years)	Males		Females	
		Total in 1981	Australia born 1981	Total in 1981	Australia born 1981
1948 /1949	32	116254	84231.84	116035	84919.05
1949 /1950	31	117525	85152.74	115322	84397.25
1950 /1951	30	120578	87364.79	120373	88093.78

[#] Census year as defined in Chapter 5, Table 5.8, e.g. 1919/1920 census year is July 1919 to June 1920.

Supplementary births-registration data

Summary births-registration data (provided by J. Wall, ABS, pers. commn) were utilised to overcome the main limitation of the census data, namely the absent month-of-birth and place-of-birth data (see Table 6.2). Using these data, estimated numbers born by month, year, state/division and sex in Australia were able to be derived to fill column 7 in Table 6.3. That is, population births in Australia recorded by state (and SD) from 1860 to 1969 and available also by sex and month over most of the years required were able to be used to derive the required births estimates by direct proportion from the census by-year and by-sex totals, as detailed in the following subsection.

Month and place of birth estimates for males and females for each census year of birth

From the ABS Australian births-registration summary tables, the numbers of births registered by month and state (and SD for Hobart, TAS) were tabulated for each census year and sex separately, over the period 1920 to 1950, as shown in the example for male births during the 1920/1921 census year (Table 6.8). The total number of registered births *over all states and months* for that census year was then calculated for each sex.

For example, in Table 6.8 for the 1920/1921 census year, 2,655 male births were registered in July 1920 in NSW (including ACT) and the total male births over the July 1920 to June 1921 year in Australia was 71,290. The corresponding Australia-born census total estimated for this (census) year of birth for males was 49,087.5 (Table 6.7) and this number represents the male *survivors* (to 1981) of that initial 1920/1921 birth cohort of 71,290.

The *proportion* of these *surviving* male (or female) persons born that year in Australia (and in Australia in 1981) relative to the total *initial* number of male (or female) births in Australia that year (i.e. 49,087.5 / 71,290 for males) was then applied to each *by-month* and *by-state* (or division) *births* number in example Table 6.8, resulting in the new by-month and by-state/area estimates in example Table 6.9.

For example, in Table 6.8, 2,655 initial male births in NSW/ACT in July 1920 were estimated to become 1,828.13 *still alive in 1981 (and still residing in Australia)* male births in Table 6.9. That is, each initial births-registration value has now been adjusted (downwards) to take account of both the survival of each birth cohort to 30 June 1981 and their residence still in Australia:

$$\frac{\text{1981 male resident survivors of Australian 1920/1921 birth cohort}}{\text{Total Australian male births 1920/1921}} \times 2{,}655$$

$$\text{i.e.} \quad \frac{49{,}087.5}{71{,}290} \times 2{,}655 = \mathbf{1{,}828.13}$$

Thus, each monthly regional (initial) birth rate has been multiplied by a proportion expressing the survival of that age group or census year group to 1981 and their continued residence in Australia.

In other words, the 1981 census total of Australian residents (for each census year and sex) has been subdivided into by-month and by-state/area values, as required for analysis; these values are now *in the same by-month and by-state/area proportions* as existed in the corresponding initial birth cohorts. This can be seen by re-arranging the above equation for the same example NSW/ACT July 1920 male births. That is,

$$\mathbf{1{,}828.13} = \frac{2{,}655}{71{,}290} \times 49{,}087.5$$

Or,

$$\text{NSW July 1920 male births (\textbf{New estimate})} = \frac{2{,}655}{\text{Total male 1920/1921 births \textbf{(both numerator and denominator from births register)}}} \times \text{Resident males born 1920/1921 and surviving to 1981 \textbf{(from 1981 Census)}}$$

The 'add-up-check total' shown in parentheses in the last column of Table 6.9; that is, 49,088.19, is within just 0.001% 'rounding error' of the original census total, 49,087.5, showing that the new month of birth and place of birth estimates correctly add up to the original totals for each census year.

As depicted diagrammatically in Figure 6.1, the 1981 yearly census totals have now been apportioned into *by-month* and *by-state/area* births estimates using the same by-month and by-state/area proportions that existed in their initial Australian birth cohorts, solving the problem of the absence of these data in the 1981 Australian census.

To derive month of birth and place of birth as described, it was necessary to assume that survival from birth to 1981 of this 1920- to 1950-born restricted group (i.e. 30 to 61 year olds in 1981), and their continued residence in Australia until 1981, were not dependent on either month or place of birth (state) within Australia. These and other necessary assumptions made will be discussed further in Section 6.4.

These month- and place-of-birth estimated values for each sex and year would then become the required population denominators (of MS rates) for column 7 in Table 6.3, after further slight adjustment to account for the effects of incomplete case ascertainment in the MS survey—this final adjustment is discussed in the following section (Section 6.2.3).

Estimates from 1928 onwards

Since the ABS summarised registered births separately by sex only up until 1928, the remaining 1929 to 1950 period required an additional estimation step. For this period, only totals (males and females combined) were available by month, but male and female data were given by state/division. Therefore, after first establishing for the earlier period for which full data were available (i.e. up to 1928) that male to female ratios at birth did not vary significantly by month (within state/division and year), annual *sex ratios* were applied, within each state/division and year, to divide the *monthly* totals into male and female subtotals to give the required denominators by sex.

Table 6.8: ABS births-registration data (from summary tables for 1920 and 1921) for males born in Australia, by state/territory and by month from July 1920 to June 1921 (census year 1920/1921 only). (TAS births shown separately for Hobart SD and remaining non-Hobart area of state, because Hobart SD denominator is required for TAS sample; specific numeric example given in text shown in bold)

Year/month	Registered births of Australian males born 1920/1921 census year								
	NSW/ACT	VIC*	QLD	SA	WA	TAS		NT*	Total
						Hobart SD	Non-Hobart*		
1920 July	**2655**	1737	977	504	378	63	170	2	6486
Aug.	2558	1610	1006	577	394	68	185	0	6398
Sept.	2425	1679	837	568	351	87	212	1	6160
Oct.	2382	1581	866	500	326	82	187	2	5926
Nov.	2326	1576	829	523	343	78	189	0	5864
Dec.	2398	1534	910	508	333	69	164	1	5917
1921 Jan.	2326	1464	906	489	279	63	174	4	5706
Feb.	2144	1422	858	526	355	61	159	6	5531
Mar.	2220	1482	955	535	340	67	178	2	5779
Apr.	2284	1599	881	505	390	64	179	3	5905
May	2291	1502	861	522	336	65	191	0	5768
June	2298	1591	880	497	337	73	165	9	5850
Total	28308	18777	10766	6254	4162	840	2153	30	**71290**

* VIC, NT and non-Hobart TAS necessarily included for apportioning at-birth denominators but not included in final dataset (see Section 6.2.3).

Table 6.9: Reference-population denominators given by estimated births of 1981 survivors from 1920/1921 Australian birth cohort, by month and birthplace. (TAS births shown separately for Hobart SD and non-Hobart areas, because Hobart SD denominator is required for TAS sample; specific numeric example given in text shown in bold)

Year/month	Estimated births of 1920/1921-born Australian males still alive 30 June 1981 in Australia								
	NSW/ACT	VIC*	QLD	SA	WA	TAS		NT*	(Check Total)
						Hobart SD	Non-Hobart*		
1920 July	**1828.13**	1196.03	672.72	347.03	260.28	43.38	117.06	1.38	(4466.01)
Aug.	1761.34	1108.58	692.69	397.30	271.29	46.82	127.38	0.00	(4405.41)
Sept.	1669.76	1156.09	576.33	391.10	241.68	59.90	145.97	0.69	(4241.53)
Oct.	1640.15	1088.61	596.29	344.28	224.47	56.46	128.76	1.38	(4080.41)
Nov.	1601.59	1085.17	570.82	360.12	236.18	53.71	130.14	0.69	(4038.41)
Dec.	1651.17	1056.25	626.59	349.79	229.29	47.51	112.92	0.69	(4074.21)
1921 Jan.	1602.28	1008.05	623.84	336.71	192.11	43.38	119.81	2.75	(3928.93)
Feb.	1476.27	979.13	590.79	362.18	244.44	42.00	109.48	4.13	(3808.43)
Mar.	1528.60	1020.45	657.58	368.38	234.11	46.13	122.56	1.38	(3979.19)
Apr.	1572.67	1101.01	606.62	347.72	268.54	44.07	123.25	2.07	(4065.95)
May	1577.49	1034.22	592.85	359.43	231.36	44.76	131.52	0.00	(3971.62)
June	1582.31	1095.50	605.93	342.21	232.04	50.26	113.61	6.20	(4028.08)
Total	19491.78	12929.10	7413.05	4306.26	2865.79	578.39	1482.47	21.35	(49088.19)

* VIC, NT and non-Hobart TAS necessarily included for apportioning at-birth denominators but not included in final dataset (see Section 6.2.3).

Census total (in 1981, Australia born, by census year of birth, by sex)

↓

Apportioned by birthplace and birth-month proportions

NSW/ACT born, by birth month					
Jul	Aug	Sep	Oct	Nov	Dec
Jan	Feb	Mar	Apr	May	Jun

VIC born					
J	A	S	O	N	D
J	F	M	A	M	J

QLD born					
J	A	S	O	N	D
J	F	M	A	M	J

SA-		
J	-	D
J	-	J

WA	
J	D
J	J

TAS NT born

12 by-month denominators for NSW/ACT birthplace (each sex, census year of birth)

(12 denoms)

J	A	S	O	N	D
J	F	M	A	M	J

Hobart SD born

J	A	S	O	N	D
J	F	M	A	M	J

Non-Hobart TAS

J	A	S	O	N	D
J	F	M	A	M	J

NT born

Figure 6.1: Diagram showing apportioning of 1981 census totals (top of diagram), for each census year of birth and sex, into birthplace and birth-month subtotals for required denominators, using births proportions from ABS registration summary tables. Boxed areas shown for birthplaces across centre of diagram indicate their approximate population sizes. Twelve by-month denominators result for each birthplace but for clarity are shown for NSW/ACT and VIC only. (Shaded cells indicate birthplace denominators not used in final MS-rates dataset [i.e. VIC, NT and non-Hobart TAS, see Section 6.2.3], but necessarily included for apportioning.)

Figure 6.2 summarises the process of derivation of MS-rates denominators described so far. That is, the 1981 Census (indicated by blue areas of pie-chart and text boxes) provides the total *surviving Australia-born* numbers by year of birth and by sex, and the ABS births-registration information (light blue/green text box) provides the supplementary data required to derive the denominator estimates by month and place of birth from these census totals.

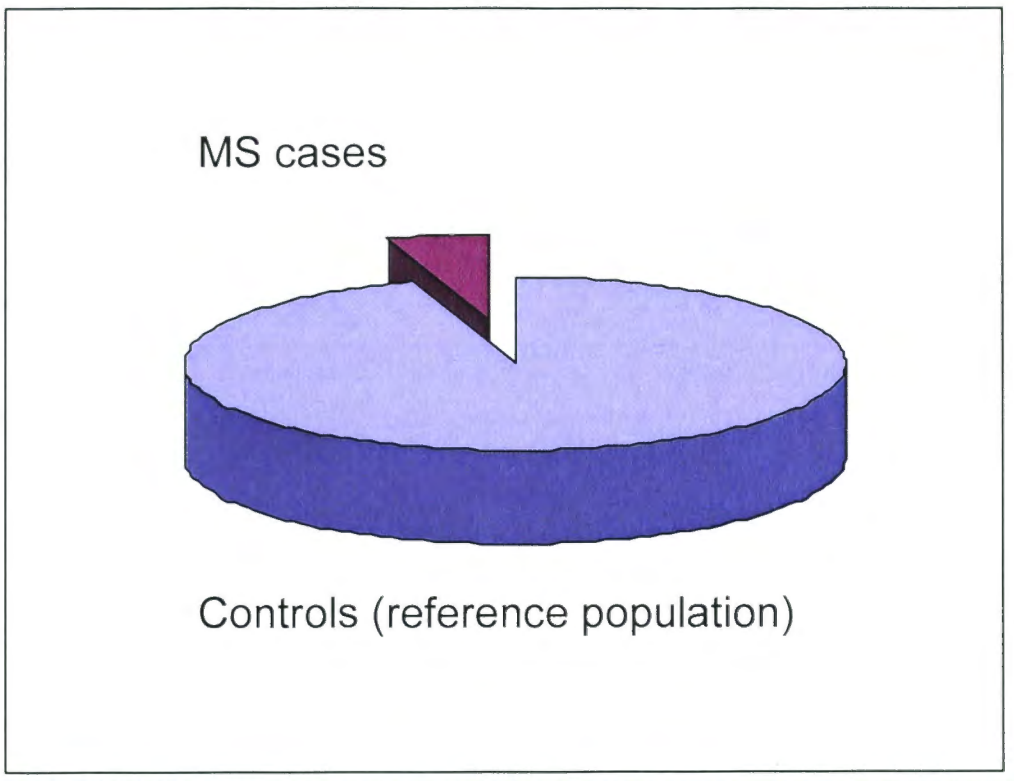
6.2.3 Birthplace and place in 1981—further adjustment for 1981 case ascertainment losses

This section will describe how and why the required dataset will be adjusted to overcome the effects of the incomplete ascertainment of cases in the 1981 MS Survey. The 1981 MS Survey did not satisfactorily survey the following areas of Australia: the state of VIC, which has the second highest population in Australia; the lowest-populated state/territory area, the NT; and the area of TAS outside Hobart SD (denoted as 'non-Hobart TAS') (J. McLeod, pers. commn). This has potential consequences for the required dataset, particularly for highly populated VIC, because some MS cases may have migrated from their original birthplace state to another Australian state during the period between birth and when both the MS survey and the census were taken in 1981. Depending on whether their 'new' state was surveyed sufficiently or not, these cases migrating interstate may or may not have been ascertained (Table 6.10).

Table 6.10 shows birthplace and place in 1981 (i.e. survey area) data (taken from the 1981 Australian MS Survey) for the surveyed MS cases born between 1920 and 1950, and summarises the net interstate post-birth migration for the MS cases. It can be seen for the surveyed states that a high proportion of cases ascertained in each state or survey area were, in fact, born in that state (last column of Table 6.10). At the same time, lower numbers born in other states or territories have also migrated to each surveyed state by 1981.

MS CASE DATA

MS Cases—1981 MS Survey (cross-sectional survey 30 June 1981 by state; unit-record data: year, month of birth, birthplace (state) in Australia, sex



REFERENCE POPULATION DATA

Reference (census) population—census date 30 June 1981 (cross-sectional: census age (years), country of birth) → totals by census year, 1920 to 1950 Australia born

+

Australian births registrations (by state, sex, year and month → supplementary data for 1920 to 1950 born

⇩ **Apportioning census totals**

Denominator estimates by month, year, birthplace (state), sex, for 1981 census population born 1920 to 1950

Figure 6.2: Initial sources and treatment of data for analysis of timing of birth.

Table 6.10: 1981 survey area and birthplace of MS cases born 1920 to 1950 (data from 1981 Australian MS Survey)

1981 MS Survey area* /residence	Birthplace (state) in Australia*									% born in survey state
	NSW /ACT	VIC	QLD	SA	WA	TAS	NT	Unknown	Total	
NSW/ACT	**875**	50	28	10	5	3	1	6	978	**89.5**
VIC	-	-	-	-	-	-	-	-	0	-
QLD	32	16	**183**	1	3	6	0	0	241	**75.9**
SA	11	11	1	**181**	3	1	0	0	208	**87.0**
WA	7	10	3	4	**112**	2	0	0	138	**81.2**
non-H TAS	-	-	-	-	-	-	-	-	0	-
Hobart SD	3	13	1	1	0	**49**	0	0	67	**73.1**
NT	-	-	-	-	-	-	-	-	0	-
Total	928	100	216	197	123	61	1	6	1632	

*State/territory areas shown in order of decreasing population size.

The two main scenarios that need to be considered are:

1. MS cases born in the **unsurveyed areas** of VIC, NT and non-Hobart TAS will have either stayed in their 'home' area since birth and not have been ascertained, or *will have migrated to one of the surveyed state/territory areas and have been ascertained there in 1981.*
2. MS cases born in the **surveyed areas** will have either stayed in their 'home' area (or other surveyed area) and have been ascertained, or *will have migrated to one of the unsurveyed areas by 1981 and then not have been ascertained.*

In this section, only the migrating cases need be considered for adjustment, as those not migrating will have been ascertained (scenario 2), or not (scenario 1), correctly.

Those migrating cases in the first scenario, for VIC and NT born at least, will be considered directly as numerators and eventually excluded, for reasons given subsequently. TAS-born MS cases will necessarily be considered a special case. The migrating cases in the second scenario will be estimated as a proportion and used to adjust the denominators, rather than the numerators, for the five surveyed birth states.

MS cases born in unsurveyed areas VIC, NT and non-Hobart TAS

From Table 6.10, because VIC and NT (and non-Hobart TAS) were not surveyed in 1981, there are likely to be considerable proportions of VIC-*born* and NT-*born* (and possibly non-Hobart-TAS-*born*) MS cases residing in these survey areas that have not been ascertained (see scenario 1, above). The few VIC-born and NT-born cases that *have* migrated to other, surveyed areas (and, therefore, have been inadvertently ascertained in the survey), are likely to be unrepresentative of the total born in that state or territory; therefore, these VIC-*born* (n=100 cases) and NT-*born* (n=1 case) migrants to other surveyed areas should, and will, be excluded.

By the same reasoning, non-Hobart-*born* cases should also be excluded; however, these are not identifiable in the dataset from those born in Hobart SD—only 'TAS-born' cases are identifiable. Therefore, TAS has to be considered a special case. All 'TAS-born' MS cases were *considered to be Hobart SD-born* and included in timing-of-birth analysis using a reference-population 'Hobart SD' birthplace denominator. Indeed, TAS is such a small state that most TAS MS cases in 1981 would have been born in the major medical facilities in the Hobart capital (A-L. Ponsonby, pers. commn). Given TAS's high latitude (and low UVR, see Chapter 4) and the possible importance of this to the analysis, this strategy was considered preferable to the alternative strategy of considering all 'TAS born' as non-Hobart born and then excluding all TAS-born cases from the analysis. The potential effects of this strategy will be further discussed and justified in Section 6.4.

MS cases born in surveyed areas and migrated to unsurveyed areas

The effects of net post-birth migration from the surveyed states (i.e. see scenario 2, above) to VIC and NT (but not non-Hobart TAS) could be taken into account by utilising another census dataset to estimate these likely sampling losses and then adjusting the appropriate denominators.

Calculation for adjustment of denominators

To determine the amount by which the denominator data should be adjusted, by census year of birth, a single census-year birth cohort—for example, NSW/ACT-born in census year July 1920 to June 1921—can be considered.

For this cohort, there are five possibilities with respect to surveying both the whole population cohort and MS ascertainment of the cohort in 1981 (i.e. to gain the MS rate in this cohort), these categories including 'death before 1981' and 'migration out of Australia'. Since the Australian census counts only those persons alive and in Australia on census day, and since the MS survey was conducted similarly, only the remaining three possibilities need be considered. That is, the survey possibilities for those persons alive in Australia on census and prevalence day (30 June 1981) are as shown (with theoretical percentages and proportions) in Table 6.11.

Table 6.11: Theoretical ascertainment and proportions indicating post-birth migration of both MS cases ('M', 1981 MS Survey) and the reference population ('C', 1981 Census), for a single census year, 1920/1921, and a single birthplace, NSW/ACT

Place in 1981 (%)		1981 MS Survey*		1981 Census**	
1. Still in NSW/ACT	85%	Ascertained	0.85M	Counted	0.85C
2. Migrated interstate to surveyed states QLD, SA, WA, or Hobart SD (TAS)	10%	Ascertained	0.10M	Counted	0.10C
3. Migrated interstate to VIC or NT (not surveyed) (or non-Hobart TAS)	5%	**Not ascertained**	**0.05M**	Counted	0.05C
Total	100%	M		C	

*Persons with MS and born in NSW/ACT in 1920/1921 census year.
**Persons born in NSW/ACT—estimated from 1981 census and July 1920 to June 1921 births-registration table (ABS): the census counts only those persons in categories 1, 2 and 3 for all of Australia, and the births-registration table is used to determine the number estimated to have been born in NSW/ACT by month (see Section 6.2.2).

The MS rate in this 1920/1921-born cohort is given by M/C. If VIC, NT and non-Hobart TAS MS cases were fully ascertained, then the (true) MS rate is easily estimated using the counts in 1, 2 and 3:

$$\frac{0.85M + 0.10M + 0.05M}{0.85C + 0.10C + 0.05C} = \frac{1.00M}{1.00C} = \frac{M}{C}$$

However, in the MS survey the '0.05M' category (shown in bold in Table 6.11), comprising those persons who migrated to the unsurveyed states or areas has not been ascertained. That is, a proportion of the MS cases (here, theoretically 0.05) has been 'lost to survey'.

Therefore, either:

a. the numerator, M, should be increased by 0.05M, or
b. the denominator, C, should be decreased by 0.05C.

However, in practice only the denominator—that is, the estimated denominator for each month and place of birth (Table 6.9)—can actually be adjusted; it would not be possible to increase the numerators when month(s) of birth cannot be allocated to these 'generated' MS cases. Therefore, in the above example, the denominator, C, should be multiplied by 0.95, this factor given by 1 – 0.05. Or, in general terms for each surveyed birthplace state:

Each by-month, by-sex and by-year denominator, C (as given in Table 6.9), will be multiplied by **(1–*x*)**, where ***x*** is defined as the *proportion* of the total number of those persons born in that state each year that are later *resident in VIC or NT in 1981*.

Thus, for the surveyed birthplace states, NSW/ACT, QLD, SA and WA, and TAS (=Hobart SD birthplace, by definition), the above (1–***x***) adjustment factor was used to account for case ascertainment losses to the unsurveyed state/territory areas of VIC and NT, as shown in Table 6.12, but not for potential case losses to non-Hobart TAS. Using census-population residence-by-birthplace data, these potential losses to non-Hobart TAS are discussed further at the end of this section (Section 6.2.3) and shown to be relatively small.

Use of 1976 census-population migration data to adjust for 1981 case losses to VIC and NT

Available residence-by-birthplace data from the 1976 Census (these data were not available in the 1981 Census but were for the previous Australian census in 1976) were used to determine the actual proportions (***x***) by which each denominator needed to be adjusted in order to account for sampling losses to the main unsurveyed areas of VIC and NT. These data (provided by the Australian Social Science Data Archive [ASSDA]) comparing *birthplace* (state/territory) and *residence* (state/territory) of the Australia-born 1976 census population, quantified net interstate movement within Australia between the time of birth and 1976 for the Australia-born census population by census age. A necessary assumption for this adjustment was that there was little change in net migration in the Australian population between the two censuses in 1976 and 1981.

Table 6.12 shows how '***x***' was derived for each birthplace state/territory and each census year of birth between 1920 and 1950, using the single census year, 1920/1921, as an example. The lower part of Table 6.12 indicates the proportions, for each birthplace, of the example Australian 1920/1921-born population that were resident in VIC and NT in 1976—that is, *had migrated to areas that were not surveyed in 1981.* Although quite small for the surveyed states, these are the proportions that were taken into account for deriving the denominator estimates for these states. For example, 3.88% of the total number (36,470) of NSW/ACT-born persons in 1920/1921 who were still alive and in Australia in 1976, would not have been included in any survey that excluded these two areas, VIC and NT, because of net interstate migration between the birth year and the survey year.

Table 6.12: Area of residence in 1976 and birthplace of Australians born in example 1920/1921 census year (ASSDA 1976 census data). Lower part of table shows number, and proportion (*x*)* (bolded for surveyed states), residing in VIC or NT, of total for each birthplace state/territory

1976 area of residence	Birthplace (state/territory) in Australia								
	NSW/ACT	VIC	QLD	SA	WA	TAS	NT	Unknown	Total
NSW/ACT	31970	1985	1344	521	295	272	10	4008	40405
VIC	1351	20395	353	582	252	635	4	2681	26253
QLD	2161	1060	11878	191	108	150	9	1578	17135
SA	493	551	77	7265	137	67	14	686	9290
WA	314	397	74	219	4531	50	4	673	6262
TAS	116	182	31	35	14	2857	0	184	3419
NT	65	52	56	66	29	7	153	89	517
Total	36470	24622	13813	8879	5366	4038	194	9899	103281
VIC + NT	1416	20447	409	648	281	642	157		
Proportion* VIC+NT (*x*)	**0.038826**	0.830436	**0.029610**	**0.072981**	**0.052367**	**0.158990**	0.809278		

* Proportion resident in VIC or NT indicates proportion (*x*) of each birthplace population that would have been 'lost to survey' were VIC and NT not sampled at all (as in 1981 MS Survey).

Assuming that net interstate migration among the MS cases and among either of the reference Australian populations in 1976 and 1981 were similar, this proportion of otherwise-excluded persons (***x***) from the 1976 population was used to adjust the corresponding 1981 population *denominators* (by 1–***x***). This was to account for the losses *of cases* to the unsurveyed states, VIC and NT, in the 1981 MS Survey.

For each birthplace state, It was necessary to use the same (1–***x***) factor for the denominators for each month within each census year, and for each sex, because the 1976 migration data was available only by census age (and therefore census year) and for the total number of persons. The required additional assumption, therefore, was that net interstate migration did not depend on month of birth within each census year, or on sex.

For the example 1920/1921 census year for males, the estimated births shown in Table 6.9 have now become the final migration-adjusted values shown in Table 6.13, for the surveyed birth states NSW/ACT, QLD, SA, WA and TAS (=Hobart SD by definition).

For example, for NSW/ACT males born in July 1920, the previous example estimate of 1828.180 births in Table 6.9 has now decreased (by 3.9%) to an estimated 1757.199 births in Table 6.13. The values in Table 6.13 are now the reference-population denominators required for the final MS-rates dataset for just the 1920/1921 birth cohort, after accounting for migration of that birth cohort to the unsurveyed areas, VIC and NT, by 1976/1981.

Table 6.13: Migration-adjusted population denominators for 1920/1921 census-year-born males, given by estimated male births of 1981 survivors from 1920/1921 Australian birth cohort adjusted for migration 'losses' to unsurveyed states VIC and NT in 1981, by month and birthplace. (Specific example given in text shown in bold)

Year/month	Birthplace (state/area)[#]				
	NSW/ACT	QLD	SA	WA	Hobart SD*
1920 July	**1757.199**	652.823	321.717	246.653	36.483
Aug.	1693.000	672.201	368.315	257.093	39.379
Sept.	1604.974	559.276	362.570	229.035	50.382
Oct.	1576.515	578.654	319.163	212.722	47.486
Nov.	1539.452	553.931	333.845	223.815	45.170
Dec.	1587.104	608.054	324.270	217.290	39.958
1921 Jan.	1540.113	605.381	312.142	182.053	36.483
Feb.	1418.996	573.308	335.760	231.645	35.325
Mar.	1469.296	638.123	341.505	221.857	38.800
Apr.	1511.654	588.677	322.355	254.483	37.063
May	1516.287	575.313	333.207	219.247	37.642
June	1520.920	588.008	317.248	219.900	42.275

[#]Surveyed birthplace states/areas only shown (cases and denominators from unsurveyed VIC and NT birthplaces not included in final MS-rates dataset).
*Assuming TAS-born cases to be (all) Hobart born, the denominator required is that of Hobart SD, adjusted for migration losses to VIC and NT (assuming same *relative* proportion lost for Hobart SD born as for TAS born).

As an illustration of the data adjustments described in this subsection, the data in Table 6.12, for the Australian population born 1920/1921, can be further summarised by comparing the relative proportions residing in the total 'surveyed' area (as defined in the 1981 MS Survey) and the main 'unsurveyed' area, VIC plus NT, for each birthplace. That is, the '*surveyed*' proportion for each birthplace will include those persons remaining in their birthplace state or territory, for the 1981-surveyed states/territories only. However, for the 1981-unsurveyed VIC and NT, the relatively high proportion remaining 'home' will necessarily now be included in the '*unsurveyed*' proportion (Figure 6.3).

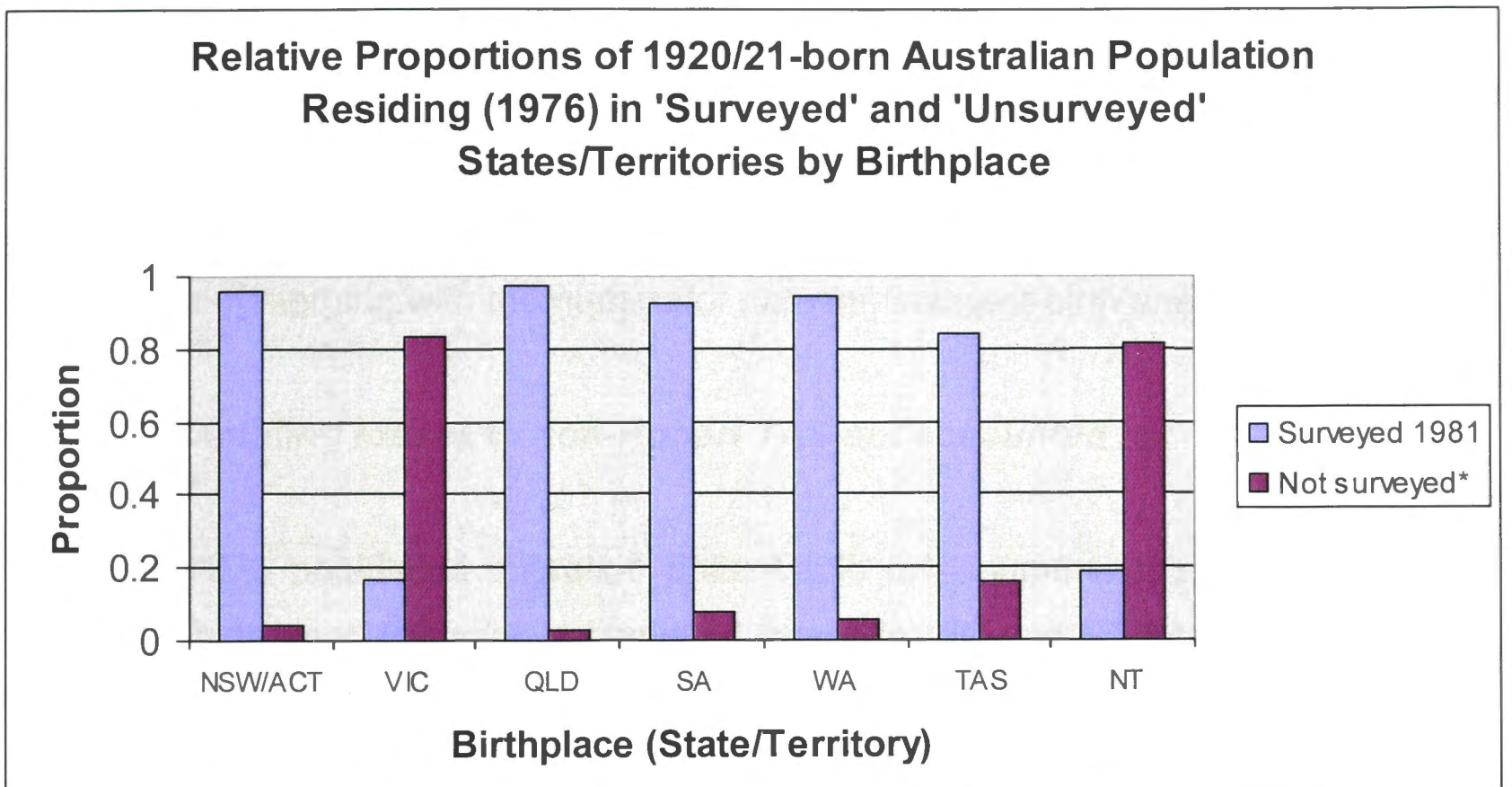


Figure 6.3: Relative proportions of 1920/1921-born Australian population residing (in 1976) in total 'surveyed' and 'unsurveyed' states/territories as defined in 1981 MS Survey, by birthplace. (Data from 1976 Census for 1920/1921-born persons [ASSDA]; *Unsurveyed states/territories [1981 MS Survey]=VIC, NT; surveyed states/territories [1981 MS Survey]=NSW/ACT, QLD, SA, WA, TAS [Hobart SD].)

Figure 6.3 shows the proportions of the state-born populations (and, by assumption, of their MS cases) that potentially would have been 'lost to survey' because of migration to the unsurveyed areas, VIC and NT, but which have now been accounted for. This figure further emphasises the relatively large proportions of VIC-*born* and NT-*born* 'losses' from the population (if sampled as in the 1981 MS Survey) and, by inference, from the MS survey itself, when these two areas were not sampled, because in these states/territories the high proportion remaining 'home' since birth have not been ascertained. As previously stated with regard to Table 6.10, because the relatively few VIC-born and NT-born migrants to other, surveyed states are probably not representative of this much larger, unsurveyed group remaining 'home' (>80% of total), Figure 6.3 shows why the migrating VIC-*born* and NT-*born* MS cases cannot be included in the final MS-rates dataset. TAS-born losses to VIC and NT, on the other hand, while greater than from other surveyed states, are still less than 20%, and these have now been accounted for in the final Hobart SD denominator estimate (by assuming the same *relative* proportion lost for Hobart SD born as for TAS born).

In conclusion, a similar table to Table 6.13, of migration-adjusted denominators by birth month and by birthplace, was constructed (in Microsoft Excel 2000) for each of the 31 census years of birth over the 1920 to 1950-born study period, for each sex. These tables of adjusted denominators were then ready for conversion to Stata 8.0 format for the purpose of merging with the numerator data for timing-of-birth analysis.

Potential sampling losses to non-Hobart TAS not accounted for

Using the 1976 population migration data, Table 6.14 summarises, for the example 1920/1921 birth year, the relative sampling losses to VIC and NT from each birthplace (=proportions from Table 6.12, shown in bold, these losses now having been taken into account). The table also summarises potential losses to non-Hobart TAS that have *not* been accounted for, as calculated from the ASSDA residence-by-birthplace population data from the 1976 Census. This table, expressed as both proportions and percentages of each birthplace total, shows that the potential unaccounted-for case losses to non-Hobart TAS from each of the other surveyed birthplace states were much smaller (~0.1 to 0.2%) than the proportions 'lost' to VIC and NT (2.96 to 15.9%) that have now been accounted for.

Table 6.14: Potential case ascertainment losses (as proportion and % of birthplace population) from each of the 1981-surveyed birthplace states because of migration from birthplaces to unsurveyed (1981) areas VIC and NT, and non-Hobart TAS, estimated from residence-by-birthplace data in 1976 Census for 1920/1921-born Australians

Unsurveyed areas (1981)	**1981-surveyed birthplace states**				
	NSW/ACT	QLD	SA	WA	TAS
VIC + NT (%)#	**0.03882 (3.88%)**	**0.02961 (2.96%)**	**0.07298 (7.30%)**	**0.05236 (5.24%)**	**0.15899 (15.90%)**
Non-Hobart TAS (%)*	0.00161 (0.16%)	0.00130 (0.13%)	0.00224 (0.22%)	0.00186 (0.18%)	---

Losses now accounted for.
* Loss unaccounted for.

Potential case losses due to net migration from Hobart SD to non-Hobart TAS were not quantifiable at all, because specific birthplace within this smallest Australian state (TAS) could not be obtained, but these losses were assumed to have also been small. As previously stated, most Tasmanians surveyed in 1981 would likely have been born in the Hobart capital.

6.2.4 Merging numerators and denominators for longitudinal MS-rates dataset

Numerator and denominator data derived as described were combined into a single year-by-year and month-by-month series in count (frequency) form, as shown in Table 6.18 (see Results, Section 6.3.3). The adjusted denominator tables in frequency format in Microsoft Excel 2000 for each census year were converted from MS Excel 2000 to Stata 8.0 using the software conversion program Stat Transfer v7.0. These were then merged with the numerator data (also in frequency format, see Section 6.2.1) in Stata 8.0, using a unique identifier for each of the now 3,720 database categories (31 years x 12 months x 2 sexes x 5 birthplace states/areas, NSW/ACT, QLD, SA, WA, Hobart SD [TAS]).

This merged MS-rates dataset with adjusted denominator estimates by birthplace and month of birth (and by year of birth and sex) will now make it possible (in Chapter 7) to derive estimates of MS rates by sex, month of birth and year of birth for timing-of-birth analysis, and by birthplace rather than by 1981 survey area as in Chapter 5.

6.3 Results

6.3.1 Final MS-case numbers (numerators)

The final MS-case sample numbers available for subsequent timing-of-birth analyses (in Chapter 7) are summarised by birthplace state and sex in Table 6.15. A total of **1,524** MS cases (i.e. 1,631 [Chapter 5] with known month of birth, less VIC born [n=100] and NT born [n=1] as well as unknown birthplace [n=6] [Table 6.10]) resulted.

Table 6.15: Final sample numbers for MS cases for MS-rates dataset. Number of MS cases (1920 to 1950 born) by Australian birthplace (state) and sex

Birthplace in Australia (state)	Number of MS cases (1920 to 1950 born)		
	Males	Females	Total
NSW/ACT	279	649	928
QLD	76	140	216
SA	54	143	197
WA	26	97	123
TAS (Hobart SD)	21	39	60
Total	456	1068	**1524**

6.3.2 Final population denominators and preliminary prevalence estimates

The final migration-adjusted denominator estimates are summarised in the next two tables and are shown together with the final MS-case numbers. The total reference-population denominator represented 2,468,779 persons born in Australia between 1920 and 1950 (i.e. aged 30 to 61 completed years in June 1981). Preliminary crude prevalence estimates have also been calculated and are shown by birthplace state and sex in Table 6.16 (all years of birth and months of birth combined, and states now shown in order of increasing south latitude), and by month of birth and sex (years of birth and birthplace states combined) in Table 6.17. The purpose of the summary crude prevalence estimates in these tables is to check the adjusted denominator estimates calculated in this chapter for obvious errors and to provide an indication of the possible relationships of interest to the analyses in the next chapter. In addition, the assumptions made particularly for TAS can also be checked (discussed in Section 6.4).

Crude prevalence by birthplace and sex

Consistent with total age-specific prevalence calculated in Chapter 5 for 30 to 59 year olds (then using the 1981 census denominators: prevalence 62.2/100,000, see Section 5.3.2, Table 5.8), Table 6.16 shows an overall MS prevalence of 61.7 per 100,000 for the approximately similarly-aged 1920 to 1950 Australia-born cohort using the now birthplace-apportioned and migration-adjusted denominators. A female to male overall

prevalence ratio of 2.27 (Table 6.16) is also comparable with the previous ratio of 2.33 evident in Table 5.5, giving confidence in the final denominators derived in this chapter.

Table 6.16 also indicates an association with increasing south latitude *of* birthplace for the total Australia-born sample with all years and months of birth combined; that is, a 4.6-fold increase in prevalence from QLD born (low mean latitude,[16] 25.1 °S) to TAS born (high mean latitude, 42.8°S) is suggested. This result is consistent with the McLeod group's findings of an overall four-fold increase in prevalence from QLD to TAS for *survey area* within Australia rather than birthplace (see Section 5.3.2), but now indicates a similarly strong relationship between latitude of *birthplace* and MS prevalence. Moreover, there appear to be no sex differences in this latitude-of-birthplace association with MS prevalence, both males and females indicating a 4.6-fold (or 3.6%) increase in prevalence from QLD to TAS birthplaces.

Crude prevalence by month of birth and sex

Importantly, for the total Australia-born sample with all years of birth and birthplaces combined, Table 6.17 indicates a possible timing-of-birth 'pattern', with lowest prevalence of MS births in May and highest in December. Whether such a pattern is real and statistically significant will be considered in the next chapter, where the now longitudinal data will be analysed in detail taking year of birth into account. As a secondary inquiry, whether this timing-of-birth pattern changes significantly with region of birth, or by sex, will also be investigated (see research question 2, Chapter 7).

[16] Mean latitude based on state population distribution in 1981, as given by Hammond et al. (2000a) [304].

Table 6.16: MS cases, final reference-population denominators and crude prevalence estimates for males, females and total persons by birthplace (1920 to 1950 Australia born)

Birthplace (state/area)[#]	Males			Females			Total		
	MS cases	Reference population	Prevalence /100,000	MS cases	Reference population	Prevalence /100,000	MS cases	Reference population	Prevalence /100,000
QLD	76	266,364.0	28.5	140	273,396.8	51.2	216	539,760.8	40.0
WA	26	117,804.8	22.1	97	121,480.0	79.8	123	239,284.8	51.4
NSW/ACT	279	676,636.4	41.2	649	697,006.1	93.1	928	1,373,642.5	67.5
SA	54	139,817.3	38.6	143	143,437.3	99.7	197	283,254.6	69.5
TAS[+]/Hobart[++]	21[+]	16,177.3[++]	129.8	39[+]	16,659.0[++]	234.1	60[+]	32,836.3[++]	182.7
Total	456	1,216,799.8	37.5	1,068	1,251,979.2	85.3	1,524	2,468,779.0	61.7
TAS to QLD*	Ratio	4.6			4.6			4.6	
	(%)	(3.6)			(3.6)			(3.6)	

[#]Birthplace states listed in order of increasing mean (based on state population distribution in 1981) south latitude.
[+]TAS-born MS cases (assumed to be Hobart SD born) ascertained in Hobart SD or other surveyed areas in 1981.
[++]Hobart SD-birthplace denominators.
*Prevalence ratio of TAS compared with QLD shown also as % increase QLD to TAS (in parentheses).

Table 6.17: MS cases, final reference-population denominators and crude prevalence estimates for males, females and total persons by month of birth (year and region of birth not distinguished; 1920 to 1950 Australia born)

Month of birth	Males			Females			Total		
	MS cases	Reference population	Prevalence /100,000	MS cases	Reference population	Prevalence /100,000	MS cases	Reference population	Prevalence /100,000
January	40	102,025.3	39.2	91	105,105.7	86.6	131	207,131.0	63.2
February	33	95,049.1	34.7	87	97,825.0	88.9	120	192,874.1	62.2
March	41	102,902.5	39.8	86	105,828.1	81.3	127	208,730.6	60.8
April	45	97,470.2	46.2	83	100,338.4	82.7	128	197,808.6	64.7
May	32	103,397.2	30.9	74	106,748.8	69.3	106	210,146.0	50.4
June	35	99,455.8	35.2	71	102,582.5	69.2	106	202,038.3	52.5
July	30	105,415.7	28.5	101	108,406.3	93.2	131	213,822.0	61.3
August	43	104,962.1	41.0	95	107,929.2	88.0	138	212,891.3	64.8
September	38	102,475.7	37.1	95	105,330.3	90.2	133	207,806.0	64.0
October	34	106,142.4	32.0	98	108,933.4	90.0	132	215,075.8	61.4
November	37	99,401.5	37.2	95	102,229.4	92.9	132	201,630.9	65.5
December	48	98,102.4	48.9	92	100,722.0	91.3	140	198,824.4	70.4
Total	456	1,216,799.9	37.5	1,068	1,251,979.1	85.3	1,524	2,468,779.0	61.7

6.3.3 Longitudinal dataset by year of birth, sex, birthplace and month of birth for timing-of-birth analyses

The complete merged MS-rates dataset included the final case dataset of 1,524 MS cases expressed in frequency form by month (and year) of birth, together with the corresponding migration-adjusted reference-population denominators also by month (and year) of birth, as shown in Table 6.18.

Table 6.18 also displays an additional data column compared with Table 6.3, this column (column 8) comprising values for 'MS rate'; these values express the number of MS births by month (column 6) *relative to* the corresponding reference-population denominator for that month (column 7), as defined in Section 6.2.1. However, this variable was not used directly for the regression analyses in Chapter 7, and is simply indicative of the relatively low MS disease rates overall; that is, the column values are often zero, or otherwise very low values, for example, ~6 x 10-4, or ~6/10,000 (shown bolded in Table 6.18). Such a 'rare disease' may be suitable for regression analysis based on, for example, a 'Poisson' distribution [654]. Regression analysis for timing of birth using Poisson and other related techniques will be detailed in the following chapter (see Chapter 7).

To sum up, Figure 6.4, which extends the previous Figure 6.2, illustrates the complete process of deriving the reference-population denominators required for construction of the longitudinal MS-rates dataset, as has been detailed in this chapter.

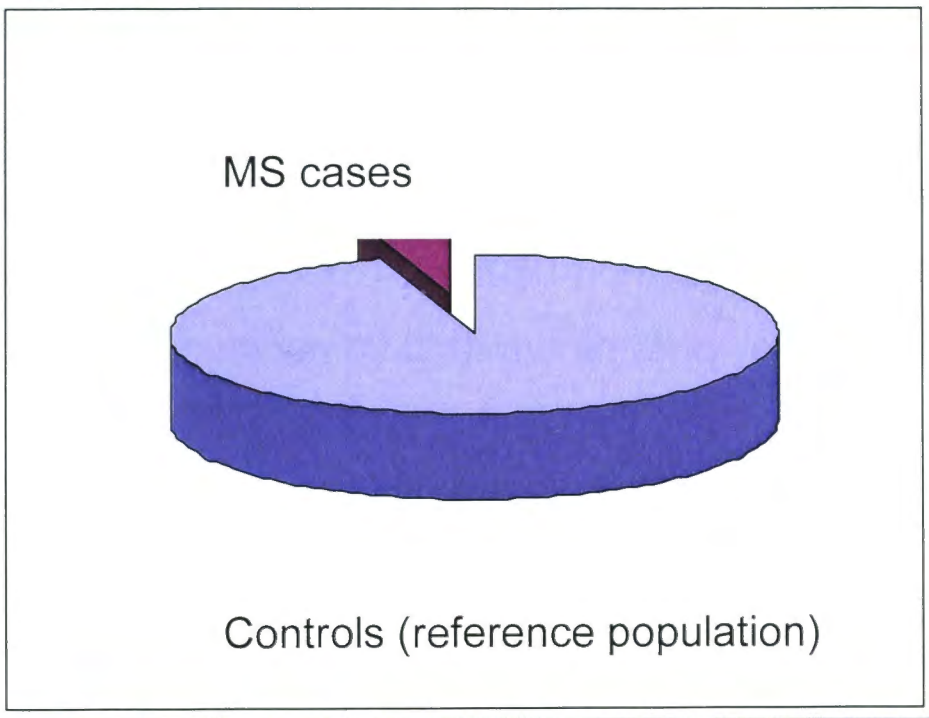
Table 6.18: Final MS-rates longitudinal dataset (portion only shown). MS-case and reference-population births, adjusted for sampling losses to unsurveyed areas, VIC and NT, by month for the first three calendar years, January 1920 to December 1922, for males (sex=1) born in NSW/ACT (place=2). (Reference-population births adjusted using population migration data from 1976; July 1920/June 1921 census-year example given in text indicated on right-hand side; MS rates other than zero are shown in bold)

	1	2	3	4	5	6	7	8	
	CensYear*	Year#	Month	Sex	Place	MS births	Adjusted pop. births	(MS rate) (E format)	
1.	20	20	1	1	2	0	1372.325	0.000e+00	
2.	20	20	2	1	2	2	1266.959	**1.579e -03**	
3.	20	20	3	1	2	1	1354.978	**7.380e -04**	
4.	20	20	4	1	2	0	1319.642	0.000e+00	
5.	20	20	5	1	2	1	1401.236	**7.137e -04**	
6.	20	20	6	1	2	0	1608.113	0.000e+00	
7.	21	20	7	1	2	2	1757.199	**1.138e -03**	**1920/21 census yr**
8.	21	20	8	1	2	0	1693.000	0.000e+00	July to June ▼
9.	21	20	9	1	2	1	1604.974	**6.231e -04**	
10.	21	20	10	1	2	0	1576.515	0.000e+00	
11.	21	20	11	1	2	0	1539.452	0.000e+00	
12.	21	20	12	1	2	0	1587.104	0.000e+00	
13.	21	21	1	1	2	2	1540.113	**1.299e -03**	
14.	21	21	2	1	2	0	1418.996	0.000e+00	
15.	21	21	3	1	2	0	1469.296	0.000e+00	
16.	21	21	4	1	2	1	1511.654	**6.615e -04**	
17.	21	21	5	1	2	0	1516.287	0.000e+00	
18.	21	21	6	1	2	0	1520.920	0.000e+00	
19.	22	21	7	1	2	0	1557.307	0.000e+00	
20.	22	21	8	1	2	2	1626.932	**1.229e -03**	
21.	22	21	9	1	2	0	1524.729	0.000e+00	
22.	22	21	10	1	2	0	1505.567	0.000e+00	
23.	22	21	11	1	2	0	1508.760	0.000e+00	
24.	22	21	12	1	2	1	1510.677	**6.620e -04**	
25.	22	22	1	1	2	1	1552.835	**6.440e -04**	
26.	22	22	2	1	2	1	1310.744	**7.629e -04**	
27.	22	22	3	1	2	1	1580.941	**6.325e -04**	
28.	22	22	4	1	2	0	1396.977	0.000e+00	
29.	22	22	5	1	2	0	1526.007	0.000e+00	
30.	22	22	6	1	2	2	1448.078	**1.381e -03**	
31.	23	22	7	1	2	0	1623.763	0.000e+00	
32.	23	22	8	1	2	1	1666.785	**6.000e -04**	
33.	23	22	9	1	2	0	1530.548	0.000e+00	
34.	23	22	10	1	2	0	1587.259	0.000e+00	
35.	23	22	11	1	2	1	1570.963	**6.366e -04**	
36.	23	22	12	1	2	0	1429.511	0.000e+00	

*CensYear '21' denotes 1920/1921 census year, i.e. July 1920 to June 1921.
#Year '20' denotes 1920 calendar year, i.e. January to December 1920.

MS CASE DATA

MS cases—1981 MS Survey (cross-sectional survey 30 June 1981 by state; unit-record data: year, month of birth, birthplace (state) in Australia, sex



REFERENCE POPULATION DATA

Reference (census) population—census date 30 June 1981 (cross-sectional: census age (years), country of birth) → totals by census year, 1920 to 1950 Australia born

+

Australian births registrations (by state, sex, year and month) → supplementary data for 1920 to 1950 born

⇩ Apportioning census totals

Denominator estimates by month, year, birthplace (state), sex for 1981 census population born 1920 to 1950

+

Migration data for 1976 census population from birth time and place

⇩ Adjustment of denominators

Denominator estimates by month, year, birthplace (state) and sex, adjusted for case ascertainment losses to unsurveyed states, VIC and NT

Numerators + denominators ⇩ Merging two datasets

MS RATES

Reconstructed longitudinal MS-rates dataset for analysis of timing of birth in Australia (by year, month, birthplace [state] and sex)

Figure 6.4: Sources and treatment of data for analysis of timing of MS births in Australia.

6.4 Discussion

This chapter has detailed the construction of a longitudinal dataset from the cross-sectional 1981 MS-case survey, in order to analyse timing of birth of the surveyed MS cases relative to the 1981-census reference population. Individual-level numerator and birthplace-denominator data have been incorporated into the one dataset, month-by-month and year-by-year for the 1920- to 1950-born age group, to give MS-births rates relative to the reference population by birth month (and year) and by birthplace.

Compared with the original cases-only survey dataset, there is now an emphasis on Australia-born MS cases, and only the Australia-born reference population has now been included. Thus, data can be expressed as 'MS rates' for these surviving Australia born in 1981, using estimates of reference-population births by month and state (or SD for Hobart) as denominators for these rates. The resulting dataset contains information on birthplace (state) within Australia by sex, for both the reference population and MS cases. Birthplace as the latitude-related exposure factor at the time of birth, instead of survey area at the later time of census, is now emphasised and this has not been investigated previously in relation to timing of birth in Australia, nor by notable northern hemisphere studies [324, 644]. Therefore, this thesis study will enable analysis at the individual level of MS risk in Australia by month of birth, taking into account year of birth, birthplace and sex (see Chapter 7).

For analysis, the year-of-birth range from the 1981 survey has been restricted to the central three decades, 1920 to 1950, where MS prevalence was highest and relatively uniform. This restriction allowed exclusion of the more obvious sampling biases pertaining to the numerators, including 'survival' bias among the older cases due to likely differential survival between cases and the general population, and the very obvious 'diagnosis' bias evident among the younger cases, where cases at this end of the year-of-birth distribution were under-represented (see Chapter 5, Section 5.3.3).

Assumptions necessary for estimation of reference-population denominators

As the original 1981 MS Survey was carried out on an Australian national census day, this allowed convenient use of the 1981 census data for derivation of the required reference-population denominators. While the 1981 Census was the appropriate starting point, representing survivors to 1981 like the surveyed MS cases, these census

data did not have all of the initial information required for the subsequent analyses, particularly month- and region-of-birth information for Australia-born persons. However, the availability of other Australian (ABS) births-registration data covering the same period enabled these shortfalls in the denominator census data to be overcome, but various assumptions were required in order to do this (Table 6.19). First, the proportion of Australia born was provided only for total Australia (and for five- to 10-year age groups), requiring an assumption of no differences in this proportion, within each age group, between regions. This assumption for region is difficult to check, but the possible effects should be considered: if northern hemisphere migrants settled in Australia in greater proportions in any region, *and* survived (to 1981) similarly to (or longer than) Australia born, *and* if they showed an opposite 'northern hemisphere timing-of-birth' pattern, this would have the effect of diluting any timing-of-birth effect for that region, even though only the denominators, and not the numerators, would be affected.

Second, the census totals, now adjusted to represent only Australia born, were apportioned into subtotals by birthplace and by month of birth, using supplementary ABS births-registration data available by state/territory and by month and year. To apportion these census totals in the same proportions as existed in the original by-year birth cohorts (see Figure 6.1), it was necessary to assume that survival from birth to 1981 (including continued residence in Australia until 1981) of the Australia-born reference population was not dependent on region of birth within Australia, nor on month of birth. In addition, the assumption that the ABS births registrations adequately reflected actual month of birth was basic here (Table 6.19). The next three subsections consider these three assumptions and possible sources of error separately.

Table 6.19: Summary of assumptions necessary for estimation and adjustment of reference-population denominators for MS-rates dataset

Required estimate	Assumptions
Proportion Australia born (from 1981 census totals)	Same proportion within five- to 10-year age groups, and among regions (states)
Month-of-birth and place-of-birth proportions (from ABS 1920 to 1950 births-registration tables and 1981 census totals)	Registrations reflect actual month (and place) of birth Survival from birth to 1981 (and lack of emigration from Australia) not dependent on place of birth or month of birth
Net interstate migration to unsurveyed states/areas (VIC, NT and non-Hobart TAS) between birth and 1981	Net interstate migration similar among MS cases and reference population, and similar for both in 1976 and 1981 Net interstate migration not dependent on month of birth within each birth year, or sex 'TAS-born' cases are all Hobart SD born (and net migration loss for Hobart SD born similar to TAS born; net migration loss from Hobart SD to non-Hobart TAS negligible)

Interval between birth and registration of birth

Measurement bias would be an important issue if, for example, month of birth was being measured differently for MS cases and the reference population. For the reference-population denominators, this thesis study was limited to using summary monthly births-*registration* data (ABS), rather than records of actual dates of birth as

were available for the MS cases. The delays before registration of births could be checked for some years—for example, 1920 and 1921—which were the earliest birth years represented in the reference population and which should indicate the maximum possible error in measurement of month of birth for the reference population (noting that time intervals between birth and registration generally decreased in the later years). For both these years 1920 and 1921, a median registration interval of 10 days was observed, with 90% of all births being registered within 31 days (one month); that is, the possible error in month of birth for most of the oldest members of the reference population was less than or equal to one month. Although only the denominators would be affected thus, and final MS rates should therefore be affected only minimally, this possible error could have a dilution effect on the *month-of-birth* variable. This provides further reason to restrict the dataset to 1920 to 1950 born to exclude earlier years when the interval between birth and registration would likely have been longer.

Region of birth within Australia and survival to 1981

Given Australia's nationwide health system, major differences in survival by region of birth within Australia would not be expected, particularly for the restricted dataset of just 30 to 61 year olds (in 1981). Nor would different rates of emigration overseas from the birth regions be expected. Nevertheless, the ASSDA residence-by-birthplace data in Table 6.12 for the oldest members of the reference population (1920/1921 census year of birth) can be used to indicate the likely maximum differences in regional birthplace proportions since the time of birth, at least for the 1976 census population for which these data were available. That is, the proportion from each birthplace state, of the total 1920/1921-born persons with known birthplace states who were present in Australia in 1976 (see Table 6.12), can be compared with the original proportion known to have been *born* in each state that year (from the ABS births-registration data). This indicates any major differences in survival (and emigration from Australia) between the states. Table 6.20 shows this comparison as regional percentages of total number of persons, for both the oldest reference-population members born in the 1920/1921 census year (left-hand side of table) and the overall 1920- to 1950-born persons in the final dataset (right-hand side).

Table 6.20: Distribution of total number of persons by region of birth, comparing proportions (as percentages) at time of birth with those in the 1976 Census, for persons born in the 1920/1921 census year and for all persons born between 1920 and 1950

Birthplace (state/territory)	Born in 1920/1921 census year		Born between 1920 and 1950	
	At birth (%)*	In 1976 (%)**	At birth (%)*	In 1976 (%)**
NSW/ACT	39.8	39.1	39.5	39.0
VIC	26.3	26.4	25.8	25.7
QLD	15.0	14.8	15.4	15.5
SA	8.8	9.5	8.4	8.7
WA	5.8	5.7	6.9	6.9
TAS	4.2	4.3	3.9	4.0
NT	0.1	0.2	0.1	0.2
Total	100.0	100.0	100.0	100.0

* Regional proportions from ABS births registrations.
** Regional proportions from 1976 Census, representing the surviving and non-emigrating persons from the initial regional birth cohorts.

Minimal differences between the regional percentages at birth and in 1976 (and presumably still in 1981) are indicated in both halves of this table, particularly for the 1920 to 1950 born (right-hand half of table). That is, only minimal regional differences in survival (and emigration) for the whole 1920- to 1950-born reference population over the period from birth to 1976/1981 are apparent, giving confidence in the at-birth proportions used to allocate the 1981 census-total denominators. To confirm this, a sensitivity analysis using the 1976 birthplace proportions, rather than the 'at-birth' calculated proportions, to derive the denominators for 1920 to 1950 born was also conducted (data not shown). This analysis showed that timing of birth (as MS risk by period of birth for the overall 1920 to 1950 born, see Chapter 7, Section 7.3) was unaffected (the MS risk IRRs for each two-month period, shown to two decimal places in Table 7.8, Chapter 7, were unchanged), confirming the validity of this assumption.

Month of birth and survival to 1981

The initial restriction of the dataset to just the 30 to 61 year-old age group (in 1981) will have limited any differences in survival with month of birth to a certain extent. However, any such differences could potentially affect a timing-of-birth pattern for MS, because

the surviving (and non-emigrating) persons to 1981 are being used as proxy for the original birth cohorts of the reference population, in order to apportion the denominator subtotals by month (and region) of birth. There is some evidence for season- or month-of-birth effects on adult life expectancy past the age of 50 years for Australians born between the 1890s and 1947, Doblhammer and Vaupel (2001) observing a decreased post-50 lifespan in spring and early summer born that they attributed to early infant health [655]. However, these authors acknowledge that this seasonal variation is relatively small compared with, for example, social factors and gender effects on longevity; they further show (using Danish data) that later-born cohorts within this time-span exhibit many fewer longevity differences because of significantly improved maternal and infant health in more recent years [655]. The 1920- to 1950-born group used in this thesis study falls into the later-born category within this period, and the possible variation in survival between different birth months should be limited; in addition, any differences would again affect only the denominators and thereby limit changes in MS rates. Indeed, the preliminary prevalence estimates given in Table 6.17 show the opposite pattern to that which would be expected if seasonal survival differences as seen by Doblhammer and Vaupel (2001) were contributing here (over-estimated spring-summer denominators would reduce spring-summer MS rates), giving further confidence in the denominators as estimated assuming no survival differences by month of birth.

Adjustment for migration of cases to unsurveyed states or areas

Because of the lack of 1981 MS-case survey data for some regions, notably VIC and NT, MS cases *born* in either of these two areas and ascertained in other, surveyed areas were necessarily excluded from the final dataset (see Section 6.2.3); denominators were then not required for these two regions. Final adjustment of the remaining denominator estimates was required to compensate for the lack of survey data for the regions VIC and NT, because cases born in surveyed areas would have been 'lost to survey' if they had migrated interstate to either VIC or NT by 1981. Available residence-by-birthplace data from the 1976 census was used to gain these adjustment factors for each surveyed region and each year of birth, necessarily assuming that net migration between regions did not change significantly in the final five unrecorded years from 1976 to 1981 (see Table 6.19). This assumption appears reasonable given that the 1920- to 1950-born study cohort were 25 to 56 years old in 1976 and only 30 to 61 years old five years later in 1981. As these interstate migration data were available only by (census) year of birth and for all persons born in Australia,

the additional assumption required was that net interstate migration did not depend on month of birth or sex; this additional assumption is not testable with available data but again appears reasonable.

While lack of case survey data for VIC and NT and migration to these areas from surveyed regions could be satisfactorily adjusted for, the 1981 survey of Hobart SD rather than the whole of TAS was more problematic, because the 'TAS-born' MS cases could not be differentiated into those born in Hobart SD and those born in non-Hobart TAS. Therefore, 'TAS-born' cases ascertained in any of the surveyed areas, including Hobart SD, were considered to have been all born in Hobart SD, in order to be able to include the 'TAS-born' cases in the analysis. Fortunately, the areas 'Hobart SD' and 'TAS' as birthplace exposure factors are relatively similar in mean latitude and UVR received, and it was considered preferable to include, rather than exclude, the TAS cases given that high-latitude (and lower UVR) TAS was potentially able to have important effects on MS risk. (The alternative to this was to consider all 'TAS-born' cases as 'non-Hobart-born' and then exclude all from analysis, as was done for VIC and NT.) The potential effect of the assumption that all 'TAS-born' cases were 'Hobart SD-born' is discussed further in the next subsection, where preliminary prevalence estimates are used to justify this strategy.

The TAS-birthplace denominators, conversely, were able to be differentiated into 'Hobart-born' and 'non-Hobart-born' (in the ABS register), this allowing use of the more relevant 'Hobart SD' birthplace denominator, rather than one simply for 'TAS' (which would have underestimated the MS rate in the 'TAS-birthplace' area). Then, for migration adjustments, the Hobart SD-born denominator could be adjusted to account for potential (case) losses to VIC and NT survey areas, assuming the same *relative* proportion lost for 'Hobart SD-born' cases as for 'TAS-born' cases (see Table 6.19) (TAS-born case losses calculated from 1976 census data).

However, 'losses to survey' by migration of cases from the other (surveyed) birthplace states to the unsurveyed portion of TAS—that is, non-Hobart TAS—could not be easily estimated, but these were shown (using 1976 census proportions) to be very small in comparison with the now accounted-for case-migration losses to VIC and NT (see Table 6.14). As well, potential losses of cases from the surveyed Hobart SD birthplace to unsurveyed non-Hobart TAS could not be estimated at all but were assumed to be negligible given the relatively small area and population size of TAS.

Summing up this subsection, all of the case ascertainment losses due to migration are, in reality, quite small and could be expected to have minimal effect on MS rates when only the denominators are involved. However, the majority of these potential errors have now been taken into account in estimating the final reference-population denominators to be used for detailed analysis of the MS-rates dataset.

Prevalence by birthplace and month of birth—preliminary estimates

Preliminary to Chapter 7, this chapter has included some simple exploratory analysis of crude prevalence rates to indicate some of the possible expected effects on MS risk, but also to allow a check of assumptions made for TAS-born cases. The increase in prevalence indicated in *birthplace* regions of higher latitude (see Table 6.16) recalls that seen with survey area (and its latitude) in Chapter 5. Perhaps this is not surprising given that both the MS cases and the Australia-born reference population appear to have not migrated from their birthplace states/territories to any great extent by 1976/1981 (see Tables 6.10 and 6.12). Most importantly, the increase in prevalence for both sexes from QLD to TAS birthplaces of 4.6-fold seen here is only slightly greater than that seen for survey area in Chapter 5 (four-fold in TAS compared with QLD). It appears that making the assumption for TAS-born cases (that these were all 'Hobart SD-born' and then using the Hobart SD-born denominator) is not greatly overestimating the number of cases allocated to the Hobart SD birthplace, nor underestimating the true denominator (i.e. by using 'Hobart SD-born' rather than the 'TAS-born' denominator), either of which would have increased the TAS-(/Hobart SD-)born prevalence more. Moreover, if latitude/UVR effects on the prevalence of MS are initiated early in life as postulated, then a latitude-prevalence association might be expected to be somewhat stronger at birth, compared with later in life when migration to other areas would dilute this association. Therefore, these preliminary results in Table 6.16 give additional confidence not only in the estimation techniques for the place-of-birth and month-of-birth denominators but also for the necessary assumption made for 'TAS born' surveyed in 1981—that is, that all 'TAS-born' cases were effectively Hobart SD born, justifying use of the Hobart SD denominator.

With regard to the main research question for this thesis concerning timing of birth in Chapter 7, Table 6.17 gives preliminary estimates for prevalence rates by month of birth, but for all years of birth from 1920 to 1950 and all regions of birth combined. For total MS cases (males and females considered together), these indicate possibly higher MS rates in December (early Australian summer) and lower rates in May and

possibly June (early winter). Regression analysis techniques will be applied in Chapter 7 to further define this suggested timing-of-birth pattern using the now longitudinal dataset. Thus, Chapter 7 will examine whether a timing-of-birth pattern exists in MS cases in Australia, using longitudinal case and reference-population data for the (birth-year) cohorts born between 1920 and 1950. Effects of other factors, including birthplace and latitude of birthplace, on any timing-of-birth pattern, and on MS risk per se, will also be considered.

In conclusion, the construction of a longitudinal dataset of MS frequency from the original cross-sectional data means that every month of every year over the study period can now be analysed for MS risk by regression modelling, rather than simply collapsing all the year-of-birth data and considering only a month-of-birth variable, as in the preliminary prevalence results indicated in this chapter (see Table 6.17). A complete *timing-of-birth*, *gender* and *region-of-birth* comparison between MS cases and the reference population in the 1920 to 1950 Australia-born cohort is now achievable from the original cross-sectional survey data, rather than simply a disease prevalence study.

CHAPTER 7

TIMING OF BIRTH AND MS RISK IN AUSTRALIA: ANALYSIS OF TIMING OF BIRTH, BIRTHPLACE AND PRENATAL UVR PATTERNS IN MS CASES

'Whoever wishes to pursue the science of medicine in a direct manner must first investigate the seasons of the year and what occurs in them'

Hippocrates, 460 BC, in *On Endemic Diseases: Airs, Waters and Places* [656].

7.1 Introduction

The investigation of seasonal birth patterns for specific diseases has been a classical approach in the epidemiological study of disease since the time of Hippocrates in 460BC. In the 1900s, such studies began to become accepted as having serious importance, especially in the field of psychiatric disorders such as schizophrenia and bipolar disorder. In the late 1960s, schizophrenia was one of the first disorders, apart from infectious diseases, to be shown to have a distinct seasonal timing-of-birth cycle; E. Hare in England and P. Dalen in Sweden were then regarded as the founders of modern psychiatric studies in this field, because they were able to use large sample numbers, adequate controls and more sophisticated statistical analyses to add credence to their findings [642]. Significantly, the presence of a seasonal birth pattern gives evidence of early life origins of disease that can provide important clues to disease aetiology [7].

As discussed in Chapters 1 to 3, the precise aetiology of MS is unknown. However, evidence has been accumulating for a number of years that causation of MS is multifactorial and most certainly involves a combination of genetic and environmental exposure factors.

Rothman and Greenland (1998) use a 'component cause model' to describe disease causation, wherein 'sufficient cause' for a disease to occur is defined as a 'set of minimal conditions and events that finally produce disease' (see Chapter 2). Each 'sufficient cause' comprises a set, or series, of 'component causes', the latter being

defined as 'an antecedent event, condition or characteristic that is *necessary* for the occurrence of disease, but which may *not be sufficient by itself* to produce disease' [1].

As discussed in Chapters 2 and 3 also, component causes acting in the same sufficient cause may be thought of as interacting biologically to produce disease. Further, this interaction or joint action need not be simultaneous: one component cause could act *many years before the other*, the condition being that the first component cause leaves some effect that interacts with the later component [1]. A similar concept of interaction of factors in a 'two-hit hypothesis' has been proposed for schizophrenia, in which this disorder 'might be predisposed by a seasonal factor occurring during the prenatal period and then precipitated by another factor, not necessarily seasonal, many years later' [642]. Such a causal scenario comprising a series of interacting components also appears likely for MS; this thesis study aims to investigate some of the possible environmental causal components of MS through their apparent timing (see Chapter 3, Section 3.2).

The present chapter will focus on timing of birth in MS cases, in order to examine the perinatal period for a possible component disease cause acting around the time of birth. That is, a seasonal or other temporal birth pattern for a specific disorder that differs significantly from the general birth pattern of the population would suggest a component causal factor operating very early in life; such a pattern may then provide information about the disorder's aetiology [7]. Particularly for a late-onset disorder such as MS, a difference in the *timing* of birth of cases compared with the general population could be a useful risk indicator, or 'proxy', for determining the *nature* of a causal component factor acting around the time of birth. Such specific early life environmental factors would otherwise be too difficult to recall when onset of a disorder such as MS typically occurs decades after birth.

Only a few published studies have investigated timing-of-birth patterns in MS patients, and all of these have been located in the northern hemisphere. In Denmark, excess MS births (that is, births of people who later developed MS) were seen in the March to June period, particularly May and June, and deficits were evident in the other eight months [321]. Swedish data showed similar excesses in the same four months, particularly in May, and deficits in the remaining eight months [322]. Data from British Columbia, Canada, revealed a similar peak in MS births in May, with excesses from April to July, and in January (Sadovnick and Yee, 1994 [657], re-interpreted by James, 1995 [320]). However, Salemi and co-workers (2000) showed a pattern of excess MS

births in the June to November period in Sicily, perhaps indicating some seasonal differences at lower latitudes [323].

More recently, northern hemisphere work on larger MS datasets from Canada and the UK by the Canadian Collaborative Study Group (CCSG) has shown a significant deficit of births in the month of November and an excess in May. These data pooled with Danish and Swedish data by CCSG showed an overall 13% increase (95% CI 5% to 22%) in MS risk in May-born compared with November-born persons [324]. Tremlett and Devonshire (2006) further confirmed this May/November peak-and-trough pattern in MS-case births in British Columbia, Canada [644]. Consistent with these studies, a November deficit in MS births was also seen in France [658], while a spring (March to May) excess and autumn (September to November) deficit occurred in Scotland [659].

Southern hemisphere analyses on month of birth of MS cases are now required. Of particular interest would be similar excesses and/or deficits in MS births, relative to that of the reference Australian population, in the corresponding seasons in the southern hemisphere. For example, excess MS births may be found to occur in or close to November in the southern hemisphere (or December, as suggested by the preliminary prevalence data in Chapter 6, Section 6.3.2, Table 6.17), in comparison with minimum MS births in or around May. Such a temporal MS-case birth pattern acting similarly in both parts of the world would suggest causal environmental factors that are 'time-locked' to the time of birth [229]. Seasonally fluctuating UVR levels, possibly mediated through vitamin D (see Chapter 2), may be contributing to such a birth pattern in MS cases. Alternatively, or additionally, seasonal infection cycles also may contribute.

As noted in Chapters 4 and 5, the continent of Australia covers a broad latitude range between approximately 10 and 44 degrees south (see Figures 4.3 and 5.2). The seasons in Australia vary from two main 'dry' and 'wet' seasons in the tropical far north (the NT and northern QLD) to a pattern of four seasons (based on temperature: summer in December to February, autumn in March to May, winter in June to August, spring in September to November) in the temperate mid- and southern regions (NSW, SA, VIC and TAS). (The State of WA covers the broadest latitude range (see Figure 5.2) but most of its population is centred in the southern, temperate, half.) As noted in Chapter 2, MS exhibits a prevalence gradient of more than six-fold between low-latitude northern QLD and high-latitude TAS, and an approximately four-fold gradient between subtropical southern QLD (Brisbane) and TAS [301].

It is possible that the effect size of any timing-of-birth pattern found in Australia may also vary regionally with latitude, as suggested by recent northern hemisphere work. In particular, the CCSG MS study suggested a trend in effect size with prevalence, the magnitude of the May-born to November-born ORs for incident cases increasing from lower-prevalence, lower-latitude Canada to higher-latitude Scotland where prevalence was highest. This work suggested that the seasonal birth effect may be linked with environmental factors determining prevalence rates [324]. Such a finding in the southern hemisphere may further implicate fluctuating UVR and/or seasonal infections as component causal factors, given that both of these potential factors may cycle more strongly at higher latitudes. For example, in Australia, ambient winter effective UVR (UVR_{eff}) falls relatively lower, compared with summer UVR, in southernmost TAS than in the northern state of QLD. That is, the seasonal variation in UVR in Australia as given by the *ratio* of maximum (summer) to minimum (winter) UVR_{eff} is 3.5-fold higher in higher-latitude (and higher-prevalence) TAS than in lower-latitude Brisbane (QLD) (Table 7.1).

Table 7.1: Monthly averages of daily total UVR_{eff} (MED)[+], 1996 to 2000, for Australian capital cities of states included in 1981 MS Survey. (Unpublished data for Brisbane, Perth, Sydney and Adelaide provided by P. Gies, Australian Radiation Protection and Nuclear Safety Agency [ARPANSA], pers. commn; latitude of capitals[17] in decimal degrees south shown in parentheses)

	Brisbane (QLD) (27.5 °S)	Perth (WA) (31.9 °S)	Sydney (NSW) (33.9 °S)	Adelaide (SA) (34.9 °S)	Hobart[#] (TAS) (42.9 °S)
January	24.7	30.4	22.2	28.1	20.4
February	22.3	27.9	21.7	24.5	18.6
March	19.0	18.9	16.2	13.7	12.4
April	12.5	11.8	9.8	9.4	6.5
May	8.6	7.1	5.6	5.2	3.7
June	6.6	4.9	3.9	3.6	1.7
July	7.5	5.4	4.1	3.7	1.6
August	10.4	7.9	6.9	6.3	3.9
September	14.6	12.1	10.4	10.0	7.0
October	18.9	18.8	15.3	15.5	12.0
November	22.7	23.8	19.1	21.5	15.7
December	24.4	29.2	21.9	26.1	18.1
Max : Min*	3.7	6.2	5.7	7.8	12.8

[+] UVR_{eff}[18] in units of minimum erythemal dose (MED), defined as the cumulative UVR exposure required to induce erythema or sunburn in humans [660].
[#] Data for Hobart for 1991 from Gies et al. (1994) [660].
* Seasonal variation as given by a ratio of summer maximum to winter minimum UVR_{eff}.

As discussed in Chapter 4, Australia also has the advantage of a relatively genetically homogeneous population distributed over the latitude range. Fortunately, most of this broad area has been included in the 1981 MS Survey dataset, giving an opportunity for this additional inquiry (see research question 2, this chapter).

[17] Latitude values for the state capital cities used in this chapter were shown in Chapter 5 to be little different from mean state latitude values based on 1981 population distribution (see Section 5.1, Figure 5.2).
[18] UVR measurements are validated ground measurements supplied by the ARPANSA, rather than UV Index forecasts from satellite monitoring as used in Chapter 4.

The main research questions for this chapter are:

1. **Is there a timing-of-birth pattern for Australia-born MS cases relative to the general Australian population?**

   For example, is there an excess (or deficit) of MS cases born in any particular month, or months, of the year in Australia, relative to the reference Australia-born population? That is, is there an association between timing of birth and MS risk? Specifically, an excess of (i.e. relatively more) MS cases born in November compared with May is posited.

2. **Within Australia, does region of birth modify an association between timing of birth and MS risk?**

   If a timing-of-birth pattern for MS cases relative to the reference population exists, does this pattern differ according to the birthplace (state) of MS cases in Australia? Specifically, MS births in November, for example, relative to those in May are posited to be greater at higher latitudes; that is, there would be a greater *relative* difference between early summer and early winter MS births (i.e. a higher early summer to early winter ratio) in TAS compared with QLD.

Additionally, as shown in Table 7.1, there are monthly ambient UVR data available that have been reliably measured over a number of years for the regional state capitals of Australia by the ARPANSA (P. Gies, pers. commn) [660]. Such quantitative UVR data can be used directly as an additional exposure factor, and independently of timing of birth, the latter being proxy for a number of possible exposure factors that include UVR and/or vitamin D, but may also include such factors as temperature, diet/nutrition and infections. Further, the UVR data conveniently encompass both the regional ambient UVR variation and the monthly (seasonal) variation in a single variable. This wide variation in ambient UVR with latitude and between summer and winter (Table 7.1) may be useful to explore the effects of (maternal) ambient UVR levels during the perinatal period, including the prenatal period in particular. For example, Sayers and co-workers have shown that 'erythemal UV' during pregnancy can be used as an instrumental variable that is strongly linked to maternal serum vitamin D during pregnancy, and that such UV data can be used to investigate subsequent vitamin D-dependent health outcomes in offspring [661]. Thus, research question 3 is:

3. **Is prenatal UVR associated with MS risk? If so, can this account for a timing of birth risk pattern?**

The aims of this chapter are to investigate, using individual-level data, the timing of birth of MS cases relative to the reference 1981 Australian population (research question 1). The MS-rates dataset constructed for this purpose in Chapter 6 will be used. Other factors potentially affecting MS risk directly, such as sex and birthplace, will also be considered. As well, the possible modifying effects of factors such as birthplace on any temporal birth pattern found will be investigated (research question 2). Finally, a possible link between maternal UVR exposure during the perinatal period and MS risk will be independently explored using available regional and seasonal UVR data (research question 3).

7.2 Methods

7.2.1 MS-rates dataset by month and year of birth

As detailed in Chapters 5 and 6, a longitudinal dataset in frequency (or count) format spanning every (birth) month of every year over the chosen study period has been constructed from the original unit-record cross-sectional 1981 MS Survey data. This MS-rates dataset comprises numerator data taken directly from the 1981-surveyed MS cases (see Chapter 5), together with reference-population denominators derived from both the 1981 Australian Census and supplementary Australian births-registration data covering the same birth years (see Chapter 6). Necessary assumptions for derivation of these denominator estimates have been examined, and final adjustment of denominators to account for the main case losses from surveyed to unsurveyed state regions between the time of birth and time of survey (1981) have also been made (see Chapter 6).

For timing-of-birth analysis, a restricted dataset of those MS cases born in Australia between 1920 and 1950 (inclusive) has been chosen to minimise potential biases arising from the use of the whole year-of-birth range in the original cross-sectional dataset (see Chapter 5, Section 5.3.3). The resulting MS-rates dataset comprises numerator and denominator data for a total of 1,524 MS cases born in any of the five surveyed states, QLD, WA, NSW/ACT, SA and TAS, during this 1920 to 1950 period

(see Chapter 6, Section 6.3.1, Tables 6.15 and 6.16). Female cases numbered 1,068 (70%) and the mean age at onset was 32.9 years (s.d. 9.0). A relapsing course from onset (which included relapsing-remitting and those who had entered a secondary progressive phase) comprised 85%, and 15% had a primary progressive course from onset (see Chapter 5).

7.2.2 Statistical analysis: MS incidence rate

The main outcome measure is the number of MS births in each period (e.g. month) relative to births in the same period in the general population, expressed as an MS-births incidence rate. Regression analysis was used to explore the possible association between the period as the independent variable, and the *number*, or frequency, of *MS births* as the dependent variable, taking into account the estimated population births (i.e. estimated denominator) as the *offset,* or statistical *exposure*,[19] variable. The constructed MS-rates dataset was different from the individual-level unit records in the original MS survey in that each record now represented a 'cell' or unique combination of factors comprising year of birth, month of birth, state and sex (see Chapter 6, Table 6.3). The number (or frequency) of MS cases and the corresponding population denominator for each 'cell' record then pertained to that particular combination of factors. The population-denominator (i.e. offset) variable thus expressed the maximum number of times that the event (=birth of a person who would later become an MS case) could have occurred in the relevant population—that is, if there was 100% probability of the disease—by month and year, and by state and sex.

Poisson and negative binomial regression models

To explore MS risk, a Poisson regression model was first investigated, the Poisson distribution often being suitable to model discrete events, such as count data, that occur infrequently in time or space. Although the data arise from a binomial situation in which there are two distinct outcomes for every birth—disease or no disease—in cases such as this, where the population is large and the event is rare, it is appropriate to use a Poisson distribution, which is sometimes called the 'distribution of rare events' [654].

---

[19] The term 'exposure' is used here with its statistical meaning, rather than an epidemiological risk 'exposure'.

The low probability of MS in the Australian population is indicated by the overall prevalence of around 30 per 100,000 (see Chapter 5, Section 5.3.2, Table 5.5) and by the very low and sparse MS frequencies shown in Tables 6.4 and 6.5 (see Chapter 6, Section 6.2.1) and Table 6.18 (see Section 6.3.3). Given the underlying assumptions of the Poisson distribution (that is, if events occur independently and with constant probability), then the counts of events (such as MS outcome) over a given period of time are expressed by the incidence rate, r_j [662]:

$$r_j = \frac{\text{count of events (MS births)}}{\text{no. of times event could have occurred (population births)}}$$

The denominator is the offset variable and is given by the number of reference-population births for each period (i.e. each month of each year, in each state and for each sex, as described above). Under a Poisson regression model, the logarithm of the incidence rate is then modelled as a linear function of one or more predictor (x) variables, which represent the epidemiological exposure (or risk) factors of interest:

$$\ln(r_t) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

Assuming that a Poisson distribution underlies the events of interest (MS births), Poisson regression finds maximum-likelihood estimates of the β parameters. Estimates of relative incidence rates, or IRRs, rather than regression coefficients can also be gained (i.e. estimated coefficients transformed to IRRs defined as exp(β) rather than β) [662] and will be used here.

For example, to ask the question: 'Does MS-births incidence rate differ between the sexes?' (i.e. 'Does the incidence rate of MS births depend on the single predictor, sex?' as in univariate model (2), Section 7.2.3), using the Stata 8.0 statistical analysis program (release 8.0, 2003; Statacorp, College Station, Texas) and the MS-rates dataset constructed in Chapter 6, the following Stata command for the Poisson regression model is used:

```
.poisson MSfreq sex, nolog exposure(Denom) irr
```

where *MSfreq* is defined as the dependent variable (MS births, by month), *sex* is the single independent variable (male or female) and *Denom* is the Poisson exposure (or offset) variable (reference-population births, by month); 'nolog' requests that iterations

for convergence of the model not be shown; 'irr' requests IRRs to be calculated rather than regression coefficients. This gives the following Stata output:

Poisson regression — Number of obs= 3720

LR chi2(1)= 235.70

Prob > chi2= 0.0000

Log likelihood=-2627.3256 — Pseudo R2= 0.0429

MSfreq	IRR	Std. Err.	z	P> | z |	[95% Conf. Interval]	
sex	**2.276294**	.1273365	14.70	0.000	2.039915	2.540065
Denom	(exposure)					

The interpretation of this regression analysis is that the MS-births rate for females (data coded female=2, male=1) is 2.28 times (95% CI 2.04, 2.54) higher than for males (significant p-value <0.001), though the fit of this Poisson model with only a single predictor, as shown by pseudo R^2, is poor. Additional predictor variables can then be added to such a model to examine their effects and improve model fit [662].

A characteristic of the Poisson distribution is that the mean of the distribution is equal to the variance, both being able to be represented by a single parameter [654]. However, as count data are commonly over-dispersed—that is, the variance of the count data is greater than the mean—a negative binomial regression model, which is a more general type of binomial regression but still of the Poisson structure, often fits these types of data better [663]. Summarising the MS-frequency data in Stata 8.0 (i.e. .summ MSfreq, detail) showed that the variance (0.65) was, indeed, some 50% greater than the mean (0.41).

Therefore, a negative binomial regression model was investigated as an alternative to the Poisson model, and in most instances the negative binomial model was more appropriate for analysing these MS data. This was shown by likelihood ratio tests for departure of this 'gamma' type of distribution from a Poisson model, these tests being automatically carried out by the statistical software package (Stata 8.0) used during negative binomial regression analysis. For example, for the same univariate model as previous—examining the effect of sex on MS-births incidence rate—the Stata command:

.nbreg MSfreq sex, nolog exposure(Denom) irr

gives the following output:

Negative binomial regression — Number of obs=3,720

LR chi2(1)= 208.89

Prob > chi2= 0.0000

Log likelihood=-2625.1494 — Pseudo R2= 0.0383

MSfreq	IRR	Std. Err.	z	P> | z |	95% Conf. Interval	
Sex	**2.275949**	.1309311	14.30	0.000	2.033267	2.547596
Denom	(exposure)					
/lnalpha	-2.452673	.5243685			-3.480416	-1.424929
alpha	.0860633	.0451289			.0307946	.2405255

Likelihood ratio test of alpha=0: chibar2(01)=4.35 Prob>=chibar2=**0.018**

The negative binomial regression model adds a parameter, alpha, that reflects unobserved heterogeneity among the observations leading to the over-dispersion [663]. The likelihood ratio test shown in the last line of the output (for whether the over-dispersion parameter alpha=0) is significant here (p=0.018), providing evidence that the MS-frequency distribution was indeed different from the Poisson type. These tests within the negative binomial regression analyses generally confirmed the over-dispersion of the MS frequencies, often at highly significant levels (p<0.0005); the negative binomial model is therefore preferable to the Poisson model for these data.

7.2.3 MS risk by month of birth

Negative binomial regression models were used to provide an estimate of the MS incidence rate for each period of birth, expressed as the IRR for each period relative to a single reference period. The independent month-of-birth variable was eventually chosen as six two-monthly categories; that is, the year was considered as six two-monthly periods: January to February, March to April, May to June, July to August, September to October, and November to December, with the May to June period then chosen as the reference period.

Initial analysis using single months of birth, with May as a reference, indicated that the sample size of MS cases available (n=1,524 cases) was inadequate for such analysis by monthly intervals—there were few IRRs significantly different from the reference level and most months showed marginal effects. Indeed, Hare (1975) emphasises that for a seasonal birth distribution to show a statistically significant 8% deviation (i.e. excess or deficit in births, albeit using different estimation techniques from those used here), 1,500 'subjects' (presumably cases) would be needed for seasonal (four per year) periods and 4,500 for monthly intervals [664]. However, Torrey and colleagues indicate in their methodological review that quarterly periods (i.e. seasons) are unsatisfactory if a birth excess or deficit occurs in an individual month or in two contiguous months in different quarters, when statistical significance may not be reached and the birth excess/deficit missed [642]. Therefore, two-monthly periods (six per year) were chosen for analysis in order to increase the power to detect a significant excess or deficit, but still achieve sufficient discrimination with respect to the periods.

May to June was chosen as the reference period because this two-monthly period showed the lowest UVR level in our region (Table 7.2). In addition, a 'trough' in MS rates in May to June was suggested by preliminary prevalence estimates shown in Table 6.17 (see Chapter 6, Section 6.3.2). MS incidence rate, or MS risk, for each of the other five two-monthly periods of birth was then expressed as the IRR relative to the reference period, May to June, for which IRR=1.0.

Table 7.2: Two-monthly averages of daily total UVR_{eff} (MED)$^{+}$ for Australian state capitals in 1981 MS Survey, 1996 to 2000 (two-monthly values obtained by averaging monthly values in Table 7.1). (Latitude of capitals in decimal degrees south in parentheses)

	Brisbane (QLD) (27.5 °S)	Perth (WA) (31.9 °S)	Sydney (NSW) (33.9 °S)	Adelaide (SA) (34.9 °S)	Hobart# (TAS) (42.9 °S)
January to February	23.50	29.15	21.95	26.30	19.50
March to April	15.75	15.35	13.00	11.55	9.45
May to June	7.60	6.00	4.75	4.40	2.70
July to August	8.95	6.65	5.50	5.00	2.75
September to October	16.75	15.45	12.85	12.75	9.50
November to December	23.55	26.50	20.50	23.80	16.90

$^{+}$ UVR_{eff} in MED units defined in Table 7.1.
Monthly data for Hobart for 1991 from Gies et al. (1994) [660].

Because the two-monthly periods were a set of indicator (or categorical, i.e. dummy) variables—the five test periods each being compared pair-wise with the sixth May to June (winter) reference period—any pattern could be considered. That is, any departure from a uniform MS incidence rate by month of birth compared to the reference period could be investigated. In particular, no pre-conceptions of any temporal pattern shape, such as sinusoidal, were imposed.

Further, because MS is an irreversible disorder once onset occurs, the number of cases reflected the numerator of a cumulative incidence rate. Analysis using negative binomial (or Poisson) models provided an estimate of *lifetime MS risk*, in terms of the cumulative incidence of cases for each two-monthly period of the year relative to the reference period, May to June. Stata 8.0 statistical software (release 8.0, 2003; Statacorp, College Station, Texas) was used for all analyses.

Basic regression model

The basic univariate timing-of-birth 'model' was:

- MS risk depends on month of birth Model (1)

the *month* (of birth) factor being a categorical variable comprising six periods, January to February ... November to December, (May to June being the reference period, IRR=1.0).

Other univariate effects on MS risk—sex and birthplace

Because preliminary prevalence estimates in Chapter 6 have indicated direct effects on MS risk per se by both sex and birthplace state (or survey area, see Chapter 5), negative binomial regression was conducted to estimate MS incidence rates for each of these factors, *sex* and *birthplace,* irrespective of any temporal birth patterns.

The factor *birthplace* was considered a categorical variable comprising the five states: QLD, WA, NSW/ACT, SA and TAS (refer to map in Chapter 5, Figure 5.2); IRRs were derived for the four states, QLD, WA, SA and TAS, relative to the most-highly populated state, NSW/ACT, as reference (IRR=1.0).

For the factor *sex*, IRRs were calculated for females relative to males as the reference category (IRR=1.0).

These univariate regression models were thus:

- MS risk depends on sex Model (2)

(reference=males, IRR=1.0), and

- MS risk depends on birthplace (state) Model (3)

birthplace comprising five states, QLD, WA, NSW/ACT, SA and TAS as indicator variables (reference=NSW/ACT; IRR=1.0).

Decade of birth within 1920 to 1950

A further univariate analysis was conducted to check the 1920 to 1950 year-of-birth data restriction period chosen for analysis. Although such restriction is a method of controlling confounding [1], it was still possible that MS risk was varying within the

chosen 1920 to 1950 year-of-birth range. Three 'decades', 1920 to 1929, 1930 to 1940 and 1941 to 1950, were therefore considered as categories of the *decade* factor (reference=1930 to 1940) to check this possibility, using the following model:

- MS risk depends on decade (of birth) within 1920 to 1950 Model (4)

decade comprising 1920 to 1929, 1930 to 1940 and 1941 to 1950 (calendar) years of birth (reference category=1930 to 1940; IRR=1.0)

7.2.4 Effects of other possible factors on timing-of-birth pattern

Subgroup analyses

Regression analyses of timing of birth using model (1) were initially conducted separately for each state region of birth, QLD, WA, NSW/ACT, SA and TAS, in an attempt to investigate any regional variation in a temporal birth pattern (i.e. research question 2). Separate timing-of-birth regression analyses were also conducted for each sex, because some characteristics of MS are known to differ by gender. For example, males often show differences in clinical course of disease and in disease severity and prognosis compared with females [665]. Sample size for each birthplace and each sex was quite low in these analyses, particularly for the smaller states (refer to Table 6.15, Section 6.3.1).

Multivariate analyses

Modelling MS risk by multiple negative binomial regression was then undertaken to investigate the possible effects of other main factors—particularly *sex* and *birthplace*—on a timing-of-birth pattern. That is, the basic regression model (1) was employed and the additional factors added to the model. The advantage of this technique over subgroup analyses was that the effects of several factors could be considered at the same time, without compromising the overall sample size [1]. In particular, the possible effects of the additional factors as confounders of any observed seasonal pattern of birth could be determined, as well as their potentially important effects as modifiers of any such pattern.

Possible confounders

A confounding variable may be described as 'a variable of little immediate interest that is correlated with the risk (or exposure) factor and is independently related to the outcome variable of interest' [666]. The following factors were first added to the basic MS-risk regression model (1) comprising just the main study exposure (risk) factor, *month* of birth, to determine any large effects on the two-monthly IRRs that might indicate possible confounding of observed results by those factors:

- sex (two categories, reference=males)
- birthplace (five states, QLD, WA, NSW/ACT, SA and TAS, as indicator variables; reference=NSW/ACT).

The full multivariate regression model, including these possible confounders as covariates, was then as follows:

- MS risk depends on month of birth, sex and birthplace Model (5)

IRR values obtained were adjusted for each of the other factors, or covariates, in the model. For example, the IRR values for the two-monthly periods were adjusted for both sex and birthplace. Importantly, by comparing the unadjusted IRRs in the univariate models (1), (2) and (3) with the adjusted IRRs in model (5) and determining the extent of each difference in IRR value, it could be judged whether the additional factors were confounding any timing-of-birth pattern.

Also, decade of birth within the restricted 1920 to 1950 year-of-birth range was checked as another possible confounder of a seasonal birth pattern. The full multivariate regression model then was:

- MS risk depends on month of birth, sex, birthplace and decade Model (6)

Potential effect modifiers

A further advantage of multivariate analysis was that statistical interactions between the factors could also be investigated. Such interaction between a factor and the risk exposure of interest (e.g. month of birth) on MS risk may indicate the presence of an important effect-measure modifier that might provide more information about the nature

of the actual study exposure factor [666]. For example, a statistical interaction between month of birth and birthplace, defined by a product-term *month*birthplace* added to a baseline regression model [1], may indicate that MS risk varied more widely between, say, summer- and winter-born persons when their birthplace was in higher-latitude TAS compared with lower-latitude QLD (research question hypothesis 2).

Therefore, the following product terms were added, in turn, to a baseline model comprising the three factors *month* (of birth), *sex* and *birthplace* (model (5)) and the effects of each product term evaluated statistically:

- *month*sex*
- *month*birthplace*
- *birthplace*sex*.

The above three interactions were tested in turn by comparing the following three models, which included the above product terms, with the baseline main-effects model (5):

- MS risk depends on month of birth, sex, birthplace and *month*sex* Model (7)
- MS risk depends on month of birth, sex, birthplace and *month*birthplace* Model (8)
- MS risk depends on month of birth, sex, birthplace and *birthplace*sex* Model (9)

Chi-square likelihood ratio tests were used to evaluate the 'nested' negative binomial models with and without each of these interactions—that is, models (7), (8) or (9) compared with model (5)—in order to determine whether any of these interactions were statistically significant ($p<0.05$) and should be retained in the final model.

Finally, the addition of a *month*decade* (of birth) product term was similarly tested to determine whether there was any significant effect modification by decade within the restricted 1920 to 1950 year-of-birth range chosen for analysis. For this check, the following model (10) was compared with the baseline main-effects model (6):

- MS risk depends on month of birth, sex, birthplace, decade and *month*decade*

Model (10)

It should be noted that the aim in this section was simply model specification rather than precise model fitting or prediction. That is, the main concern was which covariates and product terms were required, as confounders and effect modifiers, in a basic timing-of-birth model. In other words, the final model was required to provide only approximately valid *summary* estimates or trends for a few key relationships rather than (approximately) valid *exposure-specific predictions* of outcomes [1].

Regional seasonal UVR as independent risk factor

Figure 7.1 illustrates the wide seasonal and regional variation in daily ambient effective UVR in Australia, ranging from 1.6 MED units per day in Hobart, TAS in July to 30.4 MED per day in Perth, WA in January (data as monthly averages for the capitals of the states surveyed in 1981, from Table 7.1). An approximate six-month periodicity is evident between the maximum (summer) UVR period (>15 MED per day) in November to February and the minimum (winter) UVR period (<10 MED per day) in May to August in all states (Figure 7.1).

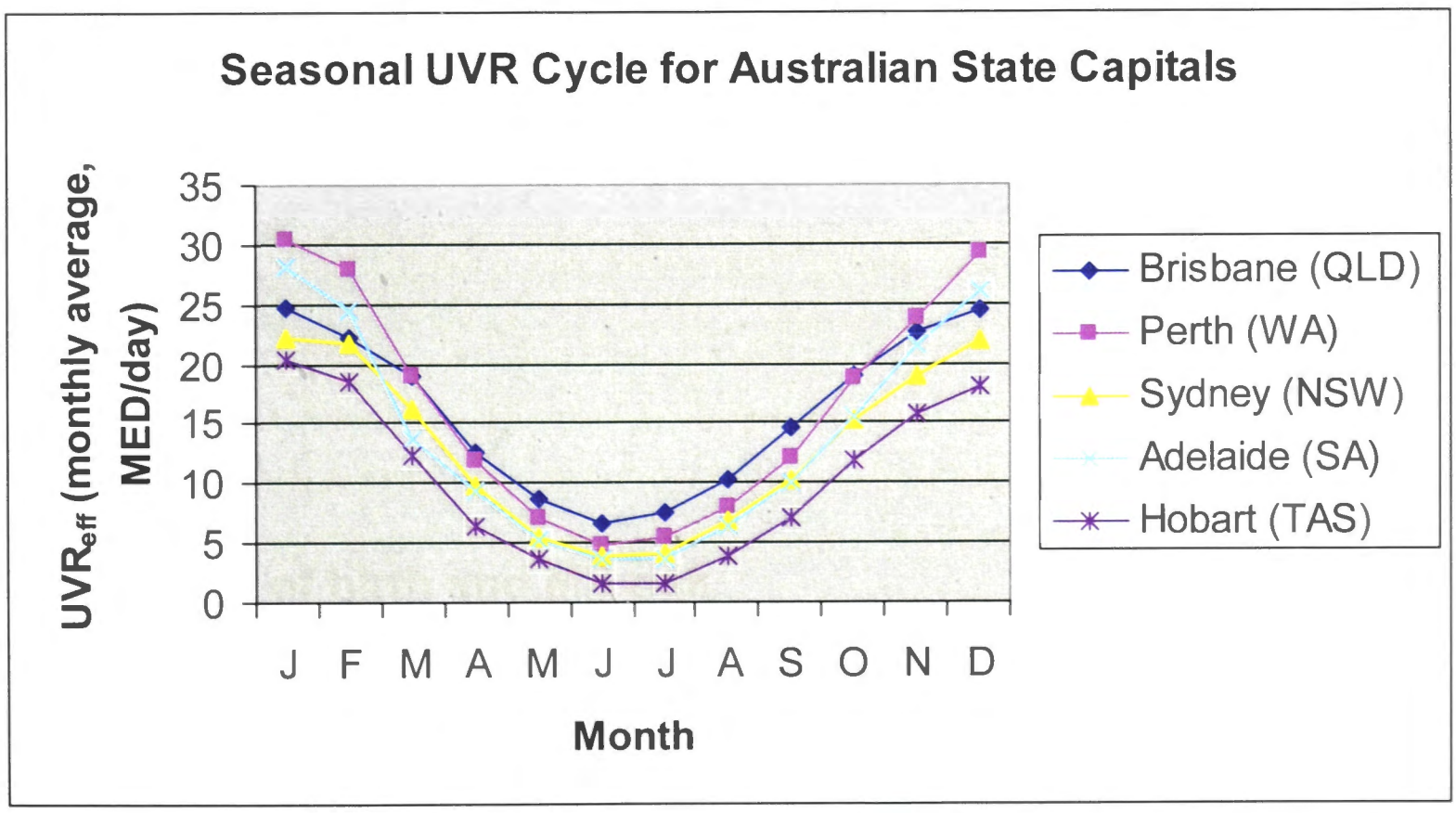


Figure 7.1: Monthly averages of daily total UVR_{eff} (MED) for Australian state capitals included in 1981 MS Survey, 1996 to 2000. (Data from P. Gies [pers. commn] and Gies et al., 1994 for Hobart [660].)

Because ambient UVR may be an important (maternal) exposure or risk factor acting around the time of birth of the offspring, this additional exposure factor was used, independently of month of birth, in regression models similar to those already described. Further, because ambient UVR may be acting even before the time of birth, for example, through maternal serum vitamin D levels, this UVR factor was able to be successively 'lagged' to represent various periods before birth; these separate UVR variables—one for each period before birth—could then be tested in regression analyses for effects on MS risk.

A continuous regional-UVR_{eff} variable was generated from the monthly ambient UVR values of the state capital cities (Table 7.1) to represent the ambient UVR at the time (month) and place (state) of birth of each individual. Similar variables were then generated for various periods before birth, from one to nine months prior, by lagging the UVR values. That is, the UVR_{eff} monthly averages *at birth* were successively lagged a number of months so that they then expressed the ambient UVR_{eff} pertaining to a particular time *before birth,* for each individual and for each region of birth. For example, the continuous variable '*UV-6mth*' (see Table 7.11, Section 7.3.4) denoted the daily ambient UVR_{eff} level at six months prior to the birth month of each individual at each respective birthplace. Nine such (continuous) variables, representing one to nine months before the birth month of an individual, were generated and included in separate negative binomial regression models, as previously described for the categorical month-of-birth variable in model (1) (Section 7.2.3). Different periods before birth could then be assessed for their ability to account for the observed variation in (lifetime) MS risk.

7.3 Results

7.3.1 Month of birth and MS risk

Figure 7.2 shows the MS-risk pattern by month of birth (two-monthly periods) expressed as the IRR for MS births, and 95% CI, relative to the reference IRR (= 1.0) for the May to June period. All (male and female) MS cases born in Australia during the 1920 to 1950 period (n=1,524 cases), together with their corresponding by-month and

by-year reference-population denominators, have been used here in negative binomial regression model (1).

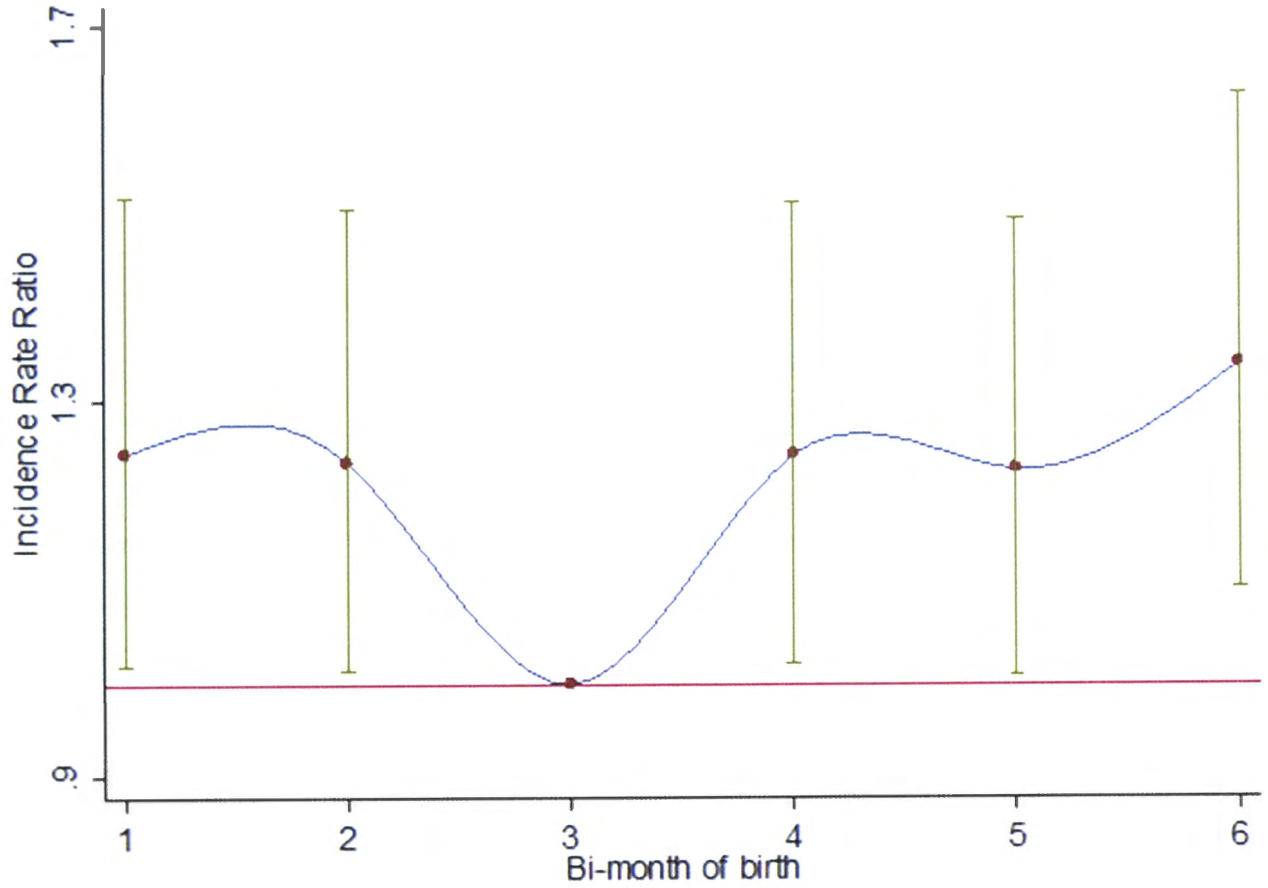


Figure 7.2: MS risk by month of birth (data for 1920 to 1950 Australia born; n=1,524 cases).

IRRs for all of the other two-monthly periods compared with the May to June reference period were statistically significant ($p<0.05$; 95% CI excluding the IRR value 1.0), indicating increased MS risk in any of these periods (ranging from 1.22 to 1.34 times) compared with the minimum-risk period in May to June. In particular, the November to December period showed a maximum and highly significant ($p<0.01$) excess of MS births relative to those in May to June (shown in bold in Table 7.3). These results indicate that the MS risk if born in the early summer months of November to December is significantly higher than the risk if born in the early winter months of May to June, by 34% (IRR 1.34; 95% CI 1.10, 1.63) (Table 7.3).

Table 7.3: IRRs and 95% confidence intervals for MS risk for two-monthly periods of birth in Australia (1920 to 1950 born; n=1,524 cases)

Factor	Category (months)	IRR [95% CI]	p-value
Month of birth	1 (January toFebruary)	1.24 [1.02, 1.52]	0.031*
	2 (March to April)	1.23 [1.01, 1.51]	0.037*
	3 (May to June)	1.00 (Reference)	-
	4 (July to August)	1.24 [1.02, 1.51]	0.030*
	5 (September to October)	1.23 [1.01, 1.49]	0.041*
	6 (November to December)	**1.34 [1.10, 1.63]**	**0.003****

* IRR statistically significant (p<0.05).
** IRR statistically significant (p<0.01).

7.3.2 Direct effects of sex and birthplace factors on MS risk

Table 7.4 shows MS risk as IRR and 95% CI for females compared with males, estimated using negative binomial regression model (2) (univariate analysis). MS risk for females, as IRR, was 2.28 times that for males, this difference being consistent with the general overall female to male prevalence ratios for MS of approximately two-fold [301].

MS risk for each birthplace state, estimated from model (3), is also shown in Table 7.4. All birthplace states except SA showed significantly different (p<0.01) IRRs from that of the reference state, NSW/ACT. Further, the results indicate increased MS risk with increase in latitude from QLD to TAS (Table 7.4).

Table 7.4: MS risk as IRR + 95% CI for univariate factors, sex and birthplace, from negative binomial regression models (2) and (3), respectively (data for 1920 to 1950 Australia born)

Factor	Category	IRR [95% CI]	p-value
Sex	Male	1.00 (Reference)	-
	Female	2.28 [2.03, 2.55]	<0.001**
Birthplace state (latitude of capital)#	QLD (27.5°S)	0.59 [0.51, 0.69]	<0.001**
	WA (31.9°S)	0.76 [0.62, 0.92]	0.005**
	NSW/ACT (33.9°S)	1.00 (Reference)	-
	SA (34.9°S)	1.03 [0.88, 1.21]	0.723
	TAS (42.9°S)	2.70 [2.06, 3.51]	<0.001**

Latitude of state capital cities in parentheses.
** IRR statistically significant (p<0.01).

This trend of increasing MS risk with southern latitude *at birth* is illustrated in Figure 7.3, where the IRR estimates from Table 7.4 are shown regressed against mean latitude of the birthplace states. The figure indicates that this association with latitude at birth is adequately explained by a linear model (p=0.031 for linear regression of mean IRR on mean latitude of birthplace states).

A latitude relationship of MS risk with *birthplace* is thus evident rather than simply with survey area. Importantly, the data in Table 7.4 and Figure 7.3 allow simple quantitative comparison of the IRR means for the birthplace states and confirm, as indicated by the prevalence ratios for QLD and TAS derived in Chapter 6 (see Section 6.3.2), that overall MS risk if *born* in TAS is some *4.6 times higher than if born* in QLD.

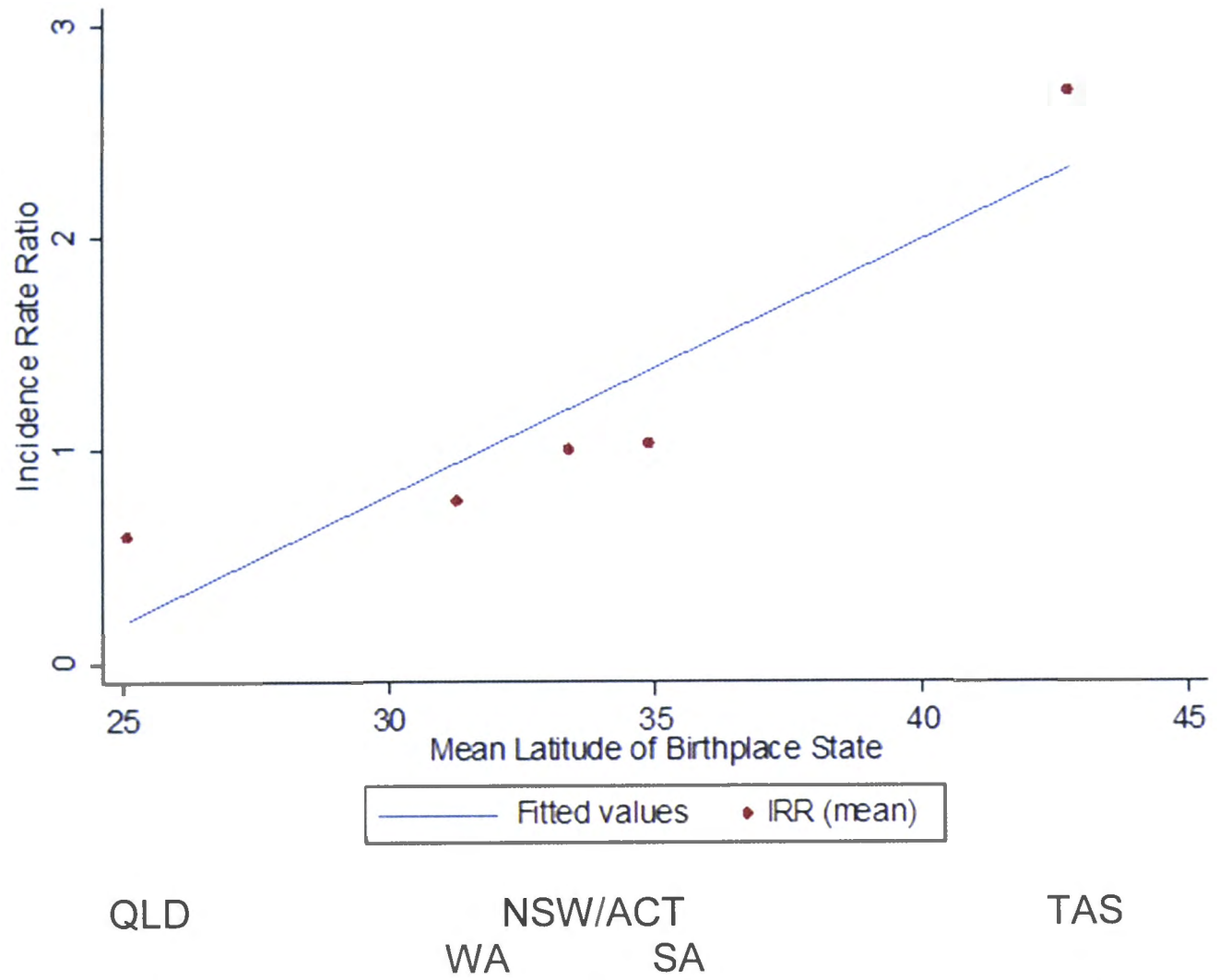


Figure 7.3: MS risk (IRR estimates from Table 7.4) by latitude of birthplace state (mean latitude of state in decimal degrees south; 1920 to 1950 Australia born; n=1,524 cases).

Decade of birth within 1920 to 1950 restricted range

As a check of the 1920 to 1950 year-of-birth restriction period chosen for analysis (see Chapter 5, Section 5.3.3, Figure 5.12), three decades within this period were investigated. MS risk was found to be significantly different in only the later-born 1941-50 decade group compared with the central decade reference group (Table 7.5), the IRR for this later-born decade group being less than 1.0. This result indicated that there were probably some as yet undiagnosed (and therefore not ascertained) MS cases in this age group compared with the other two groups. However, this was not unexpected given the result shown in Chapter 5 (see Section 5.3.3, Figure 5.11) for the mid-1940 to mid-1950 decade using the total survey sample. In that figure, the number of MS cases relative to the census population was visibly lower than in the two earlier-born decades. Whether this lowered MS risk per se in the 1941- to 1950-born decade group might have any confounding effect on an observed timing-of-birth pattern will be considered in the following section.

Table 7.5: MS risk as IRR + 95% CI for the univariate factor, decade of birth, from negative binomial regression model (4) (data for 1920 to 1950 Australia born)

Factor	Category	IRR [95% CI]	p-value
Decade of birth	1 (1920–1929)	1.09 [0.95, 1.24]	0.222
	2 (1930–1940)	1.00 (Reference)	-
	3 (1941–1950)	0.73 [0.64, 0.83]	<0.001**

** IRR statistically significant (p<0.01).

7.3.3 Effects of other factors on timing-of-birth risk pattern—sex, birthplace and decade of birth

Subgroup analyses—sex and birthplace

MS risk (as IRR and 95% CI, using negative binomial regression model (1)) is shown by month of birth (two-monthly periods) for the sexes separately (n=456 and 1,068 for males and females, respectively) in Table 7.6. IRR estimates were statistically significant (p<0.05) only for females, four of the five test periods showing increased MS risk compared with the reference May to June period (Table 7.6).

Table 7.6: IRR + 95% CI for MS risk for two-monthly periods of birth in Australia (1920 to 1950 born) by sex (males, n=456; females, n=1,068)

Month of birth	Males		Females	
	IRR [95%CI]	p-value	IRR [95%CI]	p-value
Jan–Feb	1.12 [0.80, 1.57]	0.493	1.28 [1.02, 1.61]	0.036*
Mar–Apr	1.30 [0.94, 1.80]	0.106	1.18 [0.94, 1.49]	0.150
May–Jun	1.00 (Reference)	-	1.00 (Reference)	-
Jul–Aug	1.05 [0.75, 1.47]	0.769	1.32 [1.05, 1.65]	0.016*
Sep–Oct	1.05 [0.75, 1.47]	0.785	1.30 [1.04, 1.63]	0.022*
Nov–Dec	1.31 [0.94, 1.81]	0.104	1.33 [1.06, 1.67]	0.013*

* IRR statistically significant ($p<0.05$).

MS risk by month of birth is shown separately by birthplace (state) in Australia in Table 7.7. While most of the state regions showed at least some significantly increased IRRs relative to the May to June reference, confidence intervals were also wider due to the smaller sample sizes and, overall, the results were not interpretable as either a consistent or a predictably changing pattern. The lack of any meaningful pattern of MS births in these subgroup analyses is likely due to inadequate sample size. Instead, the effect of factors such as sex, birthplace and decade of birth on the temporal pattern of birth were investigated further by multivariate analysis, considered next.

Table 7.7: IRR + 95% CI for MS risk for two-monthly periods of birth in Australia (1920 to 1950 born) by birthplace (states, QLD [n=216], WA [n=123], NSW/ACT [n=928], SA [n=197] and TAS [n=60])

Month of birth	**QLD** (27.5 °S)#		**WA** (31.9 °S)		**NSW /ACT** (33.9 °S)		**SA** (34.9 °S)		**TAS** (42.9 °S)	
	IRR [95%CI]	p	IRR [95%CI]	p	IRR [95%CI]	p	IRR [95%CI]	p	IRR [95%CI]	p
Jan–Feb	1.39 [0.83, 2.32]	0.204	1.63 [0.81, 3.31]	0.172	0.99 [0.76, 1.27]	0.924	1.90 [1.05, 3.43]	0.033*	2.81 [1.09, 7.24]	0.032*
Mar–Apr	1.35 [0.81, 2.25]	0.248	1.70 [0.84, 3.41]	0.138	1.09 [0.85, 1.40]	0.506	1.68 [0.93, 3.06]	0.086	1.39 [0.48, 4.07]	0.543
May–Jun	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-
Jul–Aug	1.00 [0.58, 1.72]	0.992	2.16 [1.11, 4.21]	0.024*	1.06 [0.83, 1.36]	0.647	2.05 [1.15, 3.64]	0.015*	1.56 [0.56, 4.35]	0.394
Sep–Oct	1.43 [0.86, 2.37]	0.168	1.34 [0.65, 2.78]	0.425	1.07 [0.84, 1.37]	0.587	2.00 [1.12, 3.56]	0.019*	0.91 [0.29, 2.86]	0.875
Nov–Dec	1.67 [1.01, 2.75]	0.045*	1.56 [0.76, 3.17]	0.222	1.09 [0.85, 1.40]	0.489	2.03 [1.14, 3.61]	0.016*	2.09 [0.78, 5.58]	0.140

Latitude of state capital cities in parentheses.
* IRR statistically significant ($p<0.05$).

Multivariate analyses—month of birth, sex, birthplace and decade of birth

Having shown in the previous section that the factors *sex* and *birthplace,* and *decade*, have a direct effect on MS risk, it was now important to consider whether these factors might confound the observed timing-of-birth pattern shown in Figure 7.2. At the same time, potentially important statistical interactions between the covariate factors could be investigated without compromising the available sample size.

Possible confounders of timing-of-birth pattern

The variables *sex* and *birthplace* were considered together with *month* in a 'full' multivariate model (model (5)). Table 7.8 shows MS risk by month of birth (two-monthly periods), the IRR values now adjusted for the covariate factors, *sex* and *birthplace.* The unadjusted IRR values (model (1), Table 7.3), are given again here for direct comparison, as are those for *sex* and *birthplace*. The close similarity between the unadjusted and adjusted estimates of MS risk by month of birth (<1.5% change in IRRs adjusted for both *sex* and *birthplace* factors together) indicate that neither of these additional factors were confounding the timing-of-birth MS-risk pattern (Table 7.8).

Table 7.8: MS risk as IRR + 95% CI for month of birth, shown both unadjusted (model (1)) and adjusted for both sex and birthplace state (model (5); data for 1920 to 1950 Australia born)

Factor	Category	Unadjusted IRR [95% CI]	p-value	Adjusted* IRR [95% CI]	p-value
Month of birth	Jan–Feb	1.24 [1.02, 1.52]	0.031	1.23 [1.02, 1.48]	0.033
	Mar–Apr	1.23 [1.01, 1.51]	0.037	1.23 [1.02, 1.48]	0.031
	May–Jun	1.00 (Reference)	-	1.00 (Reference)	-
	Jul–Aug	1.24 [1.02, 1.51]	0.030	1.23 [1.02, 1.48]	0.029
	Sep–Oct	1.23 [1.01, 1.49]	0.041	1.22 [1.01, 1.46]	0.039
	Nov–Dec	1.34 [1.10, 1.63]	0.003	1.32 [1.10, 1.58]	0.003
Sex	Male	1.00 (Reference)	-	1.00 (Reference)	-
	Female	2.28 [2.03, 2.55]	<0.001	2.28 [2.04, 2.55]	<0.001
Birthplace (state)#	QLD	0.59 [0.51, 0.69]	<0.001	0.59 [0.51, 0.69]	<0.001
	WA	0.76 [0.62, 0.92]	0.005	0.76 [0.63, 0.92]	0.004
	NSW/ACT	1.00 (Reference)	-	1.00 (Reference)	-
	SA	1.03 [0.88, 1.21]	0.723	1.03 [0.88, 1.20]	0.714
	TAS	2.70 [2.06, 3.51]	<0.001	2.70 [2.07, 3.51]	<0.001

Birthplace states listed in order of increasing south latitude (see Table 7.4)
* IRR adjusted for both other factors in multivariate model (5); for example, month of birth adjusted for both sex and birthplace

The *decade* (of birth) group within the 1920 to 1950 year-of-birth range, when considered as another possible confounder of the timing-of-birth pattern, also did not influence MS risk by month of birth to any extent—IRRs adjusted for *decade* as well as *sex* and *birthplace* (model (6)) differed from unadjusted values by <1.6% (Table 7.9).

Table 7.9: MS risk as IRR + 95% CI for month of birth, shown both unadjusted (model (1)) and adjusted for sex, birthplace and decade of birth (model (6); data for 1920 to 1950 Australia born)

Factor	Category	Unadjusted IRR [95% CI]	p-value	Adjusted* IRR [95% CI]	p-value
Month of birth	Jan–Feb	1.24 [1.02, 1.52]	0.031	1.22 [1.01, 1.47]	0.035
	Mar–Apr	1.23 [1.01, 1.51]	0.037	1.22 [1.02, 1.47]	0.033
	May–Jun	1.00 (Reference)	-	1.00 (Reference)	-
	Jul–Aug	1.24 [1.02, 1.51]	0.030	1.23 [1.02, 1.47]	0.029
	Sep–Oct	1.23 [1.01, 1.49]	0.041	1.22 [1.01, 1.46]	0.036
	Nov–Dec	1.34 [1.10, 1.63]	0.003	1.32 [1.10, 1.58]	0.003
Sex	Male	1.00 (Reference)	-	1.00 (Reference)	-
	Female	2.28 [2.03, 2.55]	<0.001	2.27 [2.03, 2.53]	<0.001
Birthplace (state)#	QLD	0.59 [0.51, 0.69]	<0.001	0.60 [0.51, 0.69]	<0.001
	WA	0.76 [0.62, 0.92]	0.005	0.77 [0.64, 0.93]	0.006
	NSW/ACT	1.00 (Reference)	-	1.00 (Reference)	-
	SA	1.03 [0.88, 1.21]	0.723	1.04 [0.89, 1.21]	0.630
	TAS	2.70 [2.06, 3.51]	<0.001	2.73 [2.10, 3.55]	<0.001
Decade of birth	1920–29	1.09 [0.95, 1.24]	0.222	1.06 [0.94, 1.20]	0.347
	1930–40	1.00 (Reference)	-	1.00 (Reference)	-
	1941–50	0.73 [0.64, 0.83]	<0.001	0.73 [0.64, 0.82]	<0.001

Birthplace states listed in order of increasing south latitude (see Table 7.4).
* IRR adjusted for all other factors in multivariate model (6); e.g. month of birth adjusted for sex, birthplace and decade of birth.

This means that although *birthplace*, *sex* and *decade* would be necessary terms in a predictive regression model (because of independent effects of these factors on MS risk), these covariates were not confounding the main timing-of-birth study factor. The adjusted model in Table 7.9 in graphical form (Figure 7.4) is therefore essentially similar to the unadjusted timing-of-birth model in Figure 7.2. This adjusted model thus indicates significantly increased MS risk, approximately 1.3-fold, if born in the early

summer November to December period compared with the lowest MS risk if born in the early winter May to June period.

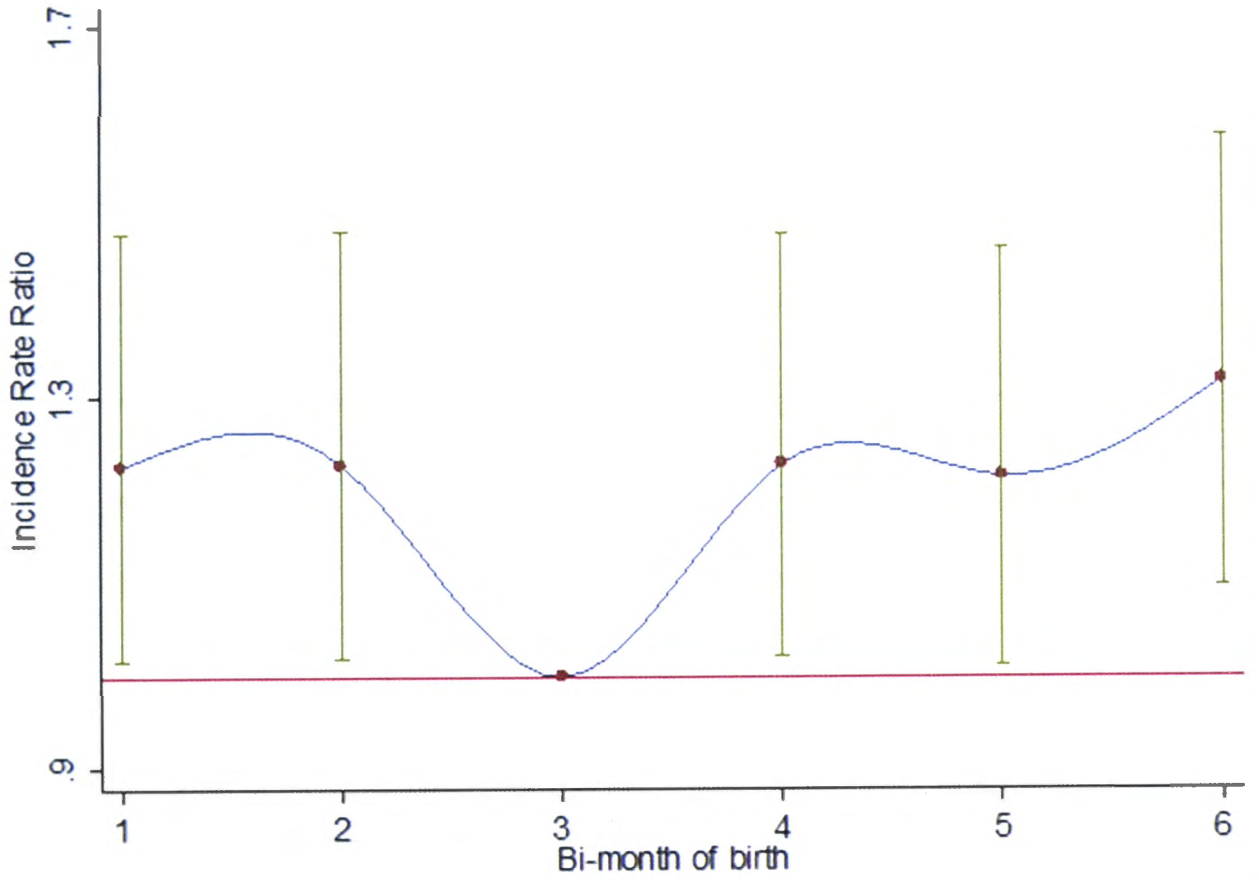


Figure 7.4: MS risk by month of birth, adjusted for sex, birthplace and decade of birth (1920 to 1950 Australia born; n=1,524 cases).

Potential interactions between month of birth, sex, birthplace and decade of birth

The product terms *month*sex* and *month*birthplace* were each considered additional terms in the 'full' main-effects model (5), to determine whether any important modification of the timing-of-birth pattern by either factor was evident. However, chi-square likelihood ratio tests of nested regression models with and without these terms showed that there was no significant interaction between the factors *month* (of birth) and *sex*, and none between *month* (of birth) and *birthplace* (Table 7.10).

Table 7.10: Multivariate modelling—results of statistical tests for interaction between factors in negative binomial regression analyses using nested models

Baseline model and factors	New model and additional product terms	p-value[#] /significance of product term
(5) Month of birth, sex, birthplace	(7) *month*sex* (8) *month*birthplace* (9) s*ex*birthplace*	0.506 NS 0.247 NS 0.066 NS
(6) Month of birth, sex, birthplace, decade	(10) *month*decade*	0.687 NS

[#]p-value for chi-square likelihood ratio test of nested regression models with and without the additional product term; e.g. baseline model (5) nested in models (7), (8) or (9) containing the additional product terms.

The product term *sex*birthplace* was also considered an additional term in main-effects model (5). Again, chi-square likelihood ratio tests of nested regression models with and without this product term showed that there was no significant interaction between the factors *birthplace* and *sex* (Table 7.10), so that this term could also be omitted from the final model. Similarly, addition of a *month*decade* product term to model (6) showed no significant modification of the month-of-birth effect by decade of birth within the data restriction range chosen (Table 7.10).

The results in Table 7.10 indicate that the observed overall month-of-birth pattern of birth relating to MS risk was not significantly different between males and females and, most importantly, did not appear to be significantly affected by the birthplace region of the cases. That is, there was no detectable effect modification of the timing-of-birth risk pattern by either birthplace or sex in these data. This means that the early summer (November to December) versus early winter (May to June) *ratio* of MS births (~1.3) did not appear to differ over the birthplace regions studied (research question 2).

Resulting MS risk model for timing of birth

A relatively simple, unadjusted, main-effects-only model with no significant interactions appears adequate to specify MS risk (as IRR) as dependent on timing of birth, using these Australian data. This model (1) is shown in Figure 7.2. The model shows

minimum MS risk for births in the (reference) early winter May to June period and higher risk for births over the remainder of the year, the increased risk rising in the early summer November to December period to approximately 1.3 times that for early winter-born.

This increase in risk appears to be additional to other direct effects on MS risk by the factors *sex* and *birthplace*. That is, while

- MS risk overall is approximately 2.3 times higher for females than for males
- MS risk overall is approximately 4.6 times higher if born in TAS than if born in QLD (Section 7.3.2, Table 7.4),

there now appears to be a further risk, for either sex and regardless of where born in Australia, of some 1.3 times if born in November to December than if born in May to June. That is,

- MS risk is approximately 1.3 times higher if born in the early summer months of November to December than if born in the early winter months of May to June (research question 1)
- the November-to-December to May-to-June risk ratio (1.3) does not differ by birthplace (research question 2) or sex.

7.3.4 Perinatal UVR as independent exposure factor

Ambient (maternal) UVR at and before time of birth

To examine possible associations between ambient UVR around the time of birth and MS, incidence of MS was modelled against perinatal ambient UVR as a continuous variable generated from monthly regional ambient UVR_{eff} values. Advantages of these quantitative UVR variables were that not only were they monthly (compared with the previous two-monthly periods), but they comprised a regional (place of birth) as well as a seasonal (month of birth) component; that is, the monthly UVR values for each perinatal period differed for each (individual) birthplace (Table 7.1 and Figure 7.1).

Ten different UVR variables, representing the birth month and one to nine months prior to birth for each individual, were tested in separate negative binomial regressions, in order to assess different periods at and before birth for possible associations between regional perinatal UVR and (lifetime) MS risk (Table 7.11, IRRs unadjusted and

adjusted for age (year of birth) and sex by multivariate analysis as described in Section 7.2.4).

There was no association between regional daily ambient UVR at the time of birth and subsequent risk of MS. Similarly, lags of one to four months before birth (late second to third trimesters) were not associated with MS risk. However, for lags of five to nine months (first to early second trimesters) there were significant inverse associations between prenatal regional ambient UVR levels and lifetime MS risk (unadjusted IRR ranging from 0.74 (95% CI 0.63, 0.85) to 0.81 (95% CI 0.7, 0.94); $p<0.01$) (Table 7.11).

The strongest associations between regional ambient UVR and MS risk were at seven and eight months before birth—this period before birth corresponds to the mid-period of the first trimester of gestation. By averaging the ambient UVR values for seven and eight months prior to birth, UVR values representing 'first trimester UVR' for each region could then be obtained and analysed similarly by negative binomial regression. The result was an IRR of 0.72 (95% CI 0.62, 0.84; $p<0.001$) (Table 7.11). Interpreted quantitatively, this means that *for an increase in ambient UVR_{eff} exposure of 20 MED units* (approximately equivalent to the regional average increase in daily total effective UVR from winter to summer in Australia, Figure 7.1) *during the first foetal trimester*, MS risk *decreased* to 0.72 of the reference (winter) risk level, or by approximately 28%.

Table 7.11: Risk of MS for regional ambient UVR at and before time of birth

Ambient UVR (monthly regional, maternal)[+]		Unadjusted IRR [95% CI]	p-value	Adjusted IRR* [95% CI]	p-value
At birth	UV birth	1.01 [0.87, 1.17]	0.906	1.01 [0.88, 1.16]	0.851
Time before birth (prenatal)	UV-1mth	1.03 [0.89, 1.19]	0.706	1.03 [0.90, 1.19]	0.630
	-2mth	1.00 [0.86, 1.16]	0.991	1.01 [0.88, 1.16]	0.894
	-3mth	0.94 [0.81, 1.10]	0.448	0.96 [0.83, 1.10]	0.525
	-4mth	0.88 [0.76, 1.02]	0.100	0.89 [0.78, 1.03]	0.117
	-5mth	0.81 [0.70, 0.94]	0.007	0.83 [0.72, 0.95]	0.007
	-6mth	0.77 [0.67, 0.90]	0.001	0.78 [0.68, 0.90]	0.001
	-7mth	**0.74 [0.63, 0.85]**	**<0.001**	**0.75 [0.65, 0.86]**	**<0.001**
	-8mth	**0.74 [0.63, 0.85]**	**<0.001**	**0.75 [0.65, 0.86]**	**<0.001**
	-9mth	0.79 [0.68, 0.92]	0.002	0.80 [0.70, 0.92]	0.002
	UV 1st trimester[#]	**0.72 [0.62, 0.84]**	**<0.001**	**0.73 [0.63, 0.84]**	**<0.001**

[+] Based on composite month and region-specific values for each individual (see Table 7.1) and expressed in units of 20 MED/day in order to gain meaningful IRR in terms of UVR difference between summer and winter (20 MED/day is approximate average difference between summer and winter UVR levels for Australian state regions, see Figure 7.1).

[#] First trimester UVR variable obtained by averaging monthly values for seven and eight months prior to birth (in bold).

* IRR adjusted for age (year of birth) and sex.

Shape of association between first trimester UVR and MS risk

The shape of the relationship between (inversed) first trimester UVR exposure and MS risk was then investigated in greater detail. The UVR_{eff} variable was considered a categorical variable of six levels of effective UVR (in units of MED per day) and each of the five lowest levels were compared with the highest (Table 7.12, IRRs adjusted for age [year of birth] and sex). The association was non-linear and suggested a UVR 'threshold' at 20 MED/day, the four UVR categories below this level showing significantly increased MS risk ($p<0.01$, Table 7.12). That is, there was a particular increase in risk for UVR levels below a monthly average of 20 MED/day (Figure 7.5).

Table 7.12: MS risk as IRR + 95% CI for six levels of regional ambient effective UVR during first trimester of gestation[#]. (Reference UVR level=highest)

Factor	Category	Ambient UVR_{eff} (range, MED/day)	IRR* [95% CI]	p-value
UVR exposure (regional ambient, maternal)	Level 1	≥25	1.00 (Reference)	-
	2	20 - <25	1.36 [0.98, 1.86]	0.062
	3	15 - <20	1.57 [1.13, 2.17]	0.007
	4	10 - <15	1.58 [1.14, 2.19]	0.006
	5	5 - <10	1.61 [1.17, 2.21]	0.003
	6	<5	1.86 [1.33, 2.59]	<0.001

[#] First trimester UVR values obtained by averaging values for seven and eight months prior to birth.
* IRR adjusted for age (year of birth) and sex.

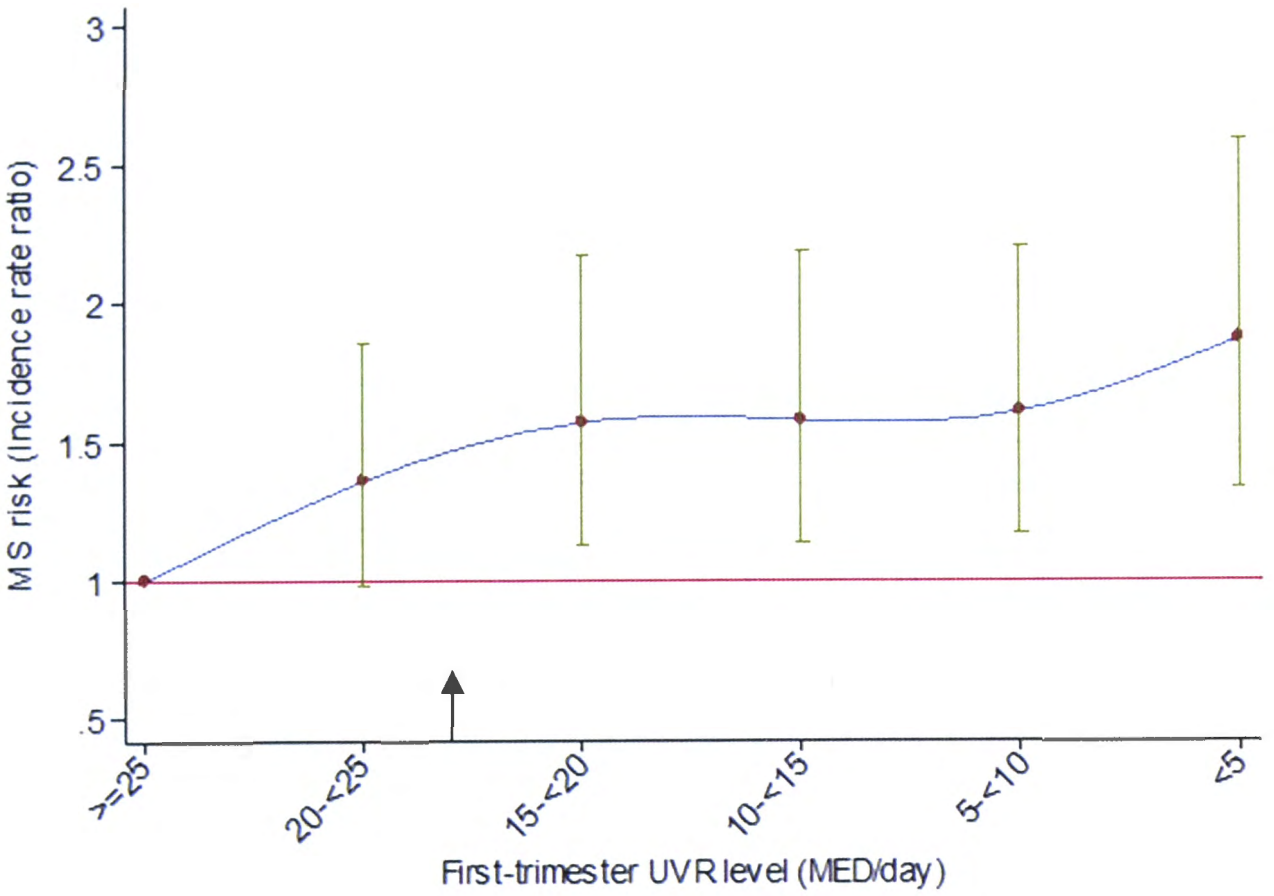


Figure 7.5: MS risk by regional ambient effective UVR level during first trimester of gestation (IRR + 95% CI adjusted for age [year of birth] and sex, from Table 7.5). (Arrow on inverse UVR axis indicates threshold at monthly average of 20 MED/day for effect of first trimester UVR on MS risk.)

Effects of other factors on first trimester UVR and MS risk

Table 7.13 shows the association between first trimester UVR (as six categories of ambient exposure) and MS risk, both unadjusted and adjusted for sex, region of birth and decade of birth. A multivariate negative binomial regression model was used, similar to model (6) (see Methods, Section 7.2.4) but with *first trimester UVR* as the main study factor instead of *month* (of birth). The association between first trimester UVR and MS persisted after adjustment for sex, birthplace region and decade of birth (Table 7.13); that is, these additional factors can again be discounted as possible confounders of the main UVR-MS association. Further, chi-square likelihood ratio tests of nested regression models, with and without additional product terms between first trimester UVR and the other factors in Table 7.13, confirmed that there was also no interaction between any of these factors and first trimester UVR. That is, there was no effect modification of the association between first trimester UVR and MS risk by these factors (p-values 0.71 and 0.80 for product terms with sex and birthplace, respectively). Thus, an inverse association between low UVR in the first trimester of gestation and increased risk of subsequent, post-birth MS has been demonstrated.

Table 7.13: MS risk as IRR + 95% CI for six levels of regional ambient effective UVR during the first trimester of gestation, with and without other factors

Factor	Category	Unadjusted IRR [95% CI]	p-value	Adjusted IRR* [95% CI]	p-value
Ambient UVR_{eff} exposure in 1st trimester+ (MED/d)	≥25	1.00 (Reference)	-	1.00 (Reference)	-
	20 - <25	1.35 [0.97, 1.87]	0.071	1.54 [1.10, 2.16]	0.013
	15 - <20	1.58 [1.14, 2.20]	0.007	1.58 [1.12, 2.22]	0.009
	10 - <15	1.58 [1.13, 2.20]	0.008	1.65 [1.17, 2.33]	0.004
	5 - <10	1.62 [1.17, 2.23]	0.004	1.65 [1.18, 2.29]	0.003
	<5	1.90 [1.35, 2.67]	<0.001	1.67 [1.18, 2.37]	0.004
Sex	Male	1.00 (Reference)	-	1.00 (Reference)	-
	Female	2.28 [2.03, 2.55]	<0.001	2.27 [2.03, 2.53]	<0.001
Birthplace (state)#	QLD	0.59 [0.51, 0.69]	<0.001	0.60 [0.52, 0.70]	<0.001
	WA	0.76 [0.62, 0.92]	0.005	0.85 [0.70, 1.04]	0.113
	NSW/ACT	1.00 (Reference)	-	1.00 (Reference)	-
	SA	1.03 [0.88, 1.21]	0.723	1.10 [0.94, 1.29]	0.237
	TAS	2.70 [2.06, 3.51]	<0.001	2.71 [2.08, 3.52]	<0.001
Decade of birth	1920–29	1.09 [0.95, 1.24]	0.222	1.06 [0.94, 1.20]	0.346
	1930–40	1.00 (Reference)	-	1.00 (Reference)	-
	1941–50	0.73 [0.64, 0.83]	<0.001	0.73 [0.64, 0.82]	<0.001

+ Ambient UVR_{eff} in first trimester based on composite month- and region-specific UVR values for each individual (see Table 7.1).
# Birthplace states listed in order of increasing south latitude (see Table 7.4).
* IRR adjusted for all other factors in a multivariate model; for example, first trimester UVR adjusted for sex, birthplace region and decade of birth.

Relationship between first trimester UVR and timing-of-birth pattern

Could low first trimester UVR now account for the observed timing of birth pattern? The relationship between the two exposure factors, prenatal *first trimester UVR* and *month* of birth, with regard to MS risk could be explored by testing nested multivariate negative binomial regression models similar to model (6) (see Methods, Section 7.2.4) with and without each of these factors. These results showed that after adjustment for prenatal UVR in the first trimester, there was no residual association between timing of birth and risk of MS. That is, once first trimester UVR was included in the model (as in Table 7.13), there was no improvement in model fit by also including month of birth (likelihood ratio chi^2 for difference between models [5 d.f.] =3.79; p=0.58). In contrast, region of birth remained significantly associated with risk of MS even after adjustment for UVR in the first trimester.

Summing up, these results mean that the risk pattern of timing of birth found here in southern hemisphere Australia among MS cases appears to be accounted for by the month- and region-specific ambient UVR during the first trimester of foetal development. That is, lower average daily levels of ambient UVR during the first trimester of pregnancy predicted a higher subsequent risk of MS in the offspring *independently of timing of birth*. This suggests that ambient UVR prior to birth might be an important (maternal) exposure factor influencing subsequent MS risk of the offspring.

The results also show that there was an independent, residual association between birthplace region and lifetime MS risk after adjustment for UVR in the first trimester. This means that not all of the regional variation in MS risk has been accounted for by the prenatal first-trimester-UVR factor. This suggests that postnatal UVR exposure in childhood or adolescence may be a further component cause of MS in later life; that is, postnatal UVR exposure might also be important in reducing overall MS risk.

7.4 Discussion

Timing-of-birth pattern

The analyses reported in this chapter show a timing-of-birth pattern for MS risk in Australia (research question 1). The risk of developing MS post-birth is some 30%

higher (approximately 1.3 times) for Australians born in the southern hemisphere early summer period, November to December, than for those born in the early winter period, May to June. This pattern was also not different by sex or birthplace (state) within Australia (research question 2).

A relatively simple model of timing-of-birth effects on lifetime MS risk was developed using data from the 1981 Australian MS Survey (see Chapter 5) on 1,524 MS cases born in Australia, by sex and state of birth for QLD, WA, NSW/ACT, SA and Hobart (TAS). These data were reconstructed into a longitudinal dataset spanning every (birth) month of every year over the chosen study period, 1920 to 1950, together with derived and adjusted reference-population denominator estimates (total n=2,468,779; see Chapter 6).

Strengths of this study include the ability to analyse data at the individual level (rather than the population level, as in Chapter 4 for other autoimmune disorders) and this was enabled for this chapter by the comprehensive and standardised case ascertainment of MS in Australia by the McLeod group's national survey in 1981 (see Chapter 5). Because this MS survey was also carried out on an Australian national census day, census totals were able to be used to derive appropriate reference-population denominators for the analyses in this chapter by employing supplementary births-registration data to apportion the census totals into month- and region-of-birth data by (birth) year (see Chapter 6, Section 6.2.2). In addition, other (1976) census data on interstate migration were able to be accessed to allow this study to make a final adjustment of the population denominators to account for potential case losses to the unsurveyed states post-birth and up to the survey date in 1981 (see Chapter 6, Section 6.2.3).

A further strength of the present study is the inclusion of individual birthplace region, rather than survey (or residence) area, as an exposure factor in the final MS-frequency dataset. This factor has not often been emphasised in other timing-of-birth studies [324, 644] (although McDowell and colleagues (2010) have recently reported earlier MS onset for US veterans born in winter in US birthplaces with low solar radiation [667]). Place of birth, rather than place of (later) residence should therefore be an exposure that is more relevant and appropriate to the time of birth.

Potential limitations include the assumptions necessarily made to derive the population denominators; however, these assumptions do not appear to be having major effects

on the final dataset, as discussed in Chapter 6 (see Section 6.4). The restriction of the year-of-birth range to the 1920 to 1950 calendar-year period further allowed exclusion of the more significant 'survival' and 'diagnosis' sampling biases for the MS cases (see Chapter 5, Section 5.3.3), as well as limitation of at least some of the 'assumption effects' for the reference-population denominators (see Chapter 6, Section 6.4). For example, potential measurement bias from differences in month-of-birth classification between cases and census-population 'controls' could be minimised by exclusion of earlier years when delays between birth and registration were greatest (see Chapter 6, Section 6.4). In the present chapter, although MS risk was found to vary by decade within this restriction range, the finding that the factor *decade* was not confounding or interacting with the timing-of-birth pattern gave further confidence in the final sample used.

The importance of maintaining adequate sample size was emphasised in this chapter in requiring two-monthly rather than monthly periods for analysis and during the attempted subgroup analyses in Section 7.2.4. Subsequent multivariate analysis allowed the effects of all principal factors to be considered at the same time while, importantly, not compromising statistical power.

Multiple regression modelling further enabled investigation of possible confounders, including birthplace and sex, that may have been influencing the subtle temporal pattern. By multivariate analysis, both *sex* and *birthplace* were confirmed as contributory factors to MS risk per se, but adjustment for these factors made little difference to the MS-risk estimates for timing of birth. The lack of substantial difference between adjusted and unadjusted IRR values (i.e. <<10%) indicated that confounding of the timing-of-birth pattern by these factors was not occurring. Data on other possible confounders were not available, especially data on personal exposure to seasonally related factors (e.g. viruses, UVR), including those possibly acting during pregnancy.

To summarise, independent effects of both sex and birthplace on MS risk in Australia have been confirmed in this study, together with a newly found timing-of-birth MS-risk pattern for the southern hemisphere. This appears to be the first recorded finding of a timing-of-birth pattern for MS risk in the southern hemisphere and the present results are broadly consistent with those for MS in the northern hemisphere (with seasons reversed). The analyses for the 1920- to 1950-born cases show a significant excess risk for those born in the early summer months of November to December, compared with the risk for those born in early winter May to June. Indeed, all other birth periods

show excess MS risk compared to the early winter May to June period, but risk is highest—approximately 1.3 times—for births in the early summer November to December period.

These results are comparable to those found in the northern hemisphere for MS from a much larger pooled sample of register data from four countries, Canada, UK, Denmark and Sweden (n=42,000 cases) [324]. In this northern hemisphere study, a significant *deficit* in MS births occurred in November and a significant *excess 'peak'* occurred in May (pattern reproduced here in Figure 7.6). For direct comparison, Figure 7.7 shows the resulting pattern from the data in this chapter when single months of birth rather than two-monthly periods of birth are used, and May is the single-month reference. In this analysis by month, November shows an increased IRR relative to May that is only of borderline significance (IRR 1.32; 95% CI 1.00, 1.74; p=0.052) while December shows statistically significant (p<0.05) 'peak' excess MS risk (IRR 1.41; 95% CI 1.07, 1.86; p=0.014). Although the confidence intervals in this analysis by month are wider because of lower sample size (Figure 7.7), these results again suggest an overall pattern of minimum MS risk in the early winter months of May and June and increased risk for all birth months other than May to June, with greatest risk if born over the November to December (and particularly December) early summer period.

While the analysis techniques in this chapter are necessarily different from those of Willer and colleagues, who show both a (November) *deficit* and a (May) *excess* MS risk compared with a baseline ('observed=expected') value of 1.0 (Figure 7.6) [324], the important comparison parameter is their overall *May-born to November-born* OR, which represents a '13% higher MS risk (95% CI 5 to 22%) if born in May compared with November'. The results in this chapter show a 30% higher risk for those born in the corresponding months of November to December compared with May to June (Figure 7.2); however, the 95% confidence interval of 10 to 63% for this ratio largely overlaps that of Willer and colleagues [324], and this result is, therefore, not inconsistent with theirs. The months of maximum (peak) births and minimum (trough) births also are remarkably similar, given that the November and May months are reversed for the southern hemisphere seasons.

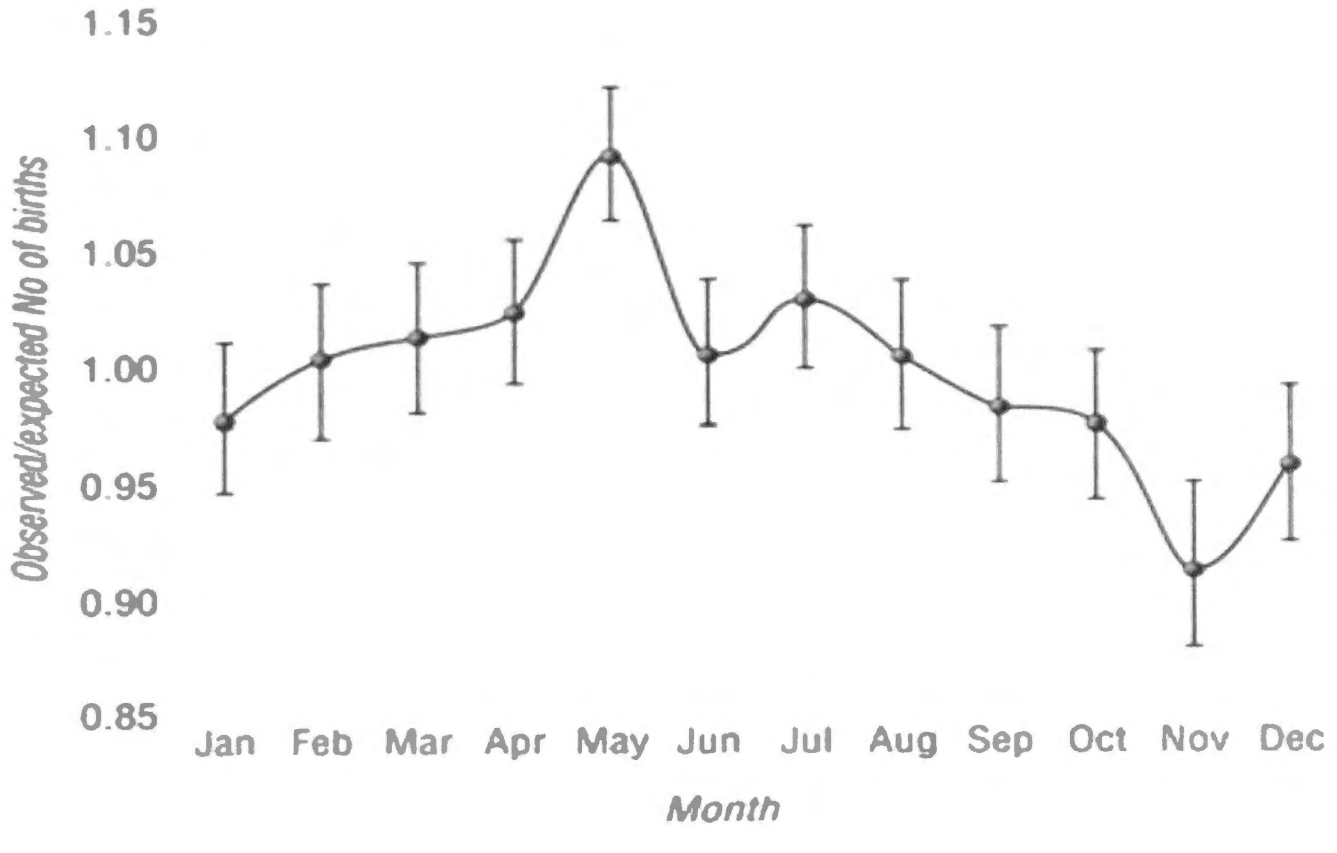


Figure 7.6: Timing of birth for northern hemisphere MS. Pooled analysis of observed/expected births in people with MS in Canadian, British, Danish and Swedish studies (n=42,045) with 95% confidence intervals (CCSG, Willer et al., 2005 [324]).

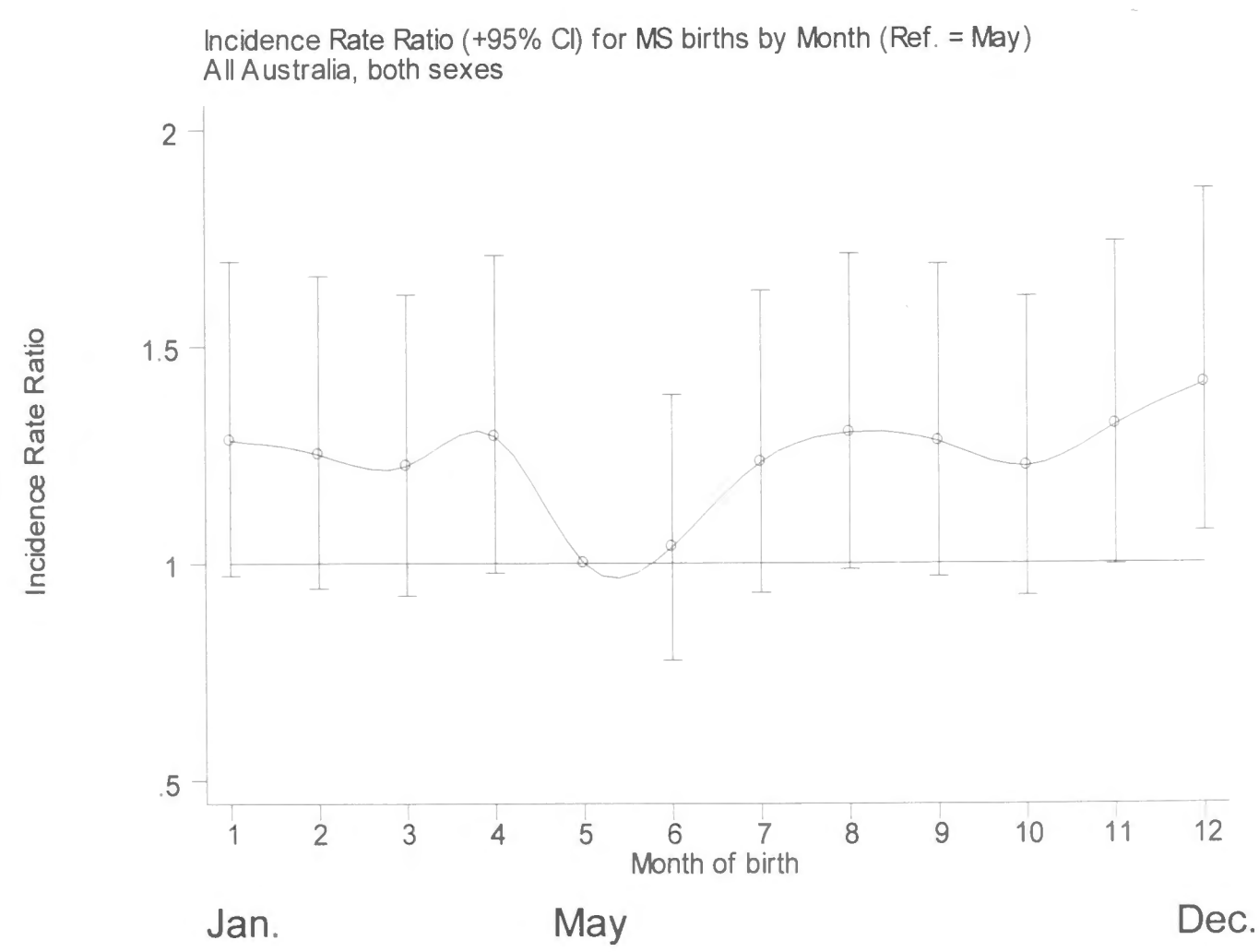


Figure 7.7: Timing of birth for southern hemisphere MS. MS risk by month of birth in Australia (1920 to 1950 Australia born; n=1,524 cases).

Interestingly, Willer and colleagues' data for the UK showed a significant deficit in December as well as in November (the same two months as shown in this chapter for excess births), although their December deficit lost significance when corrected for multiple comparisons [324]. The mean IRR values shown in this chapter thus indicate an early summer versus early winter MS-risk pattern that is quite similar to that in the northern hemisphere, in effect size and in the approximately six months periodicity between the trough and peak months.

Of note here, in a large southern hemisphere (QLD, Australia) study of schizophrenia that differentiated between southern hemisphere-born and northern hemisphere-born cases, a southern hemisphere season-of-birth pattern that similarly 'mirrored' the northern hemisphere pattern was found [668].

In this chapter's southern hemisphere results for MS, there was no measurable effect modification of the association between timing of birth and MS risk by birthplace region (research question 2), as evidenced here by the lack of statistical interaction between birthplace and the timing-of-birth pattern. The effect size of this timing-of-birth pattern appears to be not significantly affected by birthplace state from QLD to TAS, even though this factor has a significant independent effect on MS risk. Intriguingly, therefore, the ratio of early summer to early winter risk in this pattern does not change with increase in southern latitude in Australia. Such interaction might have been expected from the northern hemisphere work on MS and, perhaps, from the seasonal UVR pattern in Australia (see Table 7.1). Willer and co-workers, using incident MS cases in Scotland, Canada and Scandinavia, also reported that the magnitude of the May to November ORs appeared to increase with regional prevalence and, by implication, with regional latitude, the May to November OR for Scotland being highest. However, as in Australia, the data *within* the latitudinal breadth of Canada failed to show any difference by ascertainment area, or by sex or decade of birth [324].

For schizophrenia also, a northern hemisphere meta-analysis of several season-of-birth studies showed a small but significant positive correlation between latitude and the ORs for winter-spring versus summer-autumn births [639]. However, a southern hemisphere meta-analysis by the same study group did not show a consistent season-of-birth association for schizophrenia in this region, these authors suggesting that southern hemisphere risk-modifying environmental factors 'may be weaker, less prevalent, less regular and/or may be modified by other confounding variables' [640].

The present results for southern hemisphere MS appear to fall somewhere between these two extremes—a timing-of-birth effect consistent with that found in the northern hemisphere has been shown, but the MS-risk ratio of maximum to minimum MS births (November-to-December born to May-to-June born) of around 1.3 appears to be unaffected over the latitude and prevalence range in Australia.

It should be noted again that, in comparison with these other studies, the effect of region *at birth* has been measured here, whereas the northern hemisphere studies such as those of Willer et al. (2005) for MS and Davies et al. (2003) for schizophrenia have used the study site or ascertainment area of the study [324, 639]. The previous studies on latitude and MS prevalence in Australia by Hammond, McLeod and colleagues (the 'McLeod group') were also based on region of ascertainment (current residence) rather than birthplace [299, 301, 302, 304, 645, 650], as were results by van der Mei and colleagues on ambient UVR and MS prevalence in Australia that were based on the McLeod group's prevalence data [9]. However, the present study has further shown that a relatively high proportion of the 1981-surveyed MS cases (up to 94%), and of the census (1976) population (up to 87%), still remained in their birthplace state in 1981 or 1976, respectively (see Chapter 6, Section 6.2.3). Therefore, major differences between the results here, using region of birth, and these other studies, using region of ascertainment, should perhaps not be expected for either prevalence or timing-of-birth patterns.

Prenatal UVR and MS risk

Is prenatal UVR a candidate component causal factor affecting MS risk early in life and resulting in a timing-of-birth pattern? As outlined in Section 7.1, the finding of a timing-of-birth pattern in MS, now in both global hemispheres, strongly suggests a component causal factor acting very early in life. Evident also is that this component has a lengthy induction period (some 20 to 40 years before MS onset), defined as the time between action of the factor and onset of disease [1].

A timing-of-birth pattern is simply a risk indicator (month of birth actually being proxy for the solar zenith angle affecting solar radiation and UVR [661]) and can result from environmental influences acting at any time around the time of birth, from conception to shortly after birth. Pre-birth influences, in particular, may be especially important to this type of disorder and may be mediated through vitamin D status [221, 451, 452, 477,

484] (see Chapter 2). Pregnancy is a vulnerable time for vitamin D deficiency because of increased physiological needs and reduced maternal outdoor activity [477, 669].

A possible mechanism by which higher maternal levels of vitamin D might be protective for MS in offspring is through its immunomodulatory action, vitamin D enhancing regulatory T-cell function (see Chapter 2). Another mechanism may be that *in utero* vitamin D deficiency affects foetal brain development; experimental data on animal foetal development indicate that cerebral white matter is responsive to vitamin D and that oligodendrocytes in the brain and spinal cord have VDRs [383, 479] (see Chapter 2). Further, maternal vitamin D depletion alters neurogenesis in the developing rat brain [485], with subsequent altered gene expression in adult life [482].

As reviewed in Chapter 2, recent genetic studies in humans have further implicated vitamin D by showing direct functional interaction with the major locus that determines susceptibility to MS [188]. This risk allele, moreover, has been shown to be associated with timing of birth of MS cases in Canada, Sweden and Norway, indicating that this gene-environment interaction occurs during gestation or shortly after birth [670]. Although human evidence pertaining to foetal development has been difficult to obtain, the body of related evidence to date has led some to recommend antenatal supplementation with vitamin D specifically to prevent MS [289, 477].

Sayers and colleagues further emphasise that UVR is a relatively 'direct' variable for examining maternal vitamin D effects on subsequent offspring health and may be validly considered a causal exposure for any health outcome affected by maternal vitamin D [661]. UVR itself has also been directly implicated in autoimmune mechanisms considered to be underlying initiation of MS disease without necessarily involving vitamin D (reviewed in Chapter 2).

In this chapter, regional and seasonal variation in ambient UVR during the first foetal trimester has been shown to account (inversely) for the observed variation in MS risk with month of birth. That is, when the two exposure factors, *month of birth* and *first trimester UVR*, were considered together in nested regression models, there was no residual association between timing of birth and risk of MS once first trimester UVR was accounted for (Section 7.3.4). The timing-of-birth pattern found here has therefore now been linked, for the first time, to prenatal UVR during the first trimester of pregnancy (published as Staples et al., 2010, Appendix II).

The overall correspondence between these two exposure factors can be seen in Figure 7.8, where the month-of-birth pattern of MS risk is shown together with the UVR pattern (on an inverse scale) lagged seven to eight months to represent the ambient UVR pertaining to the first trimester. Two annual cycles are shown (data repeated for the second cycle). In addition, for diagrammatic simplicity, lagged-UVR has been averaged over all regions (however, data for the individual regions were used in all regression modelling). This figure illustrates the broad similarity between the two exposure factors with respect to MS risk, and emphasises the possible temporal link between them. Figure 7.8 thus shows a specific *in utero* link to MS risk and suggests that the higher MS risk evident in the November-to-December born in the timing-of-birth pattern may be consequent to suboptimal maternal UVR exposure seven to eight months earlier during the first trimester of pregnancy; that is, in the preceding autumn to early winter.

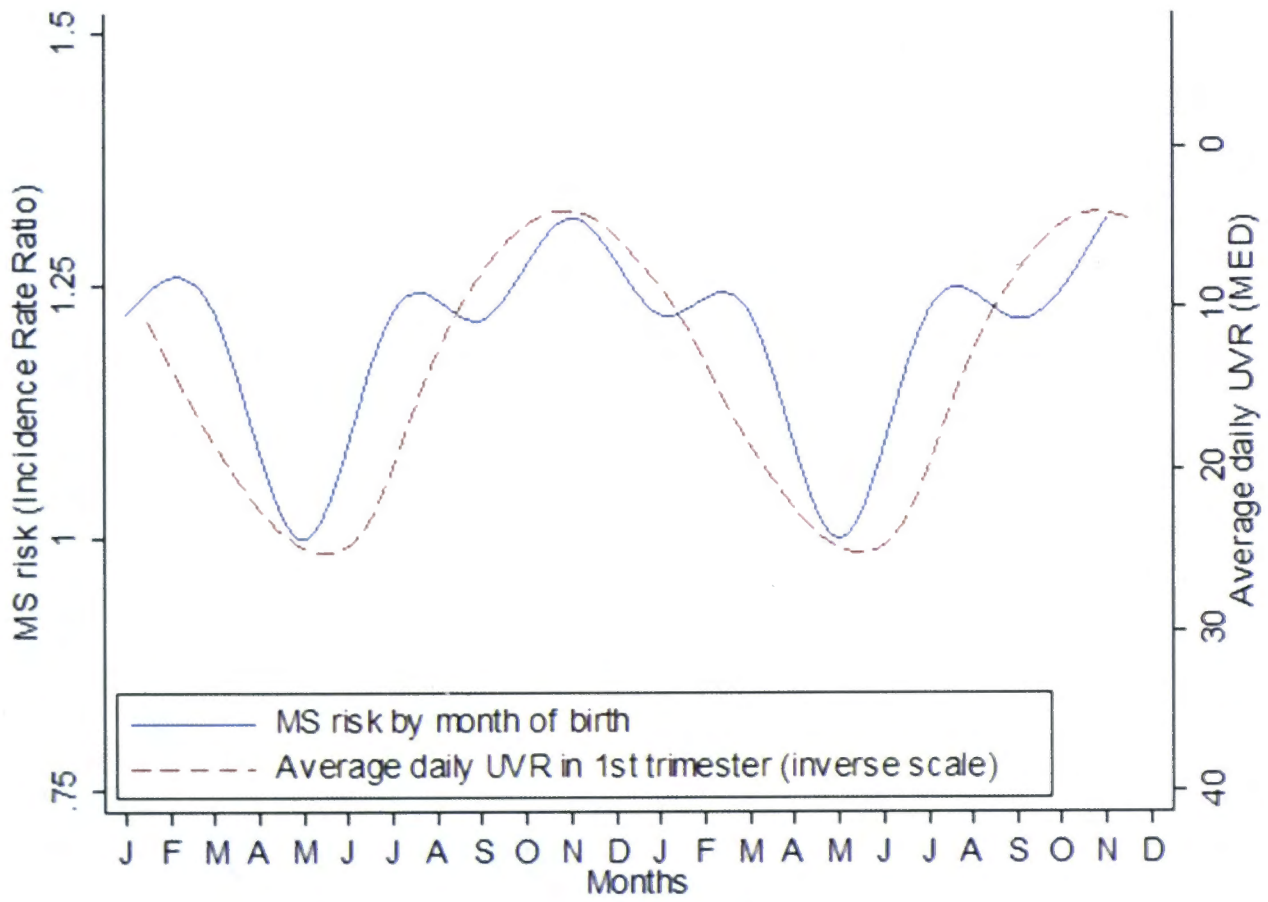


Figure 7.8: MS risk by month of birth (two-monthly periods), shown together with monthly averages of daily ambient UVR in first trimester of gestation on an inverse scale. Time interval is two annual cycles.

These results are consistent with the stronger maternal over paternal 'parent-of-origin' effect seen in familial MS [225, 226], and consistent also with Barker's 'foetal origins hypothesis' for subsequent human disease (see Chapter 2). The findings add to other work showing that maternal exposure to ambient UVR during pregnancy might influence subsequent health in both human and animal offspring [484, 671-673]. For example, in a NZ birth cohort, human infants whose mothers were exposed to peak

sunshine during the first trimester were significantly heavier at birth than infants whose mothers experienced trough levels of sunshine during the same trimester [673]. In the UK, maternal ambient UVB exposure in the third trimester has been positively related to bone mineral density and content at age nine years in children [674].

There is now growing interest in the role of maternal vitamin D deficiency in pregnancy and the development of CNS and immune disorders, particularly schizophrenia [484], type 1 diabetes and other disorders [451, 452] (see Chapter 2). A recent study in the US of gestational vitamin D and risk of MS in offspring suggested that MS risk was lower among women born to mothers with high (fortified) milk or vitamin D intake during pregnancy [675]. Although active-vitamin D concentrations in the mother increase substantially during pregnancy, early foetal supplies are lower [669, 676] and directly dependent on the mother throughout gestation [677]. By the time of delivery, maternal and infant cord serum 25(OH)D concentrations are highly correlated [469]. However, much remains unknown, leading to large international variations in maternal vitamin D monitoring and supplementation during pregnancy [451, 678, 679].

It has been previously proposed [477] that maternal vitamin D deficiency, a problem for some dark-skinned women migrating to regions with low ambient UVR, such as the UK, might explain the increase of disease seen among second generation migrants in such locations [454]. The findings in this chapter are consistent with this explanation. Because season of birth has previously also been related to the clinical course and disease progression of MS [643, 644, 680], it is possible that early life exposures determine not only onset of disease but also resistance to the demyelinating process of MS; the mechanisms involved could include neurological or immunological factors.

A strength of using prenatal UVR exposure in this study, further to those study strengths already discussed, is that it was a prospective exposure whose levels were heterogeneously distributed among the MS cases because of the large variation in UVR linked to month and region of birth across Australia (UVR levels do not change significantly year to year, even over decades, P. Gies, pers. commn). This quantitative exposure factor had both a seasonal and a regional component relating to the individual MS-case data; the UVR parameter was also by monthly, rather than two-monthly, periods as was necessary for timing of birth, increasing analytic power and sensitivity.

However, the regional ambient UVR levels were not necessarily those experienced by each mother—a limitation was that this exposure factor could not take into account individual behaviour or skin pigmentation, nor could it account for concurrent dietary vitamin D intake. Nevertheless, as recently shown for a longitudinal birth cohort in the UK, ambient erythemal UVR levels during pregnancy can be validly used to indicate maternal vitamin D status, and ambient UVR can be used as an instrumental variable for causal analysis of subsequent vitamin D-dependent disorders in offspring [661].

Similarly to the timing-of-birth pattern, the inverse association between first trimester UVR and MS risk persisted after adjustment for region of birth, sex and decade of birth, and effect modification by these factors was also not present. However, other factors, such as nutrition or physical activity, that could be associated with prenatal exposure to UVR during pregnancy and that could also determine MS risk could not be similarly controlled for, even though these would likely obscure the observed patterns rather than create them.

Importantly, both here and in the northern hemisphere CCSG MS study, the pattern of timing of birth was not smoothly sinusoidal but showed a few months of particularly altered risk of MS. This may suggest an underlying seasonal factor that is required to reach a threshold before effects can be measured [324]. Consistent with this, when different levels of first trimester UVR were examined in this chapter, a threshold effect was clearly observed; that is, risk was specifically increased in the lower levels of exposure, below a monthly average of 20 MED units per day. Interestingly, this threshold level is reached from November to February in Brisbane, QLD, but only in January in Hobart, TAS.

Taken together, all of these results provide evidence that maternal UVR exposure and/or vitamin D status are likely to be a contributory determinant of MS and linked to a critical period in or close to the prenatal first trimester. Whether the critical *foetal developmental* period is as early as this (first trimester) is not clear—if the action of UVR is vitamin D mediated, a lag period would be involved for the production of vitamin D. Recent Australian work has shown a lag of one and a half months, for example, between higher ambient UVR levels and higher vitamin D concentrations at the population level [327]. This would equate to a critical period for vitamin D in the second trimester. First or second trimester vitamin D concentrations might be particularly important in the development of the CNS because during early embryonic development, VDRs are expressed in the neuroepithelium and later in the subventricular zone [479].

Myelination occurs later; even in mid-gestation (19 to 24 weeks) cortical axonal tracts are not yet myelinated [681], with major myelination of several areas occurring as late as 29 to 39 weeks [682].

The *in utero* development of immune central tolerance occurs in the first trimester (see Chapter 1) and vitamin D has immunomodulating properties (see Chapter 2). The first trimester is also a sensitive period with regard to prenatal thymocyte differentiation, with animal studies showing that chemicals such as dioxin can alter this process, disrupt the development of central tolerance, and lead to increased auto-reactive peripheral T cells [683]. Further, indirect effects of vitamin D should be considered. For example, vitamin D can down-regulate the cytokine IL-6, an important mediator of the adverse effect of maternal infection during pregnancy on neural development in the foetus [400, 684, 685]. Further work to confirm timing of the observed effects during the prenatal period is now needed.

The independent effect of region of birth on MS risk, as shown by the strong residual association with birthplace after adjusting for first trimester UVR (see Section 7.3.4), is consistent with birthplace possibly acting as an indicator of *postnatal* exposure to UVR. Indeed, given that many Australians apparently remain in their birthplace region for many years (see Chapter 6), this factor may be a good marker for postnatal sun exposure linked to long-term residence.

Taken together, the results in this chapter indicate that the possible beneficial effect of UVR exposure may occur at more than one life stage; this can be illustrated using Goodin's (2009) multi-stage causal cascade hypothesis for MS pathogenesis as a basis, wherein perinatal and postnatal factors suggested by the results in this chapter are shown highlighted in green in Figure 7.9. As this life-course causation model proposes for MS pathogenesis, a series of environmental factors seems most likely to precipitate eventual MS in genetically susceptible people, including vitamin D deficiency very early in life [229]. The low prenatal maternal UVR exposure found in this thesis to be associated with MS risk is consistent with this proposed 'very early environmental influence' with a significant 'maternal effect' near the time of birth. Infections, or other factors, may act subsequently in childhood or adolescence, while other as yet unidentified environmental factors may act prior to onset of MS symptoms [229] (see Chapter 3); the latter may now include lower postnatal UVR also (Figure 7.9).

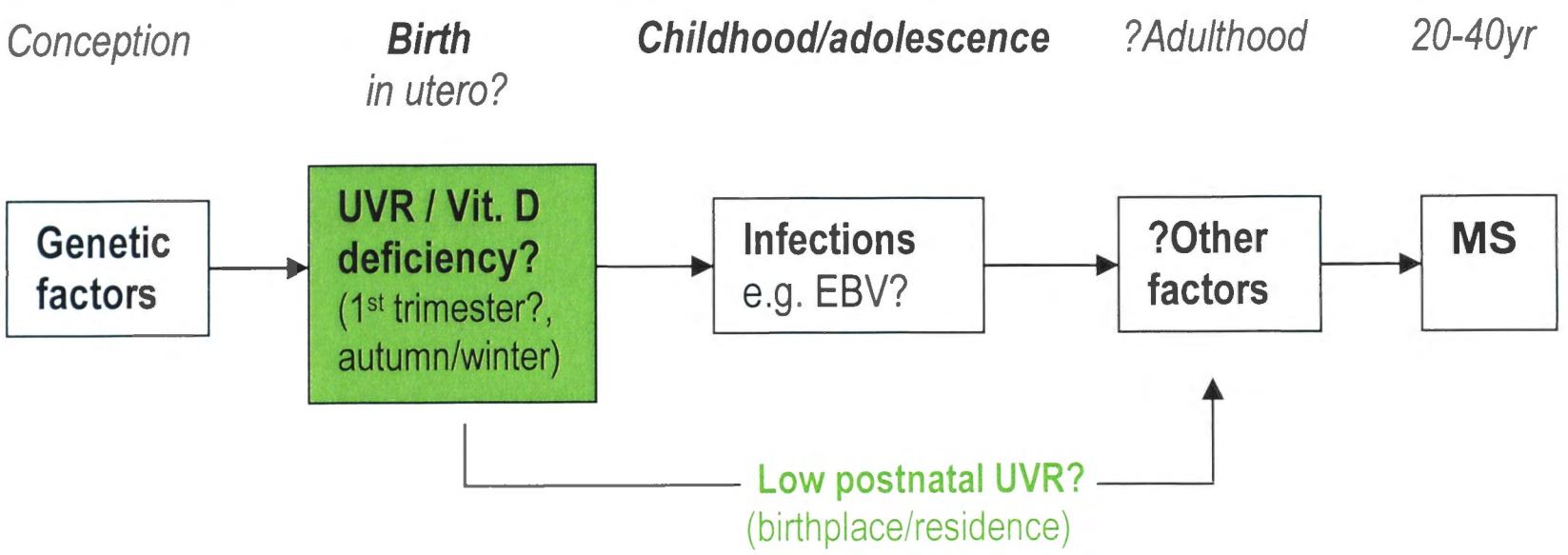


Figure 7.9: Possible causal pathway leading to MS, showing timing of putative environmental factors. (Figure based on 'causal cascade' model of Goodin, 2009 [229] [see Chapter 3, Figure 3.5] and results in this chapter.)

7.5 Conclusions

The finding of a timing-of-birth pattern among MS cases suggests that at some time during the perinatal period, a critical window of time may exist when an environmental factor that is also fluctuating seasonally may be acting to influence risk of MS onset later in life. Just when this critical period may exist, and for how long, cannot be determined directly from the timing of birth pattern, but the presence of such a pattern suggests an environmental factor, which is also seasonally fluctuating, acting during this critical developmental period. Symmetrical patterns for MS risk in both global hemispheres now reinforce this suggestion and indicate a need for further detailed analysis of possible underlying exposures and mechanisms.

Additional independent analyses undertaken with first trimester UVR in this chapter underline that prenatal UVR, and particularly first trimester UVR, is a possible candidate as an early life, component causal, environmental factor in the pathogenesis of MS. Regional ambient levels of first trimester UVR accounted for the timing-of-birth pattern found in Australia but not for the total variation in MS risk; contributory low postnatal levels of UVR/vitamin D in the birth and/or residence region may be indicated by results in this chapter. The action of first trimester UVR may be mediated through lack of vitamin D during autumn and early winter, consistent with accumulating evidence for the importance of vitamin D during the prenatal period.

Low prenatal ambient UVR exposure, perhaps mediated by vitamin D deficiency, thus appears to be a good candidate for the early life determinant (indicated by timing of birth) in MS aetiology in Australia, and low postnatal UVR exposure may also be important (see Figure 7.9).

7.6 Postscript

Chapters 5 to 7 have examined a timing-of-birth pattern in MS in Australia using a longitudinal dataset constructed from the available MS case survey conducted in 1981. Associations between timing of birth, prenatal (maternal) ambient UVR exposure, region of birth and MS risk have been shown and a putative determinant of MS identified and discussed.

The next chapter, Chapter 8, will investigate another possible influence on MS pathogenesis that may act subsequently to low prenatal UVR exposure (see Figure 7.9) and increase (or decrease) MS risk to ultimately produce sufficient cause to precipitate MS [1, 6, 189]. In Chapter 8, the possible role of childhood infections in MS will be investigated by independent analysis of birth order and sibling relationships in MS cases, using the original individual-level 1981 Australian MS survey (unit-record) case data.

CHAPTER 8

ORDER OF BIRTH AND MS RISK IN AUSTRALIA: ANALYSIS OF BIRTH ORDER OF MS CASES AS POSSIBLE EVIDENCE FOR EFFECTS OF CHILDHOOD INFECTIONS ON MS RISK

8.1 Introduction

As outlined in Chapter 1, MS is a chronic inflammatory demyelinating disorder in which the immune system is activated to attack the white matter of the CNS, leading to severe disability. Despite numerous population studies, MS aetiology is unknown but is thought to be a complex, multifactorial interplay of both genetic and environmental factors [186, 189, 193] (see Chapters 2 and 3). Temporal changes in MS incidence in many parts of the world suggest significant environmental influences on MS causation, as do also geographic latitude gradients in prevalence or incidence, and seasonal variation. The relatively high rate of non-concordance between monozygotic twins, at least 60%, further emphasises the probable environmental contribution to MS aetiology [196, 202] (see Chapter 2).

MS is generally believed to be an immune-mediated disorder (see Chapter 1); however, the sequence of environmental events that initiates the disorder remains largely unknown [150]. A possible 'cascade' model of MS causation in terms of timing of different factors has been proposed by Goodin (2009) [229], in which a lack of vitamin D around the time of birth has been postulated as one factor. Consistent with this, Chapters 5 to 7 of this thesis have presented evidence that environmental UVR at or before birth may explain the timing-of-birth cycles in MS now evident in both northern and southern hemispheres.

However, other strong evidence points to other environmental factors also being implicated in MS aetiology, particularly infectious agents. Infections may act to trigger the autoimmune processes leading to either initial disease onset or to ongoing recurrences, particularly for the most common type of MS showing a relapsing-remitting clinical pattern [487, 490]. Alternatively, and paradoxically, childhood infections may *protect* against subsequent autoimmune disease such as MS [524]. In this case, the timing of such infections appears to be crucial to subsequent effects on

MS onset. That is, when infection has been found to be protective for such disease, it has often occurred early in childhood [3]. A later-than-normal exposure to an otherwise common infectious agent, such as a virus, may predispose the child or adolescent to subsequent MS [487, 545] (see Chapter 3).

Birth-order and family-size studies can be used to measure the timing and intensity of early life exposure to childhood infections, particularly for a late-onset disease such as MS, when recall of past infections would be unreliable. The fact that childhood infections are often asymptomatic may also preclude their accurate measurement. Sibship structure of the case families, in terms of birth order of the case in relation to the total sibship size, can be used as a surrogate, or proxy, measure of timing and degree of exposure of childhood infections. For example, early birth order (i.e. being one of the older siblings in the sibship) and/or small family size are considered to lessen and delay exposure to the hypothesised infectious agents from early childhood to a later age [556, 686], thereby possibly increasing subsequent MS risk.

However, evidence from birth-order studies for MS is conflicting; most of the previous studies have also been conducted in the northern hemisphere. An *early* birth-order pattern in cases, suggesting a *lack* of early infectious exposure, has been found by some [687-690], while others have found a *late* birth-order pattern [691, 692]. Hernan and colleagues (2001) found an excess of first-born MS cases (i.e. first born had higher MS risk) only in larger families (four or more siblings) in a large nested case-control study [616], while several other studies found no evidence of any association between birth order and MS risk [686, 693-700].

A more recent large longitudinal cohort study by the CCSG also concluded 'no support for the (hygiene) hypothesis that having older siblings protects against MS' [554]—that is, no support for an early birth-order effect in MS cases. In that study, an opposite, slightly higher MS risk was shown if born in a later birth-order position, for sibships greater in size than seven siblings (i.e. a late birth-order effect in MS cases in larger sibships). However, this result was deemed by this research group to be 'more likely due to a cohort effect resulting from increasing MS incidence' rather than being any real effect. Another large population-based cohort study of MS in Denmark also found no association with the number of older siblings or any other sibship characteristic [556].

In one of the few southern hemisphere studies, MS risk in relation to the number of younger siblings, rather than birth-order position, was investigated in TAS, Australia. In this case-control study, a protective effect of longer exposure to a greater number of younger infant siblings, independent of birth order, was shown—adjusted ORs 0.57 (95% CI 0.33, 0.98), 0.40 (0.19, 0.92), and 0.12 (0.02, 0.88)—for, respectively, 1-<3, 3-<5 and ≥5 infant-years contact, compared with <1 infant-year, in the first six years of life [553]. However, in a large case-control study of both younger and older siblings in Sweden, a protective effect of having siblings per se was reported. The ORs for developing MS for people with siblings, compared with those with none, were 0.80 (95% CI 0.70, 0.92) for three or more younger siblings, and 0.83 (0.72, 0.96) for three or more older siblings [555]. Thus, the possible effect of older and/or younger siblings on subsequent MS risk is undecided, but these latter studies that include younger siblings may still suggest a protective effect of early microbial exposure by contact with both older and younger siblings, broadly consistent with the hygiene hypothesis (see Chapter 3).

With regard to methodology, most previous birth-order studies can be categorised into two main types—those employing population controls (e.g. case-control studies matching for age and sex) and those studying cases only and employing theoretical techniques such as 'Greenwood-Yule'. This technique measures observed departures from an expected, random birth-order distribution per family size [701, 702] and is the method used by many workers, including Sadovnick and colleagues for the CCSG group [554]. In this technique, the expected birth-order position of the MS case within a sibship of given size is calculated under the assumption of equal probability of the birth-order positions (i.e. uniform distribution); the null hypothesis is that there is no relationship between birth-order position and occurrence of MS within the sibship [693]. A few studies, reviewed by Ahlgren and Andersen (2005) [693], have used both techniques [686, 690, 693, 696-697, 700]; all of these studies showed agreement between the two techniques (i.e. no association between birth order and MS risk) except for the study by Zilber et al. (1988)—this showed no association by the Greenwood-Yule technique but an association between low (i.e. early) birth order and MS risk in the case-control study [690]. James (1984) also reviewed several earlier MS birth-order studies and the biases then inherent in the theoretical technique, pointing out that decreases in population family sizes over time would bias towards later birth-order positions, if family (i.e. sibship) size was not controlled for when using the theoretical technique. He concluded that the unbiased early MS studies, mainly case-

control studies, all showed an early birth-order effect in MS cases (i.e. earlier birth-order position than expected), consistent with the hygiene hypothesis [688, 689].

As described in Chapter 5, the 1981 Australian MS Survey was a comprehensive population-based, cross-sectional data collection from identified MS patients in Australia on the national census day, 30 June 1981. This cases-only dataset comprised patients born in Australia or overseas between 1897 and 1969 (ages 12 to 84 years in 1981). Therefore, theoretical techniques for birth-order analysis of MS cases were applicable in order to investigate the possibility that childhood infections played a part in MS aetiology in Australia. As in Chapters 5 to 7 of this thesis, the data were restricted to those MS cases born between 1920 and 1950 (inclusive) for analysis in order to minimise possible selection biases.

In this chapter, birth-order position among the MS cases in the available dataset will be investigated, to determine whether a lack of early microbial exposure could apply. That is: Does MS occur more frequently in those exposed relatively *less or later* in life to common childhood infections because of being one of the *older* children in their sibship? The research question is:

- **Is there an early birth-order effect in Australian MS cases?**

That is, is there an association between birth-order position and MS risk? And, specifically, are earlier-born siblings (i.e. older children in the sibship) at higher MS risk? For example, are MS cases more likely to be born earlier in their sibships than would be expected by chance? That is, is there a *shift* in the mean birth-order position in MS cases compared with the expected, by chance, mean birth-order position of each sibship size? Further:

- **Are MS cases more likely to be in the *first-born* position than expected by chance? And/or less likely to be in the *last-born* position?**

Previous studies, both case-only and case-control, have varied widely in methodological rigour and many have suffered from small sample sizes. In addition, few of the previous studies (with the exception of Sadovnick et al., 2005 [554]) appear to have taken sufficient account of differing birth-order distributions among different-sized sibships. That is, most of the case-only studies have given an estimate of mean birth order over all sibship sizes combined, without, for example, any weighting for

possible effects of sibship size, despite the importance of sibship size being recognised by some early workers in birth-order studies in other disease fields (e.g. Gregory, 1958 [703]).

In the present chapter, all analyses are conducted separately at first for each sibship size (usually up to 11, and sometimes up to 16, siblings in total) to account for possible differences in birth-order distributions. Both parametric and non-parametric tests are initially employed to analyse the birth-order position among MS cases. The first-born position, in particular, is also analysed independently to provide an additional birth-order parameter. Next, to explore the effect of sibship size on birth order, these data are pooled over different-sized sibships, but only after taking account of possibly differing birth-order distributions for each sibship size; for the parametric tests, two different weighting models are employed to weight each sibship size before pooling. Finally, the independent results of all tests (including the last-born birth-order position) are compared and an attempt is made to interpret the relationships between the different parameters measured.

8.2 Methods

8.2.1 Source data

As described in Chapter 5, a nationwide MS prevalence survey, coordinated by Professor J. McLeod, Discipline of Medicine, University of Sydney, was carried out on 30 June 1981 to coincide with a national Australian census. MS cases were recorded by state; procedures for inclusion in the survey dataset were standardised by neurologists coordinating the survey. Clinical classification and validation of cases was described in Hammond, McLeod and colleagues' various publications [301, 646-648] (see Chapter 5).

The unit-record MS-case dataset received from the McLeod research group was checked, cleaned and further verified by this candidate, as described in Chapter 5. Of the MS cases, 70% were Australia born, most of the remainder having been born in the UK and Ireland [304]. The data were then restricted to those MS cases born between 1920 and 1950 (inclusive) (see Section 5.3.3). Importantly for this chapter, restriction of the year of birth to this three-decade period also achieved minimal variation in sibship

size over time—an assumption required for the theoretical analysis technique [556, 688, 693]. That is, there was no significant difference in median sibship size over the three decades between 1920 and 1950 ($p=0.052$; Kruskall-Wallis non-parametric chi^2 test). In addition, restriction to 1920- to 1950-born cases would have ensured that sibships were complete by the survey date in 1981 (at least 30 years after birth of the index MS case), a further requirement for this technique.

Relevant demographic data variables for birth-order analysis of the MS cases included date of birth, sex, place in sibship (i.e. '1', '2', '3' ...=first born, second born, third born ...), sibship size (i.e. total number of siblings in the case family), and whether the MS case was a twin. Single-child sibships (8.2% of cases) for which the theoretical technique is not applicable [693, 703] were excluded from birth-order analysis (these cases being both 'first born' and 'last born' and providing no data for this analysis). Twins (2.5% of cases) were also necessarily excluded because a meaningful birth-order position could not be assigned.

The resulting unit-record dataset for birth-order analysis comprised 1,840 MS cases in sibships of two or more siblings who were born between 1920 and 1950 and present in Australia on 30 June 1981. These cases comprised 1,275 females and 565 males, and 82% of the cases were of the relapsing-remitting clinical course type at onset. The average age of onset was 32.3 $\pm$ 8.8 (s.d.) years. Thus, the 1,840 MS cases represented 1,840 sibships, each with a known total number of siblings in the sibship and the known order of birth of the MS case in the sibship. By tabulating birth order (of the case) by the total number of siblings in the sibship (=sibship size) as in Table 8.1, the data frequencies for the statistical analyses could be gained. The dataset for analysis thus comprised sibship sizes ranging from two to 16 siblings in total; MS cases in these sibships were born in birth-order positions ranging from first to 15^{th}, these birth-order positions being distributed among the sibships in the frequencies shown in Table 8.1.

Table 8.1: Frequencies of birth-order positions of MS cases by sibship size (1981 MS Survey cases born 1920 to 1950; n=1,840)

Sibship size	Order of birth (position in sibship) of MS case																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
2	221	223															444
3	184	174	143														501
4	97	100	76	69													342
5	43	40	51	42	39												215
6	30	23	18	16	19	15											121
7	13	15	8	10	8	12	22										88
8	2	2	5	6	6	6	9	13									49
9	1	0	3	4	3	6	0	3	2								22
10	0	5	3	2	1	2	4	2	0	2							21
11	2	0	1	3	1	0	1	2	2	2	4						18
12	0	0	1	0	1	2	1	1	0	0	4	1					11
13	0	1	0	0	0	1	0	1	0	0	1	0	0				4
14	0	0	0	0	0	2	0	0	0	0	0	0	0	0			2
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0
16	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	2
Total	593	583	309	152	78	46	37	22	4	5	9	1	0	0	1	0	1840

8.2.2 Statistical analysis

Mean birth order

The data consist of a number of sibships, each with an MS case born first, second, third …, or last (i.e. any position up to the total number of siblings in the sibship). If there is no association between birth order and MS, the birth-order positions of the various MS cases in the sibships should be distributed randomly. A statistical test for an association between birth order and MS can be constructed by comparing the observed birth-order distribution with the expected birth-order distribution.

For any particular sibship size, the mean birth-order position of the MS cases (mean birth order [MBO]) is calculated from the frequencies of each birth-order position in that sibship size. That is, for a sibship size s (=total number of siblings in the sibship):

$$MBO_{MSCASES} = \frac{\sum_{x=1}^{s} f(x)x}{\sum_{x=1}^{s} f(x)}$$

where x=birth-order position of the cases 1, 2, 3 … s, and *f* (x)=frequency of each MS case with birth order x in that sibship size. The observed MBO may thus be less (i.e. cases are earlier born), equal to, or more (i.e. cases are later born) than would be expected by chance.

Analysis by sibship size

The observed MBO was compared with an expected MBO for each sibship size, thus 'conditioning on sibship size'. The tests were carried out separately over the sibship-size range of two to 11 siblings (n=1,821 sibships). Expected MBOs for a given sibship size were calculated by assuming, under the null hypothesis, an equal chance of the MS case being born in any birth-order position in the sibship (i.e. equal frequencies of each MS-case birth-order position in the sibship). The observed distribution of MBO was then compared with the expected distribution using the central limit theorem (CLT) [654].

For example, for a sibship of four siblings, there are four birth-order positions and the expected MBO of the four birth-order positions—one, two, three and four—is 2.5 (Table 8.2). This means that an MBO of 2.5 would be expected if the birth-order positions of all of the MS cases (of sibship size four) were randomly distributed. The variance of this expected birth-order distribution (Variance $_{EXPD}$) is 1.25, which can be calculated from the general formula:

$$\text{Variance}_{EXPD} = \frac{(s^2-1)}{12} \qquad (1)$$

where s=total number of siblings in the sibship. For the example sibship of four siblings, the variance of the expected birth-order distribution is 15/12 (=1.25) (Table 8.2).

Comparing the observed sampling distribution of the means with the expected sampling distribution under the CLT, the test statistic, Z, follows a standard normal distribution and is given by:

$$Z = \frac{\text{Observed MBO} - \text{Expected MBO}}{\sqrt{(\text{Variance}_{EXPD}/n)}} \qquad (2)$$

where n is the total number of cases of that sibship size. For example, for the sibship size of four siblings (n=342 sibships), the difference between the observed and expected MBO is 2.34 – 2.50 (= –0.16), resulting in a *negative* Z-score of –2.61 (Table 8.2). This means that for this sibship size there was an observed 'shift' in MS cases toward *earlier* birth-order position than expected. For the sibship size of eight siblings (n=49 sibships), conversely, there was a *positive* Z-score of +3.58 and thus a shift toward *later* birth-order position than expected.

The test statistic, Z, was calculated similarly for sibship sizes of up to 11 siblings, since the frequencies of sibships larger than this fell below the considered limit of accuracy (n=15) of the test [654] (Table 8.2). The 5% level of significance was used to assess each test statistic, Z-scores between $\pm$1.96 being insignificant (two-tailed test).

Table 8.2: Calculation of Z-scores for MBO of MS cases in sibships of two to 11 siblings (n=1,821; calculated values shown to two decimal places only). (Sibship examples used in text shown in bold)

Sibship size (s)	Observed MBO	Expected MBO	Variance $_{EXPD}$ (equation 1)	Z-score (equation 2)	Total cases of sibship size (n)
2	1.50	1.5	0.25	+0.09	444
3	1.91	2.0	0.67	–2.24	501
4	**2.34**	**2.5**	**1.25**	**–2.61**	**342**
5	2.97	3.0	2.00	–0.29	215
6	3.13	3.5	2.92	–2.37	121
7	4.24	4.0	4.00	+1.12	88
8	**5.67**	**4.5**	**5.25**	**+3.58**	**49**
9	5.41	5.0	6.67	+0.74	22
10	5.14	5.5	8.25	–0.57	21
11	7.06	6.0	10.00	+1.42	18
12	8.27	6.5	Statistical limit of test*		11
13	6.75	7.0	-	-	4
14	6.00	7.5	-	-	2
15	-	8.0	-	-	0
16	12.50	8.5	-	-	2

* Frequencies less than 15 are below the generally considered limit of accuracy of the Z-test [654].

Pooled sibship sizes using two weighting models

The statistics for sibship sizes between two and 11 siblings (inclusive) were then pooled by weighting the {Observed MBO – Expected MBO} difference (=delta, ∂), for each sibship size, to take account of differing distributions for each sized sibship. Two models for ∂ were considered.

In model (1), the {Observed MBO – Expected MBO} difference (=∂_1) was assumed to be the same for each sibship size. In this model, the weight for each sibship size was given simply by the reciprocal of the expected birth-order variance:

$$Wt_1 = 1 / \text{Variance}_{EXPD} \qquad \text{Model (1)}$$

In model (2) the difference, ∂_2, was allowed to vary linearly with sibship size, s, thus:

$$\partial_2 = \frac{\text{Observed MBO} - \text{Expected MBO}}{s}$$

The weight for each ∂_2 for each sibship size was then given by:

$$Wt_2 = s^2 / \text{Variance}_{EXPD} \qquad \text{Model (2)}$$

For each model, the overall, weighted difference between observed and expected MBO was then converted to a Z-score given by:

$$Z = \frac{\text{Weighted estimate of } \partial}{1/\sqrt{(\Sigma \text{ weights})}}$$

for comparison with a standard normal distribution, as previously.

As for the individual sibship-size calculations, only those sibships with frequencies of at least n=15 (Table 8.2) were considered to be within the limits of accuracy of the test; therefore, sibship sizes of up to 11 siblings only were pooled.

Median birth order—non-parametric sign test

Analysis by sibship size

A second method used for testing for birth-order differences was a non-parametric sign test. This test required no assumptions about the birth-order distributions over the different sibship sizes. In the test, for each sibship size, the number of cases where the birth order was above, or below, the expected *median* birth-order position was determined. The null hypothesis here predicted an equal number of cases above and below the median, assuming an equal chance of being born in a higher or lower position than the median birth-order position.

For example, for the three-siblings sibship size (Table 8.1), where the expected median birth order is 2.0, there are more MS cases (=184) born below the expected median

than above (=143); this would result in a *lower* observed median birth order than would be expected by chance.

Because cases where the observed birth order was *equal* to this median did not contribute to the test (='ties', where difference between observed and expected medians was zero), the total frequencies first needed to be adjusted to exclude these cases. For the sibship of three siblings, the 174 cases of observed birth order of 2.0 exactly equal the expected birth order (Table 8.1); the adjusted n for this sibship size is then 327 (=501 – 174).

The remaining cases were then either higher or lower than the median, and the number of 'positives' (=frequencies of cases *above* the median birth order) for each sibship size constituted a sample from a binomial distribution with n samples; the probability of being higher than the mean was 0.5 under the null hypothesis [654]. The null hypothesis could then be evaluated using the test statistic, Z_+ , which followed a standard normal distribution providing that n was sufficiently large [654].

The test statistic, Z_+ , for a given sibship was given by:

$$Z_+ = \frac{D - (n/2)}{\sqrt{(n/4)}}$$

where D is the total number of cases with birth position above the median order, and n is the adjusted total number of cases with this sibship size. Sibship sizes up to 11 siblings (i.e. where n >15) could be evaluated individually by this technique.

Pooled sibship sizes

With this non-parametric technique, the 'positives' frequencies of *all* sibships up to 16 siblings could simply be pooled *prior* to calculating the test statistic, thus using all of the available case data in the pooled groups.

First-born fraction

Analysis by sibship size

As another way of investigating birth order of the MS cases and to support the findings so far, the specific 'first-born effect' among these cases was investigated by the parametric CLT technique as used for MBO, again initially stratifying by sibship size. In these tests, 'first-born' MS cases were compared with 'non-first-born' (i.e. with the remaining birth-order categories in each sized sibship). Again, as for MBO, the null hypothesis was that there was an equal chance of the MS case being born in any birth-order position in the sibship, the expected mean probability of being 'first born', for each sibship size, s, being given by 1/s. The expected proportion, or fraction, of first-born cases was then calculated assuming a mean of 1/s, and variance:

$$\frac{1}{s}(1 - 1/s)$$

Denoting the 'fraction of first-born cases' as FFb, the test statistic, Z_F, could be calculated, using the CLT under the same assumptions as previously for MBO:

$$Z_F = \frac{\text{Observed FFb} - \text{Expected FFb}}{\sqrt{(\text{Variance}_{EXPD} / n)}}$$

where n is the total number of cases with this sibship size. The Z-score was then compared against the standard normal distribution as previously.

For example, for the sibship of four siblings again, the fraction of first born to non-first born, FFb, is given by 97/342 (Table 8.1) (=0.28), which is *greater* than the expected fraction of 0.25, thereby resulting in a *positive* Z-score. A positive Z-score here indicates that there is an *excess* of first-born MS cases, relative to non-first born, in that sibship size. For the sibship of eight siblings, the FFb is 2/49 (Table 8.1) (=0.04), which is *less* than the expected fraction, 0.125, resulting in a *negative* Z-score and indicating a *deficit* of first-born MS cases in this sibship size.

Pooled sibship sizes

Pooling of sibship sizes up to 11 siblings was again carried out using two different weighting models. Model (1) was similar to that previously used for MBO. That is, the difference between Observed FFb and Expected FFb (∂_1) was assumed to be the same for each sibship size. Model (2) assumed that the difference between Observed FFb and Expected FFb (∂_2) increased with family size, s, thus:

$$\partial_2 \quad = \quad s \text{ (Observed FFb – Expected FFb)}$$

The weight for each ∂_2 for each sibship size (s) was then given by:

$$Wt_2 \quad = \quad 1 / s^2 \text{ Variance}_{EXPD} \qquad \text{Model (2)}$$

Interrelationships between parameters

The second analysis, specifically on first-born probability rather than mean or median birth-order positions, allowed later comparison of 'first-born' effects in particular, versus the more general 'early-born' effects investigated in the preceding subsections. That is, an *excess of first born* may be found to be consistent with a general *early-born* effect in some sibship sizes, supporting and adding to the overall birth-order information. In addition, subsequent independent analyses for the 'last-born' birth-order position were also included; these are detailed in Results, Section 8.3.3.

MS Excel 2000 software was used for all statistical analyses.

8.3 Results

8.3.1 Birth order—shifts in observed distributions

Mean birth order

Table 8.3 shows observed MBO and expected MBO, together with Z-scores indicating negative shifts (i.e. observed MBO < expected MBO) and positive shifts (i.e. observed

MBO > expected MBO) in birth order, by sibship size over the range two to 11 siblings (n=1,821).

Negative birth-order shifts (i.e. shifts toward *earlier* birth order of MS cases) were evident more often in the smaller-sized sibships; these were statistically significant for sibships of three, four and six siblings at the 5% level. Positive birth-order shifts (i.e. shifts toward *later* birth order of MS cases) were evident more often in the larger sibships of at least seven siblings, the sibship size of eight siblings reaching statistical significance (Figure 8.1).

Table 8.3: Observed MBO and expected MBO, together with Z-scores and p-values, for MS cases in sibships of two to 11 siblings, indicating significant shifts in birth order, by sibship size

Sibship size	Observed MBO[#]	Expected MBO	Z-score[#] for shift	p-value	n
2	1.50	1.5	+0.09	0.928	444
3	1.92	2.0	–2.24	0.025*	501
4	2.34	2.5	–2.61	0.009**	342
5	2.97	3.0	–0.29	0.772	215
6	3.13	3.5	–2.37	0.018*	121
7	4.24	4.0	+1.12	0.263	88
8	5.67	4.5	+3.58	<0.001**	49
9	5.41	5.0	+0.74	0.459	22
10	5.14	5.5	–0.57	0.569	21
11	7.06	6.0	+1.42	0.156	18
Total					1821

[#]Observed MBO and Z-score each shown to two decimal places only.
*Significant shift at p<0.05.
**Significant shift at p<0.01.

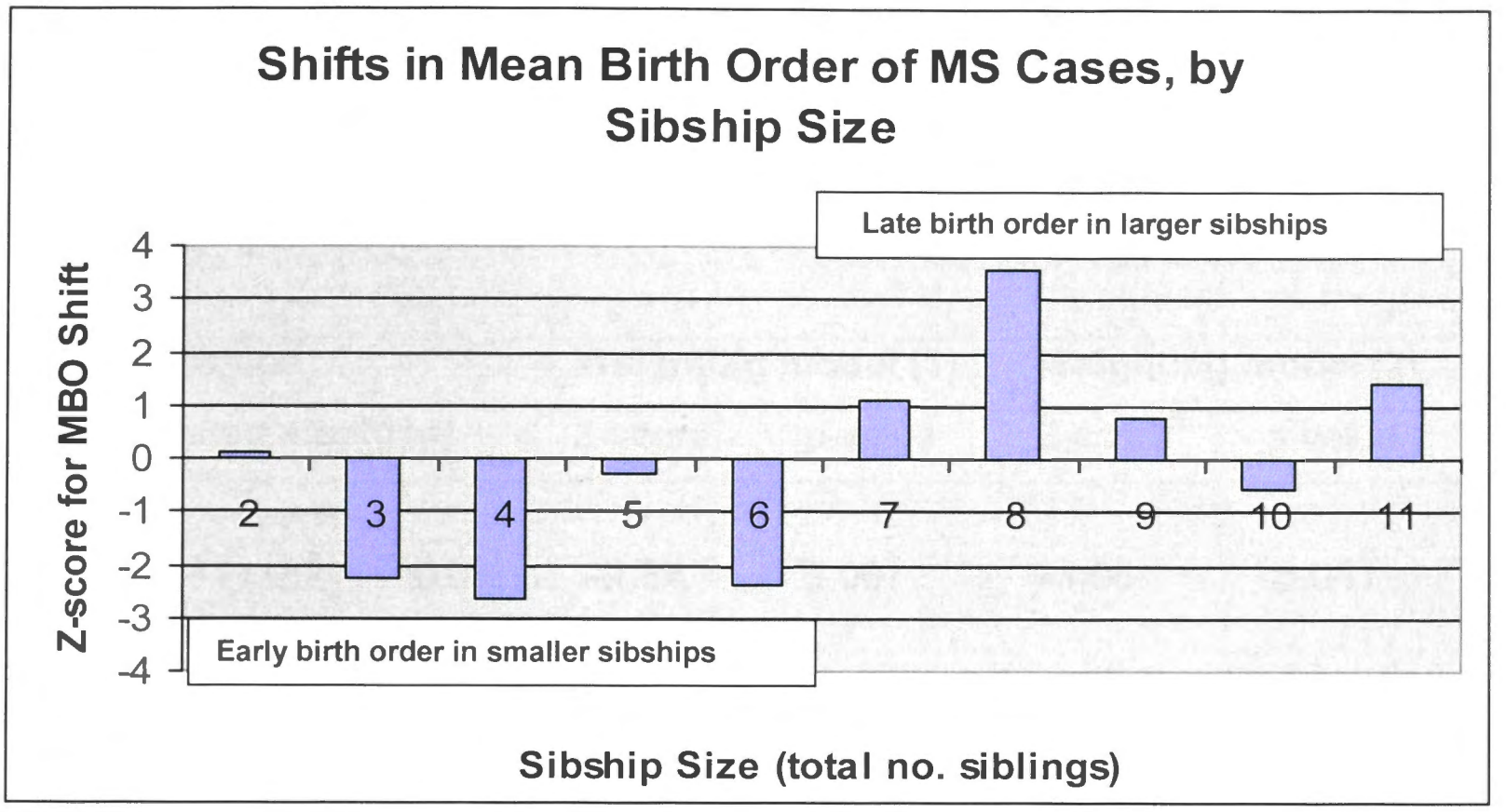


Figure 8.1: Shifts, as Z-scores, in MBO of MS cases in sibships of two to 11 siblings.

Pooled sibships

Pooling results over all sibship sizes from two to 11 siblings (n=1,821 sibships) gave non-significant or marginally significant shifts towards earlier birth order, depending on which weighting model was used for pooling (Table 8.4). That is, there appeared to be an *overall* trend toward *earlier* birth order of the MS cases in their sibships compared with expected; however, this was only of borderline significance.

Table 8.4: Shifts in MBO of MS cases for pooled sibships overall (two to 11 siblings, n=1,821), together with pooled smaller sibships (two to six siblings, n=1,623) and larger sibships (seven to 11 siblings, n=198), using two weighting models

Pooled sibships (n, % of tested sibships)	Weighting model (1)		Weighting model (2)	
	Z-score[#]	p-value	Z-score[#]	p-value
Overall 2–11 (1821, 100.0%)	–1.84	0.067	–1.95	0.051
Smaller 2–6 (1623, 89.1%)	–2.19	0.029*	–3.04	0.002**
Larger 7–11 (198, 10.9%)	+2.97	0.003**	+3.02	0.002**

[#]Z-score shown to two decimal places only.
*Significant shift at p<0.05.
**Significant shift at p<0.01.

When smaller sibships of two to six siblings were pooled (89.1% of tested sibships) and tested separately from larger sibships of seven or more siblings (10.9% of sibships), significant birth-order shifts in opposite directions resulted. In *smaller* sibships (two to six siblings; n=1,623 sibships) there was a significant shift in birth order of MS cases towards the *earlier-born* sibship positions, while in *larger* sibships (seven to 11 siblings; n=198 sibships), there was a statistically significant shift in birth order of MS cases towards the *later-born* positions (Table 8.4).

Median birth order

The non-parametric sign-test method also showed several negative birth-order shifts in the smaller sibship sizes, the three- and four-person sibships being statistically significant and the six-person sibship approaching significance (Table 8.5, sibship sizes tested separately up to 11-person sibships). The larger sibships again showed mainly positive birth-order shifts (Figure 8.2), the eight-person sibship reaching significance at the 1% level.

Table 8.5: Non-parametric sign test for differences between medians, and resulting Z-scores and p-values for birth order of MS cases, by sibship size (n=1,602). (Frequencies in italics have been adjusted for ties)

Sibship size	Expected median BO	Number +ve differences	Z-score[#]	p-value	n (*adjusted for ties*)
2	1.5	223	+0.09	0.928	444
3	2.0	143	–2.27	0.023*	*327*
4	2.5	145	–2.81	0.005**	342
5	3.0	81	–0.16	0.873	*164*
6	3.5	50	–1.91	0.056	121
7	4.0	42	+0.68	0.497	*78*
8	4.5	34	+2.71	0.007**	49
9	5.0	11	+0.69	0.490	*19*
10	5.5	10	–0.22	0.826	21
11	6.0	11	+0.94	0.347	*18*
12	6.5	7	-	-	11
13	7.0	2	-	-	4
14	7.5	0	-	-	2
15	8.0	0	-	-	0
16	8.5	2	-	-	2
Total					1602

[#]Z-score shown to two decimal places only.
*Significant shift at p<0.05.
**Significant shift at p<0.01.

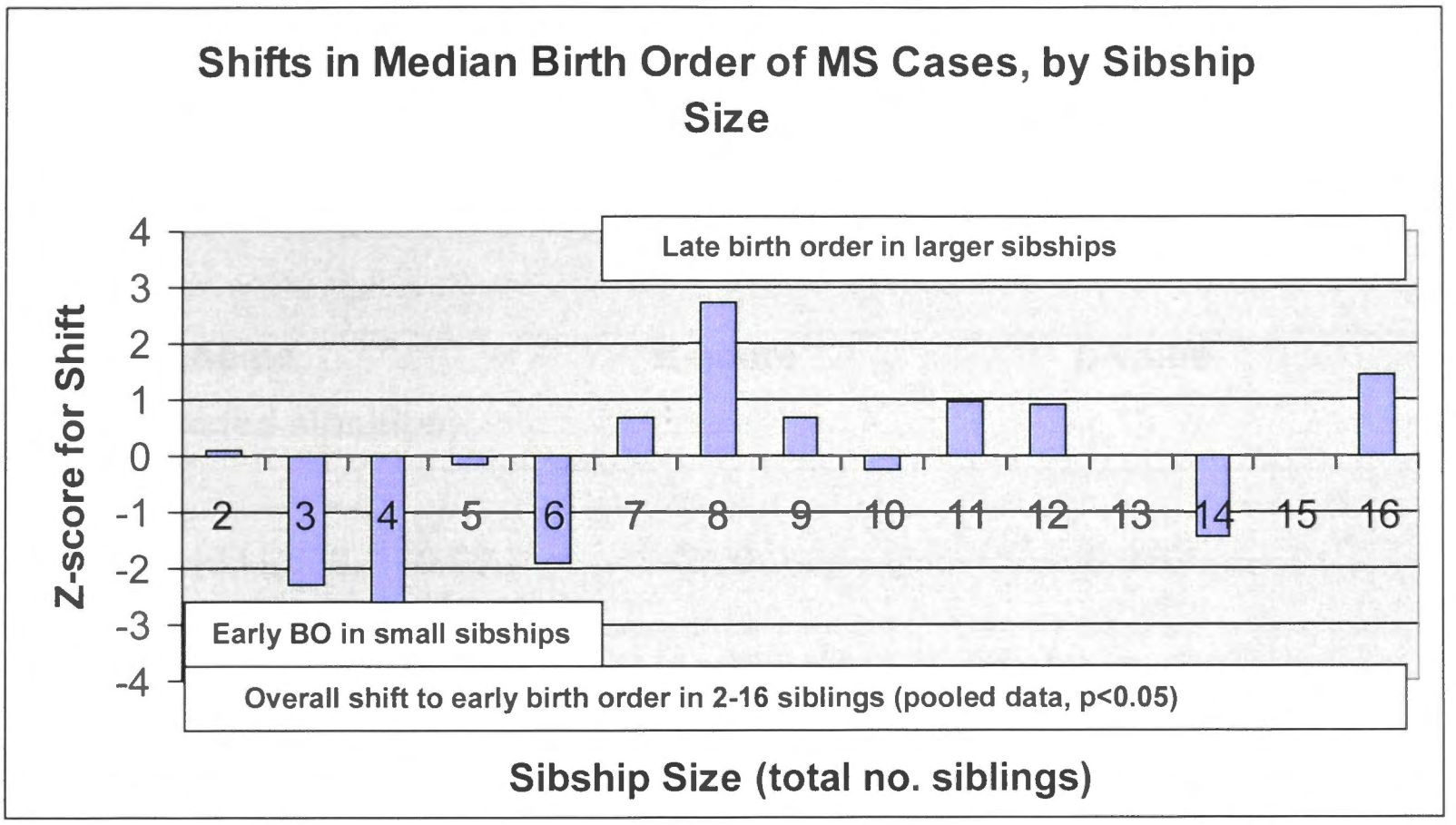


Figure 8.2: Shifts, as Z-scores, in median birth order of MS cases in sibships of two to 16 siblings (non-parametric test; n=1,602). (Statistically significant results of overall pooling of two to 16 siblings also indicated.)

Pooled sibships

Pooling over the entire range of two- to 16-person sibships (adjusted n=1,602 sibships) showed a statistically significant *overall* shift towards *earlier* birth order for MS cases (Table 8.6 and Figure 8.2). Separate pooling for smaller sibships (two to six siblings; adjusted n=1,398 [87.3% of tested sibships]) and larger sibships (seven to 16 siblings; adjusted n=204 [12.7% of sibships]) again showed significant birth-order shifts in opposite directions—towards lower (i.e. earlier) birth order in smaller sibships and towards higher (i.e. later) birth order in large sibships (Table 8.6). Using this non-parametric test, the early-born effect in smaller sibships was considerably stronger than the late-born effect in large sibships, resulting in an overall statistically significant *early birth-order* shift for the whole range of two- to 16-person sibships.

Table 8.6: Overall shift in median birth order for MS cases in pooled sibships of two to 16 siblings (n=1,602 sibships), and shifts in median birth order for pooled smaller sibships (two to six siblings, n=1,398 sibships) and larger sibships (seven to 16 siblings, n=204 sibships), using the non-parametric sign test

Pooled sibships (n†, % of tested sibships)	Z-score#	p-value
Overall 2–16 (1,602, 100.0%)	–2.00	0.046*
Smaller 2–6 (1,398, 87.3%)	–3.05	0.002**
Larger 7–16 (204, 12.7%)	+2.38	0.017*

#Z-score shown to two decimal places only.
† n adjusted for ties.
*Significant shift at p<0.05.
**Significant shift at p<0.01.

Comparison of parametric and non-parametric tests for birth-order shifts

Each of the pooled results for shifts in birth order, by both parametric and non-parametric tests, are shown together in Figure 8.3, where it can be seen that the three techniques delivered similar birth-order shifts in both effect size and direction. The non-parametric test was also able to use all of the available sibship data (i.e. up to and including sibships of 16 siblings).

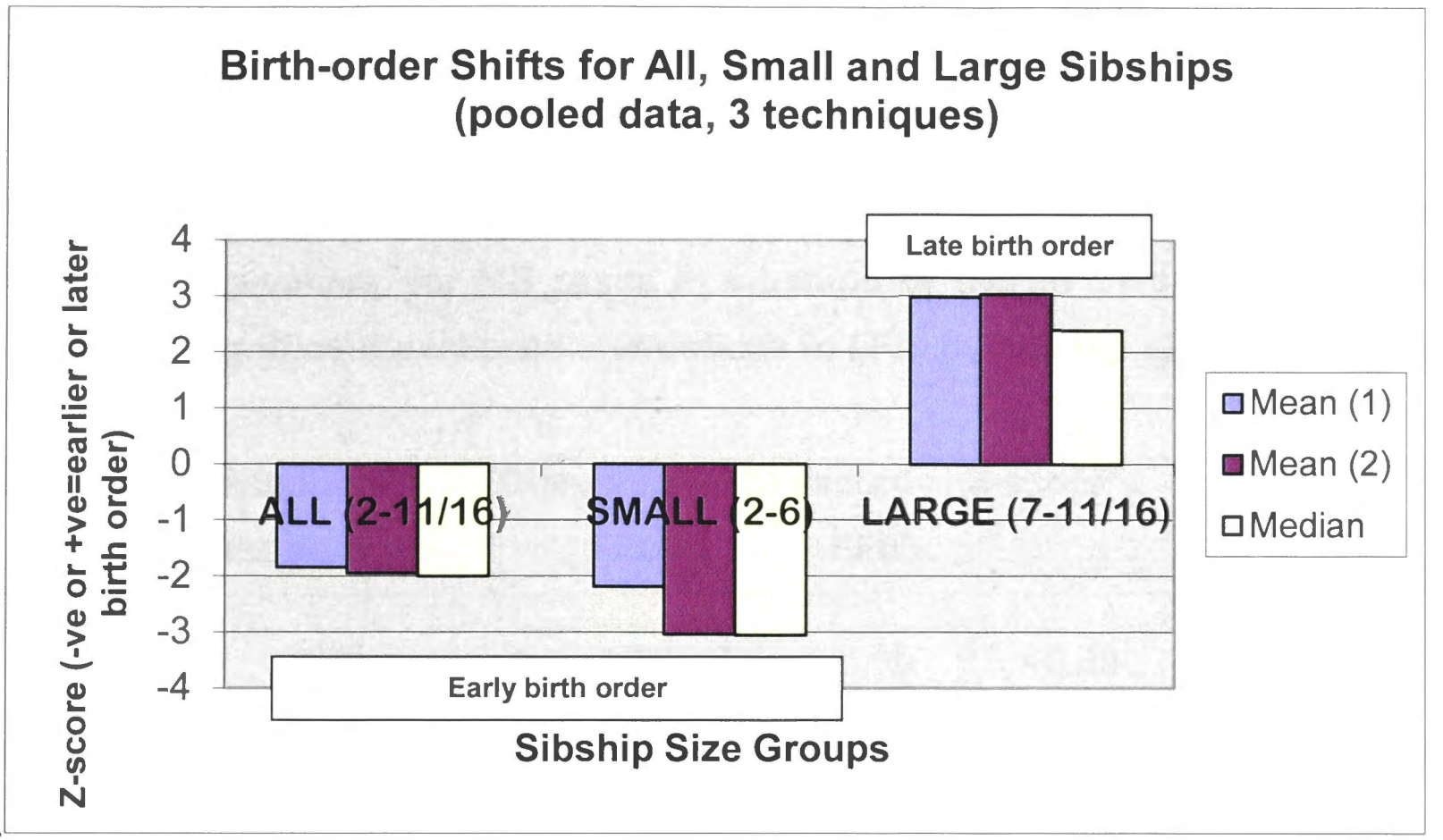


Figure 8.3: Results of all techniques for birth-order shifts in MS cases, for all-sized sibships (two to 11 [MBO] or two to 16 siblings [median birth order]); smaller sibships (two to six siblings [87 to 89% of tested sibships]); and larger sibships (seven to 11 or seven to 16 siblings [11 to 13% of tested sibships]). The figure compares the MBOs calculated using the CLT with two weighting models for pooling separate sibships (Mean [1] and Mean [2]) and the median birth order model calculated using a non-parametric sign test with 'positive' frequencies pooled before testing (Median).

8.3.2 First-born fraction—excess or deficit

Observed and expected first-born fractions (FFb) by sibship size, together with Z-scores for the differences, and their corresponding p-values, are given in Table 8.7. Here, a *positive* Z-score indicates an *excess* of *first-born* MS cases compared with the remaining sibship positions, because the observed FFb is greater than expected by chance. Similarly, a *negative* Z-score indicates a *deficit* of first born compared with the remaining birth-order positions for MS cases.

A significant excess in first-born MS cases was evident for the six-person sibship; however, other sibship sizes with positive Z-scores (e.g. three- and four-person sibships) only approached significance at the 5% level. Deficits in first-born cases featured mainly in the larger sibships, although these again only approached

significance at the 5% level (e.g. eight- and 10-person sibships) (Table 8.7 and Figure 8.4).

Table 8.7: Observed and expected first-born fractions (FFb), together with Z-scores and p-values, for MS cases in sibships of two to 11 siblings (n=1,821), indicating significant excesses and deficits in FFb, by sibship size

Sibship size	Number first born	Observed FFb[#]	Expected FFb[#]	Z-score[#]	p-value	n
2	221	0.50	0.50	–0.09	0.928	444
3	184	0.37	0.33	+1.61	0.107	501
4	97	0.28	0.25	+1.44	0.150	342
5	43	0.20	0.20	0.00	1.000	215
6	30	0.25	0.17	+2.40	0.016*	121
7	13	0.15	0.14	+0.13	0.897	88
8	2	0.04	0.12	–1.78	0.075	49
9	1	0.05	0.11	–0.98	0.327	22
10	0	0.00	0.10	–1.53	0.126	21
11	2	0.11	0.09	+0.30	0.764	18
Total						1821

[#]FFb and Z-score shown to two decimal places only.
*Significant at $p<0.05$.

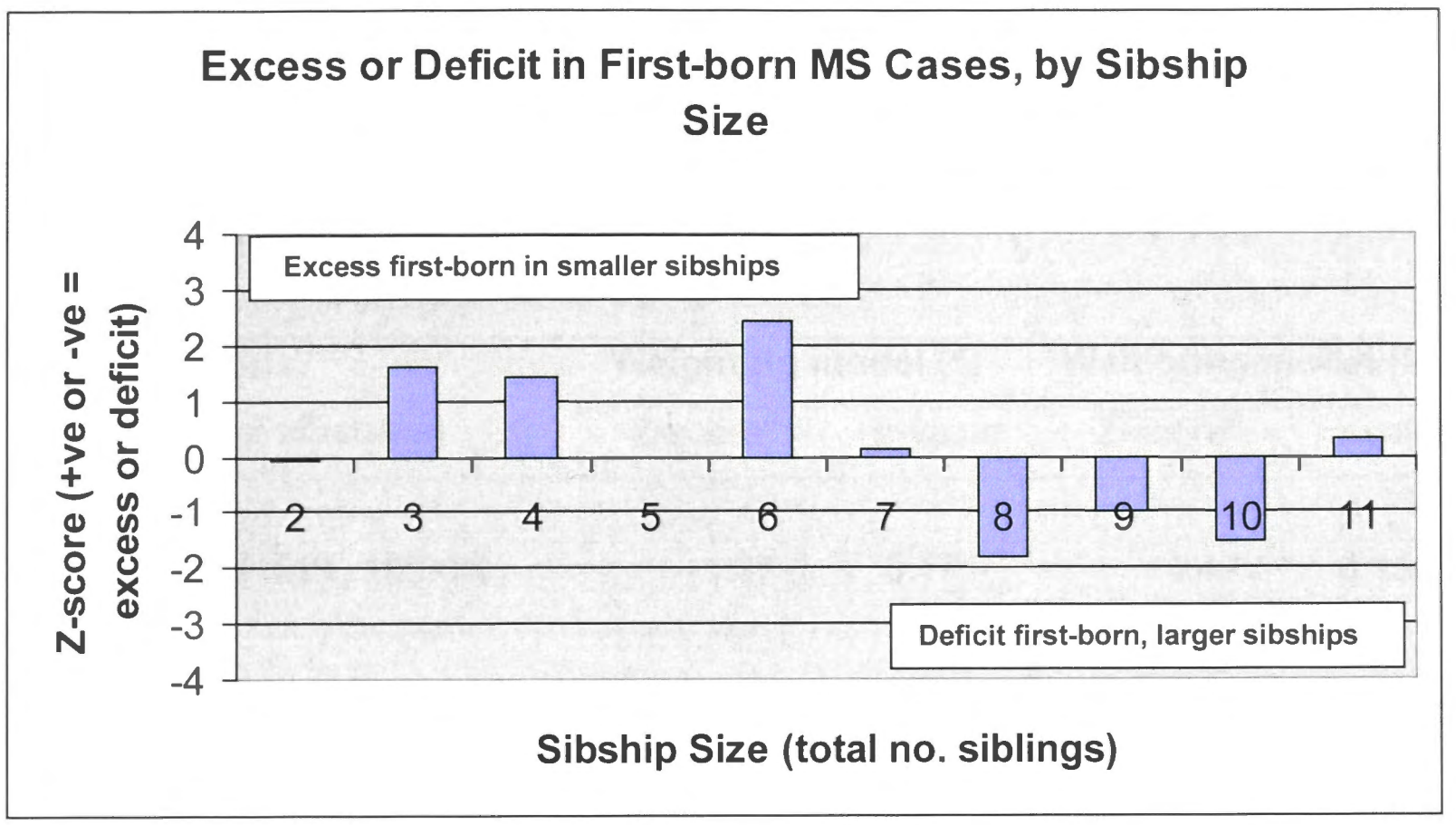


Figure 8.4: Excesses and deficits (as positive or negative Z-scores) of first-born MS cases, by sibship size (n=1,821).

Pooled sibships

Pooling over all sibship sizes from two to 11 siblings (n=1,821 sibships) gave positive Z-scores indicating *overall first-born excesses* among MS cases but these were non-significant using either weighting model (1) or (2) (Table 8.8).

Pooling separately for smaller and larger sibships using the previous pool-size 'cut-offs', again suggested differences in opposite directions, although here the suggested deficit in the 'large' sibships pool (n=198 [10.9% of tested sibships]) was non-significant. Conversely, the two- to six-sibling pool (n=1,623 [89.1% of sibships]) showed a statistically significant *excess* in first-born MS cases, using weighting model (1) (Table 8.8).

Table 8.8 Excesses and deficits in first-born MS cases in overall pooled sibships (two to 11 siblings, n=1,821), together with pooled smaller sibships (two to six siblings, n=1,623) and larger sibships (seven to 11 siblings, n=198), using two weighting models

Pooled sibships (n, % of tested sibships)	Weighting model (1)		Weighting model (2)	
	Z-score[#]	p-value	Z-score[#]	p-value
Overall 2–11 (1821, 100.0%)	+1.37	0.171	+1.42	0.156
Smaller 2–6 (1623, 89.1%)	+2.28	0.023*	+1.70	0.089
Larger 7–11 (198, 10.9%)	–1.58	0.114	–1.45	0.147

[#]Z-score shown to two decimal places only.
*Significant at $p<0.05$.

Summing up, for the entire two- to 11-person sibship pool, a trend toward an *overall excess of first-born* MS cases is evident; this may be consistent with an overall 'early-born' effect seen by the tests for mean and median birth order in preceding subsections, and is discussed further in the next sub-section. However, restricting the pool-size to two- to six-person sibships—this representing the bulk (89.1%) of the total number of sibships—showed that there was a just significant *excess of first-born* MS cases in this pool (average Z-score from both weighting models= +1.99, p=0.047).

Synthesis of results

A comparison of pooled-data Z-scores of the two main analyses of the sibship characteristics of the MS cases can be used to summarise all of these findings for the sibships overall and for the 'smaller' and 'larger' sibships separately (Table 8.9). The Z-scores for the individual MBO analyses using different weighting models and for the median birth order have been averaged for this purpose. Z-scores for the two weighting models used for the first-born fraction pools have also been averaged.

Table 8.9: Summary Z-scores for birth-order measures for MS cases in pooled overall and different-sized sibships, obtained by two independent measurement parameters—mean or median birth order and first-born fraction (Z-scores significant at 5% level shown in bold)

	All sibships (2+ siblings)	Smaller sibships (two to six siblings) [87 to 89% of sibships]	Larger sibships (7+ siblings) [11 to 13% of sibships]
Birth order (Median/MBO)	–1.93	**–2.76** **Early birth order**	**+2.79** Late birth order?
First-born fraction (FFb)	+1.40	**+1.99** **Excess first born**	–1.57

For all sibships pooled, the two parameters in Table 8.9 show a consistent pattern for MS cases. There is a trend towards an overall *excess of first-born* MS cases, consistent with an overall trend towards *earlier birth order* in these cases. That is, despite neither of these trends achieving statistical significance when all-sized sibships are pooled, there appears to be an overall trend toward MS cases being born either first, or at least earlier than expected in the sibship. Thus, in terms of MS risk, there appears to be a possible overall trend towards earlier-born people (in their sibship) having higher MS risk.

Restriction by sibship size

However, clearer results in the same respective directions are gained by restricting the sibship sizes to the smaller, more numerous sibships of just two to six siblings. Here, there is a clear *birth-order shift* towards *earlier* birth-order positions in MS cases, together with an *excess* of MS cases born first in their sibship, both of these parameters being statistically significant (indicated in bold in Table 8.9). The magnitudes of the bolded Z-scores in Table 8.9 can also be directly compared with each other; a comparison shows that the 'early-born' effect in two- to six-person sibships is statistically stronger than the 'first-born' effect. This may indicate that a heightened risk of MS extends beyond the first born in the family.

Summing up, the 'overall' trend in Table 8.9 is thus similar to that in the restricted sibship-size pool (i.e. two- to six-person sibships) for both of the measured parameters, mean/median birth order and first-born fraction, but this trend appears to be weakened because of the presence in the dataset of the larger, less numerous sibships acting in the opposite direction. Whether these effects suggested in the larger sibships are real, or possibly confounded by other factors, will be discussed in Section 8.4.

8.3.3 Additional measurement parameters

As independent data for 'last born' were also available in the dataset and because these might further support the results so far, additional analyses for the 'last-born' fraction, and for the 'first-born relative to last-born' fractions, were conducted. These supplementary tests are shown henceforth only for the restricted two- to six-person sibship pool, because in these tests this group constitutes 89 to 94.5% of all sibships (Table 8.10).

Last-born fraction

The fraction of last-born (FLb) MS cases was calculated for each sibship size in exactly the same way as for the first-born fraction, with '*last born*' being compared with '*non-last born*' (i.e. with the remaining birth-order categories in each sized sibship). The expected fraction of last born in each sibship was the same as that for first born. For the example sibship of four siblings, the FLb was 69/342 (see Table 8.1) (=0.20), which was *less* than the expected fraction of 0.25, thus resulting in a *negative* Z-score and indicating a *deficit* of last-born MS cases in this sibship size.

It can be seen that when information from these two independently derived parameters is considered together, *both* an *excess of first-born* MS cases (Section 8.2.2) and a *deficit of last-born* cases in the same sibship is possible, as in the four-person sibship example.

Pooling of sibship sizes up to six-person sibships was then carried out using two weighting models as for the first-born analyses. Results showed a *significant deficit in last-born* MS cases for these pooled sibships using either weighting model (average Z-score -2.55, p=0.011; n=1,623 sibships). Considered together with the first-born results,

this means that there is now both a significant *excess of first born* and a *deficit of last born* in this main sibship-size pool (89.1% of the total 1,821 sibships in the testable two to 11 sibship-size range).

Relative fraction—first born to last born

As a final supporting but still independent approach, the fraction of first born *relative to* last born (FFb_2) could be calculated. In this analysis, the fraction of first-born MS cases was compared directly with the fraction of last-born cases, excluding all other cases from the analysis. Here, the null hypothesis was that there was an equal chance of being born *either* first or last in those sibships *where the case was born in one of these two positions*.

The observed fraction of first-born cases relative to last-born cases, FFb_2, for each sibship size, could again be tested using the CLT, the test statistic, Z_{F2}, being given by:

$$Z_{F2} = \frac{\text{Observed } FFb_2 - \text{Expected } FFb_2}{\sqrt{(\text{Variance}_{EXPD} / n)}}$$

where the expected fraction of cases had a mean of 0.5, and variance 0.5 (1 – 0.5)= 0.25. The value n was now an adjusted value, sibships with neither a first-born or last-born case being excluded. For the example four-sibling size, there were 166 (n= 97 + 69) sibships where the MS case was *either* first- or last-born (see Table 8.1); the FFb_2 was 97/69 (=0.584), which is *greater* than the expected 0.5, resulting in a *positive* Z-score and indicating a significant *excess* of first-born relative to last-born MS cases in this sibship size.

As in the preceding analyses, the two- to six-person sibships were then pooled. As all cases that were neither first nor last born were excluded from this analysis, every remaining case had an equal chance of being first or last born under the null hypothesis, regardless of sibship size. Therefore, the data were able to be pooled prior to the final CLT analysis and no assumption models were required. A second advantage of this analysis was that all available sibship sizes could be included in the pooled analysis—that is, all sibships where the MS case was born first or last—which included up to 12-person sibships (see Table 8.1) (n=1,126 sibships).

Results from this analysis for the two- to six-person sibship pool were stronger than for the independent first-born fraction and showed a highly significant *relative excess of first born* (Z-score +2.64, p=0.008; n=1,064 sibships). Further, the two- to six-person sibship pool of 1,064 sibships in this analysis represented 94.5% of the total number of sibships in which the MS case was born either first or last.

8.3.4 Birth order and MS risk—collation of all analyses

All of these measures for the two- to six-person sibships can now be considered together (Table 8.10); the various effect sizes are indicated by the Z-score magnitudes.

Table 8.10: Summary Z-scores for birth-order measures for MS cases in pooled two- to six-person sibships, obtained by four independent measurement parameters

Test parameter	Z-score (p-value)	Interpretation of test	% sibships[#]
Birth order (median/MBO)	–2.76 (0.006)**	Early birth order	87-89.1
First-born fraction (FFb)	+1.99 (0.047)*	Excess first born	89.1
Last-born fraction (FLb)	–2.55 (0.011)*	Deficit last born	89.1
First born to last born relative fraction (FFb_2)	+2.64 (0.008)**	Excess first born (relative last born)	94.5

[#] % of sibships in testable pool.
*Significant at p<0.05.
**Significant at p<0.01.

In summary, the results show all of the following in MS cases:

- an early birth-order shift
- excess first born
- a deficit in last born
- excess first born relative to last born (where MS case either first or last born).

All of these shifts or differences were compared with that expected by chance. Each of these measured parameters constitute and contribute to an overall 'early-born' effect evident in MS cases, particularly in these two- to six-person MS-case sibships (87 to 94.5% of testable sibships).

8.4 Discussion

An association between birth order and MS risk that is dependent on sibship size has been shown by these analyses. In smaller sibships (up to a total of six siblings), shifts in mean, or median, birth order of the index MS case towards being earlier born than expected are evident. In larger (7+) sibships, birth-order shifts of the MS case towards being later born than expected appear evident.

These differences with sibship size were seen most readily by considering the smaller sibships, forming around 90% of the available sibships data, separately from the larger sibships and then averaging the (weighted) Z-scores for each sibship-size pool. By this means, the most numerous two- to six-siblings group showed a distinct early-born birth-order effect for MS cases, while the much less numerous 7+ siblings group showed a significant late-born effect. If all sibship sizes were considered together in a single (2+ siblings) pool, the overall effect was generally consistent with that for the two- to six-person sibships, though weaker.

The two- to six-person sibships further showed a significant excess of, in particular, first-born MS cases, using an additional test parameter that was independent of the mean (or median) birth-order parameter. This result supported the early-birth-order shift for this major group of sibships.

In addition, the two- to six-person sibships showed an independent, highly significant excess of first born measured relative to last-born MS cases, as well as a significant deficit of last-born MS cases by yet another independent test parameter. Together, all of these results strongly suggest a significant, non-random, early-born effect in MS cases that is different from that of the general population as expected by chance. That is, the results suggest higher MS risk if born first or early in the sibship and lower risk if born later or last, for most sibships.

This result for the majority of the sibships is consistent with shifts towards early birth order in some earlier work on MS cases [555, 616, 687-690] and may support the hygiene hypothesis concept of a lack of early infectious (or other microbial) exposure contributing to subsequent MS risk (see Chapter 3).

Conversely, the shifts toward later-born cases for the less numerous sibships of seven or more are consistent with the (opposite) findings in the large Canadian study of Sadovnick and colleagues specifically for 7+ siblings [554]. Is the late-born effect in only 7+ -person sibships real? A confounding factor which might contribute to this finding is parental age at birth of the MS case. That is, if MS risk is associated with maternal (or paternal) age at birth, the chance of an MS case being born later in the sibship would be increased. Parental age was not adjusted for in this study; however, average maternal age in the two main sibship-size groups differed by less than four years (mean maternal age for two- to six-person and 7+ -person sibships was, respectively, 28.4 $\pm$ 5.8 and 32.0 $\pm$ 7.2 [s.d.] years). Further, whereas Antonovsky and co-workers reported a significantly higher percentage of MS patients born to mothers aged 40 or more years (but not 30 or more years) compared with controls in a small study [704], and Montgomery and colleagues reported an effect with paternal, rather than maternal, age in their Swedish case-control study [555], subsequent larger longitudinal studies in Denmark and Canada did not find any association between maternal or paternal age at birth and MS risk [556, 705].

The late-born MS-case effect for 7+ siblings, found in both the Canadian study and the present study, if real, is also not inconsistent with the findings of younger infant siblings being protective for MS, independent of birth order, as found in the case-control study in TAS, Australia [553]. That is, if later birth order can be 'equated' with fewer younger siblings (as might generally be the case), then both of these parameters could conceivably result in higher MS risk. The findings of Montgomery and colleagues of protective effects of higher number of siblings (both older and younger) and of having a twin [555] are further broadly consistent here with this more general interpretation of the hygiene hypothesis—that is, that siblings per se in early childhood may be protective for MS risk.

Alternatively, the late-born trends in just the larger sibships found here and in Canada may be due to a period-of-birth cohort effect resulting from increasing MS incidence over time, as suggested by Sadovnick and colleagues for similar results in their study [554]. That is, if incidence increases and/or age of onset of MS decreases over time,

the chance of becoming an MS case would be higher in those born later in the study period. In the present dataset the period between first born and last born of the largest sibships—for example, for those with 10 to 15 siblings—may span some 20 to 30 years (assuming a two-year spacing between siblings), allowing ample time for such a cohort effect. However, it is also worth noting that if a period-of-birth cohort effect exists in the data, it will have weakened the overall results. Indeed, if the cohort effect is strong enough to produce a spurious result in the 7+ -siblings group, it may also be masking an even stronger effect in the two- to six-siblings group. Interestingly, the shift toward later birth order in the 7+ group was present only in females (subgroup analysis data by sex not shown), whereas both males and females exhibited the shift to earlier birth order in the two- to six-person sibships; this may, however, be due to inadequate sample size for the 7+ male group.

Certainly, the early birth order result in MS cases for the majority of sibship sizes, up to six siblings, appears to be the strongest and most consistent finding in the present study, and is probably also the most relevant, given that family sizes of more than six children are now so rare as to be of less population health significance. The result suggests that infection(s) may be acting protectively during early childhood to reduce subsequent MS risk.

Strengths of the present study include the relatively large, comprehensive, population-based survey of clinically identified MS patients in most regions of Australia on the national census day, 30 June 1981, comprising sibship sizes up to 16 siblings. Restriction of the data in the final sample to year of birth between 1920 and 1950 (inclusive) then not only minimised likely selection biases of the cross-sectional survey (see Chapter 5), but, importantly, also limited possible effects of changes in sibship size over time. Thus, this fundamental requirement of the theoretical analysis technique was fulfilled, together with the certainty that the sibships would most likely have been complete by the time of survey. The results should also have not been affected by the children in the sibships attending day-care facilities and thereby being exposed earlier to infections, since an older pre-1950-born dataset of cases was used.

While the theoretical 'observed versus expected' analysis technique for a cases-only dataset (rather than comparison with controls) has undoubtedly been misused in the past in some studies, the stringent use here by first conditioning (or stratifying) on sibship size, as used also by Sadovnick and co-workers [554], avoided major error. Then, for the parametric tests in the present study, weighting of these sibship sizes by

two different 'extreme-case' assumption models, to take account of different distributions for each sibship size, was carried out before any of the sibships were pooled. This does not appear to have been done by any previous such studies (Sadovnick and colleagues do not describe how they obtained their pooled value for seven or more siblings) [554].

While the correct weighting to be applied is unknown—that is, whether the difference between 'observed' and 'expected' varies linearly or not at all with sibship size—the answer is probably 'somewhere in between'. In fact, the Z-score results by these two weighting methods were not dissimilar and in most cases could finally be averaged together with results of additional non-parametric tests that did not require such assumptions and that further enabled all-sized sibships to be included (see Section 8.3.2). Moreover, four different, independent test parameters were able to be employed in the present analyses, particularly for the more numerous smaller sibships. Because the final results of these parameters were consistent, being in the same expected direction, the overall result was strengthened.

Limitations include the available cases-only dataset, for which only the theoretical analysis technique was possible; suitable population controls may have enabled a case-control study to be attempted for comparison. (Sibling controls are not suitable for even theoretical analysis as their birth-order position is 'conditioned' by that of the index MS case.) Sibling ages and inter-birth intervals of the sibships were also unavailable; these might have enabled any period-of-birth effects to have been more easily determined.

MS cases born elsewhere than Australia were also included in the present analysis to retain sufficient sample size (as other workers, including Sadovnick et al., 2005 [554], also appear to have done). However, subsequent analyses for just Australia-born MS cases showed similar birth-order shifts to those presented here for the smaller (up to six siblings) and larger (7+) sibship pools (data not shown). Further, the included migrants mostly from the UK and Ireland came from similarly-developed countries in terms of hygiene and early infectious exposure.

Summing up, MS risk in terms of birth-order position appears to depend on sibship size (i.e. whether <7 or ≥7 siblings) in these Australian data, but the strongest and most relevant result for most sibships (two to six siblings) is that an earlier birth-order

position (i.e. being one of the older siblings in the sibship) increases subsequent MS risk. This is consistent with the broadest form of the hygiene hypothesis, which posits that there is a lack of microbial exposure early in childhood for the older, earlier-born siblings in a sibship, and that this lack of beneficial immune stimulation predisposes toward subsequent autoimmune disease such as MS, as detailed in Chapter 3.

8.5 Conclusion

MS risk in Australia appears to be associated with birth-order position. That is, in most sibship sizes—up to six siblings in total—MS cases appear more likely to be born in an earlier birth-order position than expected by chance (i.e. be an older sibling in the sibship than expected). This may suggest the protective action of infectious or other microbial agents early in childhood (as one causal component in a multi-causal model) to ultimately prevent later MS disease onset.

CHAPTER 9

KEY FINDINGS AND IMPLICATIONS FOR FUTURE RESEARCH AND POPULATION HEALTH POLICY

9.1 Introduction

The aim of this thesis was to extend knowledge on key environmental risk factors for MS in Australia, and to relate this to some other organ-specific autoimmune disorders, including type 1 diabetes and RA, for which the specific aetiology is also presently unknown. The key findings for MS here emphasise the importance of timing of these putative factors in the period before disease onset, and support and extend northern hemisphere findings on both timing and possible nature of these factors.

The *timing-of-birth pattern* found at the individual level in MS cases (see Chapters 5 to 7) indicates a possible factor acting near the time of birth, consistent with northern hemisphere findings for reciprocal seasons. Importantly, for the first time, this factor may be possibly identified as low ambient UVR acting in the prenatal period and most critically in the first foetal trimester (see Chapter 7). These findings for UVR and MS are further supported by a finding, at the population level, of an inverse association between regional ambient UVR and prevalence of autoimmune type 1 diabetes within Australia (see Chapter 4).

In addition, the independently determined *birth-order pattern* found at the individual level in MS cases (see Chapter 8) suggests a second environmental factor in the pre-onset period of MS that may be possibly identified as lack of infection(s) or other microbial exposure early in childhood. This finding is consistent with the predictions of the hygiene hypothesis and again emphasises the importance of timing of such factors.

Thus, interaction of these suggested protective factors with genes, with each other and with subsequent unidentified factors, may constitute an important part of a sequential 'causal cascade' of determinants acting over the life course and leading to eventual onset of MS. These findings constitute new or extended knowledge on the sequential timing and nature of environmental factors influencing MS and, perhaps, other autoimmune disorders.

Data and level of evidence

Secondary analyses of existing datasets are a well-established methodology for addressing important research questions that would otherwise be expensive, time-consuming and difficult to answer directly, particularly for long-latency and/or rare disease conditions such as MS [10, 11]. Some challenges in the use of the two existing datasets analysed in this thesis have been discussed in preceding chapters, together with methods of overcoming these limitations. For example, a year-of-birth-restricted longitudinal study dataset, incorporating population denominators, was able to be constructed from the existing 1981 Australian MS Survey to optimise the data for analysis of the major exposure factors, including regional ambient UVR (see Chapters 6 to 7). A complete comparison of timing of birth, gender and region-of-birth between MS cases and the relevant Australian reference population was then achievable from the original cross-sectional survey data, rather than simply a disease prevalence study (see Chapter 7). Such secondary analyses can contribute to intellectual advancement by building on previously known findings and thereby creating new knowledge; the findings can further be generalised from these existing representative, national datasets to the broad Australian population.

1995 National Health Survey

Good supporting documentation and statistical advice for use of this comprehensive national Australian health interview (cross-sectional) survey, and a high response rate achieved by the ABS, made this dataset reliable for valid secondary analysis of the immune disorders chosen, particularly for type 1 diabetes (see Chapter 4). Because this was necessarily an ecological analysis, focusing on factors that are relevant and variable at the individual level, causal findings are not possible; however, associations found between exposures (latitude, ambient UVR) and disease outcomes at the population level may inform other research. Potential confounders not measurable here, such as infections, climate, temperature and diet, and possible interactions between, also warrant consideration at the individual level in future studies of type 1 diabetes.

1981 Australian MS Survey

Chapters 5 to 8 were based on an existing proprietary dataset comprising individual-level, original data from the comprehensive survey of MS cases in Australia by Professor J. McLeod's University of Sydney research group. This national, population-based, clinical and epidemiological survey was conducted on 30 June 1981, the date chosen to conveniently coincide with an ABS national Census of Population and Housing (see Chapter 5) for application of population denominators. This unique southern hemisphere MS survey was based on nationally standardised ascertainment and careful validation of MS-case diagnoses (see Chapter 5).

As for the 1995 National Health Survey, the relative genetic homogeneity and access to national health care in Australia, as well as the well-coordinated nature of the survey, contributed to the overall validity and usefulness of this MS dataset. Several frequently cited and internationally recognised publications on the epidemiology of MS in Australia by this research group resulted between 1987 and 2011, based on this dataset (see Chapter 5), suggesting that these data were a reliable and comprehensive source of information on Australian MS cases for the purposes of this thesis.

A particular strength of this thesis study lay in the extension of the initial timing-of-birth analysis to explore regional ambient prenatal UVR as a more direct, and independent, risk exposure factor at the individual level (see Chapter 7). Thus, prenatal UVR was a prospective exposure whose levels were heterogeneously distributed among the MS cases because of the large variation in UVR linked to month and region of birth across Australia. While the regional ambient UVR levels were not necessarily those experienced by each mother, it has recently been shown that ambient erythemal UVR levels during pregnancy can be validly used to indicate maternal vitamin D status, and that ambient UVR can be used as an instrumental variable for causal analysis of subsequent vitamin D-dependent disorders in offspring (see Chapter 7). Thus, potential new knowledge of MS aetiology was gained from this extended analysis, and the possible public health importance of the prenatal period for prevention of MS now warrants further confirmatory study.

For the birth-order analyses (see Chapter 8), the 1981 MS Survey dataset was used in its original unit-record, cross-sectional, cases-only form, and analyses included only those born between 1920 and 1950, as for timing-of-birth analysis. Because population controls with sibship information were unavailable for this study, a theoretical 'observed

versus expected' analysis technique, as used by some previous studies, was the only method applicable here. Importantly, in this thesis study, potential error due to differing birth-order distributions with total sibship size was able to be avoided by taking sibship size into account (see Chapter 8). Thus, while the findings need confirmation, this thesis study was able to improve on several previous theoretical birth-order analyses, the findings from which have lacked consistency.

In synopsis, although pre-existing datasets designed for other purposes have been utilised in this thesis, important indications of the timing and possible nature of some environmental determinants influencing autoimmune disorders such as MS have been achieved. By analysis of such existing quality datasets, an epidemiological 'bird's eye view' can be gained of trends that future primary studies can now examine [11]. The secondary analyses in this thesis provide a first evaluation of the given research questions, that may set priorities for subsequent in-depth studies. Further work to test these indications or trends, using data from specifically designed prospective cohort and RCT studies (as considered in the following section), is now needed.

9.2 Key findings of this thesis and implications for future research

9.2.1 Latitude, UVR and type 1 diabetes prevalence

The finding of an association between southern latitude and type 1 diabetes prevalence over the north-south breadth of Australia for the first time (see Chapter 4) is now consistent with similar gradients for such autoimmune disorders in the northern hemisphere. The further finding in Chapter 4 that this latitude gradient was accounted for by an inverse regional ambient UVR gradient over Australia (published as Staples et al., 2003, Appendix I) is also consistent with that found previously for MS in Australia (see Chapter 2). This finding supports the specific prediction that autoimmune diseases other than MS, such as type 1 diabetes, should show latitude and/or UVR gradients similar to those seen for MS if these immune disorders are similarly influenced by UVR exposure [5]. The likelihood of ambient UVR, and/or vitamin D, being an influential environmental factor for organ-specific autoimmune disorders generally has been strengthened by the finding of an inverse ambient UVR gradient of type 1 diabetes prevalence in Australia.

These findings have implications for understanding the possible environmental causes for the observed global distribution of such autoimmune disorders. That is, lack of solar UVR exposure, perhaps mediated through vitamin D deficiency, may be a contributing factor (in both hemispheres) to the observed geographic disease gradients as well as to the increase in incidence of such disease over time (see Chapter 2). Using summary estimates at the population level, the present findings can only be hypothesis generating but should inform future studies on aetiology of type 1 diabetes and other autoimmune disorders. For example, it would be important to establish whether there is a seasonal timing-of-birth risk pattern at the individual level in such disorders in the southern hemisphere, and whether this might be related to regional ambient UVR[20]. This may indicate whether a factor such as UVR might be influencing disease risk around the time of birth, and possibly when, in particular, the most critical period might be.

9.2.2 Prenatal UVR and timing-of-birth risk pattern in MS cases

The findings in Chapter 7 have extended the existing knowledge for MS. A timing-of-birth risk pattern has been shown, at the individual level, for the first time in MS cases born in Australia (cases as sampled nationally in 1981). Importantly, this pattern also mirrors seasonally that seen for MS in the northern hemisphere (see Chapter 7), extending existing knowledge to other geographic regions and supporting the concept that an important environmental determinant of MS is 'time-locked', globally, to the period around birth. A 30% increased MS risk was evident in those born in November to December (southern hemisphere early summer) compared with those born in May to June (early winter) (see Chapter 7). This indicates an environmental factor acting close to the time of birth and modifying the risk of later adult onset of MS.

Most importantly, this timing-of-birth risk pattern in Australia has been shown for the first time to be fully accounted for by the regional (state) and seasonal ambient UVR levels specific to the prenatal period seven to eight months before birth (see Chapter 7). That is, reduced ambient UVR in the first trimester of gestation appears to be associated with subsequent higher risk of MS post-birth. Thus, not only is the timing-of-

---

[20] Elliott and colleagues' study on incidence of type 1 diabetes in Australian children, conducted subsequently to this thesis study, resulted in equivocal findings, including a bi-directional association between regional ambient UVR and type 1 diabetes incidence that was dependent on population density, and no season-of-birth pattern [351].

birth risk pattern in MS cases found in this thesis now suggested to be linked to ambient UVR as a candidate disease determinant, but the particular time window for the main effects on MS risk is proposed to be prenatal and during the first trimester of gestation. This possibly critical prenatal timing is new information for MS and for autoimmune disease generally (published as Staples et al., 2010, Appendix II). The result further supports the UVR hypothesis of McMichael and Hall (1997) and contributes an answer to their subsequent (2001) question: 'Does UVR act early in life?' [5, 6].

Prenatal UVR—possible prevention of MS?

This new finding suggests that ambient UVR prior to birth, and particularly during the first foetal trimester, might be an important (maternal) exposure factor influencing subsequent MS risk in the offspring. That is, lower average daily levels of ambient UVR during the first trimester of gestation, rather than simply the proxy factor of when MS cases are born during the year, now predict a higher subsequent risk of MS. Or, more simply, *lack of* UVR during this 'critical window' of time during foetal development appears to be associated with increased MS risk later in life. Thus, as explored in Chapter 7 (see Figure 7.8), a temporal link between the timing-of-birth risk pattern and ambient UVR levels seven to eight months before birth is suggested by the work in this thesis. The higher MS risk evident in the proxy exposure of November to December-born MS cases may in fact result from the more direct exposure factor of low maternal ambient UVR exposure during the first trimester of gestation.

These findings can only be hypothesis generating, given the nature of the case data utilised and the assumptions required in constructing the longitudinal study dataset (see Chapter 6). However, the finding, if confirmed, is an important one, with real possibilities for having an effect on prevention of MS, and perhaps also other autoimmune disorders. For example, a prospective cohort study could examine both personal and ambient individual sun exposure levels before, during and after pregnancy, and their possible effects on reducing subsequent MS risk in offspring. However, a long follow-up period (>20 years because of adult MS onset) and very large cohort numbers (because of low MS incidence) would be required, making this strategy challenging for MS (but perhaps feasible for type 1 diabetes, with earlier-age onset and higher incidence). Use of a high-risk group such as first-degree relatives of MS cases might aid such future MS studies.

An alternative, or an additional measure together with sun exposure, would be to investigate vitamin D supplementation in pregnancy and beyond, because vitamin D appears to protect against all three autoimmune disorders, MS, type 1 diabetes and RA (see Chapter 2), particularly when administered early in life for MS and diabetes. Ideally, such studies would be RCTs, even though long follow-up times would still be required for MS. Such studies might only be feasible where, for example, large population-based centralised medical records are available, as in countries with universal health care where parents and their offspring can be readily tracked. In addition, the efficiency of identifying benefits of vitamin D supplementation might be increased by selecting for the presence of the vitamin D-regulated MS-susceptibility allele HLA-DRB1*15 [214].

The further finding that region of birth was still associated strongly with MS risk after accounting for either month of birth or first trimester UVR suggests that postnatal UVR exposure is also important in reducing overall MS risk. Given that many Australians apparently remain in their birthplace region for many years (see Chapter 6), this factor (birthplace) may be a good marker for postnatal sun exposure linked to long-term residence. Safe UVR exposure, or vitamin D supplementation, may thus be required into adulthood and future trials should also include this.

9.2.3 Birth-order pattern in MS cases

Using a stringent theoretical analysis technique, a tendency towards MS cases being one of the *older* siblings in their sibships has been shown in MS cases at the individual level in these Australian data (cases as sampled nationally in 1981), particularly in sibship sizes up to six siblings. MS cases also tended to be more likely to be first born and less likely to be last born in these sibships (see Chapter 8). Consistent with the hygiene hypothesis (Chapter 3), this means that MS cases may have been exposed less frequently, and/or later, to protective, immune-boosting, common childhood infections or other microbial exposure than their younger siblings early in life. This *lack of* early microbial exposure may have contributed to their subsequent MS risk (perhaps by combination with late exposure to EBV) (see Chapter 3).

Possible prevention of MS by early microbial exposure?

In view of conflicting results in other birth-order studies (see Chapter 8), these results need further confirmation. For example, specific infections have not been identified here, but the timing of likely microbial exposure has been indicated. Future studies should attempt to measure actual exposure to specific infections (including serology) and the timing (onset age) of these, prospectively if possible and with population controls, together with the proxy factor of birth order. Further, if general microbial exposure is important rather than, or together with, actual infections (see Chapter 3), factors affecting degree of microbial contact, such as day-care attendance, residential density, rural areas, exposure to farm animals and pets should also be considered. Other factors possibly confounding microbial exposure, such as SES or diet, should also be taken into account.

Nevertheless, this thesis has suggested a possible candidate environmental factor for MS that may be acting protectively in a specific time (i.e. age) window. The main conclusion of this section of the thesis is that MS risk appears to be modifiable by this early childhood factor that may be related to common infections or general microbial exposure at this time. Thus, a non-sterile early environment and increased contact with older children may be advantageous for preventing such autoimmune disease. Importantly, this factor may also be interacting, sequentially, with other identified (or as yet unidentified) factors in an overall multi-causal 'cascade' for eventual disease onset, considered next.

9.2.4 Proposed causal sequence for MS

A possible sequential 'causal cascade' of MS determinants, comprising the novel findings in this thesis and based on Goodin's (2009) [229] visualisation of a life-course approach (see Chapters 3 and 7), is presented in Figure 9.1. This diagrammatic model encapsulates the timing of the two major candidate environmental factors suggested to be influential for MS risk in this thesis and now includes the period before birth, as well as the possibility of *protective* infections early in childhood. That is, this thesis model proposes:

- lack of UVR and/or vitamin D near birth, but particularly in the first trimester of gestation
- possible lack of common infections or microbial exposure early in childhood,

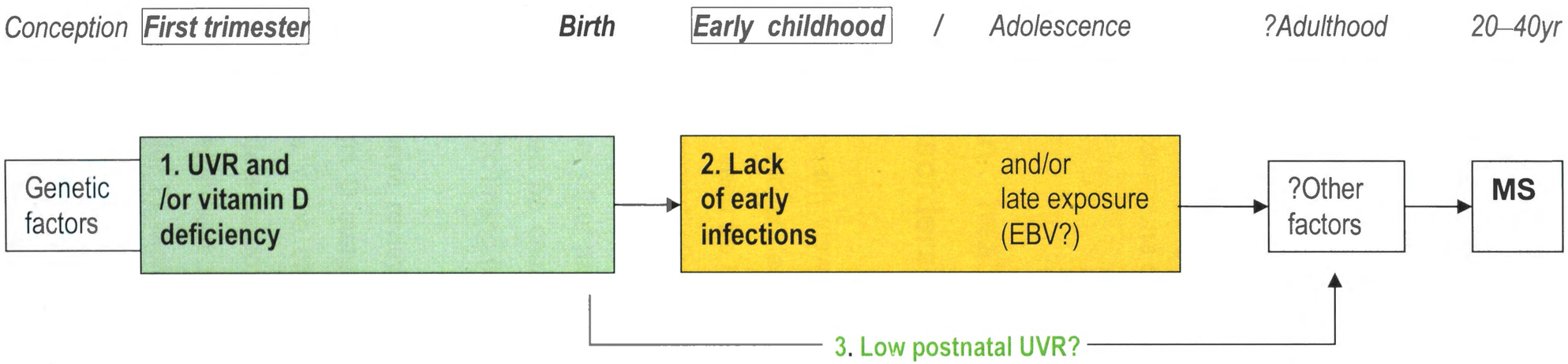
as influential in determining subsequent risk of MS.

Both of these proposed timings—pre-birth and early childhood—are consistent with evidence from age-related migration studies, of early exposure to major determinants in modifying MS risk—that is, before mid-adolescence (see Chapters 2 and 3). Both factors are also linked plausibly with necessary priming of the developing immune system for the establishment of immunological self-tolerance and prevention of autoimmune disease (see Chapter 1). Importantly, the proposed prenatal timing of the first critical factor is consistent with accumulating evidence for effects of foetal development on subsequent adult disease (see Chapters 2 and 7). Moreover, these thesis findings support the idea of an overall sequence of early life and later factors influencing subsequent MS [6].

Such a diagrammatic model further posits biological interaction between sequential factors in causation of such disease, as inherent in Rothman and Greenland's 'component cause' model (see Chapters 2 and 3). That is, each component factor is necessary but not sufficient to cause disease, each factor interacting biologically with prior factors to modify subsequent MS risk [1]. Consistent with this concept, MS-susceptibility gene loci have recently been shown to interact with later-in-life environmental factors such as vitamin D status [188, 706], and synergistically with low infant sibling exposure [707] or late EBV infection (as IM) [569, 708]. These studies indicate that both vitamin D status and hygiene hypothesis-related factors can modify functional effects of genes in MS.

Still later in the life course, environmental factors can also interact with each other. Mechanisms for synergistic interactions between vitamin D and other environmental factors have been proposed in MS and other autoimmune diseases—for example, vitamin D modulating the subsequent effect of viral infections such as EBV in MS [8, 709-711] and, vice versa, herpesviruses such as EBV possibly undermining the protective functions of vitamin D [712]. Still other, later (adult) exposures such as smoking seem to contribute further to adverse effects of vitamin D deficiency in MS patients [580], and to upper respiratory tract infections that possibly trigger MS relapses (see Chapter 3).

Possible Causal Cascade to MS Pathogenesis



Findings: **1. MS cases are more often born in November to December (early summer) in Australia and are potentially deficient in ambient UVR exposure in their first foetal trimester (Chapter 7)**

2. Australian MS cases are more often early born in their sibships (and thus may lack sufficient early microbial exposure) **(Chapter 8)**

3. MS cases are more often born in higher latitude (lower UVR) regions of Australia (perhaps receiving less UVR also postnatally) **(Chapter 7)**

4. (supporting finding) **Similar to MS, type 1 diabetes prevalence shows (positive) latitude and (inverse) ambient UVR gradients over Australia at the population level** (suggesting UVR as factor for autoimmune disease but not timing) **(Chapter 4)**

Figure 9.1: Adaptation of the diagrammatic 'causal cascade' model of Goodin (2009) [229] for MS, showing possible life-course timing (not to scale) of sequential environmental factors as identified in this thesis and based on findings (bolded) shown.

Recent RCT evidence demonstrates an improved child response to viral infection with higher vitamin D administration [713], showing that early life deficiency in vitamin D can restrict the innate and adaptive immune responses essential to a healthy, protective but self-tolerant, immune system (see Chapter 1). Further, lack of such viral or non-specific microbial exposure in early life may interact both qualitatively and quantitatively with subsequent immune response to EBV (see Chapter 3), this hygiene hypothesis mechanism potentially increasing MS risk through late EBV infection (Figure 9.1).

Summing up, the prenatal period now appears to be a possible critical period for environmental modification of MS risk. Second, a further specific critical period is indicated for MS in early childhood. Overall, interaction between the factors in these and other life-course periods most likely occurs, the specific factors converging to eventually cause autoimmune disease such as MS.

9.3 Population health implications for MS and other autoimmune disorders

From a population health viewpoint, if such a life-course causation sequence for MS (Figure 9.1) is true and all factors are 'necessary but not sufficient', then only one environmental factor need be modified to have a significant effect on MS risk [1]. The easiest, and perhaps most influential factor in terms of effects on later disease because of its early prenatal timing [714], is early life exposure to ambient UVR and/or vitamin D. Postnatal infections or other microbial exposure are less predictable, less amenable to manipulation, and can be causal or protective for the same disorders depending on conditions and specific timing (see Chapter 3).

Chapters 2 and 7 have reviewed the accumulating evidence for the role of sunlight exposure and vitamin D status during pregnancy in determining subsequent risk of autoimmune disease in the offspring. Lack of vitamin D, in particular, in early life is plausible as a candidate determinant for autoimmune disease, as evidenced by effects on early brain development and the establishment of central immunological self-tolerance (see Chapter 2). Moreover, many other conditions and long-latency diseases are now also known to be linked to vitamin D deficiency, such as several types of cancer, cardiovascular disease, type 2 diabetes, as well as some infections. Indeed, as proposed by some reviewers, the evident efficacy (Chapter 2), safety and

inexpensiveness of vitamin D supplementation may be reason enough to begin such supplementation without first completing lengthy, stringent clinical trials, particularly during pregnancy and childhood [289, 477].

Population health policy for healthy sun exposure in Australia

Exposure to UVR in sunlight is necessary for most vitamin D requirements in humans, little being derived from diet in most populations. However, there must be caution for recommendations for sunlight exposure in countries like Australia and NZ with the highest skin cancer rates in the world [450]. Current population health guidelines in Australia now attempt to balance healthy sun exposure with sun-avoidance or sun-screening advice that depends on time of day, time of year and location ('SunSmart' guidelines, 2009, revised 2006/2007 by the Australian and New Zealand Bone and Mineral Society; Osteoporosis Australia; the Australasian College of Dermatologists; and the Cancer Council of Australia[21]. The following times are advised for exposure 'of face, arms and hands to maintain healthy levels of vitamin D':

- 'a few minutes on most days in summer' throughout Australia, before 10 am and after 3 pm
- two to three hours per week during [the winter months of] June to July in Sydney, Canberra and Perth (mid-latitude Australia)
- two to three hours per week during May to August in Adelaide, Melbourne and Hobart (southern Australia).

These times and places are when and where the forecasted UV Index is likely to be below the accepted critical (for skin damage) value of 3. Vitamin D supplementation is suggested 'on doctor's advice' only for particular population groups:

- dark-skinned people
- those who cover their skin for religious or cultural purposes
- elderly, house-bound and institutionalised people
- babies (especially breast-fed) and infants of vitamin D-deficient mothers
- patients with osteoporosis.

However, Stalgis-Bilinski and co-workers, in their recent Australian study of personal sunlight exposure required for optimal vitamin D synthesis, indicate that there are few opportunities (i.e. within 30 minutes and with recommended body exposure) to

[21] Guidelines accessed at http://www.cancer.org.au/cancersmartlifestyle/SunSmart/VitaminD.htm).

synthesise 1000 IU of vitamin D during times when the UV Index is less than 3, even for fair-skinned individuals [715]. Therefore, vitamin D supplementation may be required for more individuals and specific target groups than presently recommended, in order to achieve satisfactory vitamin D levels without incurring skin damage.

Most importantly, the emphasis in these SunSmart guidelines is not particularly directed towards pregnant women, or women planning to become pregnant, whether for careful sun exposure or for vitamin D supplementation. Results from this thesis suggest a required minimum monthly average of 20 MED units of ambient UVR per day in pregnancy to reduce offspring MS risk; this is achievable from November to February in Brisbane, QLD (latitude 27.5°S), but only in midsummer January in Hobart, TAS (latitude 42.9°S) (see Chapter 7). Pregnant women also are sometimes advised by health professionals to avoid the sun because of intensified pigment changes causing irregular skin darkening. Further, screening for vitamin D deficiency in pregnant women is currently conducted only for 'at-risk' groups and not routinely in Australia [716].

The results in this thesis indicate the first trimester as possibly critical for receiving sufficient ambient UVR, suggesting a period up to mid-pregnancy that may be important for having sufficient serum vitamin D (given a lag of about one and a half months from higher ambient UVR levels to higher serum vitamin D levels at the population level in Australia, see Chapter 7). Therefore, in light of these results, population health guidelines for vitamin D supplementation should perhaps specifically include pregnant women, or all women of child-bearing age, as a special target group.

In conclusion, prevention strategies for MS and other autoimmune disorders might emulate the public health campaigns to reduce the incidence of foetal neural tube defects, such as spina bifida, in Australia and elsewhere [717]. This highly successful campaign to raise public awareness in health professionals and women of child-bearing age recommends (and heavily advertises) *pre*-pregnancy (i.e. peri-conceptional) as well as pregnancy vitamin supplements containing high levels of folic acid (e.g. 'Elevit' by Bayer, Australia, http://www.elevit.com.au/preparing-for-pregnancy), because these birth defects associated with low dietary folate levels also occur very early in pregnancy, by mid-first trimester. Peri-conceptional vitamin D supplementation for prevention of autoimmune and other disease could be similarly promoted safely to possibly reduce the rates of long-latency autoimmune and other disorders in offspring. Perhaps, like the folic acid success story, primary prevention of autoimmune and many other disorders by early pregnancy vitamin D supplementation

could become one of the 10 best public health achievements of the next decade, as folic acid fortification and promotion became over the last decade in the US [718].

9.4 Conclusions

Similar to MS, type 1 diabetes prevalence shows (positive) latitude and (inverse) regional ambient UVR gradients over Australia at the population level.

At the individual level, aetiology of MS in Australia appears to be influenced by both perinatal and early childhood environmental factors, including low prenatal ambient UVR levels linked with a pattern of increased MS risk in early summer-born people, and a possible lack of early childhood infections.

These findings for MS, in particular, provide new population-based evidence beyond timing-of-birth risk patterns to indicate that the prenatal period may be critical but also suitable for intervention, and that vitamin D supplements for prevention of this, and possibly other, autoimmune disorders might need to be considered during early *in utero* development.

BIBLIOGRAPHY

1. Rothman, K. and S. Greenland, *Modern Epidemiology*. Second ed. 1998, Philadelphia: Lippincott-Raven Publishers. 738p.
2. Davidson, A. and B. Diamond, Autoimmune diseases. *N Engl J Med*, 2001. **345**(5): p. 340-50.
3. Bach, J.F., The effect of infections on susceptibility to autoimmune and allergic diseases. *N Engl J Med*, 2002. **347**(12): p. 911-20.
4. Pugliatti, M., S. Sotgiu, and G. Rosati, The worldwide prevalence of multiple sclerosis. *Clin Neurol Neurosurg*, 2002. **104**(3): p. 182-91.
5. McMichael, A.J. and A.J. Hall, Does immunosuppressive ultraviolet radiation explain the latitude gradient for multiple sclerosis? *Epidemiology*, 1997. **8**(6): p. 642-5.
6. McMichael, A.J. and A.J. Hall, Multiple sclerosis and ultraviolet radiation: time to shed more light. *Neuroepidemiology*, 2001. **20**(3): p. 165-7.
7. Torrey, E.F., et al., Seasonal birth patterns of neurological disorders. *Neuroepidemiology*, 2000. **19**(4): p. 177-85.
8. Ascherio, A. and K.L. Munger, Environmental risk factors for multiple sclerosis. Part I: the role of infection. *Ann Neurol*, 2007. **61**(4): p. 288-99.
9. van der Mei, I.A., et al., Regional variation in multiple sclerosis prevalence in Australia and its association with ambient ultraviolet radiation. *Neuroepidemiology*, 2001. **20**(3): p. 168-74.
10. Doolan, D.M. and E.S. Froelicher, Using an existing data set to answer new research questions: a methodological review. *Res Theory Nurs Pract*, 2009. **23**(3): p. 203-15.
11. Smith, A.K., et al., Conducting high-value secondary dataset analysis: an introductory guide and resources. *J Gen Intern Med*, 2011. **26**(8): p. 920-9.
12. Sorensen, H.T., S. Sabroe, and J. Olsen, A framework for evaluation of secondary data sources for epidemiological research. *Int J Epidemiol*, 1996. **25**(2): p. 435-42.
13. *National Health Survey: Users' Guide, Australia* (ABS Catalogue No. 4363.0). 1995, Canberra: Australian Bureau of Statistics. 151p.
14. Garmon Bibb, S.C., Issues associated with secondary analysis of population health data. *Appl Nurs Res*, 2007. **20**(2): p. 94-9.
15. Steinman, R.M. and M.C. Nussenzweig, Avoiding horror autotoxicus: the importance of dendritic cells in peripheral T cell tolerance. *Proc Natl Acad Sci USA*, 2002. **99**(1): p. 351-8.
16. Ehrlich, P., On immunity with special reference to cell life. *Proc R Soc Lond*, 1900. **66**: p. 424-448.
17. Silverstein, A.M., Paul Ehrlich's passion: the origins of his receptor immunology. *Cell Immunol*, 1999. **194**(2): p. 213-21.
18. Jerne, N.K., The Natural-Selection Theory of Antibody Formation. *Proc Natl Acad Sci USA*, 1955. **41**(11): p. 849-57.
19. Burnet, F.M., A modification of Jerne's theory of antibody production using the concept of clonal selection. *Aust J Sci* 1957. **20**: p. 67-69.
20. Schwartz, R.S., Shattuck lecture: Diversity of the immune repertoire and immunoregulation. *N Engl J Med*, 2003. **348**(11): p. 1017-26.
21. Kamradt, T. and N.A. Mitchison, Tolerance and autoimmunity. *N Engl J Med*, 2001. **344**(9): p. 655-64.
22. Nossal, G.J., A purgative mastery. *Nature*, 2001. **412**(6848): p. 685-6.
23. Verhasselt, V. and M. Goldman, From autoimmune responses to autoimmune disease: what is needed? *J Autoimmun*, 2001. **16**(3): p. 327-30.
24. Parkin, J. and B. Cohen, An overview of the immune system. *Lancet*, 2001. **357**(9270): p. 1777-89.

25. Cantorna, M.T., Vitamin D and autoimmunity: is vitamin D status an environmental factor affecting autoimmune disease prevalence? *Proc Soc Exp Biol Med*, 2000. **223**(3): p. 230-3.
26. Mackay, I.R., Science, medicine, and the future: Tolerance and autoimmunity. *BMJ*, 2000. **321**(7253): p. 93-6.
27. Thompson, A.G. and R. Thomas, Induction of immune tolerance by dendritic cells: implications for preventative and therapeutic immunotherapy of autoimmune disease. *Immunol Cell Biol*, 2002. **80**(6): p. 509-19.
28. Cooper, G.S. and B.C. Stroehla, The epidemiology of autoimmune diseases. *Autoimmun Rev*, 2003. **2**(3): p. 119-25.
29. Walsh, S.J. and L.M. Rau, Autoimmune diseases: a leading cause of death among young and middle-aged women in the United States. *Am J Public Health*, 2000. **90**(9): p. 1463-6.
30. Karges, W.J., et al., Self and non-self antigen in diabetic autoimmunity: molecules and mechanisms. *Mol Aspects Med*, 1995. **16**(2): p. 79-213.
31. Kasper, L.H. and J. Shoemaker, Multiple sclerosis immunology: The healthy immune system vs the MS immune system. *Neurology*, 2010. **74** Suppl. 1: p. S2-8.
32. Afzali, B., et al., The role of T helper 17 (Th17) and regulatory T cells (Treg) in human organ transplantation and autoimmune disease. *Clin Exp Immunol,* 2007. **148**(1): p. 32-46.
33. Honkanen, J., et al., IL-17 immunity in human type 1 diabetes. *J Immunol,* 2010. **185**(3): p. 1959-67.
34. Saito, S., et al., Th1/Th2/Th17 and regulatory T-cell paradigm in pregnancy. *Am J Reprod Immunol,* 2010. **63**(6): p. 601-10.
35. Steinman, L., A brief history of T(H)17, the first major revision in the T(H)1/T(H)2 hypothesis of T cell-mediated tissue damage. *Nat Med,* 2007. **13**(2): p. 139-45.
36. Costantino, C.M., C.M. Baecher-Allan, and D.A. Hafler, Human regulatory T cells and autoimmunity. *Eur J Immunol*, 2008. **38**(4): p. 921-4.
37. Sakaguchi, S., et al., Regulatory T cells and immune tolerance. *Cell*, 2008. **133**(5): p. 775-87.
38. Shevach, E.M., From vanilla to 28 flavors: multiple varieties of T regulatory cells. *Immunity*, 2006. **25**(2): p. 195-201.
39. Wing, K. and S. Sakaguchi, Regulatory T cells exert checks and balances on self tolerance and autoimmunity. *Nat Immunol*, 2010. **11**(1): p. 7-13.
40. Focus on Tolerance Breakdown (invited forum of several contributors to focus issue). *Nat Immunol*, 2010. **11**(1): p. 1-45.
41. Goodnow, C.C., et al., Cellular and genetic mechanisms of self tolerance and autoimmunity. *Nature*, 2005. **435**(7042): p. 590-7.
42. Bach, J.F., Regulatory T cells under scrutiny. *Nat Rev Immunol,* 2003. **3**(3): p. 189-98.
43. Ribot, J., et al., Shaping of the autoreactive regulatory T cell repertoire by thymic cortical positive selection. *J Immunol,* 2007. **179**(10): p. 6741-8.
44. von Boehmer, H. and F. Melchers, Checkpoints in lymphocyte development and autoimmune disease. *Nat Immunol,* 2010. **11**(1): p. 14-20.
45. Bettini, M.L. and D.A. Vignali, Development of thymically derived natural regulatory T cells. *Ann N Y Acad Sci*, 2010. **1183**: p. 1-12.
46. Antel, J. and T. Owens, Multiple sclerosis and immune regulatory cells. *Brain,* 2004. **127**(Pt 9): p. 1915-6.
47. Mueller, D.L., Mechanisms maintaining peripheral tolerance. *Nat Immunol*, 2010. **11**(1): p. 21-7.
48. Rutella, S. and R.M. Lemoli, Regulatory T cells and tolerogenic dendritic cells: from basic biology to clinical applications. *Immunol Lett*, 2004. **94**(1-2): p. 11-26.

49. Maldonado, R.A. and U.H. von Adrian, How tolerogenic dendritic cells induce regulatory T cells. *Adv Immunol*, 2010. **108**(C): p. 111-165.
50. Steinbrink, K., et al., Myeloid dendritic cell: From sentinel of immunity to key player of peripheral tolerance? *Hum Immunol,* 2009. **70**(5): p. 289-93.
51. Turley, S.J., A.L. Fletcher, and K.G. Elpek, The stromal and haematopoietic antigen-presenting cells that reside in secondary lymphoid organs. *Nat Rev Immunol*, 2010. **10**(12): p. 813-25.
52. Adorini, L. and G. Penna, Dendritic cell tolerogenicity: a key mechanism in immunomodulation by vitamin D receptor agonists. *Hum Immunol,* 2009. **70**(5): p. 345-52.
53. Penna, G., et al., 1,25-Dihydroxyvitamin D3 selectively modulates tolerogenic properties in myeloid but not plasmacytoid dendritic cells. *J Immunol,* 2007. **178**(1): p. 145-53.
54. Mahnke, K., et al., Tolerogenic dendritic cells and regulatory T cells: a two-way relationship. *J Dermatol Sci,* 2007. **46**(3): p. 159-67.
55. Zozulya, A.L. and H. Wiendl, The role of regulatory T cells in multiple sclerosis. *Nat Clin Pract Neurol,* 2008. **4**(7): p. 384-98.
56. Schreiner, B., et al., Interferon-beta enhances monocyte and dendritic cell expression of B7-H1 (PD-L1), a strong inhibitor of autologous T-cell activation: relevance for the immune modulatory effect in multiple sclerosis. *J Neuroimmunol,* 2004. **155**(1-2): p. 172-82.
57. Rioux, J.D. and A.K. Abbas, Paths to understanding the genetic basis of autoimmune disease. *Nature*, 2005. **435**(7042): p. 584-9.
58. Jonuleit, H. and E. Schmitt, The regulatory T cell family: distinct subsets and their interrelations. *J Immunol*, 2003. **171**(12): p. 6323-7.
59. Kronenberg, M. and A. Rudensky, Regulation of immunity by self-reactive T cells. *Nature*, 2005. **435**(7042): p. 598-604.
60. Ohkura, N. and S. Sakaguchi, Regulatory T cells: roles of T cell receptor for their development and function. *Semin Immunopathol*, 2010. **32**(2): p. 95-106.
61. Brusko, T. and J. Bluestone, Clinical application of regulatory T cells for treatment of type 1 diabetes and transplantation. *Eur J Immunol,* 2008. **38**(4): p. 931-4.
62. Flores-Borja, F., C. Mauri, and M.R. Ehrenstein, Restoring the balance: harnessing regulatory T cells for therapy in rheumatoid arthritis. *Eur J Immunol,* 2008. **38**(4): p. 934-7.
63. Gregori, S., M. Battaglia, and M.G. Roncarolo, Re-establishing immune tolerance in type 1 diabetes via regulatory T cells. *Novartis Found Symp,* 2008. **292**: p. 174-83; discussion 183-6, 202-3.
64. Roncarolo, M.G. and M. Battaglia, Regulatory T-cell immunotherapy for tolerance to self antigens and alloantigens in humans. *Nat Rev Immunol,* 2007. **7**(8): p. 585-98.
65. Wraith, D.C., K.S. Nicolson, and N.T. Whitley, Regulatory CD4+ T cells and the control of autoimmune disease. *Curr Opin Immunol,* 2004. **16**(6): p. 695-701.
66. Gershon, R.K. and K. Kondo, Infectious immunological tolerance. *Immunology,* 1971. **21**(6): p. 903-14.
67. Sakaguchi, S., et al., Immunologic tolerance maintained by CD25+ CD4+ regulatory T cells: their common role in controlling autoimmunity, tumor immunity, and transplantation tolerance. *Immunol Rev*, 2001. **182**: p. 18-32.
68. Sakaguchi, S., et al., Foxp3+ CD25+ CD4+ natural regulatory T cells in dominant self-tolerance and autoimmune disease. *Immunol Rev*, 2006. **212**: p. 8-27.
69. Baecher-Allan, C., et al., CD4+CD25^{high} regulatory cells in human peripheral blood. *J Immunol*, 2001. **167**(3): p. 1245-53.

70. Stephens, L.A., et al., Human CD4(+)CD25(+) thymocytes and peripheral T cells have immune suppressive activity in vitro. *Eur J Immunol,* 2001. **31**(4): p. 1247-54.
71. Taams, L.S., et al., Antigen-specific T cell suppression by human CD4+CD25+ regulatory T cells. *Eur J Immunol,* 2002. **32**(6): p. 1621-30.
72. Cottrez, F. and H. Groux, Specialization in tolerance: innate CD(4+)CD(25+) versus acquired TR1 and TH3 regulatory T cells. *Transplantation*, 2004. **77**(1 Suppl.): p. S12-5.
73. Campbell, D.J. and M.A. Koch, Phenotypical and functional specialization of FOXP3+ regulatory T cells. *Nat Rev Immunol*, 2011. **11**(2): p. 119-30.
74. Sakaguchi, S., et al., FOXP3+ regulatory T cells in the human immune system. *Nat Rev Immunol*, 2010. **10**(7): p. 490-500.
75. Takahashi, T. and S. Sakaguchi, Naturally arising CD25+CD4+ regulatory T cells in maintaining immunologic self-tolerance and preventing autoimmune disease. *Curr Mol Med*, 2003. **3**(8): p. 693-706.
76. Vignali, D., How many mechanisms do regulatory T cells need? *Eur J Immunol,* 2008. **38**(4): p. 908-11.
77. Horwitz, D.A., et al., Critical role of IL-2 and TGF-beta in generation, function and stabilization of Foxp3+CD4+ Treg. *Eur J Immunol,* 2008. **38**(4): p. 912-5.
78. Buckner, J.H., Mechanisms of impaired regulation by CD4(+)CD25(+)FOXP3(+) regulatory T cells in human autoimmune diseases. *Nat Rev Immunol,* 2010. **10**(12): p. 849-59.
79. Viglietta, V., et al., Loss of functional suppression by CD4+CD25+ regulatory T cells in patients with multiple sclerosis. *J Exp Med,* 2004. **199**(7): p. 971-9.
80. Baecher-Allan, C.M., et al., CD2 costimulation reveals defective activity by human CD4+CD25(hi) regulatory cells in patients with multiple sclerosis. *J Immunol,* 2011. **186**(6): p. 3317-26.
81. Haas, J., et al., Reduced suppressive effect of CD4+CD25high regulatory T cells on the T cell immune response against myelin oligodendrocyte glycoprotein in patients with multiple sclerosis. *Eur J Immunol,* 2005. **35**(11): p. 3343-52.
82. Hug, A., et al., Thymic export function and T cell homeostasis in patients with relapsing remitting multiple sclerosis. *J Immunol,* 2003. **171**(1): p. 432-7.
83. Kumar, M., et al., CD4+CD25+FoxP3+ T lymphocytes fail to suppress myelin basic protein-induced proliferation in patients with multiple sclerosis. *J Neuroimmunol*, 2006. **180**(1-2): p. 178-84.
84. Venken, K., et al., Compromised CD4+ CD25(high) regulatory T-cell function in patients with relapsing-remitting multiple sclerosis is correlated with a reduced frequency of FOXP3-positive cells and reduced FOXP3 expression at the single-cell level. *Immunology*, 2008. **123**(1): p. 79-89.
85. Huan, J., et al., Decreased FOXP3 levels in multiple sclerosis patients. *J Neurosci Res*, 2005. **81**(1): p. 45-52.
86. Venken, K., et al., Natural naive CD4+CD25+CD127low regulatory T cell (Treg) development and function are disturbed in multiple sclerosis patients: recovery of memory Treg homeostasis during disease progression. *J Immunol,* 2008. **180**(9): p. 6411-20.
87. Baecher-Allan, C. and D.A. Hafler, Human regulatory T cells and their role in autoimmune disease. *Immunol Rev*, 2006. **212**: p. 203-16.
88. Cvetanovich, G.L. and D.A. Hafler, Human regulatory T cells in autoimmune diseases. *Curr Opin Immunol*, 2010. **22**(6): p. 753-60.
89. Lindley, S., et al., Defective suppressor function in CD4(+)CD25(+) T cells from patients with type 1 diabetes. *Diabetes,* 2005. **54**(1): p. 92-9.
90. Long, S.A., et al., Defects in IL-2R signaling contribute to diminished maintenance of FOXP3 expression in CD4(+)CD25(+) regulatory T cells of type 1 diabetic subjects. *Diabetes*, 2010. **59**(2): p. 407-15.

91. Lawson, J.M., et al., Increased resistance to CD4+CD25hi regulatory T cell-mediated suppression in patients with type 1 diabetes. *Clin Exp Immunol,* 2008. **154**(3): p. 353-9.
92. Schneider, A., et al., The effector T cells of diabetic subjects are resistant to regulation via CD4+ FOXP3+ regulatory T cells. *J Immunol,* 2008. **181**(10): p. 7350-5.
93. Cao, D., et al., Isolation and functional characterization of regulatory CD25brightCD4+ T cells from the target organ of patients with rheumatoid arthritis. *Eur J Immunol,* 2003. **33**(1): p. 215-23.
94. Notley, C.A. and M.R. Ehrenstein, The yin and yang of regulatory T cells and inflammation in RA. *Nat Rev Rheumatol*, 2010. **6**(10): p. 572-7.
95. Ehrenstein, M.R., et al., Compromised function of regulatory T cells in rheumatoid arthritis and reversal by anti-TNFalpha therapy. *J Exp Med,* 2004. **200**(3): p. 277-85.
96. Bettini, M. and D.A. Vignali, Regulatory T cells and inhibitory cytokines in autoimmunity. *Curr Opin Immunol*, 2009. **21**(6): p. 612-8.
97. Koch, M.A., et al., The transcription factor T-bet controls regulatory T cell homeostasis and function during type 1 inflammation. *Nat Immunol,* 2009. **10**(6): p. 595-602.
98. Balato, A., D. Unutmaz, and A.A. Gaspari, Natural killer T cells: an unconventional T-cell subset with diverse effector and regulatory functions. *J Invest Dermatol,* 2009. **129**(7): p. 1628-42.
99. Yokote, H., et al., NKT cell-dependent amelioration of a mouse model of multiple sclerosis by altering gut flora. *Am J Pathol*, 2008. **173**(6): p. 1714-23.
100. Kukreja, A., et al., Multiple immuno-regulatory defects in type-1 diabetes. *J Clin Invest*, 2002. **109**(1): p. 131-40.
101. Chrobak, P., Control of T cell responses, tolerance and autoimmunity by regulatory T cells: current concepts. *Acta Medica (Hradec Kralove)*, 2003. **46**(4): p. 131-7.
102. Curotto de Lafaille, M.A. and J.J. Lafaille, Natural and adaptive foxp3+ regulatory T cells: more of the same or a division of labor? *Immunity*, 2009. **30**(5): p. 626-35.
103. Battaglia, M. and M.G. Roncarolo, The role of cytokines (and not only) in inducing and expanding T regulatory type 1 cells. *Transplantation,* 2004. **77**(1 Suppl.): p. S16-8.
104. Roncarolo, M.G., et al., Interleukin-10-secreting type 1 regulatory T cells in rodents and humans. *Immunol Rev*, 2006. **212**: p. 28-50.
105. Barrat, F.J., et al., In vitro generation of interleukin 10-producing regulatory CD4(+) T cells is induced by immunosuppressive drugs and inhibited by T helper type 1 (Th1)- and Th2-inducing cytokines. *J Exp Med,* 2002. **195**(5): p. 603-16.
106. Astier, A.L., et al., Alterations in CD46-mediated Tr1 regulatory T cells in patients with multiple sclerosis. *J Clin Invest,* 2006. **116**(12): p. 3252-7.
107. Martinez-Forero, I., et al., IL-10 suppressor activity and ex vivo Tr1 cell function are impaired in multiple sclerosis. *Eur J Immunol*, 2008. **38**(2): p. 576-86.
108. Weiner, H.L., Induction and mechanism of action of transforming growth factor-beta-secreting Th3 regulatory cells. *Immunol Rev*, 2001. **182**: p. 207-14.
109. Chen, Y., et al., Regulatory T cell clones induced by oral tolerance: suppression of autoimmune encephalomyelitis. *Science*, 1994. **265**(5176): p. 1237-40.
110. Carrier, Y., et al., Th3 cells in peripheral tolerance. I. Induction of Foxp3-positive regulatory T cells by Th3 cells derived from TGF-beta T cell-transgenic mice. *J Immunol*, 2007. **178**(1): p. 179-85.
111. Saas, P., F. Kleinclauss, and P. Tiberghien, Immune regulation and transplantation: an exciting challenge. *Transplantation*, 2004. **77**(1 Suppl.): p. S38-40.

112. O'Garra, A., et al., Strategies for use of IL-10 or its antagonists in human disease. *Immunol Rev,* 2008. **223**: p. 114-31.
113. Couper, K.N., D.G. Blount, and E.M. Riley, IL-10: the master regulator of immunity to infection. *J Immunol,* 2008. **180**(9): p. 5771-7.
114. Saraiva, M. and A. O'Garra, The regulation of IL-10 production by immune cells. *Nat Rev Immunol,* 2010. **10**(3): p. 170-81.
115. Betelli, E., et al., IL-10 is critical in the regulation of autoimmune encephalomyelitis as demonstrated by studies of IL-10- and IL-4-deficient and transgenic mice. *J Immunol,* 1998. **161**: p. 3299-3306.
116. Asseman, C., et al., An essential role for interleukin 10 in the function of regulatory T cells that inhibit intestinal inflammation. *J Exp Med*, 1999. **190**(7): p. 995-1004.
117. Samoilova, E.B., J.L. Horton, and Y. Chen, Acceleration of experimental autoimmune encephalomyelitis in interleukin-10-deficient mice: roles of interleukin-10 in disease progression and recovery. *Cell Immunol*, 1998. **188**(2): p. 118-24.
118. Asadullah, K., W. Sterry, and H.D. Volk, Interleukin-10 therapy--review of a new approach. *Pharmacol Rev*, 2003. **55**(2): p. 241-69.
119. Rubtsov, Y.P. and A.Y. Rudensky, TGFbeta signalling in control of T-cell-mediated self-reactivity. *Nat Rev Immunol*, 2007. **7**(6): p. 443-53.
120. Mokhtarian, F., et al., Defective production of anti-inflammatory cytokine, TGF-beta by T cell lines of patients with active multiple sclerosis. *J Immunol*, 1994. **152**(12): p. 6003-10.
121. Rohowsky-Kochan, C., D. Molinaro, and S.D. Cook, Cytokine secretion profile of myelin basic protein-specific T cells in multiple sclerosis. *Mult Scler,* 2000. **6**(2): p. 69-77.
122. Mahon, B.D., et al., Cytokine profile in patients with multiple sclerosis following vitamin D supplementation. *J Neuroimmunol,* 2003. **134**(1-2): p. 128-32.
123. Caramalho, I., et al., Regulatory T cells selectively express toll-like receptors and are activated by lipopolysaccharide. *J Exp Med,* 2003. **197**(4): p. 403-11.
124. Pasare, C. and R. Medzhitov, Toll pathway-dependent blockade of CD4+CD25+ T cell-mediated suppression by dendritic cells. *Science,* 2003. **299**(5609): p. 1033-6.
125. Baecher-Allan, C., V. Viglietta, and D.A. Hafler, Inhibition of human CD4(+)CD25(+high) regulatory T cell function. *J Immunol*, 2002. **169**(11): p. 6210-7.
126. Belkaid, Y., Role of Foxp3-positive regulatory T cells during infection. *Eur J Immunol*, 2008. **38**(4): p. 918-21.
127. Christen, U. and M.G. von Herrath, Initiation of autoimmunity. *Curr Opin Immunol,* 2004. **16**(6): p. 759-67.
128. Anderson, M.S. and M.A. Su, Aire and T cell development. *Curr Opin Immunol,* 2011. **23**(2): p. 198-206.
129. Pitkanen, J. and P. Peterson, Autoimmune regulator: from loss of function to autoimmunity. *Genes Immun,* 2003. **4**(1): p. 12-21.
130. Gardner, J.M., et al., AIRE in the thymus and beyond. *Curr Opin Immunol,* 2009. **21**(6): p. 582-9.
131. Klein, L., et al., Shaping of the autoreactive T-cell repertoire by a splice variant of self protein expressed in thymic epithelial cells. *Nat Med,* 2000. **6**(1): p. 56-61.
132. Ueda, H., et al., Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature*, 2003. **423**(6939): p. 506-11.
133. Dosch, H., et al., Persistent T cell anergy in human type 1 diabetes. *J Immunol*, 1999. **163**(12): p. 6933-40.
134. Bidwell, J., et al., Cytokine gene polymorphism in human disease: on-line databases. *Genes Immun,* 1999. **1**(1): p. 3-19.
135. Hafler, D.A., Multiple sclerosis. *J Clin Invest,* 2004. **113**(6): p. 788-94.

136. Richer, M.J. and M.S. Horwitz, Coxsackievirus infection as an environmental factor in the etiology of type 1 diabetes. *Autoimmun Rev*, 2009. **8**(7): p. 611-5.
137. Horwitz, M.S., et al., Diabetes induced by Coxsackie virus: initiation by bystander damage and not molecular mimicry. *Nat Med*, 1998. **4**(7): p. 781-5.
138. Waldner, H., M. Collins, and V.K. Kuchroo, Activation of antigen-presenting cells by microbial products breaks self tolerance and induces autoimmune disease. *J Clin Invest,* 2004. **113**(7): p. 990-7.
139. Banks, W.A., The blood-brain barrier in psychoneuroimmunology. *Immunol Allergy Clin North Am*, 2009. **29**(2): p. 223-8.
140. Kerfoot, S.M., et al., TLR4 contributes to disease-inducing mechanisms resulting in central nervous system autoimmune disease. *J Immunol*, 2004. **173**(11): p. 7070-7.
141. Wucherpfennig, K.W. and J.L. Strominger, Molecular mimicry in T cell-mediated autoimmunity: viral peptides activate human T cell clones specific for myelin basic protein. *Cell,* 1995. **80**(5): p. 695-705.
142. Kukreja, A. and N.K. Maclaren, Current cases in which epitope mimicry is considered as a component cause of autoimmune disease: immune-mediated (type 1) diabetes. *Cell Mol Life Sci*, 2000. **57**(4): p. 534-41.
143. Ebringer, A., S. Khalafpour, and C. Wilson, Rheumatoid arthritis and Proteus: a possible aetiological association. *Rheumatol Int,* 1989. **9**(3-5): p. 223-8.
144. Veldhoen, M., The role of T helper subsets in autoimmunity and allergy. *Curr Opin Immunol,* 2009. **21**(6): p. 606-11.
145. Gibbon, C., et al., Early infection and subsequent insulin dependent diabetes. *Arch Dis Child,* 1997. **77**(5): p. 384-5.
146. Goodin, D.S., et al., The relationship of MS to physical trauma and psychological stress: report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. *Neurology*, 1999. **52**(9): p. 1737-45.
147. Steinman, L., Multiple sclerosis: a two-stage disease. *Nat Immunol,* 2001. **2**(9): p. 762-4.
148. Sadovnick, A.D. and G.C. Ebers, Epidemiology of multiple sclerosis: a critical overview. *Can J Neurol Sci*, 1993. **20**(1): p. 17-29.
149. Noseworthy, J.H., Progress in determining the causes and treatment of multiple sclerosis. *Nature*, 1999. **399**(6738 Suppl.): p. A40-7.
150. Noseworthy, J.H., et al., Multiple sclerosis. *N Engl J Med*, 2000. **343**(13): p. 938-52.
151. Charcot, J., Histologie de la sclerose en plaque. *Gazette des Hopitaux,* 1868. **41**: p. 554-566.
152. Kabat, E.A., D.A. Freedman, and et al., A study of the crystalline albumin, gamma globulin and total protein in the cerebrospinal fluid of 100 cases of multiple sclerosis and in other diseases. *Am J Med Sci,* 1950. **219**(1): p. 55-64.
153. Lassmann, H., Multiple Sclerosis pathology, in *McAlpine's Multiple Sclerosis*, A. Compston, Editor. 1998, Churchill Livingstone: Edinburgh. p. 323-358.
154. Lassmann, H. and R.M. Ransohoff, The CD4-Th1 model for multiple sclerosis: a critical [correction of crucial] re-appraisal. *Trends Immunol,* 2004. **25**(3): p. 132-7.
155. Miller, A. and Y. Galboiz, Multiple sclerosis: from basic immunopathology to immune intervention. *Clin Neurol Neurosurg,* 2002. **104**(3): p. 172-6.
156. McDonald, W.I., et al., Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol,* 2001. **50**(1): p. 121-7.
157. Polman, C.H., et al., Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". *Ann Neurol*, 2005. **58**(6): p. 840-6.
158. Poser, C.M., et al., New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann Neurol,* 1983. **13**(3): p. 227-31.

159. Rose, A.S., et al., Criteria for the clinical diagnosis of multiple sclerosis. *Neurology,* 1976. **26**(6 PT 2): p. 20-2.
160. Lassmann, H., Mechanisms of demyelination and tissue destruction in multiple sclerosis. *Clin Neurol Neurosurg*, 2002. **104**(3): p. 168-71.
161. Lassmann, H., What drives disease in multiple sclerosis: Inflammation or neurodegeneration? *Clin Exp Neuroimmunol*, 2010. **1**: p. 2-11.
162. McFarland, H.F. and R. Martin, Multiple sclerosis: a complicated picture of autoimmunity. *Nat Immunol*, 2007. **8**(9): p. 913-9.
163. Gold, R., C. Linington, and H. Lassmann, Understanding pathogenesis and therapy of multiple sclerosis via animal models: 70 years of merits and culprits in experimental autoimmune encephalomyelitis research. *Brain*, 2006. **129**(Pt 8): p. 1953-71.
164. Frohman, E.M., M.K. Racke, and C.S. Raine, Multiple sclerosis--the plaque and its pathogenesis. *N Engl J Med*, 2006. **354**(9): p. 942-55.
165. Lucchinetti, C., et al., Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. *Ann Neurol,* 2000. **47**(6): p. 707-17.
166. Barnett, M.H. and J.W. Prineas, Relapsing and remitting multiple sclerosis: pathology of the newly forming lesion. *Ann Neurol,* 2004. **55**(4): p. 458-68.
167. Prat, A. and J. Antel, Pathogenesis of multiple sclerosis. *Curr Opin Neurol,* 2005. **18**(3): p. 225-30.
168. Kebir, H., et al., Human TH17 lymphocytes promote blood-brain barrier disruption and central nervous system inflammation. *Nat Med*, 2007. **13**(10): p. 1173-5.
169. Linker, R.A., et al., Functional role of brain-derived neurotrophic factor in neuroprotective autoimmunity: therapeutic implications in a model of multiple sclerosis. *Brain,* 2010. **133**(Pt 8): p. 2248-63.
170. Axtell, R.C., et al., T helper type 1 and 17 cells determine efficacy of interferon-beta in multiple sclerosis and experimental encephalomyelitis. *Nat Med*, 2010. **16**(4): p. 406-12.
171. Bermel, R.A. and J.A. Cohen, Multiple sclerosis: advances in understanding pathogenesis and emergence of oral treatment options. *Lancet Neurol*, 2011. **10**(1): p. 4-5.
172. Kappos, L., et al., Ocrelizumab in relapsing-remitting multiple sclerosis: a phase 2, randomised, placebo-controlled, multicentre trial. *Lancet*, 2011. **378**(9805): p. 1779-87.
173. Chun, J. and H.P. Hartung, Mechanism of action of oral fingolimod (FTY720) in multiple sclerosis. *Clin Neuropharmacol*, 2010. **33**(2): p. 91-101.
174. Gold, R. and X. Montalban, Multiple sclerosis: more pieces of the immunological puzzle. *Lancet Neurol,* 2012. **11**(1): p. 9-10.
175. Connick, P., et al., Autologous mesenchymal stem cells for the treatment of secondary progressive multiple sclerosis: an open-label phase 2a proof-of-concept study. *Lancet Neurol*, 2012. **11**(2): p. 150-6.
176. Freedman, M.S. and A. Uccelli, Neurorepair with mesenchymal stem cells: hope or hype? *Lancet Neurol*, 2012. **11**(2): p. 123-5.
177. McLeod, J.G., et al., Long-term prognosis of multiple sclerosis in Australia. *J Neurol Sci*, 2007. **256**(1-2): p. 35-8.
178. Ragonese, P., et al., Mortality in multiple sclerosis: a review. *Eur J Neurol*, 2008. **15**(2): p. 123-7.
179. Confavreux, C., S. Vukusic, and P. Adeleine, Early clinical predictors and progression of irreversible disability in multiple sclerosis: an amnesic process. *Brain,* 2003. **126**(Pt 4): p. 770-82.
180. Koepsell, T.D. and N.S. Weiss, *Epidemiologic Methods: Studying the Occurrence of Illness.* 2003, New York: Oxford University Press. 513p.
181. Sadovnick, A.D., The genetics of multiple sclerosis. *Clin Neurol Neurosurg*, 2002. **104**: p. 199-202.

182. Carton, H., et al., Risks of multiple sclerosis in relatives of patients in Flanders, Belgium. *J Neurol Neurosurg Psychiatry*, 1997. **62**(4): p. 329-33.
183. Robertson, N.P., et al., Age-adjusted recurrence risks for relatives of patients with multiple sclerosis. *Brain,* 1996. **119 (Pt 2)**: p. 449-55.
184. Ebers, G.C., A.D. Sadovnick, and N.J. Risch, A genetic basis for familial aggregation in multiple sclerosis. Canadian Collaborative Study Group. *Nature,* 1995. **377**(6545): p. 150-1.
185. Sadovnick, A.D., et al., Evidence for genetic basis of multiple sclerosis. The Canadian Collaborative Study Group. *Lancet,* 1996. **347**(9017): p. 1728-30.
186. Willer, C.J. and G.C. Ebers, Susceptibility to multiple sclerosis: interplay between genes and environment. *Curr Opin Neurol,* 2000. **13**(3): p. 241-7.
187. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat Genet*, 2009. **41**(7): p. 824-8.
188. Ramagopalan, S.V., et al., Expression of the multiple sclerosis-associated MHC class II Allele HLA-DRB1*1501 is regulated by vitamin D. *PLoS Genet*, 2009. **5**(2): p. e1000369.
189. Ebers, G.C., Environmental factors and multiple sclerosis. *Lancet Neurol,* 2008. **7**(3): p. 268-77.
190. Ramagopalan, S.V. and G.C. Ebers, Genes for multiple sclerosis. *Lancet,* 2008. **371**(9609): p. 283-5.
191. Buckner, J.H. and G.T. Nepom, Genetics of rheumatoid arthritis: is there a scientific explanation for the human leukocyte antigen association? *Curr Opin Rheumatol*, 2002. **14**(3): p. 254-9.
192. Laron, Z., Interplay between heredity and environment in the recent explosion of type 1 childhood diabetes mellitus. *Amer J Med Genet*, 2002. **115**(1): p. 4-7.
193. Rosati, G., The prevalence of multiple sclerosis in the world: an update. *Neurol Sci*, 2001. **22**(2): p. 117-39.
194. Symmons, D.P., Epidemiology of rheumatoid arthritis: determinants of onset, persistence and outcome. *Best Pract Res Clin Rheumatol,* 2002. **16**(5): p. 707-22.
195. Rothwell, P.M. and D. Charlton, High incidence and prevalence of multiple sclerosis in south east Scotland: evidence of a genetic predisposition. *J Neurol Neurosurg Psychiatry,* 1998. **64**(6): p. 730-5.
196. Weinshenker, B.G., Epidemiology of multiple sclerosis. *Neurol Clin*, 1996. **14**(2): p. 291-308.
197. Handunnetthi, L., et al., Regulation of major histocompatibility complex class II gene expression, genetic variation and disease. *Genes Immun*, 2010. **11**(2): p. 99-112.
198. Knip, M., et al., Environmental triggers and determinants of type 1 diabetes. *Diabetes,* 2005. **54** Suppl. 2: p. S125-36.
199. Sospedra, M. and R. Martin, Immunology of multiple sclerosis. *Annu Rev Immunol*, 2005. **23**: p. 683-747.
200. Salvetti, M., et al., Twins: mirrors of the immune system. *Immunol Today*, 2000. **21**(7): p. 342-7.
201. Sadovnick, A.D., et al., A population-based study of multiple sclerosis in twins: update. *Ann Neurol,* 1993. **33**(3): p. 281-5.
202. Willer, C.J., et al., Twin concordance and sibling recurrence rates in multiple sclerosis. *Proc Natl Acad Sci USA,* 2003. **100**(22): p. 12877-82.
203. Mumford, C.J., et al., The British Isles survey of multiple sclerosis in twins. *Neurology,* 1994. **44**(1): p. 11-5.
204. Bach, J.F., Insulin-dependent diabetes mellitus as an autoimmune disease. *Endocrin Rev,* 1994. **15**(4): p. 516-42.
205. Hawa, M.I., et al., Impact of genetic and non-genetic factors in type 1 diabetes. *Amer J Med Genet,* 2002. **115**(1): p. 8-17.

206. Redondo, M.J., et al., Heterogeneity of type I diabetes: analysis of monozygotic twins in Great Britain and the United States. *Diabetologia*, 2001. **44**(3): p. 354-62.
207. Kaprio, J., et al., Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. *Diabetologia,* 1992. **35**(11): p. 1060-7.
208. Kumar, D., et al., North-American twins with IDDM. Genetic, etiological, and clinical significance of disease concordance according to age, zygosity, and the interval after diagnosis in first twin. *Diabetes*, 1993. **42**(9): p. 1351-63.
209. Aho, K., et al., Occurrence of rheumatoid arthritis in a nationwide series of twins. *J Rheumatol,* 1986. **13**(5): p. 899-902.
210. Bush, W.S., et al., *Evidence for polygenic susceptibility to multiple sclerosis--the shape of things to come. Am J Hum Genet,* 2010. **86**(4): p. 621-5.
211. Handel, A.E., et al., Type 1 diabetes mellitus and multiple sclerosis: common etiological features. *Nat Rev Endocrinol,* 2009. **5**(12): p. 655-64.
212. Sawcer, S., et al., Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature,* 2011. **476**(7359): p. 214-9.
213. Baranzini, S.E., et al., Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature,* 2010. **464**(7293): p. 1351-6.
214. Handunnetthi, L., S.V. Ramagopalan, and G.C. Ebers, Multiple sclerosis, vitamin D, and HLA-DRB1*15. *Neurology,* 2010. **74**(23): p. 1905-10.
215. Rossini, A.A., Autoimmune diabetes and the circle of tolerance. *Diabetes,* 2004. **53**(2): p. 267-75.
216. Utz, U., et al., Skewed T-cell receptor repertoire in genetically identical twins correlates with multiple sclerosis. *Nature,* 1993. **364**(6434): p. 243-7.
217. Barker, D.J., et al., Growth in utero, blood pressure in childhood and adult life, and mortality from cardiovascular disease. *BMJ*, 1989. **298**(6673): p. 564-7.
218. Barker, D.J., The fetal and infant origins of disease. *Eur J Clin Invest,* 1995. **25**(7): p. 457-63.
219. Barker, D.J., The fetal origins of coronary heart disease. *Acta Paediatr Suppl,* 1997. **422**: p. 78-82.
220. Levin, B.E., Metabolic imprinting: critical impact of the perinatal environment on the regulation of energy homeostasis. *Philos Trans R Soc Lond. Series B: Biol Sci,* 2006. **361**(1471): p. 1107-21.
221. McGrath, J., Does 'imprinting' with low prenatal vitamin D contribute to the risk of various adult disorders? *Med Hypotheses,* 2001. **56**(3): p. 367-71.
222. Bateson, P., et al., Developmental plasticity and human health. *Nature*, 2004. **430**(6998): p. 419-21.
223. Dolinoy, D.C. and R.L. Jirtle, Environmental epigenomics in human health and disease. *Environ Mol Mutagen*, 2008. **49**(1): p. 4-8.
224. Chao, M.J., et al., Epigenetics in multiple sclerosis susceptibility: difference in transgenerational risk localizes to the major histocompatibility complex. *Hum Mol Genet,* 2009. **18**(2): p. 261-6.
225. Ebers, G.C., et al., Parent-of-origin effect in multiple sclerosis: observations in half-siblings. *Lancet*, 2004. **363**(9423): p. 1773-4.
226. Herrera, B.M., et al., Parent-of-origin effects in MS: observations from avuncular pairs. *Neurology,* 2008. **71**(11): p. 799-803.
227. Chao, M.J., et al., Parent-of-origin effects at the major histocompatibility complex in multiple sclerosis. *Hum Mol Genet,* 2010. **19**(18): p. 3679-89.
228. Sadovnick, A.D., The multiple sclerosis trait: a disease waiting to happen? *Clin Neurol Neurosurg,* 2004. **106**(3): p. 172-4.
229. Goodin, D.S., The causal cascade to multiple sclerosis: a model for MS pathogenesis. *PLoS One*, 2009. **4**(2): p. e4565.

230. von Herrath, M., Immunology: insulin trigger for diabetes. *Nature,* 2005. **435**(7039): p. 151-2.
231. Franklin, G.M. and L. Nelson, Environmental risk factors in multiple sclerosis: causes, triggers, and patient autonomy. *Neurology*, 2003. **61**(8): p. 1032-4.
232. Diez Roux, A.V., The study of group-level factors in epidemiology: rethinking variables, study designs, and analytical approaches. *Epidemiol Rev*, 2004. **26**: p. 104-11.
233. Hennekens, C.H. and J.E. Buring, *Epidemiology in Medicine*. First ed. 1987, Boston/Toronto: Little, Brown and Company. 383p.
234. Casetta, I. and E. Granieri, Prognosis of multiple sclerosis: environmental factors. *Neurol Sci,* 2000. **21**(4 Suppl. 2): p. S839-42.
235. Lauer, K., Environmental associations with the risk of multiple sclerosis: the contribution of ecological studies. *Acta Neurol Scand Suppl,* 1995. **161**: p. 77-88.
236. Riise, T., M.W. Nortvedt, and A. Ascherio, Smoking is a risk factor for multiple sclerosis. *Neurology*, 2003. **61**(8): p. 1122-4.
237. Ponsonby, A.L., R.M. Lucas, and I.A. van der Mei, UVR, vitamin D and three autoimmune diseases--multiple sclerosis, type 1 diabetes, rheumatoid arthritis. *Photochem Photobiol,* 2005. **81**(6): p. 1267-75.
238. Ascherio, A. and K. Munger, Epidemiology of multiple sclerosis: from risk factors to prevention. *Semin Neurol,* 2008. **28**(1): p. 17-28.
239. Pugliatti, M., et al., Environmental risk factors in multiple sclerosis. *Acta Neurol Scand Suppl,* 2008. **188**: p. 34-40.
240. Dahlquist, G., The aetiology of type 1 diabetes: an epidemiological perspective. *Acta Paediatr Suppl,* 1998. **425**: p. 5-10.
241. Haverkos, H.W., Could the aetiology of IDDM be multifactorial? *Diabetologia,* 1997. **40**(10): p. 1235-40.
242. Norris, J.M., Can the sunshine vitamin shed light on type 1 diabetes? *Lancet*, 2001. **358**(9292): p. 1476-8.
243. Midgard, R., et al., Incidence of multiple sclerosis in More and Romsdal, Norway from 1950 to 1991. An age-period-cohort analysis. *Brain,* 1996. **119** (Pt 1): p. 203-11.
244. Pugliatti, M., et al., Multiple sclerosis epidemiology in Sardinia: evidence for a true increasing risk. *Acta Neurol Scand*, 2001. **103**(1): p. 20-6.
245. Wikstrom, J., et al., Multiple sclerosis in Finland: evidence of uneven geographic distribution, increasing frequency and high familial occurrence., in *Multiple Sclerosis in Europe: An Epidemiological Update.*, W. Firnhaber and K. Lauer, Editors. 1994, Leuchtturm-Verlag/LTV Press: Darmstadt. p. 73-77.
246. Barnett, M.H., et al., Progressive increase in incidence and prevalence of multiple sclerosis in Newcastle, Australia: a 35-year study. *J Neurol Sci,* 2003. **213**(1-2): p. 1-6.
247. Simpson, S., Jr., et al., Trends in the epidemiology of multiple sclerosis in Greater Hobart, Tasmania: 1951 to 2009. *J Neurol Neurosurg Psychiatry*, 2011. **82**(2): p. 180-7.
248. de Graaf, A.S., Multiple sclerosis in northern Norway. *Eur Neurol*, 1974. **11**(5): p. 281-95.
249. Groenlie, S.A., et al., Multiple sclerosis in North Norway and first appearance in an indigenous population. *J Neurol*, 2000. **247**: p. 129-33.
250. Edland, A., et al., Epidemiology of multiple sclerosis in the county of Vestfold, eastern Norway: incidence and prevalence calculations. *Acta Neurol Scand*, 1996. **93**(2-3): p. 104-9.
251. Celius, E.G. and B. Vandvik, Multiple sclerosis in Oslo, Norway: prevalence on 1 January 1995 and incidence over a 25-year period. *Eur J Neurol*, 2001. **8**(5): p. 463-9.

252. Grytten, N., et al., A 50-year follow-up of the incidence of multiple sclerosis in Hordaland County, Norway. *Neurology*, 2006. **66**(2): p. 182-6.
253. Sundstrom, P., L. Nystrom, and L. Forsgren, Incidence (1988-97) and prevalence (1997) of multiple sclerosis in Vasterbotten County in northern Sweden. *J Neurol Neurosurg Psychiatry,* 2003. **74**(1): p. 29-32.
254. Dahl, O.P., et al., Multiple sclerosis in Nord-Trondelag County, Norway: a prevalence and incidence study. *Acta Neurol Scand,* 2004. **109**(6): p. 378-84.
255. Sarasoja, T., et al., Occurrence of multiple sclerosis in central Finland: a regional and temporal comparison during 30 years. *Acta Neurol Scand*, 2004. **110**(5): p. 331-6.
256. Sumelahti, M.L., et al., Regional and temporal variation in the incidence of multiple sclerosis in Finland 1979-1993. *Neuroepidemiology,* 2000. **19**(2): p. 67-75.
257. Bentzen, J., et al., Prevalence of multiple sclerosis in Denmark 1950--2005. *Mult Scler,* 2010. **16**(5): p. 520-5.
258. Koch-Henriksen, N., The Danish Multiple Sclerosis Registry: a 50-year follow-up. *Mult Scler*, 1999. **5**(4): p. 293-6.
259. Phadke, J.G. and A.W. Downie, Epidemiology of multiple sclerosis in the north-east (Grampian region) of Scotland--an update. *J Epidemiol Community Health*, 1987. **41**(1): p. 5-13.
260. Allison, R.S. and J.H. Millar, Prevalence of disseminated sclerosis in Northern Ireland. *Ulster Med J,* 1954. **23**(Suppl. 2): p. 1-27.
261. Hawkins, S.A. and F. Kee, Updated epidemiological studies of multiple sclerosis in Northern Ireland. *J Neurol,* 1988. **235**(Suppl. 1): p. S86.
262. Gray, O.M., G.V. McDonnell, and S.A. Hawkins, Factors in the rising prevalence of multiple sclerosis in the north-east of Ireland. *Mult Scler*, 2008. **14**(7): p. 880-6.
263. McDonnell, G.V. and S.A. Hawkins, An epidemiologic study of multiple sclerosis in Northern Ireland. *Neurology*, 1998. **50**(2): p. 423-8.
264. Hirst, C., et al., Increasing prevalence and incidence of multiple sclerosis in South East Wales. *J Neurol Neurosurg Psychiatry,* 2009. **80**(4): p. 386-91.
265. Debouverie, M., et al., Increasing incidence of multiple sclerosis among women in Lorraine, Eastern France. *Mult Scler*, 2007. **13**(8): p. 962-7.
266. Fromont, A., et al., Geographic variations of multiple sclerosis in France. *Brain*, 2010. **133**(Pt 7): p. 1889-99.
267. Pugliatti, M., et al., Increasing incidence of multiple sclerosis in the province of Sassari, northern Sardinia. *Neuroepidemiology*, 2005. **25**(3): p. 129-34.
268. Granieri, E., et al., The increasing incidence and prevalence of MS in a Sardinian province. *Neurology*, 2000. **55**(6): p. 842-8.
269. Secular trends in incidence of childhood IDDM in 10 countries. Diabetes Epidemiology Research International Group. *Diabetes,* 1990. **39**(7): p. 858-64.
270. Variation and trends in incidence of childhood diabetes in Europe. EURODIAB ACE Study Group. *Lancet,* 2000. **355**(9207): p. 873-6.
271. Incidence and trends of childhood Type 1 diabetes worldwide 1990-1999. *Diabet Med*, 2006. **23**(8): p. 857-66.
272. Bingley, P.J. and E.A. Gale, Rising incidence of IDDM in Europe. *Diabetes Care,* 1989. **12**(4): p. 289-95.
273. Dahlquist, G. and L. Mustonen, Childhood onset diabetes--time trends and climatological factors. *Int J Epidemiol*, 1994. **23**(6): p. 1234-41.
274. Gardner, S.G., et al., Rising incidence of insulin dependent diabetes in children aged under 5 years in the Oxford region: time trend analysis. The Bart's-Oxford Study Group. *BMJ,* 1997. **315**(7110): p. 713-7.
275. Harjutsalo, V., L. Sjoberg, and J. Tuomilehto, Time trends in the incidence of type 1 diabetes in Finnish children: a cohort study. *Lancet,* 2008. **371**(9626): p. 1777-82.

276. Joner, G. and O. Sovik, Increasing incidence of diabetes mellitus in Norwegian children 0-14 years of age 1973-1982. *Diabetologia*, 1989. **32**(2): p. 79-83.
277. Maahs, D.M., et al., Epidemiology of type 1 diabetes. *Endocrinol Metab Clin North Am*, 2010. **39**(3): p. 481-97.
278. Nystrom, L., et al., The Swedish childhood diabetes study. An analysis of the temporal variation in diabetes incidence 1978-1987. *Int J Epidemiol,* 1990. **19**(1): p. 141-6.
279. Onkamo, P., et al., Worldwide increase in incidence of Type I diabetes--the analysis of the data on published incidence trends. *Diabetologia,* 1999. **42**(12): p. 1395-403.
280. Patterson, C.C., et al., Incidence trends for childhood type 1 diabetes in Europe during 1989-2003 and predicted new cases 2005-20: a multicentre prospective registration study. *Lancet*, 2009. **373**(9680): p. 2027-33.
281. Schober, E., et al., Increasing incidence of IDDM in Austrian children. A nationwide study 1979-1993. Austrian Diabetes Incidence Study Group. *Diabetes Care,* 1995. **18**(9): p. 1280-3.
282. Schoenle, E.J., et al., Epidemiology of type I diabetes mellitus in Switzerland: steep rise in incidence in under 5 year old children in the past decade. *Diabetologia*, 2001. **44**(3): p. 286-9.
283. Shamis, I., et al., Ethnic differences in the incidence of childhood IDDM in Israel (1965-1993). Marked increase since 1985, especially in Yemenite Jews. *Diabetes Care*, 1997. **20**(4): p. 504-8.
284. Soltesz, G., et al., Rising incidence of type 1 diabetes in Hungarian children (1978-1987). Hungarian Childhood Diabetes Epidemiology Group. *Diabet Med,* 1990. **7**(2): p. 111-4.
285. Tuomilehto, J., et al., Increasing trend in type 1 (insulin-dependent) diabetes mellitus in childhood in Finland. Analysis of age, calendar time and birth cohort effects during 1965 to 1984. *Diabetologia,* 1991. **34**(4): p. 282-7.
286. Craig, M.E., et al., The rising incidence of childhood type 1 diabetes in New South Wales, Australia. *J Pediatr Endocrinol Metab*, 2000. **13**(4): p. 363-72.
287. Kelly, H.A., et al., Dramatic increase in incidence of insulin dependent diabetes mellitus in Western Australia. *Med J Aust*, 1994. **161**(7): p. 426-9.
288. Willis, J.A., et al., Incidence of type 1 diabetes mellitus diagnosed before age 20 years in Canterbury, New Zealand over the last 30 years. *J Pediatr Endocrinol Metab,* 2002. **15**(5): p. 637-43.
289. Handel, A.E., et al., Environmental factors and their timing in adult-onset multiple sclerosis. *Nat Rev Neurol,* 2010. **6**(3): p. 156-66.
290. Kurtzke, J.F., Multiple sclerosis in time and space--geographic clues to cause. *J Neurovirol*, 2000. **6** Suppl. 2: p. S134-40.
291. Green, A. and C.C. Patterson, Trends in the incidence of childhood-onset diabetes in Europe 1989-1998. *Diabetologia,* 2001. **44** Suppl. 3: p. B3-8.
292. Tobon, G.J., P. Youinou, and A. Saraux, The environment, geo-epidemiology, and autoimmune disease: Rheumatoid arthritis. *J Autoimmun*, 2010. **35**(1): p. 10-4.
293. Pugliatti, M., et al., The epidemiology of multiple sclerosis in Europe. *Eur J Neurol,* 2006. **13**(7): p. 700-22.
294. Ranzato, F., et al., Increasing frequency of multiple sclerosis in Padova, Italy: a 30 year epidemiological survey. *Mult Scler,* 2003. **9**(4): p. 387-92.
295. Hernan, M.A., M.J. Olek, and A. Ascherio, Geographic variation of MS incidence in two prospective studies of US women. *Neurology*, 1999. **53**(8): p. 1711-8.
296. Alonso, A. and M.A. Hernan, Temporal trends in the incidence of multiple sclerosis: a systematic review. *Neurology*, 2008. **71**(2): p. 129-35.
297. Ascherio, A., K.L. Munger, and K.C. Simon, Vitamin D and multiple sclerosis. *Lancet Neurol,* 2010. **9**(6): p. 599-612.

298. Acheson, E.D., Multiple sclerosis in British Commonwealth countries in the Southern Hemisphere. *Br J Prev Soc Med,* 1961. **15**: p. 118-25.
299. Hammond, S.R., et al., The contribution of mortality statistics to the study of multiple sclerosis in Australia. *J Neurol Neurosurg Psychiatry*, 1989. **52**(1): p. 1-7.
300. Compston, A., Risk factors for multiple sclerosis: race or place? *J Neurol Neurosurg Psychiatry*, 1990. **53**(10): p. 821-3.
301. McLeod, J.G., S.R. Hammond, and J.F. Hallpike, Epidemiology of multiple sclerosis in Australia. With NSW and SA survey results. *Med J Aust*, 1994. **160**(3): p. 117-22.
302. Miller, D.H., et al., Multiple sclerosis in Australia and New Zealand: are the determinants genetic or environmental? *J Neurol Neurosurg Psychiatry,* 1990. **53**(10): p. 903-5.
303. Poser, C.M. and V.V. Brinar, Multiple sclerosis 2001. *Clin Neurol Neurosurg,* 2002. **104**(3): p. 165-7.
304. Hammond, S.R., D.R. English, and J.G. McLeod, The age-range of risk of developing multiple sclerosis: evidence from a migrant population in Australia. *Brain*, 2000. **123 (Pt 5)**: p. 968-74.
305. Taylor, B.V., et al., Latitudinal variation in incidence and type of first central nervous system demyelinating events. *Mult Scler,* 2010. **16**(4): p. 398-405.
306. Gale, C.R. and C.N. Martyn, Migrant studies in multiple sclerosis. *Progress in Neurobiology,* 1995. **47**(4-5): p. 425-48.
307. Ponsonby, A.L., A.M. Hughes, and R.M. Lucas, The 'hygiene hypothesis' and the development of multiple sclerosis. *Neurodegener Dis Manag,* 2011. **1**(4): p. 285-294.
308. Ahlgren, C., et al., High risk of MS in Iranian immigrants in Gothenburg, Sweden. *Mult Scler*, 2010. **16**(9): p. 1079-82.
309. Guimond, C., et al., Prevalence of MS in Iranian immigrants to British Columbia, Canada. *J Neurol,* 2010. **257**(4): p. 667-8.
310. Kahana, E., et al., Environmental factors determine multiple sclerosis risk in migrants to Israel. *Mult Scler,* 2008. **14**(Suppl. 1): p. S69-S70.
311. Wallin, M.T., W.F. Page, and J.F. Kurtzke, Migration and multiple sclerosis in Alaskan military veterans. *J Neurol,* 2009. **256**(9): p. 1413-7.
312. McLeod, J.G., S.R. Hammond, and J.F. Kurtzke, Migration and multiple sclerosis in immigrants to Australia from United Kingdom and Ireland: a reassessment. I. Risk of MS by age at immigration. *J Neurol,* 2011. **258**(6): p. 1140-9.
313. Geographic patterns of childhood insulin-dependent diabetes mellitus. Diabetes Epidemiology Research International Group *Diabetes,* 1988. **37**(8): p. 1113-9.
314. Karvonen, M., et al., A review of the recent epidemiological data on the worldwide incidence of type 1 (insulin-dependent) diabetes mellitus. World Health Organization DIAMOND Project Group. *Diabetologia*, 1993. **36**(10): p. 883-92.
315. Yang, Z., et al., Childhood diabetes in China. Enormous variation by place and ethnic group. *Diabetes Care*, 1998. **21**(4): p. 525-9.
316. Karvonen, M., et al., Incidence of childhood type 1 diabetes worldwide. Diabetes Mondiale (DiaMond) Project Group. *Diabetes Care,* 2000. **23**(10): p. 1516-26.
317. Mohr, S.B., et al., The association between ultraviolet B irradiance, vitamin D status and incidence rates of type 1 diabetes in 51 regions worldwide. *Diabetologia*, 2008. **51**(8): p. 1391-8.
318. Nystrom, L., et al., Risk of developing insulin-dependent diabetes mellitus (IDDM) before 35 years of age: indications of climatological determinants for age at onset. *Int J Epidemiol,* 1992. **21**(2): p. 352-8.

319. Udelhofen, P.M., et al., Surface UV radiation over Australia, 1979-1992: Effects of ozone and cloud cover changes on variations of UV radiation. *J Geophys Res*, 1999. **104**(D16): p. 19135-19159.
320. James, W.H., Season of birth in multiple sclerosis. *Acta Neurol Scand*, 1995. **92**(5): p. 430.
321. Templer, D.I., et al., Season of birth in multiple sclerosis. *Acta Neurol Scand*, 1992. **85**(2): p. 107-9.
322. Wiberg, M. and D.I. Templer, Season of birth in multiple sclerosis in Sweden: replication of Denmark findings. *J Orthomol Med*, 1994. **9**: p. 71-74.
323. Salemi, G., et al., Is season of birth associated with multiple sclerosis? *Acta Neurol Scand,* 2000. **101**(6): p. 381-3.
324. Willer, C.J., et al., Timing of birth and risk of multiple sclerosis: population based study. *BMJ*, 2005. **330**(7483): p. 120-124.
325. Jin, Y., et al., Seasonal patterns in optic neuritis and multiple sclerosis: a meta-analysis. *J Neurol Sci*, 2000. **181**(1-2): p. 56-64.
326. Gray, O., et al., Onset of relapses in multiple sclerosis: the effect of seasonal change in both the northern and southern hemisphere. *Mult Scler,* 2009. **15**: p. S158.
327. Tremlett, H., et al., Monthly ambient sunlight, infections and relapse rates in multiple sclerosis. *Neuroepidemiology,* 2008. **31**(4): p. 271-9.
328. Auer, D.P., et al., Seasonal fluctuations of gadolinium-enhancing magnetic resonance imaging lesions in multiple sclerosis. *Ann Neurol,* 2000. **47**(2): p. 276-7.
329. Embry, A.F., L.R. Snowdon, and R. Vieth, Vitamin D and seasonal fluctuations of gadolinium-enhancing magnetic resonance imaging lesions in multiple sclerosis. *Ann Neurol,* 2000. **48**(2): p. 271-2.
330. Rovaris, M., et al., Effects of seasons on magnetic resonance imaging--measured disease activity in patients with multiple sclerosis. *Ann Neurol,* 2001. **49**(3): p. 415-6.
331. Meier, D.S., et al., Seasonal prevalence of MS disease activity. *Neurology*, 2010. **75**(9): p. 799-806.
332. Jongbloet, P.H., et al., Seasonality of birth in patients with childhood diabetes in The Netherlands. *Diabetes Care*, 1998. **21**(1): p. 190-1.
333. Rothwell, P.M., et al., Seasonality of birth of patients with childhood diabetes in Britain. *BMJ,* 1996. **312**(7044): p. 1456-7.
334. Samuelsson, U., C. Johansson, and J. Ludvigsson, Month of birth and risk of developing insulin dependent diabetes in south east Sweden. *Arch Dis Child,* 1999. **81**(2): p. 143-6.
335. Songini, M., et al., Seasonality of birth in children (0-14 years) and young adults (0-29 years) with type 1 diabetes mellitus in Sardinia differs from that in the general population. The Sardinian Collaborative Group for Epidemiology of IDDM. *J Pediatr Endocrinol Metab*, 2001. **14**(6): p. 781-3.
336. Ursic-Bratina, N., et al., Seasonality of birth in children (0-14 years) with type 1 diabetes mellitus in Slovenia. *J Pediatr Endocrinol Metab,* 2001. **14**(1): p. 47-52.
337. McKinney, P.A., Seasonality of birth in patients with childhood Type I diabetes in 19 European regions. *Diabetologia,* 2001. **44** Suppl. 3: p. B67-74.
338. Rothwell, P.M., et al., Seasonality of birth in children with diabetes in Europe: multicentre cohort study. European Diabetes Study Group. *BMJ,* 1999. **319**(7214): p. 887-8.
339. Vaiserman, A.M., et al., Seasonality of birth in children and young adults (0-29 years) with type 1 diabetes in Ukraine. *Diabetologia,* 2007. **50**(1): p. 32-5.
340. Laron, Z., et al., Month of birth and subsequent development of type I diabetes (IDDM). *J Pediatr Endocrinol Metab,* 1999. **12**(3): p. 397-402.
341. Kida, K., et al., Incidence of Type 1 diabetes mellitus in children aged 0-14 in Japan, 1986-1990, including an analysis for seasonality of onset and month of

birth: JDS study. The Data Committee for Childhood Diabetes of the Japan Diabetes Society (JDS). *Diabet Med*, 2000. **17**(1): p. 59-63.
342. Kahn, H.S., et al., Association of type 1 diabetes with month of birth among U.S. youth: The SEARCH for Diabetes in Youth Study. *Diabetes Care,* 2009. **32**(11): p. 2010-5.
343. Laron, Z., et al., Seasonality of month of birth of children and adolescents with type 1 diabetes mellitus in homogenous and heterogeneous populations. *Isr Med Assoc J,* 2005. **7**(6): p. 381-4.
344. Willis, J.A., et al., Seasonality of birth and onset of clinical disease in children and adolescents (0-19 years) with type 1 diabetes mellitus in Canterbury, New Zealand. *J Pediatr Endocrinol Metab,* 2002b. **15**(5): p. 645-7.
345. Levy-Marchal, C., C. Patterson, and A. Green, Variation by age group and seasonality at diagnosis of childhood IDDM in Europe. The EURODIAB ACE Study Group. *Diabetologia*, 1995. **38**(7): p. 823-30.
346. Moltchanova, E.V., et al., Seasonal variation of diagnosis of Type 1 diabetes mellitus in children worldwide. *Diabet Med*, 2009. **26**(7): p. 673-8.
347. Karvonen, M., et al., Comparison of the seasonal pattern in the clinical onset of IDDM in Finland and Sardinia. *Diabetes Care,* 1998. **21**(7): p. 1101-9.
348. Fleegler, F.M., et al., Age, sex, and season of onset of juvenile diabetes in different geographic areas. *Pediatrics,* 1979. **63**(3): p. 374-9.
349. Verge, C.F., M. Silink, and N.J. Howard, The incidence of childhood IDDM in New South Wales, Australia. *Diabetes Care*, 1994a. **17**(7): p. 693-6.
350. Chong, J.W., et al., Marked increase in type 1 diabetes mellitus incidence in children aged 0-14 yr in Victoria, Australia, from 1999 to 2002. *Pediatr Diabetes,* 2007. **8**(2): p. 67-73.
351. Elliott, J.C., et al., Population density determines the direction of the association between ambient ultraviolet radiation and type 1 diabetes incidence. *Pediatr Diabetes*, 2010. **11**(6): p. 394-402.
352. Afoke, A., et al., Raised IgG and IgM in "epidemic" IDDM suggest that infections are responsible for the seasonality of type I diabetes. *Diabetes Res,* 1991. **16**(1): p. 11-7.
353. Acheson, E.D., C.A. Bachrach, and F.M. Wright, Some comments on the relationship of the distribution of multiple sclerosis to latitude, solar radiation, and other variables. *Acta Psychiatr Scand Suppl,* 1960. **35**(147): p. 132-47.
354. Leibowitz, U. and M. Alter, in *Multiple Sclerosis. Clues to its Cause*. 1973, North-Holland Publishing Company: Amsterdam. p. 243-348.
355. Hutter, C.D. and P. Laing, Multiple sclerosis: sunlight, diet, immunology and aetiology. *Med Hypotheses,* 1996. **46**(2): p. 67-74.
356. Goldberg, P., Multiple sclerosis: vitamin D and calcium as environmental determinants of prevalence. Part 1: Sunlight, dietary factors and epidemiology. *Int J Environ Stud,* 1974. **6**: p. 19-27.
357. Hayes, C.E., M.T. Cantorna, and H.F. DeLuca, Vitamin D and multiple sclerosis. *Proc Soc Exp Biol Med,* 1997. **216**(1): p. 21-7.
358. Ponsonby, A.L., A. McMichael, and I. van der Mei, Ultraviolet radiation and autoimmune disease: insights from epidemiological research. *Toxicology*, 2002. **181-182**: p. 71-8.
359. Beretich, B.D. and T.M. Beretich, Explaining multiple sclerosis prevalence by ultraviolet exposure: a geospatial analysis. *Mult Scler*, 2009. **15**(8): p. 891-8.
360. Sloka, S., et al., Environmental risks for multiple sclerosis: quantitative analyses and biological mechanisms. *Mult Scler,* 2009. **15**: p. S158.
361. Lucas, R.M., et al., Sun exposure and vitamin D are independent risk factors for CNS demyelination. *Neurology*, 2011. **76**(6): p. 540-8.
362. Ullrich, S.E., Mechanisms underlying UV-induced immune suppression. *Mutat Res,* 2005. **571**(1-2): p. 185-205.

363. Clydesdale, G.J., G.W. Dandie, and H.K. Muller, Ultraviolet light induced injury: immunological and inflammatory effects. *Immunol Cell Biol*, 2001. **79**(6): p. 547-68.
364. Duthie, M.S., I. Kimber, and M. Norval, The effects of ultraviolet radiation on the human immune system. *Br J Dermatol,* 1999. **140**(6): p. 995-1009.
365. Garssen, J., et al., UVB exposure-induced systemic modulation of Th1- and Th2-mediated immune responses. *Immunology*, 1999. **97**(3): p. 506-14.
366. Kripke, M.L., Ultraviolet radiation and immunology: something new under the sun--presidential address. *Cancer Res*, 1994. **54**(23): p. 6102-5.
367. Ting, W.W., C.D. Vest, and R. Sontheimer, Practical and experimental consideration of sun protection in dermatology. *Int J Dermatol,* 2003. **42**(7): p. 505-13.
368. Aubin, F., Mechanisms involved in ultraviolet light-induced immunosuppression. *Eur J Dermatol*, 2003. **13**(6): p. 515-23.
369. Schwarz, T., 25 years of UV-induced immunosuppression mediated by T cells-from disregarded T suppressor cells to highly respected regulatory T cells. *Photochem Photobiol,* 2008. **84**(1): p. 10-8.
370. Timares, L., S.K. Katiyar, and C.A. Elmets, DNA damage, apoptosis and langerhans cells--Activators of UV-induced immune tolerance. *Photochem Photobiol,* 2008. **84**(2): p. 422-36.
371. Ullrich, S.E., Photoimmune suppression and photocarcinogenesis. *Front Biosci,* 2002. **7**: p. d684-703.
372. Leitenberger, J., H.T. Jacobe, and P.D. Cruz, Jr., Photoimmunology--illuminating the immune system through photobiology. *Semin Immunopathol,* 2007. **29**(1): p. 65-70.
373. Maverakis, E., et al., Light, including ultraviolet. *J Autoimmun,* 2010. **34**(3): p. J247-57.
374. Nghiem, D.X., et al., Ultraviolet A radiation suppresses an established immune response: implications for sunscreen design. *J Invest Dermatol,* 2001. **117**(5): p. 1193-9.
375. Aubin, F. and C. Mousson, Ultraviolet light-induced regulatory (suppressor) T cells: an approach for promoting induction of operational allograft tolerance? *Transplantation,* 2004. **77**(1 Suppl.): p. S29-31.
376. Vieth, R., Why "Vitamin D" is not a hormone, and not a synonym for 1,25-dihydroxy-vitamin D, its analogs or deltanoids. *J Steroid Biochem Mol Biol*, 2004. **89-90**(1-5): p. 571-3.
377. Hayes, C.E., et al., The immunological functions of the vitamin D endocrine system. *Cell Mol Biol (Noisy-le-grand)*, 2003. **49**(2): p. 277-300.
378. Holick, M.F., Vitamin D deficiency. *N Engl J Med*, 2007. **357**(3): p. 266-81.
379. Pierrot-Deseilligny, C. and J.C. Souberbielle, Is hypovitaminosis D one of the environmental risk factors for multiple sclerosis? *Brain*, 2010. **133**(Pt 7): p. 1869-88.
380. Lin, R. and J.H. White, The pleiotropic actions of vitamin D. *Bioessays*, 2004. **26**(1): p. 21-8.
381. Holick, M.F., Sunlight and vitamin D for bone health and prevention of autoimmune diseases, cancers, and cardiovascular disease. *Am J Clin Nutr*, 2004. **80**(6 Suppl.): p. 1678S-88S.
382. Tetlow, L.C., et al., Vitamin D receptors in the rheumatoid lesion: expression by chondrocytes, macrophages, and synoviocytes. *Ann Rheum Dis,* 1999. **58**(2): p. 118-21.
383. Garcion, E., et al., New clues about vitamin D functions in the nervous system. *Trends Endocrinol Metab,* 2002. **13**(3): p. 100-5.
384. Vieth, R., Why the optimal requirement for Vitamin D3 is probably much higher than what is officially recommended for adults. *J Steroid Biochem Mol Biol*, 2004. **89-90**(1-5): p. 575-9.

385. Zittermann, A., Vitamin D in preventive medicine: are we ignoring the evidence? *Br J Nutr,* 2003. **89**(5): p. 552-72.
386. Lemire, J.M., et al., 1 alpha,25-dihydroxyvitamin D3 suppresses proliferation and immunoglobulin production by normal human peripheral blood mononuclear cells. *J Clin Invest,* 1984. **74**(2): p. 657-61.
387. Lemire, J.M., et al., 1,25-Dihydroxyvitamin D3 suppresses human T helper/inducer lymphocyte activity in vitro. *J Immunol,* 1985. **134**(5): p. 3032-5.
388. Lemire, J.M. and D.C. Archer, 1,25-dihydroxyvitamin D3 prevents the in vivo induction of murine experimental autoimmune encephalomyelitis. *J Clin Invest*, 1991. **87**(3): p. 1103-7.
389. Lemire, J.M. and J.S. Adams, 1,25-Dihydroxyvitamin D3 inhibits the passive transfer of cellular immunity by a myelin basic protein-specific T cell clone. *J Bone Miner Res,* 1992. **7**(2): p. 171-7.
390. Lemire, J.M., Immunomodulatory role of 1,25-dihydroxyvitamin D3. *J Cell Biochem,* 1992. **49**(1): p. 26-31.
391. Lemire, J.M., D.C. Archer, and G.S. Reddy, 1,25-Dihydroxy-24-OXO-16ene-vitamin D3, a renal metabolite of the vitamin D analog 1,25-dihydroxy-16ene-vitamin D3, exerts immunosuppressive activity equal to its parent without causing hypercalcemia in vivo. *Endocrinology*, 1994. **135**(6): p. 2818-21.
392. Lemire, J.M., Immunomodulatory actions of 1,25-dihydroxyvitamin D3. *J Steroid Biochem Mol Biol,* 1995. **53**(1-6): p. 599-602.
393. Lemire, J.M., et al., Immunosuppressive actions of 1,25-dihydroxyvitamin D3: preferential inhibition of Th1 functions. *J Nutr,* 1995. **125**(6 Suppl.): p. 1704S-1708S.
394. Meehan, M.A., R.H. Kerman, and J.M. Lemire, 1,25-Dihydroxyvitamin D3 enhances the generation of nonspecific suppressor cells while inhibiting the induction of cytotoxic cells in a human MLR. *Cell Immunol*, 1992. **140**(2): p. 400-9.
395. Arnson, Y., H. Amital, and Y. Shoenfeld, Vitamin D and autoimmunity: new aetiological and therapeutic considerations. *Ann Rheum Dis*, 2007. **66**(9): p. 1137-42.
396. Smolders, J., et al., Vitamin D as an immune modulator in multiple sclerosis, a review. *J Neuroimmunol,* 2008. **194**(1-2): p. 7-17.
397. Hewison, M., Vitamin D and the immune system: new perspectives on an old theme. *Endocrinol Metab Clin North Am,* 2010. **39**(2): p. 365-79.
398. Adorini, L., et al., Dendritic cells as key targets for immunomodulation by Vitamin D receptor ligands. *J Steroid Biochem Mol Biol*, 2004. **89-90**(1-5): p. 437-41.
399. Penna, G. and L. Adorini, 1 Alpha,25-dihydroxyvitamin D3 inhibits differentiation, maturation, activation, and survival of dendritic cells leading to impaired alloreactive T cell activation. *J Immunol*, 2000. **164**(5): p. 2405-11.
400. Correale, J., M.C. Ysrraelit, and M.I. Gaitan, Immunomodulatory effects of Vitamin D in multiple sclerosis. *Brain,* 2009. **132**(Pt 5): p. 1146-60.
401. Smolders, J., et al., Vitamin D status is positively correlated with regulatory T cell function in patients with multiple sclerosis. *PLoS One*, 2009. **4**(8): p. e6635.
402. Adorini, L., et al., Tolerogenic dendritic cells induced by vitamin D receptor ligands enhance regulatory T cells inhibiting allograft rejection and autoimmune diseases. *J Cell Biochem*, 2003. **88**(2): p. 227-33.
403. Gregori, S., et al., Regulatory T cells induced by 1 alpha,25-dihydroxyvitamin D3 and mycophenolate mofetil treatment mediate transplantation tolerance. *J Immunol,* 2001. **167**(4): p. 1945-53.
404. Holick, M.F., McCollum Award Lecture, 1994: vitamin D--new horizons for the 21st century. *Am J Clin Nutr*, 1994. **60**(4): p. 619-30.
405. Webb, A.R., L. Kline, and M.F. Holick, Influence of season and latitude on the cutaneous synthesis of vitamin D3: exposure to winter sunlight in Boston and

Edmonton will not promote vitamin D3 synthesis in human skin. *J Clin Endocrinol Metab,* 1988. **67**(2): p. 373-8.
406. Kimlin, M.G., Geographic location and vitamin D synthesis. *Mol Aspects Med,* 2008. **29**(6): p. 453-61.
407. Kimlin, M.G., The climatology of Vitamin D producing ultraviolet radiation over the United States. *J Steroid Biochem Mol Biol*, 2004. **89-90**(1-5): p. 479-83.
408. Nieves, J., et al., High prevalence of vitamin D deficiency and reduced bone mass in multiple sclerosis. *Neurology,* 1994. **44**(9): p. 1687-92.
409. Soilu-Hanninen, M., et al., A longitudinal study of serum 25-hydroxyvitamin D and intact parathyroid hormone levels indicate the importance of vitamin D and calcium homeostasis regulation in multiple sclerosis. *J Neurol Neurosurg Psychiatry*, 2008. **79**(2): p. 152-7.
410. Svoren, B.M., et al., Significant vitamin D deficiency in youth with type 1 diabetes mellitus. *J Pediatr*, 2009. **154**(1): p. 132-4.
411. Hillman, L., et al., Vitamin D metabolism and bone mineralization in children with juvenile rheumatoid arthritis. *J Pediatr,* 1994. **124**(6): p. 910-6.
412. Kroger, H., I.M. Penttila, and E.M. Alhava, Low serum vitamin D metabolites in women with rheumatoid arthritis. *Scand J Rheumatol*, 1993. **22**(4): p. 172-7.
413. Cantorna, M.T. and B.D. Mahon, Mounting evidence for vitamin D as an environmental factor affecting autoimmune disease prevalence. *Exp Biol Med (Maywood)*, 2004. **229**(11): p. 1136-42.
414. Tajouri, L., et al., Variation in the vitamin D receptor gene is associated with multiple sclerosis in an Australian population. *J Neurogenet*, 2005. **19**(1): p. 25-38.
415. Ponsonby, A.L., et al., Variation in associations between allelic variants of the vitamin D receptor gene and onset of type 1 diabetes mellitus by ambient winter ultraviolet radiation levels: a meta-regression analysis. *Am J Epidemiol*, 2008. **168**(4): p. 358-65.
416. Torkildsen, O., et al., Vitamin D-dependent rickets as a possible risk factor for multiple sclerosis. *Arch Neurol,* 2008. **65**(6): p. 809-11.
417. Bailey, R., et al., Association of the vitamin D metabolism gene CYP27B1 with type 1 diabetes. *Diabetes*, 2007. **56**(10): p. 2616-21.
418. Cantorna, M.T., C.E. Hayes, and H.F. DeLuca, 1,25-Dihydroxyvitamin D3 reversibly blocks the progression of relapsing encephalomyelitis, a model of multiple sclerosis. *Proc Natl Acad Sci USA,* 1996. **93**(15): p. 7861-4.
419. Cantorna, M.T., C.E. Hayes, and H.F. DeLuca, 1,25-Dihydroxycholecalciferol inhibits the progression of arthritis in murine models of human arthritis. *J Nutr,* 1998. **128**(1): p. 68-72.
420. Larsson, P., et al., A vitamin D analogue (MC 1288) has immunomodulatory properties and suppresses collagen-induced arthritis (CIA) without causing hypercalcaemia. *Clin Exp Immunol*, 1998. **114**(2): p. 277-83.
421. Mathieu, C., et al., Prevention of autoimmune diabetes in NOD mice by 1,25 dihydroxyvitamin D3. *Diabetologia,* 1994. **37**(6): p. 552-8.
422. Adorini, L., Immunomodulatory effects of vitamin D receptor ligands in autoimmune diseases. *Int Immunopharmacol*, 2002. **2**(7): p. 1017-28.
423. Freedman, D.M., M. Dosemeci, and M.C. Alavanja, Mortality from multiple sclerosis and exposure to residential and occupational solar radiation: a case-control study based on death certificates. *Occup Environ Med,* 2000. **57**(6): p. 418-21.
424. Westberg, M., et al., Occupational exposure to UV light and mortality from multiple sclerosis. *Am J Ind Med,* 2009. **52**(5): p. 353-7.
425. Goldacre, M.J., et al., Skin cancer in people with multiple sclerosis: a record linkage study. *J Epidemiol Community Health,* 2004. **58**(2): p. 142-4.
426. van der Mei, I.A., et al., Past exposure to sun, skin phenotype, and risk of multiple sclerosis: case-control study. *BMJ,* 2003. **327**(7410): p. 316.

427. Islam, T., et al., Childhood sun exposure influences risk of multiple sclerosis in monozygotic twins. *Neurology*, 2007. **69**(4): p. 381-8.
428. Kampman, M.T., T. Wilsgaard, and S.I. Mellgren, Outdoor activities and diet in childhood and adolescence relate to MS risk above the Arctic Circle. *J Neurol,* 2007. **254**(4): p. 471-7.
429. Kampman, M.T. and M. Brustad, Vitamin D: a candidate for the environmental effect in multiple sclerosis - observations from Norway. *Neuroepidemiology*, 2008. **30**(3): p. 140-6.
430. Munger, K.L., et al., Vitamin D intake and incidence of multiple sclerosis. *Neurology,* 2004. **62**(1): p. 60-5.
431. Munger, K.L., et al., Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. *JAMA*, 2006. **296**(23): p. 2832-8.
432. Goldberg, P., M.C. Fleming, and E.H. Picard, Multiple sclerosis: decreased relapse rate through dietary supplementation with calcium, magnesium and vitamin D. *Med Hypotheses*, 1986. **21**(2): p. 193-200.
433. Nordvik, I., et al., Effect of dietary advice and n-3 supplementation in newly diagnosed MS patients. *Acta Neurol Scand,* 2000. **102**(3): p. 143-9.
434. Stein, M.S., et al., A randomized trial of high-dose vitamin D2 in relapsing-remitting multiple sclerosis. *Neurology*, 2011. **77**(17): p. 1611-8.
435. Hypponen, E., et al., Intake of vitamin D and risk of type 1 diabetes: a birth-cohort study. *Lancet,* 2001. **358**(9292): p. 1500-3.
436. Vitamin D supplement in early childhood and risk for Type I (insulin-dependent) diabetes mellitus. The EURODIAB Substudy 2 Study Group. *Diabetologia*, 1999. **42**(1): p. 51-4.
437. Stene, L.C., et al., Use of cod liver oil during pregnancy associated with lower risk of Type I diabetes in the offspring. *Diabetologia,* 2000. **43**(9): p. 1093-8.
438. Stene, L.C. and G. Joner, Use of cod liver oil during the first year of life is associated with lower risk of childhood-onset type 1 diabetes: a large, population-based, case-control study. *Am J Clin Nutr,* 2003. **78**(6): p. 1128-34.
439. Fronczak, C.M., et al., In utero dietary exposures and risk of islet autoimmunity in children. *Diabetes Care,* 2003. **26**(12): p. 3237-42.
440. Marjamaki, L., et al., Maternal intake of vitamin D during pregnancy and risk of advanced beta cell autoimmunity and type 1 diabetes in offspring. *Diabetologia*, 2010. **53**(8): p. 1599-607.
441. Brekke, H.K. and J. Ludvigsson, Vitamin D supplementation and diabetes-related autoimmunity in the ABIS study. *Pediatr Diabetes,* 2007. **8**(1): p. 11-4.
442. Tenconi, M.T., et al., Major childhood infectious diseases and other determinants associated with type 1 diabetes: a case-control study. *Acta Diabetol,* 2007. **44**(1): p. 14-9.
443. Simpson, M., et al., No association of vitamin D intake or 25-hydroxyvitamin D levels in childhood with risk of islet autoimmunity and type 1 diabetes: the Diabetes Autoimmunity Study in the Young (DAISY). *Diabetologia*, 2011. **54**(11): p. 2779-88.
444. Zipitis, C.S. and A.K. Akobeng, Vitamin D supplementation in early childhood and risk of type 1 diabetes: a systematic review and meta-analysis. *Arch Dis Child*, 2008. **93**(6): p. 512-7.
445. Merlino, L.A., et al., Vitamin D intake is inversely associated with rheumatoid arthritis: results from the Iowa Women's Health Study. *Arthritis Rheum,* 2004. **50**(1): p. 72-7.
446. Cutolo, M., et al., Vitamin D and rheumatoid arthritis: comment on the letter by Nielen et al. *Arthritis Rheum*, 2007. **56**(5): p. 1719-20.
447. Patel, S., et al., Association between serum vitamin D metabolite levels and disease activity in patients with early inflammatory polyarthritis. *Arthritis Rheum,* 2007. **56**(7): p. 2143-9.

448. Andjelkovic, Z., et al., Disease modifying and immunomodulatory effects of high dose 1 alpha (OH) D3 in rheumatoid arthritis patients. *Clin Exp Rheumatol*, 1999. **17**(4): p. 453-6.
449. Becklund, B.R., et al., UV radiation suppresses experimental autoimmune encephalomyelitis independent of vitamin D production. *Proc Natl Acad Sci USA,* 2010. **107**(14): p. 6418-23.
450. Holick, M.F., High prevalence of vitamin D inadequacy and implications for health. *Mayo Clin Proc*, 2006. **81**(3): p. 353-73.
451. Lucas, R.M., et al., Future health implications of prenatal and early-life vitamin D status. *Nutr Rev*, 2008. **66**(12): p. 710-20.
452. Ponsonby, A.L., et al., Vitamin D status during pregnancy and aspects of offspring health. *Nutrients,* 2010. **2**(3): p. 389-407.
453. Gardener, H., et al., Prenatal and perinatal factors and risk of multiple sclerosis. *Epidemiology*, 2009. **20**(4): p. 611-8.
454. Datta, S., et al., Vitamin D deficiency in pregnant women from a non-European ethnic minority population--an interventional study. *BJOG*, 2002. **109**(8): p. 905-8.
455. Hollis, B.W. and C.L. Wagner, Vitamin D deficiency during pregnancy: an ongoing epidemic. *Am J Clin Nutr,* 2006. **84**(2): p. 273.
456. Nowson, C.A. and C. Margerison, Vitamin D intake and vitamin D status of Australians. *Med J Aust*, 2002. **177**(3): p. 149-52.
457. Pasco, J.A., et al., Vitamin D status of women in the Geelong Osteoporosis Study: association with diet and casual exposure to sunlight. *Med J Aust,* 2001. **175**(8): p. 401-5.
458. Bowyer, L., et al., Vitamin D, PTH and calcium levels in pregnant women and their neonates. *Clin Endocrinol (Oxf)*, 2009. **70**(3): p. 372-7.
459. Grover, S.R. and R. Morley, Vitamin D deficiency in veiled or dark-skinned pregnant women. *Med J Aust,* 2001. **175**(5): p. 251-2.
460. Nozza, J.M. and C.P. Rodda, Vitamin D deficiency in mothers of infants with rickets. *Med J Aust*, 2001. **175**(5): p. 253-5.
461. Robinson, P.D., et al., The re-emerging burden of rickets: a decade of experience from Sydney. *Arch Dis Child*, 2006. **91**(7): p. 564-8.
462. McGrath, J.J., et al., Vitamin D insufficiency in south-east Queensland. *Med J Aust*, 2001. **174**(3): p. 150-1.
463. Judkins, A. and C. Eagleton, Vitamin D deficiency in pregnant New Zealand women. *NZ Med J,* 2006. **119**(1241): p. U2144.
464. Perampalam, S., et al., Vitamin D status and its predictive factors in pregnancy in 2 Australian populations. *Aust NZ J Obstet Gynaecol* , 2011. **51**(4): p. 353-9.
465. Teale, G.R. and C.E. Cunningham, Vitamin D deficiency is common among pregnant women in rural Victoria. *Aust NZ J Obstet Gynaecol* , 2010. **50**(3): p. 259-61.
466. Zittermann, A., J. Dembinski, and P. Stehle, Low vitamin D status is associated with low cord blood levels of the immunosuppressive cytokine interleukin-10. *Pediatr Allergy Immunol,* 2004. **15**(3): p. 242-6.
467. Pichler, J., et al., 1 alpha,25(OH)2D3 inhibits not only Th1 but also Th2 differentiation in human cord blood T cells. *Pediatr Res* , 2002. **52**(1): p. 12-8.
468. Novakovic, B., et al., Placenta-specific methylation of the vitamin D 24-hydroxylase gene: implications for feedback autoregulation of active vitamin D levels at the fetomaternal interface. *J Biol Chem*, 2009. **284**(22): p. 14838-48.
469. Maghbooli, Z., et al., Vitamin D status in mothers and their newborns in Iran. *BMC Pregnancy Childbirth*, 2007. **7**: p. 1.
470. Bassir, M., et al., Vitamin D deficiency in Iranian mothers and their neonates: a pilot study. *Acta Paediatr*, 2001. **90**(5): p. 577-9.

471. Bodnar, L.M., et al., High prevalence of vitamin D insufficiency in black and white pregnant women residing in the northern United States and their neonates. *J Nutr,* 2007. **137**(2): p. 447-52.
472. Holmes, V.A., et al., Vitamin D deficiency and insufficiency in pregnant women: a longitudinal study. *Br J Nutr*, 2009. **102**(6): p. 876-81.
473. Molla, A.M., et al., Vitamin D status of mothers and their neonates in Kuwait. *Pediatr Int,* 2005. **47**(6): p. 649-52.
474. Sachan, A., et al., High prevalence of vitamin D deficiency among pregnant women and their newborns in northern India. *Am J Clin Nutr*, 2005. **81**(5): p. 1060-4.
475. van der Meer, I.M., et al., High prevalence of vitamin D deficiency in pregnant non-Western women in The Hague, Netherlands. *Am J Clin Nutr*, 2006. **84**(2): p. 350-3; quiz 468-9.
476. Sahu, M., et al., Vitamin D deficiency in rural girls and pregnant women despite abundant sunshine in northern India. *Clin Endocrinol (Oxf)*, 2009. **70**(5): p. 680-4.
477. Chaudhuri, A., Why we should offer routine vitamin D supplementation in pregnancy and childhood to prevent multiple sclerosis. *Med Hypotheses,* 2005. **64**(3): p. 608-18.
478. Neveu, I., et al., 1,25-dihydroxyvitamin D3 regulates the synthesis of nerve growth factor in primary cultures of glial cells. *Brain Res Mol Brain Res,* 1994. **24**(1-4): p. 70-6.
479. Wion, D., et al., 1,25-Dihydroxyvitamin D3 is a potent inducer of nerve growth factor synthesis. *J Neurosci Res*, 1991. **28**(1): p. 110-4.
480. McGrath, J., et al., Vitamin D supplementation during the first year of life and risk of schizophrenia: a Finnish birth cohort study. *Schizophr Res,* 2004. **67**(2-3): p. 237-45.
481. Almeras, L., et al., Developmental vitamin D deficiency alters brain protein expression in the adult rat: implications for neuropsychiatric disorders. *Proteomics*, 2007. **7**(5): p. 769-80.
482. Eyles, D., et al., Developmental vitamin D deficiency alters the expression of genes encoding mitochondrial, cytoskeletal and synaptic proteins in the adult rat brain. *J Steroid Biochem Mol Biol,* 2007. **103**(3-5): p. 538-45.
483. Feron, F., et al., Developmental Vitamin D3 deficiency alters the adult rat brain. *Brain Res Bull ,* 2005. **65**(2): p. 141-8.
484. McGrath, J.J., et al., Vitamin D3-implications for brain development. *J Steroid Biochem Mol Biol,* 2004. **89-90**(1-5): p. 557-60.
485. Cui, X., et al., Maternal vitamin D depletion alters neurogenesis in the developing rat brain. *Int J Dev Neurosci,* 2007. **25**(4): p. 227-32.
486. Veenstra, T.D., et al., 1,25-Dihydroxyvitamin D3 receptors in the central nervous system of the rat embryo. *Brain Research*, 1998. **804**(2): p. 193-205.
487. Granieri, E., et al., Multiple sclerosis: infectious hypothesis. *Neurol Sci,* 2001. **22**(2): p. 179-85.
488. Kurtzke, J.F. and K. Hyllested, Multiple sclerosis in the Faroe Islands: I. Clinical and epidemiological features. *Ann Neurol,* 1979. **5**(1): p. 6-21.
489. Kurtzke, J.F., Epidemiologic evidence for multiple sclerosis as an infection. *Clin Microbiol Rev* , 1993. **6**(4): p. 382-427.
490. Hunter, S.F. and D.A. Hafler, Ubiquitous pathogens: links between infection and autoimmunity in MS? *Neurology,* 2000. **55**(2): p. 164-5.
491. Benoist, C. and D. Mathis, Autoimmunity provoked by infection: how good is the case for T cell epitope mimicry? *Nat Immunol,* 2001. **2**(9): p. 797-801.
492. Getts, M.T. and S.D. Miller, 99th Dahlem conference on infection, inflammation and chronic inflammatory disorders: triggering of autoimmune diseases by infections. *Clin Exp Immunol,* 2010. **160**(1): p. 15-21.

493. Munz, C., et al., Antiviral immune responses: triggers of or triggered by autoimmunity? *Nat Rev Immunol,* 2009. **9**(4): p. 246-58.
494. Albert, L.J. and R.D. Inman, Molecular mimicry and autoimmunity. *N Engl J Med,* 1999. **341**(27): p. 2068-74.
495. Ercolini, A.M. and S.D. Miller, The role of infections in autoimmune disease. *Clin Exp Immunol,* 2008. **155**: p. 1-15.
496. Lunemann, J.D., et al., EBNA1-specific T cells from patients with multiple sclerosis cross react with myelin antigens and co-produce IFN-gamma and IL-2. *J Exp Med*, 2008. **205**(8): p. 1763-73.
497. Filippi, C.M. and M.G. von Herrath, 99th Dahlem conference on infection, inflammation and chronic inflammatory disorders: viruses, autoimmunity and immunoregulation. *Clin Exp Immunol*, 2010. **160**(1): p. 113-9.
498. Ascherio, A. and M. Munch, Epstein-Barr virus and multiple sclerosis. *Epidemiology*, 2000. **11**(2): p. 220-4.
499. Sundstrom, P., et al., An altered immune response to Epstein-Barr virus in multiple sclerosis: a prospective study. *Neurology*, 2004. **62**(12): p. 2277-82.
500. Levin, L.I., et al., Temporal relationship between elevation of Epstein-Barr virus antibody titers and initial onset of neurological symptoms in multiple sclerosis. *JAMA*, 2005. **293**(20): p. 2496-500.
501. Thacker, E.L., F. Mirzaei, and A. Ascherio, Infectious mononucleosis and risk for multiple sclerosis: a meta-analysis. *Ann Neurol,* 2006. **59**(3): p. 499-503.
502. Lunemann, J.D., et al., Increased frequency and broadened specificity of latent EBV nuclear antigen-1-specific T cells in multiple sclerosis. *Brain,* 2006. **129**(Pt 6): p. 1493-506.
503. Nielsen, T.R., et al., Multiple sclerosis after infectious mononucleosis. *Arch Neurol,* 2007. **64**(1): p. 72-5.
504. Serafini, B., et al., Dysregulated Epstein-Barr virus infection in the multiple sclerosis brain. *J Exp Med,* 2007. **204**(12): p. 2899-912.
505. Lunemann, J.D., et al., Broadened and elevated humoral immune response to EBNA1 in pediatric multiple sclerosis. *Neurology*, 2008. **71**(13): p. 1033-5.
506. Alspaugh, M.A., et al., Elevated levels of antibodies to Epstein-Barr virus antigens in sera and synovial fluids of patients with rheumatoid arthritis. *J Clin Invest,* 1981. **67**(4): p. 1134-40.
507. Tosato, G., A.D. Steinberg, and R.M. Blaese, Defective EBV-specific suppressor T-cell function in rheumatoid arthritis. *N Engl J Med,* 1981. **305**(21): p. 1238-43.
508. Tosato, G., et al., Abnormally elevated frequency of Epstein-Barr virus-infected B cells in the blood of patients with rheumatoid arthritis. *J Clin Invest,* 1984. **73**(6): p. 1789-95.
509. Scotet, E., et al., T cell response to Epstein-Barr virus transactivators in chronic rheumatoid arthritis. *J Exp Med*, 1996. **184**(5): p. 1791-800.
510. Balandraud, N., et al., Epstein-Barr virus load in the peripheral blood of patients with rheumatoid arthritis: accurate quantification using real-time polymerase chain reaction. *Arthritis Rheum*, 2003. **48**(5): p. 1223-8.
511. Lunemann, J.D., et al., Increased frequency of EBV-specific effector memory CD8+ T cells correlates with higher viral load in rheumatoid arthritis. *J Immunol,* 2008. **181**(2): p. 991-1000.
512. Challoner, P.B., et al., Plaque-associated expression of human herpesvirus 6 in multiple sclerosis. *Proc Natl Acad Sci USA*, 1995. **92**(16): p. 7440-4.
513. Soldan, S.S., et al., Association of human herpes virus 6 (HHV-6) with multiple sclerosis: increased IgM response to HHV-6 early antigen and detection of serum HHV-6 DNA. *Nat Med*, 1997. **3**(12): p. 1394-7.
514. Johnson, R.T., et al., Measles encephalomyelitis--clinical and immunologic studies. *N Engl J Med*, 1984. **310**(3): p. 137-41.

515. Link, H., et al., Virus-reactive and autoreactive T cells are accumulated in cerebrospinal fluid in multiple sclerosis. *J Neuroimmunol*, 1992. **38**(1-2): p. 63-73.
516. Jarius, S., et al., The intrathecal, polyspecific antiviral immune response: Specific for MS or a general marker of CNS autoimmunity? *J Neurol Sci,* 2009. **280**(1-2): p. 98-100.
517. Jones, D.B. and I. Crosby, Proliferative lymphocyte responses to virus antigens homologous to GAD65 in IDDM. *Diabetologia*, 1996. **39**(11): p. 1318-24.
518. Horwitz, M.S., et al., Presented antigen from damaged pancreatic beta cells activates autoreactive T cells in virus-mediated autoimmune diabetes. *J Clin Invest*, 2002. **109**(1): p. 79-87.
519. Ylipaasto, P., et al., Enterovirus infection in human pancreatic islet cells, islet tropism in vivo and receptor involvement in cultured islet beta cells. *Diabetologia,* 2004. **47**(2): p. 225-39.
520. Menser, M.A., J.M. Forrest, and R.D. Bransby, Rubella infection and diabetes mellitus. *Lancet*, 1978. **1**(8055): p. 57-60.
521. Ou, D., et al., Cross-reactive rubella virus and glutamic acid decarboxylase (65 and 67) protein determinants recognised by T cells of patients with type I diabetes mellitus. *Diabetologia,* 2000. **43**(6): p. 750-62.
522. Saal, J.G., et al., Persistence of B19 parvovirus in synovial membranes of patients with rheumatoid arthritis. *Rheumatol Int*, 1992. **12**(4): p. 147-51.
523. Kozireva, S.V., et al., Incidence and clinical significance of parvovirus B19 infection in patients with rheumatoid arthritis. *J Rheumatol,* 2008. **35**(7): p. 1265-70.
524. Bach, J.F., Infections and autoimmune diseases. *J Autoimmun*, 2005. **25** Suppl.: p. 74-80.
525. Leibowitz, U., et al., Epidemiological study of multiple sclerosis in Israel. II. Multiple sclerosis and level of sanitation. *J Neurol Neurosurg Psychiatry*, 1966. **29**(1): p. 60-8.
526. Strachan, D.P., Hay fever, hygiene, and household size. *BMJ*, 1989. **299**(6710): p. 1259-60.
527. Okada, H., et al., The 'hygiene hypothesis' for autoimmune and allergic diseases: an update. *Clin Exp Immunol*, 2010. **160**(1): p. 1-9.
528. Kondrashova, A., et al., A six-fold gradient in the incidence of type 1 diabetes at the eastern border of Finland. *Annals of Medicine*, 2005. **37**(1): p. 67-72.
529. Leibowitz, U., E. Kahana, and M. Alter, The changing frequency of multiple sclerosis in Israel. *Arch Neurol,* 1973. **29**(2): p. 107-10.
530. Bodansky, H.J., et al., Evidence for an environmental effect in the aetiology of insulin dependent diabetes in a transmigratory population. *BMJ*, 1992. **304**(6833): p. 1020-2.
531. Detels, R., J.A. Brody, and A.H. Edgar, Multiple sclerosis among American, Japanese and Chinese migrants to California and Washington. *J Chronic Dis*, 1972. **25**(1): p. 3-10.
532. Staines, A., et al., Incidence of insulin dependent diabetes mellitus in Karachi, Pakistan. *Arch Dis Child*, 1997. **76**(2): p. 121-3.
533. Patterson, C.C., D.J. Carson, and D.R. Hadden, Epidemiology of childhood IDDM in Northern Ireland 1989-1994: low incidence in areas with highest population density and most household crowding. Northern Ireland Diabetes Study Group. *Diabetologia*, 1996. **39**(9): p. 1063-9.
534. Hammond, S.R., et al., Multiple sclerosis in Australia: socioeconomic factors. *J Neurol Neurosurg Psychiatry*, 1996. **61**(3): p. 311-3.
535. Hempel, K., et al., Unresponsiveness to experimental allergic encephalomyelitis in Lewis rats pretreated with complete Freund's adjuvant. *Int Arch Allergy Appl Immunol* , 1985. **76**(3): p. 193-9.

536. Alyanakian, M.A., et al., Transforming growth factor-beta and natural killer T cells are involved in the protective effect of a bacterial extract on type 1 diabetes. *Diabetes*, 2006. **55**(1): p. 179-85.
537. Zaccone, P., et al., Schistosoma mansoni egg antigens induce Treg that participate in diabetes prevention in NOD mice. *Eur J Immunol*, 2009. **39**(4): p. 1098-107.
538. Correale, J. and M. Farez, Association between parasite infection and immune responses in multiple sclerosis. *Ann Neurol*, 2007. **61**(2): p. 97-108.
539. Feillet, H. and J.F. Bach, On the mechanisms of the protective effect of infections on type 1 diabetes. *Clin Dev Immunol*, 2004. **11**(3-4): p. 191-4.
540. Chatenoud, L., et al., 99th Dahlem conference on infection, inflammation and chronic inflammatory disorders: immune therapies of type 1 diabetes: new opportunities based on the hygiene hypothesis. *Clin Exp Immunol*, 2010. **160**(1): p. 106-12.
541. Belkaid, Y., O. Liesenfeld, and R.M. Maizels, 99th Dahlem conference on infection, inflammation and chronic inflammatory disorders: induction and control of regulatory T cells in the gastrointestinal tract: consequences for local and peripheral immune responses. *Clin Exp Immunol*, 2010. **160**(1): p. 35-41.
542. Belkaid, Y., et al., CD4+CD25+ regulatory T cells control Leishmania major persistence and immunity. *Nature*, 2002. **420**(6915): p. 502-7.
543. Correale, J. and M.F. Farez, The impact of parasite infections on the course of multiple sclerosis. *J Neuroimmunol*, 2011. **233**(1-2): p. 6-11.
544. Wong, F.S. and L. Wen, Toll-like receptors and diabetes. *Ann N Y Acad Sci*, 2008. **1150**: p. 123-32.
545. Poskanzer, D.C., K. Schapira, and H. Miller, Multiple Sclerosis and Poliomyelitis. *Lancet*, 1963. **2**(7314): p. 917-21.
546. Hafler, D.A., The distinction blurs between an autoimmune versus microbial hypothesis in multiple sclerosis. *J Clin Invest*, 1999. **104**(5): p. 527-9.
547. McKinney, P.A., et al., Early social mixing and childhood Type 1 diabetes mellitus: a case-control study in Yorkshire, UK. *Diabet Med*, 2000. **17**(3): p. 236-42.
548. Cardwell, C.R., D.J. Carson, and C.C. Patterson, Parental age at delivery, birth order, birth weight and gestational age are associated with the risk of childhood Type 1 diabetes: a UK regional retrospective cohort study. *Diabet Med*, 2005. **22**(2): p. 200-6.
549. Bingley, P.J., et al., Influence of maternal age at delivery and birth order on risk of type 1 diabetes in childhood: prospective population based family study. Bart's-Oxford Family Study Group. *BMJ*, 2000. **321**(7258): p. 420-4.
550. Waldhoer, T., B. Rami, and E. Schober, Perinatal risk factors for early childhood onset type 1 diabetes in Austria - a population-based study (1989-2005). *Pediatr Diabetes*, 2008. **9**(3 Pt 1): p. 178-81.
551. Cardwell, C.R., et al., Atopy, home environment and the risk of childhood-onset type 1 diabetes: a population-based case-control study. *Pediatr Diabetes*, 2008. **9**(3 Pt 1): p. 191-6.
552. Cardwell, C.R., et al., Birth order and childhood type 1 diabetes risk: a pooled analysis of 31 observational studies. *Int J Epidemiol*, 2011. **40**(2): p. 363-74.
553. Ponsonby, A.L., et al., Exposure to infant siblings during early life and risk of multiple sclerosis. *JAMA*, 2005. **293**(4): p. 463-9.
554. Sadovnick, A.D., I.M. Yee, and G.C. Ebers, Multiple sclerosis and birth order: a longitudinal cohort study. *Lancet* Neurol, 2005. **4**(10): p. 611-7.
555. Montgomery, S.M., et al., Parental age, family size, and risk of multiple sclerosis. *Epidemiology*, 2004. **15**(6): p. 717-23.
556. Bager, P., et al., Sibship characteristics and risk of multiple sclerosis: a nationwide cohort study in Denmark. *Am J Epidemiol*, 2006. **163**(12): p. 1112-7.

557. Gilden, D.H., Infectious causes of multiple sclerosis. *Lancet Neurol*, 2005. **4**(3): p. 195-202.
558. Sriram, S., et al., Chlamydia pneumoniae infection of the central nervous system in multiple sclerosis. *Ann Neurol*, 1999. **46**(1): p. 6-14.
559. Kuo, C.C., et al., Chlamydia pneumoniae (TWAR). *Clin Microbiol Rev* , 1995. **8**(4): p. 451-61.
560. Sriram, S., et al., Detection of chlamydial bodies and antigens in the central nervous system of patients with multiple sclerosis. *J Infect Dis*, 2005. **192**(7): p. 1219-28.
561. Munger, K.L., et al., Infection with Chlamydia pneumoniae and risk of multiple sclerosis. *Epidemiology*, 2003. **14**(2): p. 141-7.
562. Lunemann, J.D., et al., Epstein-Barr virus: environmental trigger of multiple sclerosis? *J Virol,* 2007. **81**(13): p. 6777-84.
563. Ascherio, A., et al., Epstein-Barr virus antibodies and risk of multiple sclerosis: a prospective study. *JAMA,* 2001. **286**(24): p. 3083-8.
564. DeLorenze, G.N., et al., Epstein-Barr virus and multiple sclerosis: evidence of association from a prospective study with long-term follow-up. *Arch Neurol,* 2006. **63**(6): p. 839-44.
565. Sundstrom, P., et al., Antibodies to specific EBNA-1 domains and HLA DRB1*1501 interact as risk factors for multiple sclerosis. *J Neuroimmunol*, 2009. **215**(1-2): p. 102-7.
566. Levin, L.I., et al., Primary infection with the Epstein-Barr virus and risk of multiple sclerosis. *Ann Neurol,* 2010. **67**(6): p. 824-30.
567. Ramagopalan, S.V., et al., Association of infectious mononucleosis with multiple sclerosis. A population-based study. *Neuroepidemiology*, 2009. **32**(4): p. 257-62.
568. Handel, A.E., et al., An updated meta-analysis of risk of multiple sclerosis following infectious mononucleosis. *PLoS One*, 2010. **5**(9).
569. Nielsen, T.R., et al., Effects of infectious mononucleosis and HLA-DRB1*15 in multiple sclerosis. *Mult Scler*, 2009. **15**(4): p. 431-6.
570. Salvetti, M., G. Giovannoni, and F. Aloisi, Epstein-Barr virus and multiple sclerosis. *Curr Opin Neurol*, 2009. **22**(3): p. 201-6.
571. Ascherio, A. and K.L. Munger, 99th Dahlem conference on infection, inflammation and chronic inflammatory disorders: Epstein-Barr virus and multiple sclerosis: epidemiological evidence. *Clin Exp Immunol,* 2010. **160**(1): p. 120-4.
572. Lucas, R.M., et al., Epstein-Barr virus and multiple sclerosis. *J Neurol Neurosurg Psychiatry*, 2011. **82**(10): p. 1142-8.
573. Goldacre, M.J., et al., Multiple sclerosis after infectious mononucleosis: record linkage study. *J Epidemiol Community Health,* 2004. **58**(12): p. 1032-5.
574. Granieri, E. and I. Casetta, Selected reviews. Common childhood and adolescent infections and multiple sclerosis. *Neurology*, 1997. **49**(2 Suppl. 2): p. S42-54.
575. Alotaibi, S., et al., Epstein-Barr virus in pediatric multiple sclerosis. *JAMA*, 2004. **291**(15): p. 1875-9.
576. Banwell, B., et al., Clinical features and viral serologies in children with multiple sclerosis: a multinational observational study. *Lancet Neurol,* 2007. **6**(9): p. 773-81.
577. Henle, W., G. Henle, and E.T. Lennette, The Epstein-Barr virus. *Sci Am* , 1979. **241**(1): p. 48-59.
578. Ascherio, A. and K.L. Munger, Environmental risk factors for multiple sclerosis. Part II: Noninfectious factors. *Ann Neurol*, 2007. **61**(6): p. 504-13.
579. Lauer, K., Environmental risk factors in multiple sclerosis. *Expert Rev Neurother*, 2010. **10**(3): p. 421-40.

580. Shirani, A. and H. Tremlett, The effect of smoking on the symptoms and progression of multiple sclerosis: a review. *J Inflamm Res,* 2010. **3**: p. 115-26.
581. Handel, A.E., et al., Smoking and multiple sclerosis: an updated meta-analysis. *PLoS One*, 2011. **6**(1): p. e16149.
582. Healy, B.C., et al., Smoking and disease progression in multiple sclerosis. *Arch Neurol*, 2009. **66**(7): p. 858-64.
583. Pittas, F., et al., Smoking is associated with progressive disease course and increased progression in clinical disability in a prospective cohort of people with multiple sclerosis. *J Neurol,* 2009. **256**(4): p. 577-85.
584. Karvonen, M., et al., Importance of smoking for Chlamydia pneumoniae seropositivity. *Int J Epidemiol*, 1994. **23**(6): p. 1315-21.
585. Mikaeloff, Y., et al., Parental smoking at home and the risk of childhood-onset multiple sclerosis in children. *Brain*, 2007. **130**(Pt 10): p. 2589-95.
586. Montgomery, S.M., et al., Maternal smoking during pregnancy and multiple sclerosis amongst offspring. *Eur J Neurol,* 2008. **15**(12): p. 1395-9.
587. Akerblom, H.K., et al., Environmental factors in the etiology of type 1 diabetes. *Amer J Med Genet,* 2002. **115**(1): p. 18-29.
588. The Environmental Determinants of Diabetes in the Young (TEDDY) Study. *Ann N Y Acad Sci*, 2008. **1150**: p. 1-13.
589. Yoon, J.W., A new look at viruses in type 1 diabetes. *Diabetes Metab Rev,* 1995. **11**(2): p. 83-107.
590. Honeyman, M.C., et al., Association between rotavirus infection and pancreatic islet autoimmunity in children at risk of developing type 1 diabetes. *Diabetes*, 2000. **49**(8): p. 1319-24.
591. Gamble, D.R., et al., Viral antibodies in diabetes mellitus. *BMJ*, 1969. **3**(5671): p. 627-30.
592. van Belle, T.L., K.T. Coppieters, and M.G. von Herrath, Type 1 diabetes: etiology, immunology, and therapeutic strategies. *Physiol Rev* , 2011. **91**(1): p. 79-118.
593. Gamble, D.R. and K.W. Taylor, Seasonal incidence of diabetes mellitus. BMJ, 1969. **3**(5671): p. 631-3.
594. Hiltunen, M., et al., Islet cell antibody seroconversion in children is temporally associated with enterovirus infections. Childhood Diabetes in Finland (DiMe) Study Group. *J Infect Dis*, 1997. **175**(3): p. 554-60.
595. Hyoty, H., et al., A prospective study of the role of coxsackie B and other enterovirus infections in the pathogenesis of IDDM. Childhood Diabetes in Finland (DiMe) Study Group. *Diabetes,* 1995. **44**(6): p. 652-7.
596. Lonnrot, M., et al., Enterovirus RNA in serum is a risk factor for beta-cell autoimmunity and clinical type 1 diabetes: a prospective study. Childhood Diabetes in Finland (DiMe) Study Group. *J Med Virol,* 2000. **61**(2): p. 214-20.
597. Lonnrot, M., et al., Enterovirus infection as a risk factor for beta-cell autoimmunity in a prospectively observed birth cohort: the Finnish Diabetes Prediction and Prevention Study. *Diabetes,* 2000. **49**(8): p. 1314-8.
598. Salminen, K., et al., Enterovirus infections are associated with the induction of beta-cell autoimmunity in a prospective birth cohort study. *J Med Virol*, 2003. **69**(1): p. 91-8.
599. Fuchtenbusch, M., et al., No evidence for an association of coxsackie virus infections during pregnancy and early childhood with development of islet autoantibodies in offspring of mothers or fathers with type 1 diabetes. *J Autoimmun*, 2001. **17**(4): p. 333-40.
600. Graves, P.M., et al., Prospective study of enteroviral infections and development of beta-cell autoimmunity. Diabetes autoimmunity study in the young (DAISY). *Diabetes Res Clin Pract*, 2003. **59**(1): p. 51-61.
601. Tauriainen, S., et al., Enteroviruses in the pathogenesis of type 1 diabetes. *Semin Immunopathol,* 2011. **33**(1): p. 45-55.

602. Lonnrot, M., et al., Enterovirus antibodies in relation to islet cell antibodies in two populations with high and low incidence of type 1 diabetes. *Diabetes Care*, 1999. **22**(12): p. 2086-8.
603. Viskari, H.R., et al., Can enterovirus infections explain the increasing incidence of type 1 diabetes? *Diabetes Care,* 2000. **23**(3): p. 414-6.
604. Dahlquist, G.G., et al., Maternal enteroviral infection during pregnancy as a risk factor for childhood IDDM. A population-based case-control study. *Diabetes,* 1995. **44**(4): p. 408-13.
605. Stene, L.C., et al., Perinatal factors and development of islet autoimmunity in early childhood: the diabetes autoimmunity study in the young. *Am J Epidemiol,* 2004. **160**(1): p. 3-10.
606. Harrison, L.C. and M.C. Honeyman, Cow's milk and type 1 diabetes: the real debate is about mucosal immune function. *Diabetes,* 1999. **48**(8): p. 1501-7.
607. Carmona, L., et al., Rheumatoid arthritis. *Best Pract Res Clin Rheumatol,* 2010. **24**(6): p. 733-45.
608. Scott, D.L., F. Wolfe, and T.W. Huizinga, Rheumatoid arthritis. *Lancet*, 2010. **376**(9746): p. 1094-108.
609. Mackenzie, A.R. and J. Dawson, Could rheumatoid arthritis have an infectious aetiology? *Drug Discov Today* , 2005. **2**(3): p. 345-49.
610. Meron, M.K., et al., Infectious aspects and the etiopathogenesis of rheumatoid arthritis. *Clin Rev Allergy Immunol* , 2010. **38**(2-3): p. 287-91.
611. Kerr, J.R., Pathogenesis of human parvovirus B19 in rheumatic disease. *Ann Rheum Dis,* 2000. **59**(9): p. 672-83.
612. Mandl, L.A., et al., Is birthweight associated with risk of rheumatoid arthritis? Data from a large cohort study. *Ann Rheum Dis,* 2009. **68**(4): p. 514-8.
613. Avgil, M. and A. Omoy, Herpes simplex virus and Epstein-Barr virus infections in pregnancy: consequences of neonatal or intrauterine infection. *Reprod Toxicol* , 2006. **21**(4): p. 436-445.
614. Pugliatti, M., et al., Evidence of early childhood as the susceptibility period in multiple sclerosis: space-time cluster analysis in a Sardinian population. *Am J Epidemiol,* 2006. **164**(4): p. 326-33.
615. Riise, T., et al., Clustering of residence of multiple sclerosis patients at age 13 to 20 years in Hordaland, Norway. *Am J Epidemiol,* 1991. **133**(9): p. 932-9.
616. Hernan, M.A., et al., Multiple sclerosis and age at infection with common viruses. *Epidemiology,* 2001. **12**(3): p. 301-6.
617. Marrie, R.A., et al., Multiple sclerosis and antecedent infections: a case-control study. *Neurology*, 2000. **54**(12): p. 2307-10.
618. Aberg, N., I. Engstrom, and U. Lindberg, Allergic diseases in Swedish school children. *Acta Paediatr Scand*, 1989. **78**(2): p. 246-52.
619. Weiland, S.K., et al., Climate and the prevalence of symptoms of asthma, allergic rhinitis, and atopic eczema in children. *Occup Environ Med* , 2004. **61**(7): p. 609-15.
620. Reynolds, N.J., et al., Narrow-band ultraviolet B and broad-band ultraviolet A phototherapy in adult atopic eczema: a randomised controlled trial. *Lancet*, 2001. **357**(9273): p. 2012-6.
621. Selgrade, M.K., M.H. Repacholi, and H.S. Koren, Ultraviolet radiation-induced immune modulation: potential consequences for infectious, allergic, and autoimmune disease. *Environ Health Perspect*, 1997. **105**(3): p. 332-4.
622. Platts-Mills, T., et al., Sensitisation, asthma, and a modified Th2 response in children exposed to cat allergen: a population-based cross-sectional study. *Lancet*, 2001. **357**(9258): p. 752-6.
623. Platts-Mills, T.A., The role of allergens in the induction of asthma. *Curr Allergy Asthma Rep*, 2002. **2**(2): p. 175-80.

624. Kero, J., et al., Could TH1 and TH2 diseases coexist? Evaluation of asthma incidence in children with coeliac disease, type 1 diabetes, or rheumatoid arthritis: a register study. *J Allergy Clin Immunol,* 2001. **108**(5): p. 781-3.
625. *National Health Survey: Data Reference Package* (ABS Catalogue No. 4363.0.001). 1995, Canberra: Australian Bureau of Statistics.
626. *National Health Survey: Summary Results, Australian States and Territories* (ABS Catalogue No. 4368.0). 1995, Canberra: Australian Bureau of Statistics. 55.
627. Gies, P., et al., Global Solar UV Index: Australian measurements, forecasts and comparison with the UK. *Photochem Photobiol,* 2004. **79**(1): p. 32-9.
628. Lemus-Deschamps, L., L. Rikus, and P. Gies, The operational Australian ultraviolet index forecast 1997. *Meteorol Appl,* 1999. **6**: p. 241-251.
629. Hersey, P., et al., Alteration of T cell subsets and induction of suppressor T cell activity in normal subjects after exposure to sunlight. *J Immunol,* 1983. **131**(1): p. 171-4.
630. McNally, N.J., et al., Is there a geographical variation in eczema prevalence in the UK? Evidence from the 1958 British Birth Cohort Study. *Br J Dermatol,* 2000. **142**(4): p. 712-20.
631. Gies, P., et al., Ambient solar UVR, personal exposure and protection. *J Epidemiol ,* 1999. **9**(6 Suppl.): p. S115-22.
632. Infections and vaccinations as risk factors for childhood type I (insulin-dependent) diabetes mellitus: a multicentre case-control investigation. EURODIAB Substudy 2 Study Group. *Diabetologia,* 2000. **43**(1): p. 47-53.
633. Wilson, C., H. Tiwana, and A. Ebringer, Molecular mimicry between HLA-DR alleles associated with rheumatoid arthritis and Proteus mirabilis as the aetiological basis for autoimmunity. *Microbes Infect*, 2000. **2**(12): p. 1489-96.
634. Blasi, F., et al., Atypical pathogens and asthma: can they influence the natural history of the disease? *Monaldi Arch Chest Dis*, 2001. **56**(3): p. 276-80.
635. Pearce, N., J. Pekkanen, and R. Beasley, How much asthma is really attributable to atopy? *Thorax*, 1999. **54**(3): p. 268-72.
636. Couper, D., A.L. Ponsonby, and T. Dwyer, Determinants of dust mite allergen concentrations in infant bedrooms in Tasmania. *Clin Exp Allergy*, 1998. **28**(6): p. 715-23.
637. Hughes, A.M., et al., The role of latitude, ultraviolet radiation exposure and vitamin D in childhood asthma and hayfever: an Australian multicenter study. *Pediatr Allergy Immunol*, 2011. **22**(3): p. 327-33.
638. Mason, R.S. and T.H. Diamond, Vitamin D deficiency and multicultural Australia. *Med J Aust,* 2001. **175**(5): p. 236-7.
639. Davies, G., et al., A systematic review and meta-analysis of Northern Hemisphere season of birth studies in schizophrenia. *Schizophr Bull* , 2003. **29**(3): p. 587-93.
640. McGrath, J.J. and J.L. Welham, Season of birth and schizophrenia: a systematic review and meta-analysis of data from the Southern Hemisphere. *Schizophr Res,* 1999. **35**(3): p. 237-42.
641. McGrath, J., J.P. Selten, and D. Chant, Long-term trends in sunshine duration and its association with schizophrenia birth rates and age at first registration--data from Australia and the Netherlands. *Schizophr Res,* 2002. **54**(3): p. 199-212.
642. Torrey, E.F., et al., Seasonality of births in schizophrenia and bipolar disorder: a review of the literature. *Schizophr Res,* 1997. **28**(1): p. 1-38.
643. Sadovnick, A.D., et al., A timing-of-birth effect on multiple sclerosis clinical phenotype. *Neurology*, 2007. **69**(1): p. 60-2.
644. Tremlett, H.L. and V.A. Devonshire, Does the season or month of birth influence disease progression in multiple sclerosis? *Neuroepidemiology*, 2006. **26**(4): p. 195-8.

645. Hammond, S.R., et al., The relationship between multiple sclerosis frequency and latitude within Australia., in *Multiple sclerosis research: proceedings of the International Multiple Sclerosis Conference - an update on Multiple Sclerosis, Rome, 15-17 September 1988*, M.A. Battaglia, Editor. 1989, Elsevier: Amsterdam. p. 171-7.
646. Hammond, S.R., et al., The epidemiology of multiple sclerosis in Queensland, Australia. *J Neurol Sci,* 1987. **80**(2-3): p. 185-204.
647. Hammond, S.R., et al., The epidemiology of multiple sclerosis in three Australian cities: Perth, Newcastle and Hobart. *Brain*, 1988. **111 (Pt 1)**: p. 1-25.
648. Hammond, S.R., et al., The epidemiology of multiple sclerosis in Western Australia. *Aust N Z J Med* 1988. **18**: p. 102-110.
649. Hammond, S.R., et al., The clinical profile of MS in Australia: a comparison between medium- and high-frequency prevalence zones. *Neurology*, 1988. **38**(6): p. 980-6.
650. Hammond, S.R., et al., Multiple sclerosis in Australia: prognostic factors. *J Clin Neurosci*, 2000. **7**(1): p. 16-9.
651. Kurtzke, J.F., Neurologic impairment in multiple sclerosis and the disability status scale. *Acta Neurol Scand,* 1970. **46**(4): p. 493-512.
652. Richards, R.G., et al., A review of the natural history and epidemiology of multiple sclerosis: implications for resource allocation and health economic models. *Health Technol Assess*, 2002. **6**(10): p. 1-73.
653. Sadovnick, A.D., et al., Life expectancy in patients attending multiple sclerosis clinics. *Neurology*, 1992. **42**(5): p. 991-4.
654. Pagano, M. and K. Gauvreau, *Principles of Biostatistics.* Second ed. 2000, Pacific Grove: Duxbury-Thomson Learning. 525p.
655. Doblhammer, G. and J.W. Vaupel, Lifespan depends on month of birth. *Proc Natl Acad Sci USA*, 2001. **98**(5): p. 2934-9.
656. Mattock, J.N. and M.C. Lyons, eds. *Hippocrates: On Endemic Diseases: Airs, Waters and Places*. Vol. 5. 1969, Heffer and Sons: Cambridge.
657. Sadovnick, A.D. and I.M. Yee, Season of birth in multiple sclerosis. *Acta Neurol Scand*, 1994. **89**(3): p. 190-1.
658. Fernandes de Abreu, D.A., et al., Season of birth and not vitamin D receptor promoter polymorphisms is a risk factor for multiple sclerosis. *Mult Scler*, 2009. **15**(10): p. 1146-52.
659. Bayes, H.K., C.J. Weir, and C. O'Leary, Timing of birth and risk of multiple sclerosis in the Scottish population. *Eur Neurol,* 2010. **63**(1): p. 36-40.
660. Gies, H.P., et al., The ARL solar-UVR measurement network: calibration and results. *Ultraviolet Tech V, SPIE (Proc)* 1994. **2282**: p. 274-284.
661. Sayers, A., et al., Predicting ambient ultraviolet from routine meteorological data; its potential use as an instrumental variable for vitamin D status in pregnancy in a longitudinal birth cohort in the UK. *Int J Epidemiol*, 2009. **38**(6): p. 1681-8.
662. Hamilton, L.C., *Statistics with Stata - Updated for Version 8*. 2004, Belmont: Brooks/Cole-Thomson Learning. 409p.
663. Long, J.S. and J. Freese, *Regression Models for Categorical Dependent Variables Using Stata*. 2001, College Station, TX: Stata Press.
664. Hare, E.H., Season of birth in schizophrenia and neurosis. *Am J Psychiatry* , 1975. **132**(11): p. 1168-71.
665. Whitacre, C.C., S.C. Reingold, and P.A. O'Looney, A gender gap in autoimmunity. *Science*, 1999. **283**(5406): p. 1277-8.
666. Dupont, W.D., *Statistical Modelling for Biomedical Researchers: A Simple Introduction to the Analysis of Complex Data*. 2002, Cambridge: Cambridge University Press. 386p.
667. McDowell, T.Y., et al., Time of birth, residential solar radiation and age at onset of multiple sclerosis. *Neuroepidemiology*, 2010. **34**(4): p. 238-44.

668. McGrath, J., J. Welham, and M. Pemberton, Month of birth, hemisphere of birth and schizophrenia. *Br J Psychiatry,* 1995. **167**(6): p. 783-5.
669. Mulligan, M.L., et al., Implications of vitamin D deficiency in pregnancy and lactation. *Am J Obstet Gynecol* , 2010. **202**(5): p. 429 e1-9.
670. Ramagopalan, S.V., et al., HLA-DRB1 and month of birth in multiple sclerosis. *Neurology*, 2009. **73**(24): p. 2107-11.
671. Eyles, D., et al., Vitamin D3 and brain development. *Neuroscience*, 2003. **118**(3): p. 641-53.
672. Eyles, D.W., et al., Distribution of the vitamin D receptor and 1 alpha-hydroxylase in human brain. *J Chem Neuroanat,* 2005. **29**(1): p. 21-30.
673. Tustin, K., J. Gross, and H. Hayne, Maternal exposure to first-trimester sunshine is associated with increased birth weight in human infants. *Dev Psychobiol*, 2004. **45**(4): p. 221-30.
674. Javaid, M.K., et al., Maternal vitamin D status during pregnancy and childhood bone mass at age 9 years: a longitudinal study. *Lancet,* 2006. **367**(9504): p. 36-43.
675. Mirzaei, F., et al., Gestational vitamin D and the risk of multiple sclerosis in offspring. *Ann Neurol,* 2011. **70**(1): p. 30-40.
676. Delvin, E.E., et al., Control of vitamin D metabolism in preterm infants: feto-maternal relationships. *Arch Dis Child*, 1982. **57**(10): p. 754-7.
677. Salle, B.L., et al., Perinatal metabolism of vitamin D. *Am J Clin Nutr*, 2000. **71**(5 Suppl.): p. 1317S-24S.
678. Vitamin D supplementation: Recommendations for Canadian mothers and infants. *Paediatr Child Health*, 2007. **12**(7): p. 583-98.
679. Munns, C., et al., Prevention and treatment of infant and childhood vitamin D deficiency in Australia and New Zealand: a consensus statement. *Med J Aust*, 2006. **185**(5): p. 268-72.
680. Koch, M., J. De Keyser, and H. Tremlett, Timing of birth and disease progression in multiple sclerosis. *Mult Scler*, 2008. **14**(6): p. 793-8.
681. Jakovcevski, I., Z. Mo, and N. Zecevic, Down-regulation of the axonal polysialic acid-neural cell adhesion molecule expression coincides with the onset of myelination in the human fetal forebrain. *Neuroscience*, 2007. **149**(2): p. 328-37.
682. Abe, S., et al., Semiquantitative assessment of myelination using magnetic resonance imaging in normal fetal brains. *Prenat Diagn*, 2004. **24**(5): p. 352-7.
683. Gogal, R.M., Jr. and S.D. Holladay, Perinatal TCDD exposure and the adult onset of autoimmune disease. *J Immunotoxicol*, 2008. **5**(4): p. 413-8.
684. Meyer, U., B.K. Yee, and J. Feldon, The neurodevelopmental impact of prenatal infections at different times of pregnancy: the earlier the worse? *Neuroscientist,* 2007. **13**(3): p. 241-56.
685. Smith, S.E., et al., Maternal immune activation alters fetal brain development through interleukin-6. *J Neurosci,* 2007. **27**(40): p. 10695-702.
686. Koch-Henriksen, N., An epidemiological study of multiple sclerosis. Familial aggregation social determinants, and exogenic factors. *Acta Neurol Scand Suppl*, 1989. **124**: p. 1-123.
687. Isager, H., E. Andersen, and K. Hyllested, Risk of multiple sclerosis inversely associated with birth order position. *Acta Neurol Scand*, 1980. **61**(6): p. 393-6.
688. James, W.H., Multiple sclerosis and birth order. *J Epidemiol Community Health,* 1984. **38**(1): p. 21-2.
689. James, W.H., Further evidence in support of the hypothesis that one cause of multiple sclerosis is childhood infection. *Neuroepidemiology*, 1988. **7**(3): p. 130-3.
690. Zilber, N., et al., Multiple sclerosis and birth order. *Acta Neurol Scand,* 1988. **78**(4): p. 313-7.

691. Alperovitch, A., P. Le Canuet, and R. Marteau, Birth order and risk of multiple sclerosis: are they associated and how? *Acta Neurol Scand*, 1981. **63**(2): p. 136-8.
692. Cendrowski, W.S., Birth order of multiple sclerotic subjects. *Psychiatr Neurol (Basel)*, 1966. **151**(3): p. 161-7.
693. Ahlgren, C. and O. Andersen, No major birth order effect on the risk of multiple sclerosis. *Neuroepidemiology*, 2005. **24**(1-2): p. 38-41.
694. Bager, P., et al., Childhood infections and risk of multiple sclerosis. *Brain*, 2004.
695. Cripps, J., A. Rudd, and G.C. Ebers, Birth order and multiple sclerosis. *Acta Neurol Scand*, 1982. **66**(3): p. 342-6.
696. Gaudet, J.P., et al., A study of birth order and multiple sclerosis in multiplex families. *Neuroepidemiology*, 1995. **14**(4): p. 188-92.
697. Gaudet, J.P., et al., Is sporadic MS caused by an infection of adolescence and early adulthood? A case-control study of birth order position. *Acta Neurol Scand*, 1995. **91**(1): p. 19-21.
698. Matias-Guiu, J., et al., Birth order in multiple sclerosis: a population-based case-control study in Alcoi, Spain. *Neuroepidemiology*, 1994. **13**(3): p. 129-30.
699. Operskalski, E.A., et al., A case-control study of multiple sclerosis. *Neurology*, 1989. **39**(6): p. 825-9.
700. Visscher, B.R., et al., Birth order and multiple sclerosis. *Acta Neurol Scand*, 1982. **66**(2): p. 209-15.
701. Greenwood, M. and U. Yule, On the determination of size of family and of the distribution of characters in order of birth from samples taken through members of sibships. *J Roy Stat Soc*, 1914. **77**: p. 179-197.
702. Hardy, A. and M.E. Magnello, Statistical methods in epidemiology: Karl Pearson, Ronald Ross, Major Greenwood and Austin Bradford Hill, 1900-1945. *Soz Praventivmed*, 2002. **47**(2): p. 80-9.
703. Gregory, I., An analysis of familial data on psychiatric patients: parental age, family size, birth order, and ordinal position. *Br J Prev Soc Med*, 1958. **12**(1): p. 42-59.
704. Antonovsky, A., et al., Epidemiologic Study of Multiple Sclerosis in Israel. I. An Overall Review of Methods and Findings. *Arch Neurol*, 1965. **13**: p. 183-93.
705. Ramagopalan, S.V., et al., No effect of parental age on risk of multiple sclerosis: a population-based study. *Neuroepidemiology*, 2010. **34**(2): p. 106-9.
706. Mkhikian, H., et al., Genetics and the environment converge to dysregulate N-glycosylation in multiple sclerosis. *Nat Commun*, 2011. **2**: p. 334.
707. van der Mei, I.A., et al., Human leukocyte antigen-DR15, low infant sibling exposure and multiple sclerosis: gene-environment interaction. *Ann Neurol*, 2010. **67**(2): p. 261-5.
708. Ramagopalan, S.V., et al., Effects of infectious mononucleosis and HLA-DRB1*15 in multiple sclerosis. *Mult Scler*, 2010. **16**(1): p. 127-8.
709. Grant, W.B., Epidemiological evaluations of sun-induced vitamin D, in *Solar Radiation and Human Health*, E. Bjertness, Editor. 2008, The Norwegian Academy of Science and Letters: Oslo. p. 167-182.
710. Grant, W.B., Latitude and multiple sclerosis prevalence: vitamin D reduces risk of Epstein-Barr virus infection. *Mult Scler*, 2010. **16**(3): p. 373; author reply 374-5.
711. Holmoy, T., Vitamin D status modulates the immune response to Epstein Barr virus: Synergistic effect of risk factors in multiple sclerosis. *Med Hypotheses*, 2008. **70**(1): p. 66-9.
712. Hayes, C.E. and E. Donald Acheson, A unifying multiple sclerosis etiology linking virus infection, sunlight, and vitamin D, through viral interleukin-10. *Med Hypotheses*, 2008. **71**(1): p. 85-90.
713. Urashima, M., et al., Randomized trial of vitamin D supplementation to prevent seasonal influenza A in schoolchildren. *Am J Clin Nutr*, 2010. **91**(5): p. 1255-60.

714. Smith, G.D., Is the first cut the deepest? Ernst Engel on the statistical imperative of embracing the lifecourse perspective. *Int J Epidemiol,* 2011. **40**(5): p. 1135-7.
715. Stalgis-Bilinski, K.L., et al., Burning daylight: balancing vitamin D requirements with sensible sun exposure. *Med J Aust,* 2011. **194**(7): p. 345-8.
716. Ebeling, P.R., Routine screening for vitamin D deficiency in early pregnancy: past its due date? *Med J Aust*, 2011. **194**(7): p. 332-3.
717. Stanley, F.J. and C. Bower, Commentary: folate and neural tube defects--the influence of Smithells et al. on research and policy in the antipodes. *Int J Epidemiol,* 2011. **40**(5): p. 1159-60.
718. CDC, Ten great public health achievements--United States, 2001-2010. *MMWR Morb Mortal Wkly Rep,* 2011. **60**(19): p. 619-23.

APPENDIX I:

Staples J.A., A-L. Ponsonby, L.L-Y Lim and A.J. McMichael (2003). Ecologic analysis of some immune-related disorders, including type 1 diabetes, in Australia: latitude, regional ultraviolet radiation, and disease prevalence. *Environ Health Perspect* **111**(4): 518-523.

Ecologic Analysis of Some Immune-Related Disorders, Including Type 1 Diabetes, in Australia: Latitude, Regional Ultraviolet Radiation, and Disease Prevalence

Judith A. Staples, Anne-Louise Ponsonby, Lynette L-Y. Lim, and Anthony J. McMichael

National Center for Epidemiology and Population Health, The Australian National University, Canberra, Australian Capital Territory, Australia

The apparent immune-suppressive effect of ultraviolet radiation (UVR) has suggested that this environmental exposure may influence the development of immune-related disorders. Self-reported prevalence rates of type 1 diabetes mellitus, rheumatoid arthritis (RA), eczema/dermatitis, and asthma, from the 1995 Australian National Health Survey, were therefore examined by latitude and ambient level of UVR. A positive association of type 1 diabetes mellitus prevalence was found with both increasing southern latitude of residence ($r = 0.77$; $p = 0.026$) and decreasing regional annual ambient UVR ($r = -0.80$; $p = 0.018$); a 3-fold increase in prevalence from the northernmost region to the southernmost region was evident. In contrast, asthma correlated negatively with latitude ($r = -0.72$; $p = 0.046$), although the change in asthma prevalence from the north to the south of Australia was only 0.7-fold. For both RA and eczema/dermatitis, there were no statistically significant associations between latitude/UVR and disease prevalence. These ecologic data provide some support for a previously proposed beneficial effect of UVR on T-helper 1–mediated autoimmune disorders such as type 1 diabetes. The inverse association of type 1 diabetes prevalence with UVR is consistent with that previously reported for another autoimmune disease, multiple sclerosis, in Australia, and also with type 1 diabetes latitudinal gradients in the Northern Hemisphere. The finding also accords with photoimmunologic evidence of UVR-induced immunosuppression and may suggest a beneficial effect of UVR in reducing the incidence of such autoimmune conditions. In light of this study, analytic epidemiologic studies investigating risk of immune disorders in relation to personal UVR exposure in humans are required. *Key words:* asthma, Australia, autoimmune disease, ecologic analysis, eczema/dermatitis, immune disorders, latitude, rheumatoid arthritis, type 1 diabetes, ultraviolet radiation. *Environ Health Perspect* 111:518–523 (2003). doi:10.1289/ehp.5941 available via *http://dx.doi.org/* [Online 19 December 2002]

Autoimmune diseases such as type 1 diabetes mellitus, multiple sclerosis (MS), and rheumatoid arthritis (RA) are immune system disorders that share common features of self-reactive T cells and the presence of autoantibodies; as a group they affect some 5% of the population (Davidson and Diamond 2001). Although their precise etiologies are unknown, these autoimmune disorders are generally agreed to reflect interactions of polygenic traits with various ill-defined environmental factors (Cantorna 2000; Dahlquist 1998; Hayes et al. 1997; Karges et al. 1995; Weinshenker 1996). Descriptive epidemiology may further elucidate the role of environmental factors in the etiology of the autoimmune diseases MS, type 1 diabetes, and RA, as well as other immune-related disturbances such as asthma and eczema/dermatitis.

MS, type 1 diabetes, and, to a lesser extent, RA in the Northern Hemisphere, particularly in Western Europe and North America, display a latitudinal gradient in disease frequency, with the prevalence of these disorders increasing at higher latitudes (Cantorna 2000; DERIG 1988; Hayes et al. 1997; Karvonen et al. 1993). MS exhibits a similar prevalence gradient in the Southern Hemisphere, in Australia and New Zealand (Hammond et al. 2000; Miller et al. 1990). In Australia, however, where the opportunity exists to study gradients in rates across a large-area population that is less ethnically and genetically diverse than across Europe, analyses for other immune-related disorders have not been done previously.

Ultraviolet radiation (UVR) reaching the earth's surface varies inversely with latitude; UVR is thus a prominent latitude-related environmental factor. Recent photoimmunologic work shows that UVR downregulates cellular immunity (Damian et al. 1998; Kelly et al. 1998), attenuating T-helper (Th)1 T-cell–mediated immune responses (Clydesdale et al. 2001). These responses are thought to be significantly involved in some autoimmune disorders such as MS, type 1 diabetes, and RA (Mackay 2000). UVR might therefore be expected to be beneficial for these disorders. Few autoimmune or other immune-related disorders, however, have been assessed ecologically with respect to UVR. A notable exception for the Southern Hemisphere is a recent analysis of MS in Australia, where the regional variation in MS prevalence was strongly inversely associated with ambient UVR levels ($r = -0.91$; $p = 0.01$) (van der Mei et al. 2001). This finding supports the possibility of UVR being a protective modulator of immune and autoimmune processes involved in the etiology of such immune disorders (McMichael and Hall 1997). Type 1 diabetes and RA were therefore chosen for ecologic analysis to determine whether these disorders showed environmental gradients similar to those observed for MS in Australia.

By downregulating Th1-mediated immunity, UVR has been thought to effect a shift from Th1- to Th2-mediated processes (Clydesdale et al. 2001). Th2 cells are responsible for immediate-type hypersensitivity to allergens such as dust mites; UVR may thus have the potential to exacerbate allergic disease (Selgrade et al. 1997). However, recent work has cast doubt on the mutual antagonism of Th1 and Th2 cytokine expression, particularly in humans (Davidson and Diamond 2001; Mackay 2000; Platts-Mills 2002; Platts-Mills et al. 2001). Ultraviolet B, particularly through interleukin 10, can inhibit both Th1- and Th2-mediated immune responses in mice (Garssen et al. 1999). In children, asthma can coexist with Th1-type disorders such as type 1 diabetes, RA, and celiac disease, suggesting a common environmental influence (Kero et al. 2001). A recent randomized controlled trial in the United Kingdom on adult atopic eczema has shown UVR, particularly narrow-band ultraviolet B, to also have a beneficial effect (Reynolds et al. 2001). Eczema/dermatitis, together with asthma, were therefore chosen to be similarly analyzed for regional prevalence gradients to compare with any associations evident for type 1 diabetes and RA.

We have examined the association between latitude and prevalence of the immune-related disorders type 1 diabetes mellitis, RA, eczema/dermatitis, and asthma in Australia, a country with a relatively genetically homogeneous population. We further considered

Address correspondence to J.A. Staples, National Center for Epidemiology and Population Health, The Australian National University, Canberra ACT 0200, Australia. Telephone: 61 2 6125 8088. Fax: 61 2 6125 5614. E-mail: judy.staples@anu.edu.au

We thank L. Lemus-Deschamps, Bureau of Meteorology Research Center, Australia, for provision of regional monthly UVR data.

The research was supported by an Australian National University (ANU) Graduate School scholarship and a supplementary scholarship from the National Center for Epidemiology and Population Health, ANU, awarded to J.A.S.

Received 16 August 2002; accepted 19 December 2002.

regional differences in ambient UVR to examine possible associations between regional and seasonal UVR levels and the prevalence of these immune disorders in Australia.

Methods

Prevalence data source. The 1995 Australian National Health Survey (NHS) was conducted by the Australian Bureau of Statistics (ABS) during the 12-month period from January 1995 to January 1996. Approximately 54,000 people from all states and territories and across all age groups provided information about their own health status. Residents of a stratified, multistage area, random sample of 23,800 private households and households in nonprivate dwellings were interviewed in person by ABS interviewers; persons within each state/territory had a known, and in the main equal, chance of selection. Residents of hospitals and other institutions were excluded, however, and the Northern Territory (NT) sample was predominantly urban. Responses were received from 91.5% of households and 97% of people from these households fully completed questionnaires. Quarterly population estimates were used by ABS for standardization, and prevalence rate estimates were adjusted for household size (ABS 1995a).

The survey interviewers recorded information on recent illness and/or long-term conditions, as reported by respondents. Specific questioning about conditions, including diabetes, arthritis, and asthma, was followed by action-based questions on recent visits to hospitals or health professionals. Information on medication usage, including insulin and medications for asthma, arthritis, and allergies, was also elicited (ABS 1995b). Disease conditions were classified by the *International Classification of Diseases, 9th Revision* (cited by ABS 1995a), categories. Information recorded in the survey was not medically verified, however.

For this study we used summary-data prevalence estimates for four immune-related disorders from the published survey results (ABS 1995c). We compared the age- and sex-standardized prevalence rates of type 1 diabetes, RA, eczema/dermatitis, and asthma, per 1,000 population, over the eight major state and territory regions of Australia, as listed in Table 1.

Latitude and ultraviolet radiation. The latitude of the regional capital city (decimal degrees south; Geoscience Australia mapping agency) was used for each region (Table 1). Three UVR measures were examined, each relating to the Australian UV index (Lemus-Deschamps et al. 1999) and expressed in milliwatts per square meter (where 1 UV index unit = 25 mW/m^2). The measures were, for each regional capital, an arithmetic mean of 12 monthly UV index averages, each monthly average calculated from single, daily (at local solar noon), erythemally weighted clear-sky UV index values derived from ozone data over the period 1979–1993 (Lemus-Deschamps L. Personal communication), a midwinter minimum (June) solar-noon UVR value, and a midsummer maximum (January) solar-noon UVR value (Table 1).

Statistical analysis. To account for the NHS sampling variability between regions, we calculated approximate standard errors (SEs) for each regional prevalence estimate, as outlined in Appendix H and "Technical Note: Sampling Variability" in the respective ABS publications (ABS 1995a, 1995c). The approximate SEs were calculated from SEs for the corresponding numerator estimates shown in Table 2, as listed by ABS for each region (ABS 1995a, 1995c). Age-standardization factors for the different regions were also applied, as recommended in the same Technical Note, and the final SEs for the prevalence rates calculated (Table 2). The reciprocal of the variance of each prevalence rate was then used to weight the relationships to account for differing sample sizes in the regions.

Associations between the environmental variables and the immune disorder prevalence rates were examined by variance-weighted least-squares regression of the prevalence rate estimates versus, first, the latitude values of the regional capital cities, and second, the three UVR measures for each of the regional capitals. The magnitude of change in disease prevalence rates over the north-south range was compared by substituting latitude values for Darwin and Hobart into the regression equations. The statistical program Stata 7.0 (Stata Corporation, College Station, Texas, USA) was used for regression analyses.

Table 1. Australian state and territory regions and approximate latitude ranges; regional capitals and their latitudes (decimal degrees south) shown together with midsummer (January) and midwinter (June) solar-noon UVR[a] for each regional capital.

State/territory region	Latitude range for region (degrees)	Regional capital	Latitude of capital (degrees)	Midsummer UVR (mW/m^2)	Midwinter UVR (mW/m^2)
Northern Territory	11–26	Darwin	12.45	339.5	205.9
Queensland	10–29	Brisbane	27.47	332.4	103.9
Western Australia	14–35	Perth	31.95	326.1	82.9
New South Wales	28–37	Sydney	33.87	306.1	66.5
South Australia	26–38	Adelaide	34.93	303.4	60.7
Australian Capital Territory	35–35.5	Canberra	35.30	302.7	55.7
Victoria	34–39	Melbourne	37.82	287.6	48.6
Tasmania	41–43.5	Hobart	42.88	256.7	31.4

[a]UVR data are monthly averages calculated from daily (at local solar noon) erythemally weighted clear-sky UV Index values derived from ozone data over the period 1979–1993 (courtesy of L. Lemus-Deschamps, Bureau of Meteorology Research Center, Australia) and expressed as mW/m^2.

Table 2. Regional immune disorder (excluding MS) prevalence rates[a] (shaded lines with SE[b] in parentheses) together with numerator data on which SE estimates were based.

State/territory	Type 1 diabetes	Rheumatoid arthritis	Eczema/dermatitis	Asthma	Regional population
Northern Territory					
No. cases	200	1,500	2,900	17,200	145,300
Prevalence	2.9 (1.80)	18.2 (5.56)	19.2 (4.43)	127.2 (13.11)	
Queensland					
No. cases	10,500	82,100	96,900	438,000	3,277,800
Prevalence	3.2 (0.62)	25.7 (1.34)	29.5 (1.40)	132.6 (2.51)	
Western Australia					
No. cases	6,900	49,000	82,400	201,500	1,732,400
Prevalence	4.2 (0.79)	29.8 (1.66)	47.1 (1.90)	115.2 (2.76)	
New South Wales					
No. cases	26,300	170,700	181,900	633,700	6,120,500
Prevalence	4.2 (0.56)	27.3 (1.24)	29.8 (1.30)	103.9 (2.10)	
South Australia					
No. cases	8,300	41,900	73,000	163,500	1,474,800
Prevalence	5.4 (0.63)	26.8 (1.15)	50.3 (1.55)	112.4 (2.16)	
Australian Capital Territory					
No. cases	1,100	5,600	14,400	35,500	304,900
Prevalence	4.6 (0.99)	21.4 (1.65)	45.5 (3.98)	111.9 (2.65)	
Victoria					
No. cases	23,800	106,400	171,500	501,500	4,503,100
Prevalence	5.2 (0.50)	23.2 (0.88)	38.2 (1.07)	111.8 (1.63)	
Tasmania					
No. cases	2,100	19,100	20,700	48,800	473,600
Prevalence	4.5 (1.14)	39.5 (2.38)	44.0 (2.53)	102.1 (3.48)	

[a]Rate per 1,000, age- and sex-standardized to the 1995/1996 Australian population. Data from ABS (1995c). [b]Approximate SE calculated from ABS-provided SE for numerator (ABS 1995a, 1995c).

Because the use of capital city latitudes may not have been representative of the regional population distribution, particularly for larger regions with the capital at either a northern or southern extreme of the region rather than medially placed, a sensitivity analysis was carried out to test the effect of using an alternative latitude value midway between the two main population centers in these regions. Results were then compared with those obtained from the regional capitals.

Results

Latitude and immune disorders. The relationships between latitude of the regional capitals and immune disorder prevalence are shown in Figure 1 as regression lines fitted to the prevalence estimates; 95% confidence intervals (CIs) are also shown, and these indicate, reciprocally, the relative weighting applied to each estimate in the regression analysis.

Prevalence of type 1 diabetes was positively correlated with latitude (Pearson $r = 0.77$; $p = 0.026$), with the prevalence increasing 2.97-fold over the north-south latitude gradient. Asthma prevalence, on the other hand, was inversely correlated with latitude (Pearson $r = -0.72$; $p = 0.046$), with the prevalence rate decreasing 0.7-fold, i.e., by approximately one-third, over the same latitude range. Although both eczema/dermatitis and, to a much lesser extent, RA showed trends of increasing prevalence with increasing latitude, these were not statistically significant (Figure 1).

Sensitivity analysis. In four of the eight regions, Western Australia (WA), South Australia (SA), Victoria (VIC), and Australian Capital Territory (ACT), at least 70% of the state or territory populations resided in the capital metropolitan area (ABS census data 1991/1996, not shown). Among the remaining regions in which the population was more dispersed outside of the capital, only NT and Queensland (QLD) were also both of wide latitudinal range and had capitals located noncentrally in terms of the regional latitude range, thus potentially biasing the associations markedly.

The effect on the initial associations of using alternative latitude values midway between the two main population centers, Darwin and Alice Springs, for NT (i.e., latitude 18.08 degrees south), and Brisbane and Townsville for QLD (i.e., latitude 23.36 degrees south), was to strengthen the statistical significance of the associations for type 1 diabetes, asthma, and eczema/dermatitis when one or both midway latitudes were substituted. For example, for type 1 diabetes and asthma, the statistical significance of the associations was raised maximally [from $p = 0.026$ to $p = 0.010$ for diabetes ($r = 0.77$ to 0.84), and from $p = 0.046$ to $p = 0.013$ ($r = -0.72$ to -0.82) for asthma] when midway latitudes were used for both regions. Both approaches unavoidably entail some exposure misclassification for dispersed regional populations. Whereas the latter (midway latitude) approach may lead to better estimation of the correlation, by using the capital latitude values for type 1 diabetes and asthma as we have done (Table 1), the estimate was somewhat biased toward the null hypothesis of no association and thus toward underestimation of r. Similarly, for eczema/dermatitis the statistical significance was raised maximally [from $p = 0.206$ to $p = 0.174$ ($r = 0.50$ to 0.54)] when only the QLD latitude was altered, but this was still nonsignificant at the 5% level. For RA, on the other hand, using both midway latitudes gave maximal change, but the statistical significance of the observed trend was lowered [from $p = 0.727$ to $p = 0.795$ ($r = 0.15$ to 0.11)]. In summary, the use of the regional-capital latitude values rather than latitude values of the midway points did not alter the overall conclusions of whether the observed associations were statistically significant at the 5% level.

Ultraviolet radiation and immune disorders. In light of the sensitivity analysis, UVR values for the regional capitals were used. A solar-noon, clear-sky measure of ambient UVR was chosen because midday, or noontime, exposure to the maximal UVR level in sunlight in Australia suppresses the systemic immune response in humans (Hersey et al. 1983). Type 1 diabetes prevalence was inversely correlated with regional average solar-noon UVR (Pearson $r = -0.80$; $p = 0.018$), whereas asthma prevalence was positively correlated with regional average solar-noon UVR (Pearson $r = 0.73$; $p = 0.040$) (Figure 2). Eczema/dermatitis and RA prevalence showed inverse trends with average solar-noon UVR, but these were not significant (Table 3). The prevalence of asthma was positively correlated with either average or midwinter solar-noon UVR but not midsummer UVR. Although type 1 diabetes prevalence was inversely correlated with all three ambient UVR measures, it was more closely correlated with average or midwinter solar-noon UVR than with midsummer UVR (Table 3).

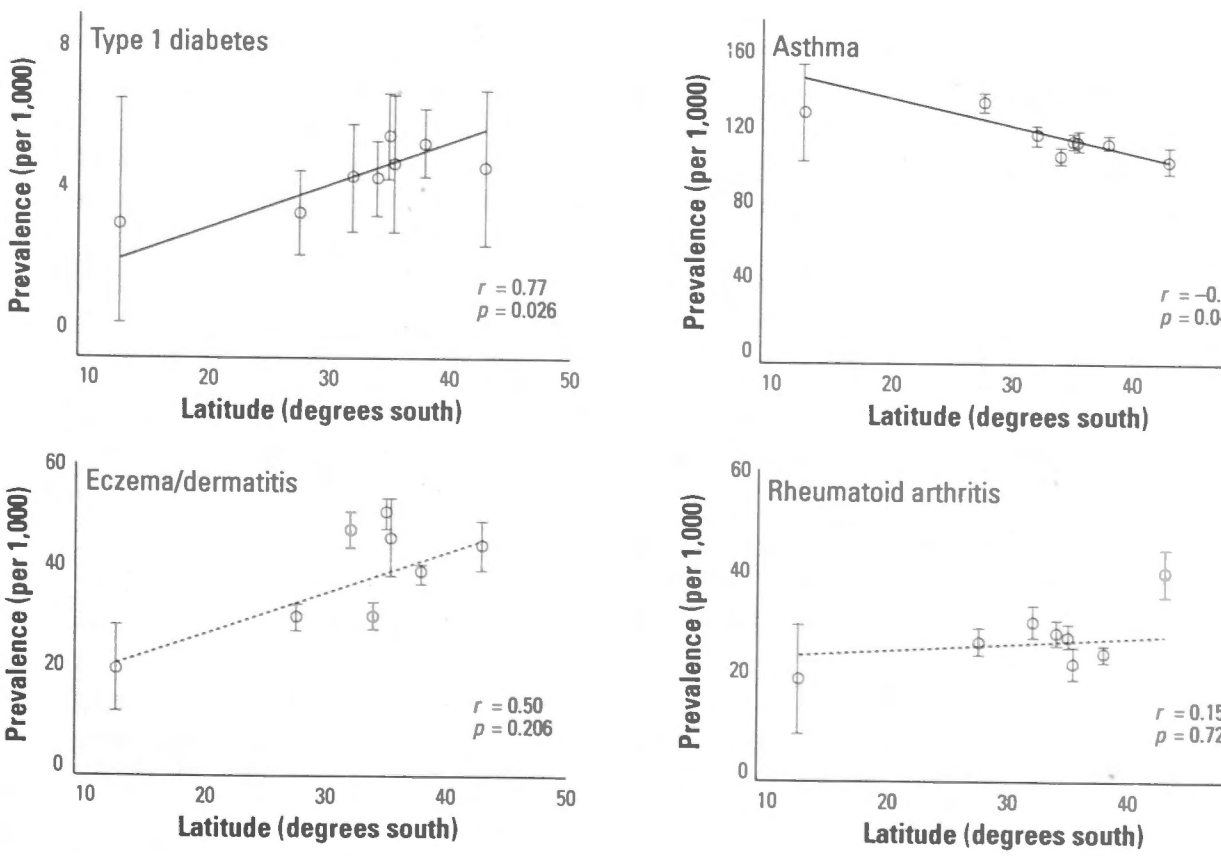


Figure 1. Associations between latitude and prevalence per 1,000 (open circle) of type 1 diabetes, asthma, eczema/dermatitis, and RA. 95% CIs of prevalence estimates shown as error bars; solid lines denote statistically significant association, $p < 0.05$; dotted lines denote nonsignificant trend.

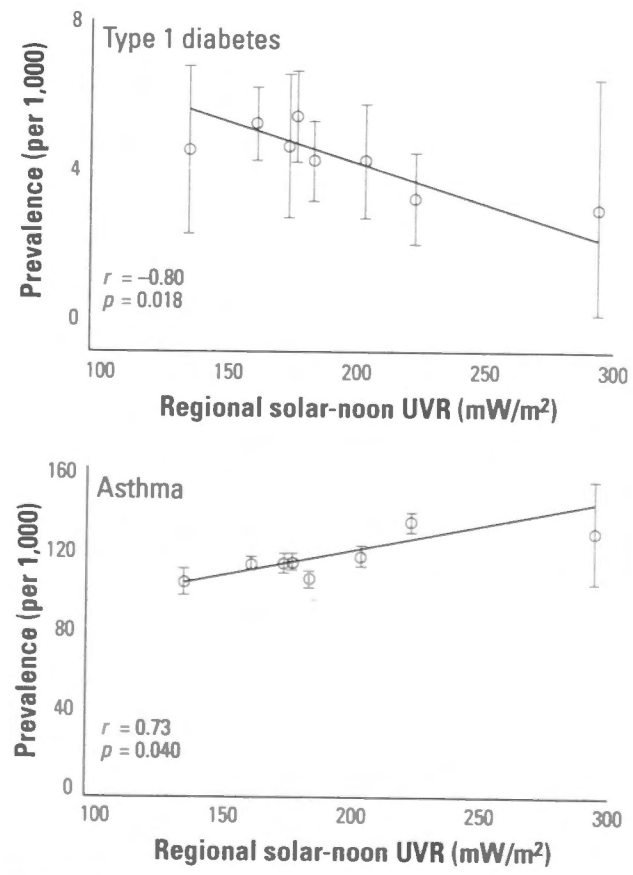


Figure 2. Associations between regional average ambient solar-noon UVR and prevalence per 1,000 (open circle) of type 1 diabetes and asthma. 95% CIs of prevalence estimates shown as error bars.

Discussion. Using summary age- and sex-standardized data on self-reported prevalence of immune disorders from the ABS 1995 NHS, strong gradients in type 1 diabetes prevalence with latitude and also, inversely, with UVR were observed. The magnitude of change for type 1 diabetes in Australia was an approximately 3-fold increase in prevalence from the northernmost region, NT, to the southernmost region, Tasmania (TAS). These ecologic data provide support for a previously proposed beneficial effect of UVR on autoimmune disorders such as type 1 diabetes (McMichael and Hall 1997). In contrast to type 1 diabetes, asthma correlated negatively with latitude and positively with regional annual or midwinter UVR, but the magnitude of change for asthma prevalence from the north to south of Australia was only 0.7-fold. For both RA and eczema/dermatitis, there were no statistically significant associations between latitude/UVR and disease prevalence.

A strength of this ecologic analysis derives from the relative genetic homogeneity of the Australian population (McLeod et al. 1994) and relative uniformity also in the standard of national health care. A further advantage is the wide latitude range and the resulting wide range of ambient UVR levels over the Australian continent. A high total response rate to the survey was achieved by the ABS over all regions.

One potential source of (probably random) error, however, lies in the self-reported, nonverified nature of the data, which could have resulted in some misclassification of disease conditions. This is particularly so for eczema/dermatitis, the classification of which appears to have been based largely on medication usage; there were no direct disease-specific questions on eczema (ABS 1995b). The classification eczema is often used as an umbrella term encompassing various dry, itchy skin conditions (McNally et al. 2000); dermatitis similarly can include contact dermatitis. Also the eczema/dermatitis category in the NHS included heat eczema and sunburn (ABS 1995a), leading to possible regional differences in prevalence. In addition, although arthritis was allocated four specific questions, misclassification between rheumatoid arthritis, osteoarthritis, and general arthritic or rheumatism conditions could have occurred. Type 1 diabetes, on the other hand, was classified on the basis of 24 direct questions for diabetes and insulin usage, including duration of use, both past and future expected use, and age at first use (ABS 1995b). The relative validity of classification of type 1 diabetes, a serious disease requiring specific treatment, should therefore have been more regionally consistent.

Exposure misclassification at the ecologic (i.e., state or territory) level compared with the individual level could have occurred because actual personal UVR exposure depends on both behavior in relation to the sun (i.e., whether sun avoidant or taking sun protection measures) and on regional ambient UVR. For example, regional UVR would not have been a good measure of personal sun exposure among sun-avoidant people, and if the proportion of such people in each local population varied markedly, then population-level exposure assessment would be biased. In fact, the proportion was small and did not vary substantially by region [percentage recording "don't go out in sun" was 3.4% for NT, 2.0% for QLD, and 2.3% for TAS (ABS 1995c)]. For most individuals, however, personal UVR exposure varies between 5 and 15% of daily total ambient UVR (Gies et al. 1999). Ambient UVR levels thus may provide a reasonable measure of average personal sun exposure at the population level.

Another possible source of measurement error—using capital city latitudes and UVR values as representative of each region—was shown by sensitivity analysis not to significantly bias the results even for the two regions having the most noncentral capitals and the widest latitude range, QLD and NT. However, we considered only current residence and could not take into account the contribution of prior residence areas, or the timing of the critical UVR exposure, which may have occurred earlier in life, or even possible migration after disease initiation, as prevalence rather than incidence data were used. In addition, we could not control for latitude when examining the relationships between the immune disorders and UVR because of collinearity between UVR and latitude.

Table 3. Correlations between immune disorder prevalence and regional UVR levels in the 1995 NHS.[a,b]

Prevalence of disorder	Regional solar-noon UVR		
	Average	Midwinter	Midsummer
Type 1 diabetes	–0.80 (0.018)*	–0.77 (0.024)*	–0.72 (0.045)*
Eczema/dermatitis	–0.47 (0.243)	–0.49 (0.215)	–0.34 (0.412)
RA	–0.08 (0.845)	–0.06 (0.880)	–0.12 (0.774)
Asthma	0.73 (0.040)*	0.72 (0.044)*	0.68 (0.060)

[a]Pearson correlation coefficients (correlations based on the eight Australian state/territory regions) between immune disorder prevalence and solar-noon UVR, for each type of immune disorder and three different measures of UV Index values—average over year, midwinter minimum, and midsummer maximum. [b]*p*-Values in parentheses. Strength of association examined by variance-weighted least-squares regression. *Statistically significant association, $p < 0.05$.

Potential ecologic confounders include other possible causal factors for immune-related disorders that vary with latitude or UVR; for example, regional infection patterns that may be associated with climatic differences. Infectious agents have been linked to the etiology of immune disorders, including diabetes (Dahlquist 1998; EURODIAB 2000; Kamradt and Mitchison 2001; Notkins and Lernmark 2001), RA (Albert and Inman 1999; Ebringer and Wilson 2000; Wilson et al. 2000), and asthma (Blasi et al. 2001). A lack of infections in early life may also adversely affect immune development—the so-called hygiene hypothesis (Bach 2002). A second important ecologic factor to consider for asthma is that there are marked regional differences in allergen levels. For example, an 11-fold higher level of mean house dust mite allergen (Der p1) concentration in homes in Sydney, New South Wales (low latitude, warm and more humid) compared with TAS (higher latitude, cool and dry) (Couper et al. 1998) may contribute to the inverse latitude gradient for asthma. Other confounding factors could include environmental temperature and dietary differences (Dahlquist 1998). These potential confounders warrant more detailed consideration in future research. There may also be interaction between some of these environmental exposures, for example, infection and nutritional factors (Haverkos 1997). However, it is not possible to determine such interactions in ecologic analyses, as there is a lack of data on joint environmental exposures at the individual level (Ponsonby et al. 2002).

UVR has an inverse, or lack of, association with the immune-related disorders analyzed except for asthma, which showed a positive association between prevalence and UVR. We are uncertain of the significance of this finding, particularly as the prevalence increase over the UVR range was only low. UVR-induced Th2 upregulation is one possibility, but UVR can also suppress Th2 responses (Garssen et al. 1999). In addition, the contribution of regional allergen levels to asthma prevalence could be important. Furthermore, much of asthma at the population level may be nonallergic (Pearce et al. 1999), with Th2 mechanisms not involved.

The stronger association found between latitude and type 1 diabetes prevalence in Australia is consistent with similar incidence gradients found in Western Europe and North America (DERIG 1988; Karges et al. 1995; Karvonen et al. 1993) and also within China (Yang et al. 1998). The corresponding inverse association between UVR levels in Australia and type 1 diabetes prevalence is also consistent with previous photoimmunologic work showing that ultraviolet B irradiation has systemic as well as local

immunosuppressive effects in humans and animals (Clydesdale et al. 2001; Duthie et al. 1999; Garssen et al. 1999; Sleijffers et al. 2001), as does also ultraviolet A irradiation (Nghiem et al. 2001). UVR exposure may be protective against Th1-mediated disorders such as type 1 diabetes by downregulating Th1 autoimmune responses by several different immunoregulatory mechanisms (Clydesdale et al. 2001; Duthie et al. 1999; Ponsonby et al. 2002).

One of these possible mechanisms for downregulation involves UVR-induced vitamin D. Thus, the proposed protective role of UVR for autoimmune type 1 diabetes may act through its important role in vitamin D synthesis in the skin. Some 90% of plasma vitamin D in humans is produced endogenously via skin exposure to UVR in sunshine (Norris 2001) and particularly so in Australia where foods are not generally fortified with vitamin D (Mason and Diamond 2001; Pasco et al. 2001). There are also recent reports of vitamin D deficiency in some Australian populations (Grover and Morley 2001), particularly in winter (McGrath et al. 2001; Pasco et al. 2001). For type 1 diabetes, this possible mechanism is consistent with recent reports of decreased risk of this disorder in offspring of mothers supplemented with cod liver oil in pregnancy (Stene et al. 2000), and decreased incidence of type 1 diabetes in children supplemented with vitamin D in infancy (EURODIAB 1999; Hypponen et al. 2001). The Finnish birth cohort study reported a rate ratio of 0.12 (95% CI, 0.03–0.51) for diagnosis of type 1 diabetes by adulthood comparing regular versus no vitamin D supplementation in the first year of life (Hypponen et al. 2001). These studies suggest that vitamin D may prevent initiation of type 1 diabetes, and thus accord with the vitamin D–protective hypothesis proposed for this and other autoimmune disorders such as MS (Cantorna 2000; Hayes et al. 1997). For MS this hypothesis is consistent with the recent finding of a strong inverse gradient between ambient UVR levels and MS prevalence in another Australian ecologic study (van der Mei et al. 2001). Our analogous finding for ambient UVR and type 1 diabetes in Australia supports the specific prediction that autoimmune diseases other than MS, such as type 1 diabetes, should show latitude and/or UVR gradients similar to those seen for MS if these immune disorders are influenced by UVR exposure (McMichael and Hall 1997).

In conclusion, our ecologic analysis of data from the 1995 NHS has demonstrated a regional gradient of type 1 diabetes prevalence within Australia that is inversely associated with regional UVR levels measured at local solar noon. The inverse association with UVR is consistent with that found for another autoimmune disease, MS, in Australia and consistent with type 1 diabetes latitudinal gradients seen in the Northern Hemisphere. The finding is also consistent with photoimmunologic evidence of UVR-induced immunosuppression and suggests a beneficial effect of UVR in preventing both these autoimmune conditions. Analytic epidemiology studies investigating risks of type 1 diabetes and other immune disorders in relation to personal UVR exposure in humans are now required (McMichael and Hall 2001).

References

ABS. 1995a. National Health Survey: Users' Guide, Australia (ABS Catalogue no. 4363.0). Canberra, Australian Capital Territory:Australian Bureau of Statistics.

———. 1995b. National Health Survey: Data Reference Package (ABS Catalogue no. 4363.0.001). Canberra, Australian Capital Territory:Australian Bureau of Statistics.

———. 1995c. National Health Survey: Summary Results, Australian States and Territories (ABS Catalogue no. 4368.0). Canberra, ACT:Australian Bureau of Statistics.

Albert LJ, Inman RD. 1999. Molecular mimicry and autoimmunity. N Engl J Med 341:2068–2074.

Bach JF. 2002. The effect of infections on susceptibility to autoimmune and allergic diseases. N Engl J Med 347:911–920.

Blasi F, Cosentini R, Tarsia P, Capone P, Allegra L. 2001. Atypical pathogens and asthma: can they influence the natural history of the disease? Monaldi Arch Chest Dis 56:276–280.

Cantorna MT. 2000. Vitamin D and autoimmunity: is vitamin D status an environmental factor affecting autoimmune disease prevalence? Proc Soc Exp Biol Med 223:230–233.

Clydesdale GJ, Dandie GW, Muller HK. 2001. Ultraviolet light induced injury: immunological and inflammatory effects. Immunol Cell Biol 79:547–568.

Couper D, Ponsonby AL, Dwyer T. 1998. Determinants of dust mite allergen concentrations in infant bedrooms in Tasmania. Clin Exp Allergy 28:715–723.

Dahlquist G. 1998. The aetiology of type 1 diabetes: an epidemiological perspective. Acta Paediatr Suppl 425:5–10.

Damian DL, Halliday GM, Taylor CA, Barnetson RS. 1998. Ultraviolet radiation induced suppression of Mantoux reactions in humans. J Invest Dermatol 110:824–827.

Davidson A, Diamond B. 2001. Autoimmune diseases. N Engl J Med 345:340–350.

DERIG. 1988. Geographic patterns of childhood insulin-dependent diabetes mellitus. Diabetes Epidemiology Research International Group. Diabetes 37:1113–1119.

Duthie MS, Kimber I, Norval M. 1999. The effects of ultraviolet radiation on the human immune system. Br J Dermatol 140:995–1009.

Ebringer A, Wilson C. 2000. HLA molecules, bacteria and autoimmunity. J Med Microbiol 49:305–311.

EURODIAB. 1999. Vitamin D supplement in early childhood and risk for Type I (insulin-dependent) diabetes mellitus. The EURODIAB Substudy 2 Study Group. Diabetologia 42:51–54.

———. 2000. Infections and vaccinations as risk factors for childhood type I (insulin-dependent) diabetes mellitus: a multicentre case-control investigation. EURODIAB Substudy 2 Study Group. Diabetologia 43:47–53.

Garssen J, Vandebriel RJ, De Gruijl FR, Wolvers DA, Van Dijk M, Fluitman A et al. 1999. UVB exposure-induced systemic modulation of Th1- and Th2-mediated immune responses. Immunology 97:506–514.

Gies HP, Roy C, McLennan A, Tomlinson D. 1999. Trends in ultraviolet radiation. In: Trends in Sun Protection Seminar. Melbourne, Australia:Anti-Cancer Council of Victoria, 1–19.

Grover SR, Morley R. 2001. Vitamin D deficiency in veiled or dark-skinned pregnant women. Med J Aust 175:251–252.

Hammond SR, English DR, McLeod JG. 2000. The age-range of risk of developing multiple sclerosis: evidence from a migrant population in Australia. Brain 123(pt 5):968–974.

Haverkos HW. 1997. Could the aetiology of IDDM be multifactorial? Diabetologia 40:1235–1240.

Hayes CE, Cantorna MT, DeLuca HF. 1997. Vitamin D and multiple sclerosis. Proc Soc Exp Biol Med 216:21–27.

Hersey P, Haran G, Hasic E, Edwards A. 1983. Alteration of T cell subsets and induction of suppressor T cell activity in normal subjects after exposure to sunlight. J Immunol 131:171–174.

Hypponen E, Laara E, Reunanen A, Jarvelin MR, Virtanen SM. 2001. Intake of vitamin D and risk of type 1 diabetes: a birth-cohort study. Lancet 358:1500–1503.

Kamradt T, Mitchison NA. 2001. Tolerance and autoimmunity. N Engl J Med 344:655–664.

Karges WJ, Ilonen J, Robinson BH, Dosch HM. 1995. Self and non-self antigen in diabetic autoimmunity: molecules and mechanisms. Mol Aspects Med 16:79–213.

Karvonen M, Tuomilehto J, Libman I, LaPorte R. 1993. A review of the recent epidemiological data on the worldwide incidence of type 1 (insulin-dependent) diabetes mellitus. World Health Organization DIAMOND Project Group. Diabetologia 36:883–892.

Kelly DA, Walker SL, McGregor JM, Young AR. 1998. A single exposure of solar simulated radiation suppresses contact hypersensitivity responses both locally and systemically in humans: quantitative studies with high-frequency ultrasound. J Photochem Photobiol B 44:130–142.

Kero J, Gissler M, Hemminki E, Isolauri E. 2001. Could TH1 and TH2 diseases coexist? Evaluation of asthma incidence in children with coeliac disease, type 1 diabetes, or rheumatoid arthritis: a register study. J Allergy Clin Immunol 108:781–783.

Lemus-Deschamps L, Rikus L, Gies HP. 1999. The operational Australian ultraviolet index forecast 1997. Meteorol Appl 6:241-251.

Mackay IR. 2000. Science, medicine, and the future: tolerance and autoimmunity. Bmj 321:93–96.

Mason RS, Diamond TH. 2001. Vitamin D deficiency and multicultural Australia. Med J Aust 175:236–237.

McGrath JJ, Kimlin MG, Saha S, Eyles DW, Parisi AV. 2001. Vitamin D insufficiency in south-east Queensland. Med J Aust 174:150–151.

McLeod JG, Hammond SR, Hallpike JF. 1994. Epidemiology of multiple sclerosis in Australia. With NSW and SA survey results. Med J Aust 160:117–122.

McMichael AJ, Hall AJ. 1997. Does immunosuppressive ultraviolet radiation explain the latitude gradient for multiple sclerosis? Epidemiology 8:642–645.

———. 2001. Multiple sclerosis and ultraviolet radiation: time to shed more light. Neuroepidemiology 20:165–167.

McNally NJ, Williams HC, Phillips DR, Strachan DP. 2000. Is there a geographical variation in eczema prevalence in the UK? Evidence from the 1958 British Birth Cohort Study. Br J Dermatol 142:712–720.

Miller DH, Hammond SR, McLeod JG, Purdie G, Skegg DC. 1990. Multiple sclerosis in Australia and New Zealand: are the determinants genetic or environmental? J Neurol Neurosurg Psychiatry 53:903–905.

Nghiem DX, Kazimi N, Clydesdale G, Ananthaswamy HN, Kripke ML, Ullrich SE. 2001. Ultraviolet A radiation suppresses an established immune response: implications for sunscreen design. J Invest Dermatol 117:1193–1199.

Norris JM. 2001. Can the sunshine vitamin shed light on type 1 diabetes? Lancet 358:1476–1478.

Notkins AL, Lernmark A. 2001. Autoimmune type 1 diabetes: resolved and unresolved issues. J Clin Invest 108:1247–1252.

Pasco JA, Henry MJ, Nicholson GC, Sanders KM, Kotowicz MA. 2001. Vitamin D status of women in the Geelong Osteoporosis Study: association with diet and casual exposure to sunlight. Med J Aust 175:401–405.

Pearce N, Pekkanen J, Beasley R. 1999. How much asthma is really attributable to atopy? Thorax 54:268–272.

Platts-Mills T, Vaughan J, Squillace S, Woodfolk J, Sporik R. 2001. Sensitisation, asthma, and a modified Th2 response in children exposed to cat allergen: a population-based cross-sectional study. Lancet 357:752–756.

Platts-Mills TA. 2002. The role of allergens in the induction of asthma. Curr Allergy Asthma Rep 2:175–180.

Ponsonby AL, McMichael A, van der Mei I. 2002. Ultraviolet radiation and autoimmune disease: insights from epidemiological research. Toxicology 181–182:71–78.

Reynolds NJ, Franklin V, Gray JC, Diffey BL, Farr PM. 2001. Narrow-band ultraviolet B and broad-band ultraviolet A phototherapy in adult atopic eczema: a randomised controlled trial. Lancet 357:2012–2016.

Selgrade MK, Repacholi MH, Koren HS. 1997. Ultraviolet radiation-induced immune modulation: potential consequences

for infectious, allergic, and autoimmune disease. Environ Health Perspect 105:332–334.

Sleijffers A, Garssen J, de Gruijl FR, Boland GJ, van Hattum J, van Vloten WA et al. 2001. Influence of ultraviolet B exposure on immune responses following hepatitis B vaccination in human volunteers. J Invest Dermatol 117:1144–1150.

Stene LC, Ulriksen J, Magnus P, Joner G. 2000. Use of cod liver oil during pregnancy associated with lower risk of Type I diabetes in the offspring. Diabetologia 43:1093–1098.

van der Mei IA, Ponsonby AL, Blizzard L, Dwyer T. 2001. Regional variation in multiple sclerosis prevalence in Australia and its association with ambient ultraviolet radiation. Neuroepidemiology 20:168–174.

Weinshenker BG. 1996. Epidemiology of multiple sclerosis. Neurol Clin 14:291–308.

Wilson C, Tiwana H, Ebringer A. 2000. Molecular mimicry between HLA-DR alleles associated with rheumatoid arthritis and Proteus mirabilis as the aetiological basis for autoimmunity. Microbes Infect 2:1489–1496.

Yang Z, Wang K, Li T, Sun W, Li Y, Chang YF et al. 1998. Childhood diabetes in China. Enormous variation by place and ethnic group. Diabetes Care 21:525–529.

APPENDIX II:

Staples, J., A-L. Ponsonby and L. Lim (2010). Low maternal exposure to ultraviolet radiation in pregnancy, month of birth, and risk of multiple sclerosis in offspring: longitudinal analysis. *BMJ* **340**: c1640.

BMJ

RESEARCH

Low maternal exposure to ultraviolet radiation in pregnancy, month of birth, and risk of multiple sclerosis in offspring: longitudinal analysis

Judith Staples, student,[1] Anne-Louise Ponsonby, professor,[2] Lynette Lim, biostatistician[1]

[1]National Centre for Epidemiology and Population Health, Australian National University, Canberra, Australia

[2]Murdoch Childrens Research Institute, Royal Children's Hospital, Melbourne, Victoria, Australia

Correspondence to: A-L Ponsonby anne-louise.ponsonby@mcri.edu.au

Cite this as: *BMJ* 2010;340:c1640
doi:10.1136/bmj.c1640

ABSTRACT

Objectives To investigate the distribution of month of birth in people with multiple sclerosis in Australia. To use the large regional and seasonal variation in ambient ultraviolet radiation in Australia to explore the association between exposure to ultraviolet radiation during pregnancy and subsequent risk of multiple sclerosis in offspring.

Design Data were gathered on birth month and year (1920-1950), sex, and state of birth for all patients surveyed in 1981 in Queensland, Western Australia, New South Wales (including Australian Capital Territory), South Australia, and Hobart (Tasmania). Population denominators were derived from the 1981 census and supplementary birth registration data. A variable for exposure to ambient ultraviolet radiation "at birth" was generated from monthly averages of daily total ambient ultraviolet radiation for each region. Negative binomial regression models were used to investigate exposure to ambient ultraviolet radiation at birth and at various intervals before birth.

Setting Patient data from multiple sclerosis prevalence surveys carried out in 1981; 1981 Australian census (giving the total number of people born in Australia and still alive and living in Australia in 1981 by year of birth 1920-50); supplementary Australian birth registration data covering the same birth years by month and state.

Participants 1524 patients with multiple sclerosis born in Australia 1920-50 from total population of 2 468 779.

Main outcome measure Cumulative incidence rate of multiple sclerosis.

Results There was a pattern of risk of multiple sclerosis with month of birth (adjusted incidence rate ratio 1.32, 95% confidence interval 1.10 to 1.58, P<0.01, for those born in November-December compared with those born in May-June). This pattern mirrored that previously reported in the northern hemisphere. Region of birth was related to risk. After adjustment for region of birth and other factors, there was an inverse association between ambient ultraviolet radiation in the first trimester and risk of multiple sclerosis (with ≥25 erythemal (skin reddening) dose units as reference (that is, adjusted incidence rate ratio=1.00), the rates were 1.54 (1.10 to 2.16) for 20-<25 units; 1.58 (1.12 to 2.22) for 15-<20 units; 1.65 (1.17 to 2.33) for 10-<15 units; 1.65 (1.18 to 2.29) for 5-<10 units; and 1.67 (1.18 to 2.37) for <5 units). After adjustment for this exposure during early pregnancy, there was no residual association between month of birth and multiple sclerosis.

Conclusion Region of birth and low maternal exposure to ultraviolet radiation in the first trimester are independently associated with subsequent risk of multiple sclerosis in offspring in Australia.

INTRODUCTION

Multiple sclerosis is a chronic demyelinating disorder that most commonly presents in the second to fourth decade of life. Studies of migrants indicate that risk is strongly associated with place of residence in early life.[1] In Australia[2] and elsewhere,[3 4] there is a latitudinal gradient with increasing prevalence of multiple sclerosis, or incidence of first demyelinating event, as one moves away from the equator. This latitudinal gradient seems environmentally related because the risk associated with latitude alters if people move after birth.[4 5] A strong environmental candidate is the level of ambient regional ultraviolet radiation, acting either directly or through the generation of vitamin D.[6] Higher exposure to ultraviolet radiation,[7] higher vitamin D intake,[8] and also higher serum vitamin D concentrations[9] seem to be associated with a reduced risk of onset of multiple sclerosis. This evidence comes from both case-control and cohort studies and indicates that age of operation for such a protective effect might include both childhood and early adulthood.[6-11]

Exposure to ultraviolet radiation in early life has not yet been formally examined but might be linked to excess risk of multiple sclerosis at birth through seasonal deficiency in maternal vitamin D concentrations.[12] A study of half siblings with multiple sclerosis has also shown that risk can be maternally mediated.[13] Pregnancy is a vulnerable time for vitamin D deficiency because of increased physiological needs and reduced maternal outdoor activity.[14] Experimental data on animal fetal development indicate that cerebral white matter is responsive to vitamin D and that oligodendrocytes in the brain and spinal cord have vitamin D receptors.[15 16] Furthermore, maternal vitamin D depletion alters neurogenesis in the developing rat brain,[17]

with subsequent altered gene expression in adult life.[18] A recent genetic study in humans has further implicated vitamin D as a strong environmental candidate by showing direct functional interaction with the major locus that determines susceptibility to multiple sclero sis.[19] Although human evidence pertaining to fetal development has been difficult to obtain, the body of related evidence to date has led some to recommend antenatal supplementation with vitamin D to prevent multiple sclerosis.[14]

As an indicator of possible perinatal environmental exposures, such as ultraviolet radiation, individual studies in the northern hemisphere have examined month of birth and risk of multiple sclerosis with varied results.[20-24] A large recent pooled analysis of births in the northern hemisphere, however, showed an excess of multiple sclerosis among people born in May and a relative deficit among those born in November; these results were stronger in familial cases and suggested interactions between genes and environment that are related to climate and that might act during gestation or shortly after birth.[12] We examined month of birth and risk of multiple sclerosis in Australia. We used the large regional and seasonal variation in ambient ultraviolet radiation across the continent to investigate the association between exposure to ambient ultraviolet radiation during pregnancy and risk of multiple sclerosis among the offspring.

METHODS

Case ascertainment

We obtained data on the number of patients with multiple sclerosis born in Australia for each birth month of every (birth) year, 1920-50, by sex and state of birth for Queensland, Western Australia, New South Wales (including Australian Capital Territory), South Australia, and Hobart (Tasmania) from prevalence surveys carried out in 1981.[2 25 26] In these surveys, cases were ascertained from hospital records, treating doctors, multiple sclerosis societies, records from the Department of Veterans' Affairs, and the Australian Bureau of Statistics.[2] In the Hobart region the State Chronic Care Hospital Register and Commonwealth Department of Health notifications were also used.[26] All patients were interviewed and examined for verification of multiple sclerosis, except in New South Wales, where only 57% of the patients were interviewed and examined because of the large number of patients notified. Almost all of the remaining patients had been examined previously by a neurologist.[2] All patients in whom the diagnosis of multiple sclerosis was considered to be correct were classified clinically according to the diagnostic criteria of Rose et al.[27]

Analysis dataset

We constructed a longitudinal dataset in frequency (or count) format, spanning every (birth) month of every year over the chosen study period, 1920-50, from the original unit record cross sectional 1981 survey data. The constructed dataset comprised numerator data taken directly from the surveyed cases of multiple sclerosis in 1981,[2] together with reference population denominators (n=2 468 779) derived from the 1981 Australian census (giving the number of people still alive and living in Australia in 1981 by year of birth and the proportion who were born in Australia) and supplementary data on 1920-50 births registered in Australia by month, year, and state provided by the Australian Bureau of Statistics. We adjusted denominators to account for the main sampling losses from surveyed to unsurveyed state regions between the time of birth and time of survey (1981). A restricted dataset of those people with multiple sclerosis born in Australia between 1920 and 1950 (inclusive) was chosen to minimise problems arising from differential survival (for those born before 1920) or age of onset (for those born after 1950) from use of the whole year of birth range (1897-1969) in the original cross sectional dataset. The resulting dataset of adjusted rates of multiple sclerosis comprised numerator and denominator data by month and by year for 1524 people with multiple sclerosis born in any of the five surveyed states during this 1920-50 period.

We generated a variable for exposure to ambient ultraviolet radiation at birth from monthly averages of daily total ambient effective ultraviolet radiation for each region (table 1).

Table 1 | Monthly averages of daily ambient ultraviolet radiation in minimum erythemal dose units*, 1996-2000, for capital cities of Australian states included in 1981 multiple sclerosis survey (data provided by H P Gies, personal communication) (capital city latitudes shown in decimal degrees in parentheses)

	Brisbane (Qld) (27.5° S)	Perth (WA) (31.9° S)	Sydney (NSW) (33.9° S)	Adelaide (SA) (34.9° S)	Hobart† (Tas) (42.9° S)
January	24.7	30.4	22.2	28.1	20.4
February	22.3	27.9	21.7	24.5	18.6
March	19.0	18.9	16.2	13.7	12.4
April	12.5	11.8	9.8	9.4	6.5
May	8.6	7.1	5.6	5.2	3.7
June	6.6	4.9	3.9	3.6	1.7
July	7.5	5.4	4.1	3.7	1.6
August	10.4	7.9	6.9	6.3	3.9
September	14.6	12.1	10.4	10.0	7.0
October	18.9	18.8	15.3	15.5	12.0
November	22.7	23.8	19.1	21.5	15.7
December	24.4	29.2	21.9	26.1	18.1
Seasonal variation‡					
Max:min	3.7	6.2	5.7	7.8	12.8

*Minimum erythemal dose is measure of ultraviolet radiation exposure required to induce erythema or sunburn.[28]
†1991 data.[28]
‡Seasonal variation as given by ratio of summer maximum to winter minimum ultraviolet radiation.

Statistical analysis

Our main outcome measure was the number of people with multiple sclerosis born in each month relative to the general population, expressed as a cumulative incidence rate of multiple sclerosis.

We used negative binomial regression models to provide an estimate of the incidence rate for each month of birth, expressed as an incidence rate ratio

for each time period relative to a single reference period. Because of the small case numbers for the month of birth analysis, we collapsed months into two monthly periods (fig 1). May-June was the reference period because the average ambient ultraviolet radiation was generally lowest then.

To examine the associations between levels of ambient ultraviolet radiation and multiple sclerosis, we modelled ambient ultraviolet radiation as a continuous variable against incidence of multiple sclerosis. Monthly averages were lagged a number of months so that they then expressed the ambient ultraviolet radiation pertaining to a particular length of time before birth for each individual and for each region of birth. Nine such (continuous) variables, representing one to nine months before the birth month of an individual, were generated and included in separate negative binomial regression models (see table 2), as previously done for the categorical variable for period of birth. We thus assessed different gestational periods, in terms of the associated month and region specific levels of ultraviolet radiation, in relation to risk of multiple sclerosis. P values assessing effect modification were derived from likelihood ratio tests of nested regression models with and without the relevant interaction term. Analyses were conducted with Stata 8.0.

RESULTS

There was a large variation in average total daily ambient ultraviolet radiation, from 1.6 minimum erythemal dose units/day in Hobart, Tasmania in July to 30.4 units/day in Perth, Western Australia in January. Overall, the seasonal variation in total daily ambient ultraviolet radiation increased with increasing latitude from Brisbane, Queensland at lowest south latitude to Hobart, Tasmania at highest latitude (table 1).

For year of birth from 1920 to 1950, we identified 1524 cases in the prevalence study from a denominator population of 2 468 779. As expected from the previous Australian surveys,[2] the incidence rate of multiple sclerosis was higher among women than men (incidence rate ratio 2.28, 95% confidence interval 2.03 to 2.55). There was also a latitudinal gradient in cumulative incidence rate by region of birth. Compared with the reference birth region, New South Wales/Australian Capital Territory, the risk was lower for those born in Queensland (0.59, 0.51 to 0.69) and higher for those born in Tasmania (2.70, 2.06 to 3.51). The capital cities of these northernmost and southernmost states are located at 27.5° south and 42.9° south, respectively (table 1).

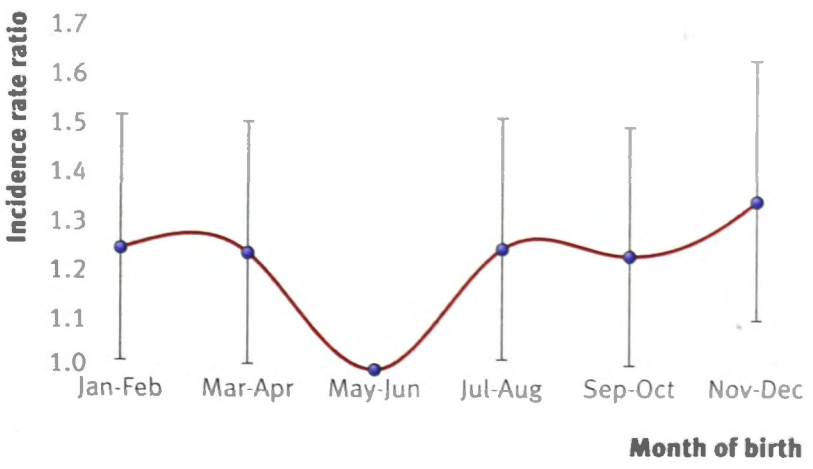


Fig 1| Risk of multiple sclerosis for each two month period of birth. May-June is reference period. July and January represent southern hemisphere mid-winter and mid-summer, respectively

Figure 1 shows the pattern of risk for multiple sclerosis for each two month period of birth, expressed as an incidence rate ratio for each period of birth relative to the reference incidence rate ratio (1.0) for May-June. The risk was 1.23-fold to 1.34-fold higher ($P<0.05$) for people born in all periods other than May-June; the highest magnitude of risk was for those born in the early summer months of November-December compared with the early winter months of May-June (1.34, 1.10 to 1.63; $P<0.01$). This pattern of month of birth persisted after adjustment for sex, age, and region of birth in Australia (1.32, 1.10 to 1.58; $P<0.01$, for November-December compared with May-June). We also examined whether the November-December to May-June risk ratio differed by region of birth; this ratio of around 1.3 was the same over the Queensland-Tasmania range, with no effect modification by region of birth ($P=0.25$).

We then examined the role of prenatal exposure to ambient ultraviolet radiation. We found no association between daily ambient ultraviolet radiation at the time of birth and subsequent risk of multiple sclerosis (table 2). Similarly, lags of one to four months before birth (late second to third trimesters) were not informative. For lags of five to nine months (first to early second trimesters), however, there were inverse associations between prenatal ambient ultraviolet radiation levels and multiple sclerosis (unadjusted incidence rate ratio ranging from 0.74 (0.63 to 0.85) to 0.81 (0.70 to 0.94); $P<0.01$). Table 2 summarises these results and shows a strong inverse association for the first trimester (0.72, 0.62 to 0.84), $P<0.001$).

We examined the shape of the association between inversed first trimester exposure and risk of multiple sclerosis in greater detail by using six categories of ultraviolet radiation (in minimum erythemal dose units/day, see table 1) and comparing each of the five lowest levels with the highest (table 3). The association was non-linear with a particular increase in risk for levels below a monthly average of 20 minimum erythemal dose units/day (fig 2). This "threshold" level is reached only in January in Hobart, Tasmania, but from November to February in Brisbane, Queensland.

Table 3 also shows the increase in risk for women and people born at higher latitudes; the result for decade of birth, showing a deficit in more recently born affected people, probably reflects ascertainment because for some in this group onset of multiple sclerosis might not have occurred by the time of the 1981 survey. The association between first trimester ultraviolet radiation and multiple sclerosis persisted, however, after adjustment for region of birth and the other factors listed in table 3.

Table 2 | Risk of multiple sclerosis as unadjusted and adjusted incidence rate ratio for ambient ultraviolet radiation at and before time of birth

Ambient ultraviolet radiation*	Unadjusted		Adjusted†	
	Ratio (95% CI)	P value	Ratio (95% CI)	P value
At birth	1.01 (0.87 to 1.17)	0.906	1.01 (0.88 to 1.16)	0.851
Time before birth (months):				
1	1.03 (0.89 to 1.19)	0.706	1.03 (0.90 to 1.19)	0.630
2	1.00 (0.86 to 1.16)	0.991	1.01 (0.88 to 1.16)	0.894
3	0.94 (0.81 to 1.10)	0.448	0.96 (0.83 to 1.10)	0.525
4	0.88 (0.76 to 1.02)	0.100	0.89 (0.78 to 1.03)	0.117
5	0.81 (0.70 to 0.94)	0.007	0.83 (0.72 to 0.95)	0.007
6	0.77 (0.67 to 0.90)	0.001	0.78 (0.68 to 0.90)	0.001
7	0.74 (0.63 to 0.85)	<0.001	0.75 (0.65 to 0.86)	<0.001
8	0.74 (0.63 to 0.85)	<0.001	0.75 (0.65 to 0.86)	<0.001
9	0.79 (0.68 to 0.92)	0.002	0.80 (0.70 to 0.92)	0.002
First trimester	0.72 (0.62 to 0.84)	<0.001	0.73 (0.63 to 0.84)	<0.001

*Based on composite month and region specific values for each individual (see table 1) and expressed in units of 20 minimum erythemal dose units/day to gain meaningful incidence rate ratio in terms of difference in minimum erythemal dose between summer and winter (20 minimum erythemal dose units/day is approximate average difference between summer and winter ultraviolet radiation levels for Australian state regions, table 1). First trimester ultraviolet radiation variable obtained by averaging monthly values for seven and eight months before birth, for each region.
†Adjusted for age (year of birth) and sex.

We examined the association between prenatal exposure to ultraviolet radiation and risk of multiple sclerosis related to month of birth. Figure 3 shows the inverse relation between first trimester ultraviolet radiation and risk related to months of birth. That is, when we lagged the ultraviolet radiation values by seven to eight months, there was overall similarity between the two curves, such that infants experiencing low levels of ultraviolet radiation in the first trimester are associated with a higher risk of multiple sclerosis by birth month.

After adjustment for ultraviolet radiation in the first trimester, we found no residual association between period of birth and risk of multiple sclerosis. That is, once we accounted for first trimester ultraviolet radiation (as in table 3) there was no improvement in model fit by also including variation in period of birth (likelihood ratio χ^2 (5 df) =3.79; P=0.58). In contrast, region of birth remained significantly associated with risk of multiple sclerosis even after adjustment for ultraviolet radiation in the first trimester (table 3).

There was no effect modification of the association between ultraviolet radiation in the first trimester and multiple sclerosis by sex (P=0.71) or region of birth (P=0.80).

DISCUSSION

In the southern hemisphere there is a relative excess of multiple sclerosis in people born in November-December compared with the May-June reference minimum. This pattern is consistent with the pattern reported by Willer et al in their larger study in the northern hemisphere,[12] given that the seasons are reversed. Our results show a trough in multiple sclerosis in people born in May-June, when a protective effect is evident, compared with a peak in those born in November-December, thus mirroring the northern hemisphere pattern of a peak associated with May births and a deficit associated with November births. We found no interaction between this pattern and region of birth within Australia, nor with sex or decade of birth; this was also consistent with findings of Willer et al for region of ascertainment, sex, and decade of birth in Canada.[12]

The higher risk of multiple sclerosis for people born in November-December is consistent with these infants having experienced lower levels of ultraviolet radiation during the first trimester. In fact, the pattern of month of birth in the southern hemisphere was accounted for by the month and region specific ambient ultraviolet radiation during the first trimester—that is, the effect of month of birth did not persist after adjustment for first trimester ultraviolet radiation. Lower average daily levels of ambient ultraviolet radiation during the first trimester predicted a higher subsequent risk of multiple sclerosis independently of month of birth. This association was non-linear, with a particular increase in risk under the (monthly average) level of 20 units of minimum erythemal dose of daily ambient ultraviolet radiation. We have thus shown an inverse association between low ultraviolet radiation in the first trimester and increased risk of multiple sclerosis in the offspring.

Strengths and weaknesses

In this longitudinal study we used a prospective exposure—prenatal ambient ultraviolet radiation—the levels of which were heterogeneously distributed among the study participants because of the large variation in ultraviolet radiation linked to month and region of birth across Australia. A further strength is the standardised case ascertainment of multiple sclerosis from the national study in 1981,[2] even though this study was smaller than the extensive northern hemisphere analysis by Willer et al[12] because of the much smaller population in Australia.

We could not examine the vitamin D status of the fetus directly but instead chose the ambient ultraviolet radiation level experienced by the mother during gestation. This proxy has some limitations in that it

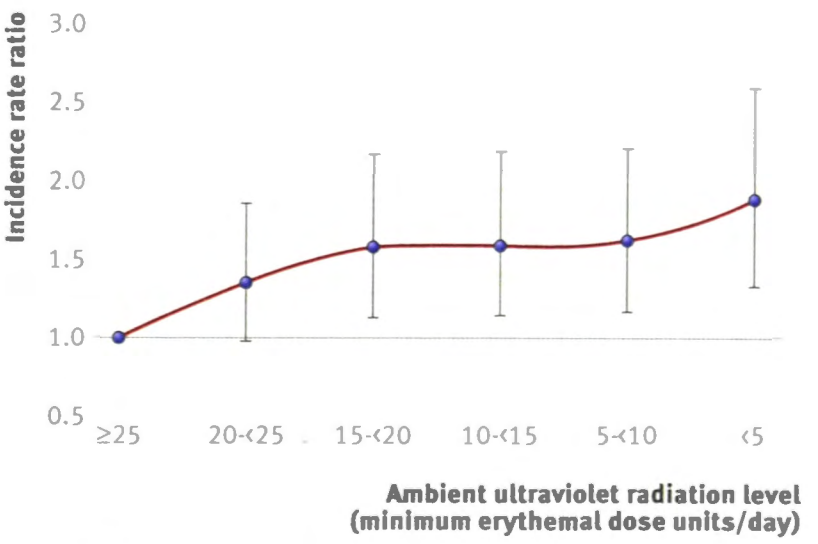


Fig 2 | Risk of multiple sclerosis by average level of daily ambient ultraviolet radiation during first trimester. Incidence rate ratio adjusted for age (year of birth) and sex

Table 3 | Risk of multiple sclerosis as incidence rate ratio for six levels of ambient ultraviolet radiation during first trimester, with and without other factors

Factor	Unadjusted		Adjusted*	
	Ratio (95% CI)	P value	Ratio (95% CI)	P value
Ambient ultraviolet radiation exposure in first trimester† (minimum erythemal doses/day):				
≥25	1.00 (reference)	—	1.00 (reference)	—
20-<25	1.35 (0.97 to 1.87)	0.071	1.54 (1.10 to 2.16)	0.013
15-<20	1.58 (1.14 to 2.20)	0.007	1.58 (1.12 to 2.22)	0.009
10-<15	1.58 (1.13 to 2.20)	0.008	1.65 (1.17 to 2.33)	0.004
5-<10	1.62 (1.17 to 2.23)	0.004	1.65 (1.18 to 2.29)	0.003
<5	1.90 (1.35 to 2.67)	<0.001	1.67 (1.18 to 2.37)	0.004
Sex:				
Men	1.00 (reference)	—	1.00 (reference)	—
Women	2.28 (2.03 to 2.55)	<0.001	2.27 (2.03 to 2.53)	<0.001
Region of birth (latitude of capital city, degrees south):				
Queensland (27.5°)	0.59 (0.51 to 0.69)	<0.001	0.60 (0.52 to 0.70)	<0.001
Western Australia (31.9°)	0.76 (0.62 to 0.92)	0.005	0.85 (0.70 to 1.04)	0.113
New South Wales/Australian Capital Territory (33.9°)	1.00 (reference)	—	1.00 (reference)	—
South Australia (34.9°)	1.03 (0.88 to 1.21)	0.723	1.10 (0.94 to 1.29)	0.237
Tasmania (42.9°)	2.70 (2.06 to 3.51)	<0.001	2.71 (2.08 to 3.52)	<0.001
Decade of birth:				
1920-29	1.09 (0.95 to 1.24)	0.222	1.06 (0.94 to 1.20)	0.346
1930-40	1.00 (reference)	—	1.00 (reference)	—
1941-50	0.73 (0.64 to 0.83)	<0.001	0.73 (0.64 to 0.82)	<0.001

*Adjusted for all other factors in multivariate model—for example, ultraviolet radiation in first trimester adjusted for sex, region of birth, and decade of birth.
†Ambient ultraviolet radiation in first trimester based on composite month and region specific values for each individual (see table 1).

does not take into account individual personal behaviour, concurrent dietary vitamin D intake, or skin pigmentation. These omissions, however, would probably obscure rather than create the patterns observed. Recent work by Sayers et al also indicates that ambient erythemal ultraviolet radiation levels during pregnancy, even in a single location, can be used to indicate vitamin D status.[29] Further work could confirm the timing of the observed effect of prenatal ultraviolet radiation.

Other considerations

Prenatal exposure to ambient ultraviolet radiation during the first trimester was probably not just a marker for postnatal exposure because the association had temporal specificity and was not evident for exposure at the time of birth. Furthermore, the association was independent of region of birth, a probable marker for postnatal sun exposure correlated with long term residence.

Here, and in the larger northern hemisphere analysis,[12] the pattern of month of birth was not smoothly sinusoidal but had a few months of particularly altered risk of multiple sclerosis. This would support an underlying seasonal factor that was not altering smoothly throughout the year but was more evident in particular months or was a threshold biological effect of a continuous variable. Consistent with this, when we examined different levels of prenatal exposure to ultraviolet radiation in the first trimester, we clearly observed that risk was specifically increased in the lower levels of exposure, below a monthly average of 20 minimum erythemal dose units a day. We could not, however, control for other factors such as nutrition or physical activity that could be associated with prenatal exposure to ultraviolet radiation in the first trimester and that could also determine risk of multiple sclerosis. In this setting, there is strong a priori evidence that sun exposure or vitamin D, or both, is likely to be the central exposure[6] and little a priori evidence for any other strong determinant of multiple sclerosis likely to be linked to maternal ambient ultraviolet radiation exposure during the first trimester.

Maternal exposure and health of offspring

This report adds to other work showing that maternal exposure to ambient ultraviolet radiation during pregnancy might influence subsequent health in the offspring.[30-33] In a New Zealand birth cohort, infants whose mothers were exposed to peak sunshine during the first trimester were significantly heavier at birth than infants whose mothers experienced trough levels of sunshine during the same trimester.[30] In the United Kingdom, maternal ambient ultraviolet B exposure in the third trimester has been positively related to bone mineral density and content at age 9 in offspring.[34] There is now growing interest in the role of maternal vitamin D deficiency in pregnancy and the development of central nervous system and immune disorders, particularly schizophrenia,[33] type 1 diabetes, and other disorders.[35] Although active vitamin D_3 concentrations in the mother increase substantially during pregnancy, early fetal supplies are lower[36] and directly dependent on the mother. By the time of delivery, maternal and infant cord serum 25-hydroxyvitamin D_3 concentrations are highly correlated.[37] Unfortunately, much remains unknown, leading to large international variations in maternal vitamin D monitoring and supplementation during pregnancy.[35 38 39]

A maternal effect operating antenatally would also be consistent with the stronger maternal than paternal "parent of origin" effect in familial multiple sclerosis.[13 40] It has been previously proposed[14] that maternal vitamin D deficiency, a problem for some dark skinned women migrating to regions with low ambient ultra-

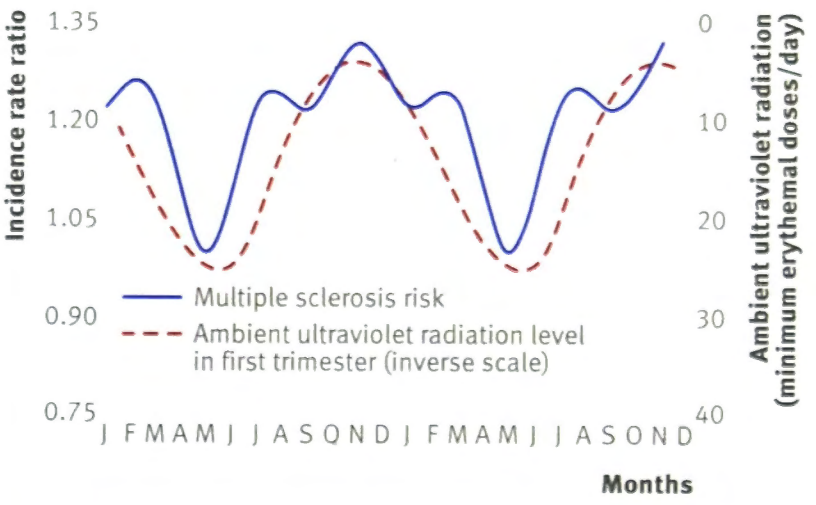


Fig 3 | Multiple sclerosis risk by two month periods of birth with monthly averages of daily ambient ultraviolet radiation in first trimester of pregnancy on inverse scale. Time interval is two annual cycles

WHAT IS ALREADY KNOWN ON THIS TOPIC

In the northern hemisphere, there are more cases of multiple sclerosis in people born in May and fewer in those born in November

Maternal exposure to ambient ultraviolet radiation in pregnancy can be used as an instrumental variable for vitamin D status during pregnancy

Low vitamin D concentrations have been associated with a higher risk of multiple sclerosis

WHAT THIS STUDY ADDS

In Australia there is a reciprocal pattern of month of birth and risk of multiple sclerosis, with a higher risk for those born in November-December compared with May-June

This pattern was accounted for by maternal exposure to ambient ultraviolet radiation in the first trimester

Low maternal exposure to ultraviolet radiation in the first trimester was inversely related to risk of multiple sclerosis in the offspring after adjustment for either month of birth or place of birth

violet radiation, such as the UK, might explain the increase in incidence of disease seen among second generation migrants in such locations.[41] Our findings are consistent with this explanation. Because season of birth has previously also been related to the clinical course of multiple sclerosis,[42-44] it is possible that early life exposures determine not only onset of disease but also resistance to the demyelinating process of multiple sclerosis. The mechanisms involved could include neurological or immunological factors. Ultraviolet radiation exposure during the first trimester would be expected to specifically influence vitamin D concentrations up to early in the second trimester, given that vitamin D has a half life of one to two months.[45] In recent Australian work, higher levels of ambient ultraviolet radiation were associated with higher vitamin D concentrations at the population level with a lag of one and a half months.[46]

Possible mechanisms

First trimester vitamin D concentrations might be particularly important in the development of the central nervous system because during early embryonic development, vitamin D receptors are expressed in the neuroepithelium and later in the subventricular zone.[15] Myelination occurs later; even in mid-gestation (19-24 weeks) cortical axonal tracts are not yet myelinated,[47] with major myelination of several areas occurring as late as 29-39 weeks.[48] In addition, the in utero development of immune central tolerance occurs in the first trimester and vitamin D has immunomodulating properties. The first trimester is also a sensitive period with regard to prenatal thymocyte differentiation, with animal studies showing that chemicals such as dioxin can alter this process, disrupt the development of central tolerance, and lead to increased autoreactive peripheral T cells.[49] Furthermore, indirect effects of vitamin D should be considered. For example, vitamin D can down-regulate interleukin 6, an important mediator of the adverse effect of maternal infection during pregnancy on neural development in the fetus.[50-52]

Prenatal and postnatal timing

Our results do not indicate that the possible beneficial effect of ultraviolet radiation exposure is confined only to the prenatal period. The finding that both first trimester ultraviolet radiation and region of birth were independent predictors of risk of multiple sclerosis is consistent with birth region also acting as an indicator of postnatal exposure to ultraviolet radiation. Overall, epidemiological studies support a protective role for vitamin D in autoimmune disease, particularly in childhood and adolescence, and vitamin D supplementation in early adulthood effectively reduces the risk of multiple sclerosis; therefore, supplementation of adolescents and young adults could be effectively used for prevention.[6] The findings here provide the first population based evidence beyond month of birth patterns to indicate that vitamin D supplementation for the prevention of multiple sclerosis might also need to be considered during in utero development.[14] They are consistent with a multi-hit causal cascade for multiple sclerosis, with putative adverse environmental factors, such as low vitamin D concentrations, acting at more than one stage of life.[53]

We gratefully acknowledge the assistance of J G McLeod, S R Hammond, and P Macaskill in providing case data and advice on the conduct of the original 1981 national multiple sclerosis survey. H P Gies, Australian Radiation Protection and Nuclear Safety Agency (ARPANSA), kindly provided the regional ultraviolet radiation data for the Australian State capital cities for 1996-2000. We also thank A J McMichael and J G McLeod for their comments on the manuscript and H Raschella and A Pezic for literature review and manuscript preparation.

Contributors: All authors were involved in the study design and analysis, participated in the preparation and writing of the manuscript, and approved the final version. JS is guarantor.

Funding: The research was supported by an Australian National University (ANU) Graduate School scholarship and a supplementary scholarship from the National Centre for Epidemiology and Population Health, ANU, awarded to JS.

Competing interests: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that (1) JS, ALP, LL have no relationships with companies that might have an interest in the submitted work in the previous 3 years; (3) their spouses, partners, or children have no financial relationships that may be relevant to the submitted work; and (4) JS, ALP, LL have no non-financial interests that may be relevant to the submitted work.

Ethical approval: This study was approved by the human research ethics committee, Australian National University.

Data sharing: No additional data available.

1 Dean G. Annual incidence, prevalence, and mortality of multiple sclerosis in white South-African-born and in white immigrants to South Africa. *BMJ* 1967;ii:724-30.

2 McLeod JG, Hammond SR, Hallpike JF. Epidemiology of multiple sclerosis in Australia. With NSW and SA survey results. *Med J Aust* 1994;160:117-22.

3 Alter M. Epidemiology of multiple sclerosis in Israel. In: Alter M, Kurtzke JF, eds. The epidemiology of multiple sclerosis. Charles C Thomas, 1968:83-109.

4 Kurtzke JF, Beebe GW, Norman JE Jr. Epidemiology of multiple sclerosis in US veterans: III. Migration and the risk of MS. *Neurology* 1985;35:672-8.

5 Hammond SR, English DR, McLeod JG. The age-range of risk of developing multiple sclerosis: evidence from a migrant population in Australia. *Brain* 2000;123:968-74.

6 Ascherio A, Munger K. Epidemiology of multiple sclerosis: from risk factors to prevention. *Semin Neurol* 2008;28:17-28.

7 Van der Mei IA, Ponsonby AL, Dwyer T, Blizzard L, Simmons R, Taylor BV, et al. Past exposure to sun, skin phenotype, and risk of multiple sclerosis: case-control study. *BMJ* 2003;327:316.

8 Munger KL, Zhang SM, O'Reilly E, Hernan MA, Olek MJ, Willett WC, et al. Vitamin D intake and incidence of multiple sclerosis. *Neurology* 2004;62:60-5.
9 Munger KL, Levin LI, Hollis BW, Howard NS, Ascherio A. Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. *JAMA* 2006;296:2832-8.
10 Islam T, Gauderman WJ, Cozen W, Mack TM. Childhood sun exposure influences risk of multiple sclerosis in monozygotic twins. *Neurology* 2007;69:381-8.
11 Kampman MT, Wilsgaard T, Mellgren SI. Outdoor activities and diet in childhood and adolescence relate to MS risk above the Arctic Circle. *J Neurol* 2007;254:471-7.
12 Willer CJ, Dyment DA, Sadovnick AD, Rothwell PM, Murray TJ, Ebers GC. Timing of birth and risk of multiple sclerosis: population based study. *BMJ* 2005;330:120.
13 Ebers GC, Sadovnick AD, Dyment DA, Yee IML, Willer CJ, Risch N. Parent of origin effect in multiple sclerosis: observations in half siblings. *Lancet* 2004;363:1773-4.
14 Chaudhuri A. Why we should offer routine vitamin D supplementation in pregnancy and childhood to prevent multiple sclerosis. *Med Hypotheses* 2005;64:608-18.
15 Wion D, MacGrogan D, Neveu I, Jehan F, Houlgatte R, Brachet P. 1,25-Dihydroxyvitamin D3 is a potent inducer of nerve growth factor synthesis. *J Neurosci Res* 1991;28:110-4.
16 Garcion E, Wion-Barbot N, Montero-Menei CN, Berger F, Wion D. New clues about vitamin D functions in the nervous system. *Trends Endocrinol Metab* 2002;13:100-5.
17 Cui X, McGrath JJ, Burne TH, Mackay-Sim A, Eyles DW. Maternal vitamin D depletion alters neurogenesis in the developing rat brain. *Int J Dev Neurosci* 2007;25:227-32.
18 Eyles D, Almeras L, Benech P, Patatian A, Mackay-Sim A, McGrath J, et al. Developmental vitamin D deficiency alters the expression of genes encoding mitochondrial, cytoskeletal and synaptic proteins in the adult rat brain. *J Steroid Biochem Mol Biol* 2007;103:538-45.
19 Ramagopalan SV, Maugeri NJ, Handunnetthi L, Lincoln MR, Orton S-M, Dyment DA, et al. Expression of the multiple sclerosis-associated MHC class II allele HLA-DRB1*1501 is regulated by vitamin D. *PLoS Genet* 2009;5:e1000369.
20 Templer DI, Trent NH, Spencer DA, Trent A, Corgiat MD, Mortensen PB, et al. Season of birth in multiple sclerosis. *Acta Neurol Scand* 1992;85:107-9.
21 Wiberg M. Season of birth in multiple sclerosis in Sweden: replication of Denmark findings. *J Orthomol Med* 1994;9:71-4.
22 Sadovnick AD, Yee IM. Season of birth in multiple sclerosis. *Acta Neurol Scand* 1994;89:190-1.
23 Salemi G, Ragonese P, Aridon P, Reggio A, Nicoletti A, Buffa D, et al. Is season of birth associated with multiple sclerosis? *Acta Neurol Scand* 2000;101:381-3.
24 Torrey EF, Miller J, Rawlings R, Yolken RH. Seasonal birth patterns of neurological disorders. *Neuroepidemiology* 2000;19:177-85.
25 Hammond SR, de Wytt C, Maxwell IC, Landy PJ, English D, McLeod JG, et al. The epidemiology of multiple sclerosis in Queensland, Australia. *J Neurol Sci* 1987;80:185-204.
26 Hammond SR, McLeod JG, Millingen KS, Stewart-Wynne EG, English D, Holland JT, et al. The epidemiology of multiple sclerosis in three Australian cities: Perth, Newcastle and Hobart. *Brain* 1988;111:1-25.
27 Rose AS, Ellison GW, Myers LW, Tourtellotte WW. Criteria for the clinical diagnosis of multiple sclerosis. *Neurology* 1976;26:20-2.
28 Gies HP, Roy CR, Toomey S, Tomlinson D. The ARL solar UVR measurement network: calibration and results. *SPIE* 1994;2282:274-284.
29 Sayers A, Tilling K, Boucher BJ, Noonan K, Tobias JH. Predicting ambient ultraviolet from routine meteorological data; its potential use as an instrumental variable for vitamin D status in pregnancy in a longitudinal birth cohort in the UK. *Int J Epidemiol* 2009;38:1681-8.
30 Tustin K, Gross J, Hayne H. Maternal exposure to first-trimester sunshine is associated with increased birth weight in human infants. *Dev Psychobiol* 2004;45:221-30.
31 Eyles DW, Smith S, Kinobe R, Hewison M, McGrath JJ. Distribution of the vitamin D receptor and 1 alpha-hydroxylase in human brain. *J Chem Neuroanat* 2005;29:21-30.
32 Eyles D, Brown J, Mackay-Sim A, McGrath J, Feron F. Vitamin D3 and brain development. *Neuroscience* 2003;118:641-53.
33 McGrath JJ, Feron FP, Burne TH, Mackay-Sim A, Eyles DW. Vitamin D3-implications for brain development. *J Steroid Biochem Mol Biol* 2004;89-90:557-60.
34 Javaid MK, Crozier SR, Harvey NC, Gale CR, Dennison EM, Boucher BJ, et al. Maternal vitamin D status during pregnancy and childhood bone mass at age 9 years: a longitudinal study. *Lancet* 2006;367:36-43.
35 Lucas RM, Ponsonby AL, Pasco JA, Morley R. Future health implications of prenatal and early-life vitamin D status. *Nutr Rev* 2008;66:710-20.
36 Delvin EE, Glorieux FH, Salle BL, David L, Varenne JP. Control of vitamin D metabolism in preterm infants: feto-maternal relationships. *Arch Dis Child* 1982;57:754-7.
37 Maghbooli Z, Hossein-Nezhad A, Shafaei AR, Karimi F, Madani FS, Larijani B. Vitamin D status in mothers and their newborns in Iran. *BMC Pregnancy Childbirth* 2007;7:1.
38 Munns C, Zacharin MR, Rodda CP, Batch JA, Morley R, Cranswick NE, et al. Prevention and treatment of infant and childhood vitamin D deficiency in Australia and New Zealand: a consensus statement. *Med J Aust* 2006;185:268-72.
39 Vitamin D supplementation: recommendations for Canadian mothers and infants. *Paediatr Child Health* 2007;12:583-98.
40 Herrera BM, Ramagopalan SV, Lincoln MR, Orton SM, Chao MJ, Sadovnick AD, et al. Parent-of-origin effects in multiple sclerosis: observations from avuncular pairs. *Neurology* 2008;71:799-803.
41 Datta S, Alfaham M, Davies DP, Dunstan F, Woodhead S, Evans J, et al. Vitamin D deficiency in pregnant women from a non-European ethnic minority population—an interventional study. *BJOG* 2002;109:905-8.
42 Tremlett HL, Devonshire VA. Does the season or month of birth influence disease progression in multiple sclerosis? *Neuroepidemiology* 2006;26:195-8.
43 Koch M, De Keyser J, Tremlett H. Timing of birth and disease progression in multiple sclerosis. *Mult Scler* 2008;14:793-8.
44 Sadovnick AD, Duquette P, Herrera B, Yee IM, Ebers GC. A timing-of-birth effect on multiple sclerosis clinical phenotype. *Neurology* 2007;69:60-2.
45 Vieth R. Vitamin D supplementation, 25-hydroxyvitamin D concentrations, and safety. *Am J Clin Nutr* 1999;69:842-56.
46 Tremlett H, van der Mei IA, Pittas F, Blizzard L, Paley G, Mesaros D, et al. Monthly ambient sunlight, infections and relapse rates in multiple sclerosis. *Neuroepidemiology* 2008;31:271-9.
47 Jakovcevski I, Mo Z, Zecevic N. Down-regulation of the axonal polysialic acid-neural cell adhesion molecule expression coincides with the onset of myelination in the human fetal forebrain. *Neuroscience* 2007;149:328-37.
48 Abe S, Takagi K, Yamamoto T, Okuhata Y, Kato T. Semiquantitative assessment of myelination using magnetic resonance imaging in normal fetal brains. *Prenat Diagn* 2004;24:352-7.
49 Gogal RM Jr, Holladay SD. Perinatal TCDD exposure and the adult onset of autoimmune disease. *J Immunotoxicol* 2008;5:413-8.
50 Meyer U, Yee BK, Feldon J. The neurodevelopmental impact of prenatal infections at different times of pregnancy: the earlier the worse? *Neuroscientist* 2007;13:241-56.
51 Smith SE, Li J, Garbett K, Mirnics K, Patterson PH. Maternal immune activation alters fetal brain development through interleukin-6. *J Neurosci* 2007;27:10695-702.
52 Correale J, Ysrraelit MC, Gaitan MI. Immunomodulatory effects of vitamin D in multiple sclerosis. *Brain* 2009;132:1146-60.
53 Goodin DS. The causal cascade to multiple sclerosis: a model for multiple sclerosis pathogenesis. *PLoS One* 2009;4:e4565.

Accepted: 12 January 2010