Multiple Sclerosis
http://msj.sagepub.com

Observational analytic studies in multiple sclerosis: controlling bias through study design and conduct. The Australian Multicentre Study of Environment and Immune Function

RM Lucas, A-L. Ponsonby, AJ McMichael, I. van der Mei, C. Chapman, A. Coulthard, K. Dear, T. Dwyer, TJ Kilpatrick, MP Pender, B. Taylor, P. Valery and D. Williams

Mult Scler 2007; 13; 827
DOI: 10.1177/1352458507077174

The online version of this article can be found at:
http://msj.sagepub.com/cgi/content/abstract/13/7/827

Published by:
SAGE Publications
http://www.sagepublications.com

ARTICLE *Multiple Sclerosis* 2007; **13**: 827–839

Observational analytic studies in multiple sclerosis: controlling bias through study design and conduct. The Australian Multicentre Study of Environment and Immune Function

RM Lucas[1], A-L Ponsonby[2], AJ McMichael[1], I van der Mei[3], C Chapman[4], A Coulthard[5], K Dear[1], T Dwyer[2], TJ Kilpatrick[6], MP Pender[5], B Taylor[7], P Valery[8] and D Williams[9]

Rising multiple sclerosis incidence over the last 50 years and geographic patterns of occurrence suggest an environmental role in the causation of this multifactorial disease. Design options for epidemiological studies of environmental causes of multiple sclerosis are limited by the low incidence of the disease, possible diagnostic delay and budgetary constraints. We describe scientific and methodological issues considered in the development of the Australian Multicentre Study of Environment and Immune Function (the Ausimmune Study), which seeks, in particular, to better understand the causes of the well-known MS positive latitudinal gradient. A multicentre, case-control design down the eastern seaboard of Australia allows the recruitment of sufficient cases for adequate study power and provides data on environmental exposures that vary by latitude. Cases are persons with an incident first demyelinating event (rather than prevalent multiple sclerosis), sourced from a population base using a two tier notification system. Controls, matched on sex, age (within two years) and region of residence, are recruited from the general population. Biases common in case-control studies, eg, prevalence-incidence bias, admission-rate bias, non-respondent bias, observer bias and recall bias, as well as confounding have been carefully considered in the study design and conduct of the Ausimmune Study. *Multiple Sclerosis* 2007; **13**: 827–839. http://msj.sagepub.com

Key words: bias; case-control; confounding; epidemiologic research design; multicentre; multiple sclerosis

Introduction

The causes of multiple sclerosis (MS) are unknown, but probably involve an interplay of genetic susceptibility and environmental exposures that may operate from early in life [1] to just before diagnosis [2]. MS incidence is increasing (1950–2000) in developed countries [3], including Australia [4], where, for example, MS incidence in Newcastle doubled from 1.2 per 100 000 in 1961 to 2.4 per 100 000 in 1996 [4].

Research into the causal roles of environmental exposures has yielded inconsistent findings which at least partly reflect problems in study design [5]. Epidemiological studies have great potential to elucidate the causes of MS, but require careful design and execution [6].

[1] National Centre for Epidemiology and Population Health, The Australian National University
[2] Murdoch Childrens Research Institute, Melbourne, Australia
[3] Menzies Research Institute, Hobart, Australia
[4] Barwon Health, Geelong, Australia
[5] Royal Brisbane and Women's Hospital and The University of Queensland, Brisbane, Australia
[6] Centre for Neuroscience, The University of Melbourne, Melbourne, Australia
[7] Otago University, Christchurch, New Zealand
[8] Queensland Institute of Medical Research, Brisbane, Australia
[9] John Hunter Hospital, University of Newcastle, Newcastle, Australia
Author for correspondence: Dr R Lucas, National Centre for Epidemiology and Population Health,The Australian National University, Canberra, ACT 0200, Australia. Email: Robyn.Lucas@anu.edu.au
Received 28 July 2006; accepted 18 December 2006

 10.1177/1352458507077174

Here we outline scientific and methodological issues considered in developing the study design of the Australian Multicentre Study of Environment and Immune Function (the Ausimmune Study) (see Box 1). We first consider the choice of study design and then address five key considerations specific to the study hypotheses. In reviewing issues in the control of bias and confounding we hope to assist others in planning future research.

Possible study designs

Ecological studies have an important role in exploring patterns of disease occurrence in relation to other population-level variables and in developing new hypotheses. Such studies have highlighted, for example, a positive association between latitude and prevalence of MS [7,8]. This latitudinal gradient could be further examined to try to determine which correlate of latitude is of primary importance, eg, ambient ultraviolet radiation (UVR) [8], temperature, dietary patterns or patterns of infection. But ecological studies cannot examine disease occurrence in relation to personal exposure to putative risk factors – are those individuals with elevated exposure at increased risk of disease?

Observational study designs such as case-control or cohort studies are commonly used to elucidate disease etiology, using individual-level data. The advantages and disadvantages of each design for the study of environmental risk factors for MS have been described in detail [6] and are briefly reviewed here (see Table 1).

The case-control design is common in MS research, as it is efficient, particularly for an uncommon disease for which a number of exposures are to be studied and where understanding of the possible role of such exposures is limited. Exposure data are collected from two groups: 'cases' who have the health outcome of interest and 'controls' who do not. Controls are, ideally, outcome-free representatives of the source population from which the cases derive. The case-control study is particularly prone to selection and measurement biases and the control of confounding also requires care.

Cohort studies, following up a large sample of disease-free persons with varied exposures, have the advantage of examining exposure and disease in the correct chronological order. However, cohort studies are expensive, take considerable time (unless done from an historical base), and are inefficient for studying uncommon diseases – a very large cohort, studied over many years, would be required to accrue sufficient cases of MS.

Intervention studies, eg, randomized controlled trials, allow much tighter control of bias and confounding, but require specification of a risk factor with a strong probability of being causative and which is also amenable to a preventive intervention. Further, there can be ethical difficulties, such as failing to provide the intervention to the control arm when there is already some confidence in the efficacy of that intervention, and the possibility of unexpected adverse outcomes of the intervention.

Case-control studies: adapting the design to the research question

Consideration 1: Location

Latitudinal gradients in MS in the United States or Europe often occur across populations with different genetic profiles, making the importance of environmental exposures difficult to ascertain. However, in Australia, previous studies note the relative cultural, socioeconomic and genetic homogeneity of populations down the eastern seaboard [9] and the preservation of a clear latitudinal gradient, even when the analyses are restricted to immigrants

Table 1 Strengths and limitations of the case-control design in the study of the role of environmental factors in the onset of MS

	Strengths	Limitations
Case-control	1) Relatively quick and inexpensive compared with other analytic designs	1) Both exposure and disease have already occurred at time of enrolment; the temporal relationship between exposure and disease may be difficult to establish
	2) Optimal for the evaluation of rare diseases	2) Particularly prone to bias compared to other analytic designs, in particular: recall bias – differential recall of exposure by disease status selection bias – differential selection of cases or controls, based on exposure status
	3) Particularly well-suited to the evaluation of diseases with long latent periods	3) Cannot directly compute incidence rates of disease in exposed and non-exposed individuals, unless the study is population-based
	4) Can examine multiple etiologic factors for a single disease	

Adapted from [54].

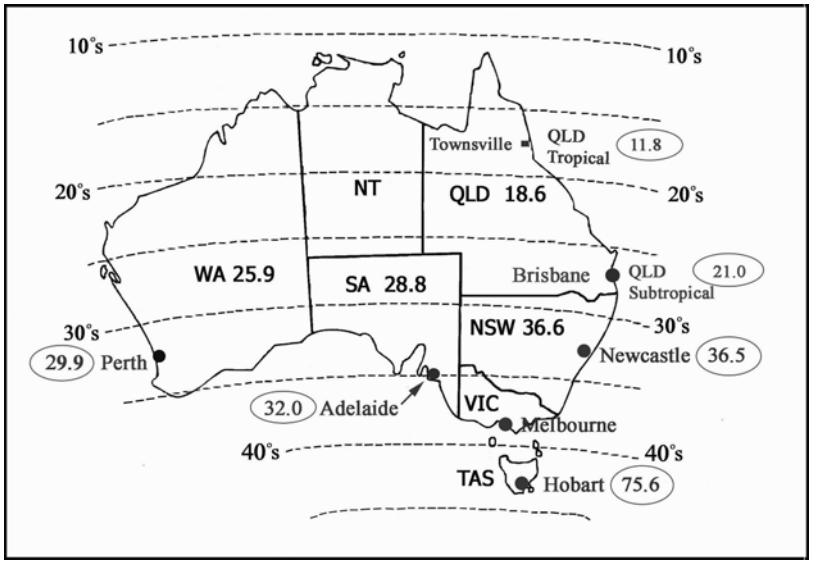

Figure 1 Age standardized MS prevalence in Australia (per 100 000 population on 30 June 1981) (McLeod JG, Hammond SR, Hallpike JF. Epidemiology of multiple sclerosis in Australia. 1994 *MJA* 1994; **160**: 117–22. © Copyright 1994. The Medical Journal of Australia – reproduced with permission [7]).

from the UK and Ireland [9]. Australia thus provides an excellent location to study MS latitudinal variation (note that blood is being collected and stored for a later genetic substudy).

Consideration 2: Multicentre

Multicentre case-control studies are often used in epidemiological MS research, particularly to achieve sufficient case sample size [10,11]. Within the Ausimmune Study there is a further reason to use the multicentre design. The Ausimmune Study seeks specifically to study the previously demonstrated Australian latitudinal MS prevalence gradient [7] (see Figure 1).

This multicentre study down the eastern seaboard of Australia encompasses a relatively wide latitudinal span, thus allowing enhanced assessment of possible environmental influences contributing to the observed variation in prevalence. Four study regions, from Brisbane [latitude 27°S, the most northern study centre with sufficient population to recruit adequate cases over three years (see Consideration 4)] to Tasmania (latitude 43°S) are participating in the Ausimmune Study.

Consideration 3: Incident versus prevalent cases

A key exposure of interest in the Ausimmune Study, based on previous research, is UVR exposure [12], and, related to that, vitamin D levels [13]. As in another multicentre case-control study examining risk factors for onset of MS [11], the Ausimmune Study is recruiting persons with a first clinical diagnosis of central nervous system (CNS) demyelination, rather than patients with established MS. Approximately six in 10 of those diagnosed with a clinically isolated syndrome will progress to clinically definite MS within 10 years [14]. This strategy aims to minimize the types of bias that can occur in studies of persons with established MS: questionnaires capture reports of current behaviours less biased by disease-related changes in behaviour, and biological parameters reflect current or recent behaviours – for example, vitamin D can be sampled with some confidence that the levels are not influenced by post-diagnostic factors (medication or altered behaviour due to a formal diagnosis of MS).

Consideration 4: Duration of study and study power

The projected statistical power of the study to test the study hypotheses depends primarily on the expected accrual of cases and matched controls. There is little flexibility in the study population from which incident cases are drawn, because of the need to span as great a latitude range as possible while including only reasonably large population centres. The power calculation therefore amounts to verifying that the planned three-year accrual period is likely to provide sufficient numbers of incident first demyelinating event (FDE) cases.

Table 2 Sample size estimates for field work and analysis in the Ausimmune Study

Study region	Study population (million)	Annual MS Incidence (per 100 000)	MRI-positive FDE cases in three years	Control: case ratio	Controls for matched case-control analysis
Brisbane city	1.6	2.0	109	2 : 1	218
Newcastle region	0.48	2.5	41	4 : 1	164
Geelong and the Western Districts of Victoria	0.475	3.0	49	3 : 1	147
Tasmania	0.46	9.0	141	1 : 1	141

To estimate the current incidence of FDE across Australia, we started with published data on regional prevalence [7]. A single estimate of local annual incidence was available for Newcastle: the average age-standardized annual incidence for MS from 1986 to 1996 in Newcastle was 2.5 cases per 100 000 population per year [4]. This established the ratio of incidence to prevalence there. After weighting for the prevalence gradient reported by McLeod *et al*. [7] and assuming that the same incidence to prevalence ratio applied across the country, this provided estimates of current projected MS incidence (per 100 000) of 2.0 for Brisbane, 3.0 for Geelong and 9.0 for Tasmania.

The expected cases per year in each region are a product of incidence and population, as shown in Table 2. However, these are incidence rates for confirmed MS: the incidence of FDE will be higher because only a proportion of FDE cases will eventually develop MS [14]. Furthermore, of all apparent FDE cases accrued, 67% are expected to show a positive magnetic resonance imaging (MRI) brain scan confirming demyelination [14]. These MRI-positive FDE cases, with their matched controls, are the core group for testing the study hypotheses, so our power calculations are based on these numbers, shown in Table 2.

We planned a different case : control ratio in each region, with more controls per case in centres where the expected accrual of cases was lower. This was intended to build a cohort of controls of similar size in each region, which would be a valuable resource in future, possibly unrelated epidemiologic studies. In the event, various practical considerations required altering these ratios during the course of the study, so that in 2006 all centres were recruiting two controls per case.

Formulae are not available for power calculation in matched case-control studies with varying case : control ratio. We therefore estimated study power using simulations, to ensure that three years' accrual would provide adequate statistical power to test all the study hypotheses.

Consideration 5: Exposures of interest

Given the observed latitudinal variation in MS, a further key consideration in the Ausimmune Study was to investigate with high priority those environmental factors that: 1) varied with latitude and 2) have been linked with MS. We next consider some MS risk factors against these two criteria.

Sun exposure

1) *UVR and latitude*: The quantity (intensity) and quality (spectrum) of incident UVR varies with latitude, season and time of day [15], with highest UVR levels occurring near the equator, during summer and around midday [15]. While all ambient UVR wavelengths affect immune function [16], only shorter-wavelength UVB radiation induces the cutaneous synthesis of vitamin D [17]. Individual-level UVR dose is determined by both ambient UVR and sun-exposure behaviour (which is highly variable) [18].
2) *UVR and MS*: Geographical patterns of MS occurrence suggest a role for relative lack of UVR exposure in the development of MS [19]. Further support comes from recent understanding of the immunosuppressive actions of UVR [16,20], and, empirically, from epidemiological evidence of the role of past childhood UVR exposure (estimated four-fold inverse gradient in risk between most and least exposed individuals) [12], cumulative UVR exposure as indicated by skin cancer as exposure proxy [21], and vitamin D supplements [13], in reducing the risk of MS onset.

Infection

1) *Infection and latitude*: Recent research confirms a latitudinal gradient in number of pathogenic species, with tropical areas typically harboring many more pathogens than temperate regions [22]. Pathogenic species found at higher latitudes are a subset of those present in equatorial areas [22]. Infection-related behaviours and environmental conditions may also vary by latitude.
2) *Infection and MS*: There is strong evidence from prospective studies of a causative role for

Epstein-Barr virus in MS [23], while other work implicates Human Herpes virus 6 infection [24]. The evidence for the involvement of other putative infectious agents, including human endogenous retroviruses [25], Chlamydia [26], canine distemper virus [27] and others [28,29] is less clear. There is mixed epidemiological support for the 'hygiene hypothesis', that the rising incidence of immune disorders, including MS, is due to reduced exposure to childhood infections and consequent primary infection at a later age [3]. In one study, older age of contracting measles and mumps were each associated with an approximate doubling of MS risk [30]. Greater exposure to infant siblings in the first six years of life was associated with reduced risk of having elevated antibodies to Epstein–Barr virus (or history of glandular fever) and was inversely associated with MS [31]. Another study, however, found no association between MS and later-age (10–14 years) viral infection [32].

Box 1 Summary: Design of the Ausimmune Study

Study design: a multicentre matched case-control study in incident cases.

Source population: all residents of four specified geographical regions (defined by postcodes and spanning 16 degrees of latitude), aged between 18 and 59 years of age. Geographical regions were chosen to achieve a source population from which sufficient cases were likely to accrue within three years to provide an adequate sample size.

Study sample: Cases – first clinical diagnosis of CNS demyelinating disease between 1 November 2003 and 31 December 2006.

Controls – randomly selected from the Australian Electoral Roll, resident in the defined geographical region and sex and age (within two years) matched to a case.

Outcome: first clinical diagnosis of CNS demyelination; subgroup analysis will examine groups defined by the presence (or not) of MRI brain scan lesions at diagnosis and groups defined by progression (or not) to MS diagnosed according to the 2005 revised McDonald criteria [48].

Measures:

1) Validated questionnaire measures of recent summer and winter sun exposure [49].
2) Serum 25(OH) D as a measure of blood vitamin D level.
3) Lifetime sun exposure and skin type.
4) Silicone-rubber skin surface casts, as a non-invasive measure of photo-ageing and cumulative sun exposure.
5) Sun exposure index using spectrophotometry (the relative increase in skin pigmentation in a UV-exposed site compared to a UV-unexposed site). Skin pigmentation increases with age in UV-exposed but not UV-unexposed sites, thus this measure provides a measure of cumulative UV exposure.
6) Spectrophotometric skin type (high correlation with histological melanin content).
7) Other environmental factors assessed by structured questionnaire, including data on the timing of exposures by age:
   a) History of childhood infection: infectious mononucleosis (glandular fever), herpes simplex, varicella, measles, rubella, mumps, any infectious illness leading to three or more school absences.
   b) Sibling structure, age and sex of siblings and half siblings, including the determination of inter-sibling birth intervals between subject and all siblings and half siblings.
   c) Exposure to pesticides, organic solvents, and other chemicals by age period.
   d) Immunization and breastfeeding history.
   e) Tobacco smoking and alcohol intake, past and current.
   f) Antenatal and infant vitamin D and cod liver oil administration; month and residence of birth [50].
   g) Demographic variables including ancestry, education, employment, past occupational history.
   h) Recent stressful life events.
   i) General medical history including other autoimmune diseases, asthma and allergic disease, head injury, history of skin cancer, family history of MS or other immune disorders.
   j) Usual dietary intake, using a self-administered 145-item semiquantitative food frequency questionnaire including dietary supplements.
   k) The International Physical Activity Questionnaire with modification to focus only on a usual week in the past year when the subject was not ill.
   l) Life course documentation of familial, social and occupational contacts with children [51].
   m) Subject's opinions on possible causes of MS.
8) A blood sample: 35 mL of venous blood is collected (15 mL serum, 10 mL whole blood in EDTA, 10 mL whole blood in ACD).

Other factors, eg, diet, physical activity

1) *Diet, physical activity and latitude*: There are clear north-south trends in dietary intake in the European Union (EU), with southern EU states having higher intake of monounsaturated fatty acids and lower intakes of saturated fatty acids, than northern EU states [33]. There is little information about geographical patterns of physical activity, but outdoor activity is likely to be influenced by ambient temperature, with perhaps lower levels of physical activity at both temperature extremes.
2) *Diet, physical activity and MS*: Ecological evidence indicates that high intake of saturated fat is associated with greater risk of MS, and meta-analysis of three randomized intervention trials indicates that unsaturated fatty acids may ameliorate the course of MS [34]. A number of studies have examined the importance of physical activity in persons already diagnosed with MS (reviewed in [35]) and physical trauma has been suggested as a risk factor for MS onset [36] (with inconclusive findings), however, there appears to be little research to support a role of either low or high levels of physical activity, as risk factors for MS onset.

There is limited support for a possible role for various other environmental factors, such as stress [37], exposure to organic solvents [38], and contact with pets [27], and the timing of such environmental exposures may be important [30]. However, while these factors may have a role in the risk of MS, they appear unlikely to contribute strongly to the latitudinal gradient in MS prevalence in Australia. Nevertheless, we are measuring these factors also. Multiple exposure-disease associations will be examined. Multiple comparison testing will not be conducted because, although this reduces the likelihood of Type 1 error, it increases the likelihood of Type 2 error [39]. We will use epidemiologic analysis to assist the assessment of whether an association has causal features [40] and to exclude non-causal explanations such as confounding [41]. Further, MS is a complex multifactorial disease and the limitation of study measures to the main exposures of interest would limit the study's ability to evaluate whether such exposures acted independently or were modified by other important disease determinants. The key hypotheses are shown in Box 2.

Reducing bias in MS case-control studies: the Ausimmune Study

The specific aims of the Ausimmune Study are to better understand the role of environmental factors, particularly past UVR exposure and past and current vitamin D intake and levels, in relation to the onset of CNS demyelinating disease (see Box 1). These aims have particular significance in relation to certain biases that are common in case-control studies, and which can misrepresent the true relationship of interest that exists within the source population.

Selection bias

Selection bias arises because of 'systematic differences in characteristics between those who take part in a study and those that do not' [42]. Selective processes may pertain only to cases or to both cases and controls.

Selection bias affecting cases only

1) *Prevalence-incidence (Neyman) bias* [43]: Prevalent cases are a subset of the complete set of incident cases that originally occurred within the study's source population. They differ from the incident cases in that none of them has died, migrated or become otherwise uncontactable – and such differences may reflect differences in exposure history. Further, the longer the prevalent cases have lived with the disease, the more likely it is that their behaviours will change with respect to the reporting of, or distribution of, hypothesized exposure factors – such as sunlight exposure and UVR-induced vitamin D levels (although, strictly, this relates to measurement bias/confounding). In the Ausimmune Study, the focus on incident CNS demyelination as the 'case' disease criterion reduces the risk of this potential source of bias and affords the opportunity to assess vitamin D levels at disease onset. Note however, that due to the sometimes insidious nature of disease onset in MS, symptom-associated behaviours, eg, lessening of symptoms by avoiding heat or sun exposure, could alter some 'exposure' parameters prior to formal diagnosis and inclusion as an incident case. In the Ausimmune Study, cases are sub-grouped: true FDE, primary progressive MS, first clinical diagnosis of CNS demyelination but a past history of suggestive symptoms, and others. This will allow the assessment of bias related to clinical course.
2) *Unmasking (detection signal) bias*: An innocent exposure, eg, stress, causing a sign or symptom, eg, numbness, tingling, that leads to detection of the disease may result in those with that exposure (stress) being more likely to be diagnosed because of this more thorough clinical assessment. While a diagnosis of FDE or MS might be

Box 2 Study Hypotheses

Table i Hypotheses for the original Ausimmune Study*

Hypothesis	Exposure	Outcome	Statistical model	Study power
There is a latitudinal gradient of increasing FDE incidence with increasing latitude in Australia in 2003–2006.	Latitude	FDE incidence in Australia in four regions, 2003–2006	Poisson regression	90% power ($\alpha = 0.05$) to detect an incidence ratio of a 3% increase per degree of latitude
There is a latitudinal gradient of increasing vitamin D insufficiency (serum 25(OH) D <40nmol/L) with increasing latitude in Australia (2003–2006) among healthy controls	Latitude of residence	Proportion of controls with vitamin D insufficiency	Poisson regression	90% power ($\alpha = 0.05$) to detect a gradient of a 0.75% increase per degree of latitude
Low personal lifetime UVR exposure, as measured by actinic damage on the dorsum of the hand and other sun exposure measures, is associated with increased risk of a FDE after controlling for potential confounding factors such as skin type.	Personal lifetime UVR exposure (high vs low)	Newly diagnosed FDE case status	Conditional logistic regression	90% power ($\alpha = 0.05$) to detect matched odds ratios (ORs) for varying prevalences of high exposure, as listed below Prevalence: 5% — Detectable matched odds ratio: 0.41 or less Prevalence: 10% — Detectable matched odds ratio: 0.50 or less Prevalence: 20% — Detectable matched odds ratio: 0.58 or less Prevalence: 50% — Detectable matched odds ratio: 0.63 or less
There is potentiation between high sun exposure behaviour and low residential latitude in lowering the risk of FDEs. That is, the protective effect of high outdoor activity (eg, outdoor occupation) will be more evident in a high residential UVR location such as Brisbane (27°S) than a lower UVR location such as Tasmania (43°S).	High sun exposure behaviour and low residential latitude	Newly diagnosed FDE case status	Conditional logistic regression	Simulation studies indicate 90% power ($\alpha = 0.05$) to detect a difference in the magnitude of the ORs for this interaction similar to that reported by Freedman et al [52].

Box 2 Study Hypotheses (Continued)

Table ii Hypotheses for the viral substudy ** of the Ausimmune Study

Hypothesis	Exposure	Outcome	Statistical model	Study power
Higher EBV or HHV-6 IgG levels are associated with increased risk of a FDE after controlling for potential confounding factors.	EBV and HHV-6 IgG levels	Newly diagnosed FDE case status	Conditional logistic regression	Simulation studies indicate 90% power ($\alpha = 0.05$) to detect a 30% increase in mean titre in MRI-positive FDE cases relative to their matched controls
Higher peripheral blood viral load of EBV or HHV-6 is associated with increased risk of a FDE after controlling for potential confounding factors.	EBV and HHV-6 peripheral blood viral load	Newly diagnosed FDE case status	Conditional logistic regression	Assuming that whole-blood viral loads are lognormally distributed with $\sigma = 0.5$, [53] we will have 80% power to detect, at $P < 0.05$, a 39% increase in viral load in cases over their matched controls
There is a synergistic interaction between low past UVR exposure and high EBV IgG titres in the development of FDEs.	Low past UVR exposure and high EBV IgG	Newly diagnosed FDE case status	Conditional logistic regression	Simulation studies indicate 90% power ($\alpha = 0.05$) to detect a 2.5-fold increase in mean IgG titre in normal-UVR cases but a five-fold increase among low-UVR cases
There is a gradient of increasing EBV and HHV-6 IgM and IgG levels with increasing latitude in Australia (2003–2006) among healthy controls.	Latitude of residence	EBV and HHV-6 antibody titres (controls only)	General linear models	90% power ($\alpha = 0.05$) to detect a linear increase in the proportion of the sample with excessive antibody levels (top quartile) of 0.75% per degree of latitude

*See Table 2 for sample size.

**Viral load sample consists of a proposed 151 cases and 295 controls, with viral serological tests on 1338 subjects (452 cases and 886 controls).

EBV = Epstein–Barr virus; HHV-6 = Human Herpes virus 6.

precipitated by an innocent exposure such as stress, ie, earlier diagnosis, it seems unlikely that the diagnosis would be completely missed in the absence of the 'innocent exposure'. When cases are subgrouped it is possible that the true FDE group may be biased towards those with higher stress (slower diagnosis in non-stressed may mean they have a second event before diagnosis as CNS demyelination). An incorrect association between stress and FDE might then be inferred.

Selection bias affecting both cases and controls

1) *Admission rate (Berkson) bias* [43]: In hospital-based case-control studies participants may be systematically different from the source population, based on their exposure status. In the Ausimmune Study, participants are recruited from the general population.
   a) Recruiting cases: Admission rate bias could still occur in the Ausimmune Study setting if only a particular subset of possible cases presented to notifying doctors. For example, if neurologists alone could notify cases to the study, then only persons referred to a neurologist, rather than managed by other doctors, would be enrolled. This would potentially exclude persons with uncomplicated optic neuritis, those resident in areas not serviced by a neurologist or those unable to access a neurologist for other reasons. To reduce this source of bias, cases in the Ausimmune Study are notified to the study by a range of clinicians including neurologists, ophthalmologists, general physicians and occasionally general practitioners. A two-tier notification system utilizing both clinicians and radiology practices performing MRI maximizes case notification. Case eligibility is then reviewed by study neurologists after notification to the study.
   b) Recruiting controls: To avoid admission rate bias in controls, the Ausimmune Study uses population-based controls, rather than hospital controls.
2) *Non-respondent bias*: If non-respondents differ systematically from respondents in their exposure or outcome history then the study results may be biased. For example, if young sun-loving potential controls were too involved in outdoor activities to take time to participate as controls in the Ausimmune Study, the resulting control group could be biased towards less active, indoor-living participants who are not typical of the source population from which the cases derived.
   a) Minimizing non-response in cases: Non-respondent bias is best avoided by maintaining a high response rate. In the Ausimmune Study, eligible cases are generally keen to participate in order to better understand the causes of the disease and factors involved in progression. The case response rate has been maintained at over 90%.
   b) Minimizing non-response in controls: Sackett recommends aiming for response rates of at least 80% [43], but few recent Australian case-control studies have been able to achieve this level. The Ausimmune Study uses a number of strategies to enhance the response rate:
      i) scheduling study interviews for a time and place convenient to the participant;
      ii) offering parking vouchers or taxi fares to participants for interviews taking place at the study premises;
      iii) offering a grocery voucher or movie tickets to participants as reimbursement for their time taken to participate;
      iv) offering an anaesthetic patch prior to venepuncture.

In addition, controls who do not wish to participate are invited to answer eight brief questions covering the key exposures of interest, to determine the extent of any systematic difference between respondents and non-respondents on these exposures. A number of factors associated with non-response in the Ausimmune Study control group are described in Table 3.

Measurement/classification bias

This type of bias occurs when there is 'systematic error arising from inaccurate measurements (or classification) of subjects on a study variable(s)' [42]. In case-control studies this may result in systematic misclassification of 'exposed' and 'non-exposed', based on the case or control status.

Bias in exposure measurement

1) *Observer (interviewer) bias*: If an interviewer seeks information on an exposure more intensely in cases than controls, cases may then tend to over-report that exposure – and the data will be biased. In the Ausimmune Study, nurse-interviewers are provided with rigorous training and standardized fieldwork protocols. This standardization is ongoing, with regular reviews and audits of study procedures.
2) *Recall bias*: Cases may recall past exposures differently from controls, because their disease

Table 3 Factors associated with non-response in the Ausimmune Study control group (matched controls contacted or unable to be located from 1 November 2003 to 28 February 2006)
(Note: all chosen possible controls are considered eligible unless out of region or out of age range. This includes those not able to be contacted and those not participating because of personal or family illness or language difficulty, as well as those refusing to participate.)
3A) Control participation by study region

Region	Participating *n* (%)	Not participating *n* (%)
Brisbane	96 (50.8)	93 (49.2)
Newcastle	57 (58.2)	41 (41.8)
Geelong	111 (60.0)	74 (40.0)
Tasmania	54 (67.5)	26 (32.5)

3B) Control participation by age group and sex

		Participating	
Age (years)	Male (%)	Female (%)	Overall *n* (%)
20–24	57.1	47.4	13 (50.0)
25–29	48.2	48.8	33 (48.5)
30–34	26.1	62.7	53 (54.1)
35–39	63.9	61.5	74 (62.2)
40–44	47.1	70.0	43 (64.2)
45–49	61.5	55.6	48 (56.5)
50–54	71.4	66.0	36 (66.7)
55–59	33.3	60.9	18 (51.4)

3C) Reasons for non-participation

	n (%)
Refuse – no reason given	84 (36.4)
Refuse – too busy	54 (23.4)
Refuse – too ill	6 (2.6)
Refuse – family difficulty	6 (2.6)
Refuse – language difficulty	3 (1.3)
Not able to be located	78 (33.8)

positive status has made them think differently about possibly etiological past exposures. Sun exposure history over the lifecourse is a key exposure of interest in the Ausimmune Study, but particularly difficult to recall accurately. The Ausimmune Study uses objective (unbiased) measures of cumulative sun exposure, ie, silicone casts of the back of the hand, as well as self-reported sun exposure utilizing a lifetime calendar, as used in other studies examining past sun exposure history [12,44]. Participants provide information, for every year of life, on, for example, their place of residence, their school or work and what pets they had. Sun exposure questions are then tied to these details, using memorable past events as guideposts to assist recall. Sun exposure history derived from the calendar is reproducible for reported sun exposure during childhood and adolescence [45] and shows moderate agreement with cumulative sun exposure as measured by the silicone casts [45]. In addition, case participants in the Ausimmune Study are asked about their thoughts on the importance of particular risk factors, in the development of their illness. Finding a similar association for a putative risk factor among those who did and did not think the factor was important, increases the confidence that bias due to pre-held beliefs did not influence the answer [12].

3) *Family information bias*: Following a diagnosis of FDE, cases are likely to seek or be given information about possible risk exposures, by family and friends. Cases may thus report exposure (or lack of exposure) from an information base which is systematically different for cases and controls. Although recruitment of incident FDE rather than prevalent MS cases should reduce family information bias, the extent of persisting bias will depend on whether the case FDE diagnosis was discussed in the family.

Bias in health outcome measurement

Bias in measurement of the health outcome (CNS demyelination) may occur if those exposed are more likely to be classified as cases because this diagnosis is sought with greater intensity. As previously noted, in the Ausimmune Study, the clinical history of each participating case is reviewed by the study neurologist team and subgroups constructed. The 'true FDE' group can be further subdivided into those with one or more lesions on MRI brain scanning (MRI-positive) and others with no lesions (MRI-negative). The collection of data on MRI-negative single demyelinating events is important for two reasons: 1) for analysis of the association between environmental factors and lesion number, the MRI-negative cases will provide an important baseline group with no lesions; 2) as an alternative case group who have been through the same selection and clinical procedures but are much less likely to have ongoing biological disease. A finding of an association between an exposure and MRI-positive cases, but not MRI-negative cases, would enhance an interpretation that the association was causal and did not reflect selection bias or recent clinical management.

Controlling possible confounding: matching, measurement and related analysis

Epidemiological studies aim to identify and quantify the effect of a specific exposure on a specified health outcome. Confounding may occur when a third

factor is statistically associated with the exposure of interest and is an independent risk factor for the outcome of interest. The association between exposure and health outcome may then appear to be stronger or weaker than it truly is, because of misattribution of part of the influence of the third factor to the exposure factor of interest. Confounding can be controlled in study design, by restriction or matching, or in data analysis, but both require accurate measurement of potential confounders. In the Ausimmune Study, cases and controls are matched on age, sex and study region of residence. Various other possible confounders are measured to enable subsequent adjustment in the analysis phase: these include lifestyle factors such as smoking status and history, physical activity practices and diet, and other factors such as skin type. The adjustment will be done using conditional logistic regression modelling, which is one of the standard models for matched case-control studies. The analysis will also allow an exploration of both interactions (effect modification) between risk factors, and whether related factors are antecedent causes of exposure, confounders or intermediates between an exposure and outcome [41].

Enhancing external validity

The points made above refer primarily to 'internal validity' – the ability of the study to provide a correct estimate, for the designated source population, of the association between the exposure and outcome of interest. However it is also important to consider external validity, or generalizability to different groups. The source population in the Ausimmune Study is predominantly Caucasian. What, then, is the external validity of any study findings on time in the sun to those of different skin type? Darker skin requires longer sun exposure to produce an equivalent amount of vitamin D: intermediate (Asian) skin types require twice as long and deeply pigmented skin types may require six times longer than fair skin types [46,47]. In order to enhance external validity, the Ausimmune Study contains study measures that allow for very careful phenotyping of skin type. With a large sample size, it may be possible to examine subgroups by skin type, allowing better generalizability to populations of various skin types.

Concluding remarks

This report has outlined key considerations in the design and conduct of the Ausimmune Study: case-control design for the efficient study of multiple risk factors for an uncommon disease; multicentre down the eastern Australian seaboard to investigate aspects of the latitudinal gradient in MS; population-based recruitment of incident cases to minimize selection bias; use of standardized fieldwork protocols to minimize classification bias; and case-control matching and measurement of a wide range of possible confounders to control possible confounding.

Well-conducted observational epidemiological studies can provide important evidence on the causes of MS. Rigorous study design – taking consideration of the exposures and outcomes of interest and the etiological questions being asked – is extremely important. However, the ability of the study to provide valid results relies on anticipating and minimizing bias, recognizing and controlling possible confounding and providing findings of relevance to other populations.

Acknowledgements

This work is supported by a grant from the National Multiple Sclerosis Society of the USA and the National Health and Medical Research Council of Australia (No. 316901) with additional support from Multiple Sclerosis Research Australia and the Multiple Sclerosis Societies of Australia.

Dr Lucas is supported by a NHMRC Capacity Building Grant, 'Environment and Population Health: Research Development from Local to Global, 2003–2007 (No. 224215)' and a Macquarie Bank Multiple Sclerosis Research Australia Fellowship.

References

1. **Willer CJ, Dyment DA, Sadovnick AD, Rothwell PM, Murray TJ, Ebers GC.** Timing of birth and risk of multiple sclerosis: population based study. *Br Med J* 2005; **330**: 120.
2. **Villoslada P, Juste C, Tintore M, Llorenc VV, Codina G, Pozo-Rosich P** *et al*. The immune response against herpesvirus is more prominent in the early stages of MS. *Neurology* 2003; **60**: 1944–48.
3. **Bach JF.** The effect of infections on susceptibility to autoimmune and allergic diseases. *N Engl J Med* 2002; **347**: 911–20.
4. **Barnett MH, Williams DB, Day S, Macaskill P, McLeod JG.** Progressive increase in incidence and prevalence of multiple sclerosis in Newcastle, Australia: a 35-year study. *J Neurol Sci* 2003; **213**: 1–6.
5. **Riise T, Wolfson C.** The epidemiological study of exogenous factors in the etiology of multiple sclerosis: Final thoughts. *Neurology* 1997; **49**(2 Suppl): S82.
6. **Wolfson C, Granieri E, Lauer K.** Case-control studies in multiple sclerosis. *Neurology* 1997; **492**(Suppl 2): S5–14.
7. **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.
8. **van der Mei IA, Ponsonby AL, Blizzard L, Dwyer T.** Regional variation in multiple sclerosis prevalence in Australia and its association with ambient ultraviolet radiation. *Neuroepidemiology* 2001; **20**: 168–74.
9. **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.
10. **Granieri E, Casetta I, Tola MR.** A multicenter study methodologic experience from a multicenter case-control study in Italy. The Italian Multiple Sclerosis Study Group. *Neurology* 1997; **49**(2 Suppl 2): S33–41.
11. **Touze E, Fourrier A, Rue-Fenouche C, Ronde-Oustau V, Jeantaud I, Begaud B** *et al*. Hepatitis B vaccination and first central nervous system demyelinating event: a case-control study. *Neuroepidemiology* 2002; **21**: 180–86.
12. **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. *Br Med J* 2003; **327**: 316.
13. **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**: 60–65.
14. **O'Riordan JI, Thompson AJ, Kingsley DP, MacManus DG, Kendall BE, Rudge P** *et al*. The prognostic value of brain MRI in clinically isolated syndromes of the CNS. A 10-year follow-up. *Brain* 1998; **121**: 495–503.
15. **Diffey BL.** Sources and measurement of ultraviolet radiation. *Methods* 2002; **28**: 4–13.
16. **Ponsonby AL, Lucas RM, van der Mei IA.** UVR, vitamin D and three autoimmune diseases—multiple sclerosis, type 1 diabetes, rheumatoid arthritis. *Photochem Photobiol* 2005; **81**: 1267–75.
17. **Holick MF.** McCollum Award Lecture, 1994: vitamin D – new horizons for the 21st century. *Am J Clin Nutr* 1994; **60**: 619–30.
18. **Gies P, Roy C, Toomey S, Tomlinson D.** Ambient solar UVR, personal exposure and protection. *J Epidemiol* 1999; **9**(6 Suppl): S115–22.
19. **Acheson ED, Bachrach CA, Wright FM.** Some comments on the relationship of the distribution of multiple sclerosis to latitude, solar radiation, and other variables. *Acta Psychiatr Scand* 1960; **35**(Suppl 147): 132–47.
20. **McMichael AJ, Hall AJ.** Does immunosuppressive ultraviolet radiation explain the latitude gradient for multiple sclerosis? *Epidemiology* 1997; **8**: 642–45.
21. **Goldacre MJ, Seagroatt V, Yeates D, Acheson ED.** Skin cancer in people with multiple sclerosis: a record linkage study. *J Epidemiol Community Health* 2004; **58**: 142–44.
22. **Guernier V, Hochberg ME, Guegan JF.** Ecology drives the worldwide distribution of human diseases. *PLoS Biol* 2004; **2**: e141.
23. **Ascherio A, Munger KL, Lennette ET, Spiegelman D, Hernan MA, Olek MJ** *et al*. Epstein-Barr virus antibodies and risk of multiple sclerosis: a prospective study. *JAMA* 2001; **286**: 3083–88.
24. **Ablashi DV, Lapps W, Kaplan M, Whitman JE, Richert JR, Pearson GR.** Human Herpesvirus-6 (HHV-6) infection in multiple sclerosis: a preliminary report. *Mult Scler* 1998; **4**: 490–96.
25. **Clausen J.** Endogenous retroviruses and MS: using ERVs as disease markers. *Int MS J* 2003; **10**: 22–28.
26. **Swanborg RH, Whittum-Hudson JA, Hudson AP.** Infectious agents and multiple sclerosis – are Chlamydia pneumoniae and human herpes virus 6 involved? *J Neuroimmunol* 2003; **136**: 1–8.
27. **Vandevelde M, Zurbriggen A.** Demyelination in canine distemper virus infection: a review. *Acta Neuropathol (Berl)* 2005; **109**: 56–68.
28. **Sondergaard HP, Theorell T.** A putative role for Toxocara species in the aetiology of multiple sclerosis. *Med Hypotheses* 2004; **63**: 59–61.
29. **Kriesel JD, Sibley WA.** The case for rhinoviruses in the pathogenesis of multiple sclerosis. *Mult Scler* 2005; **11**: 1–4.
30. **Hernan MA, Zhang SM, Lipworth L, Olek MJ, Ascherio A.** Multiple sclerosis and age at infection with common viruses. *Epidemiology* 2001; **12**: 301–306.
31. **Ponsonby AL, van der Mei I, Dwyer T, Blizzard L, Taylor B, Kemp A** *et al*. Exposure to infant siblings during early life and risk of multiple sclerosis. *JAMA* 2005; **293**: 463–69.
32. **Bager P, Nielsen NM, Bihrmann K, Frisch M, Hjalgrim H, Wohlfart J** *et al*. Childhood infections and risk of multiple sclerosis. *Brain* 2004; **127**: 2491–97.
33. **Gibney MJ.** Nutrition, physical activity and health status in Europe: an overview. *Public Health Nutr* 1999; **2**: 329–33.
34. **Schwarz S, Leweling H.** Multiple sclerosis and nutrition. *Mult Scler* 2005; **11**: 24–32.
35. **Motl RW, McAuley E, Snook EM.** Physical activity and multiple sclerosis: a meta-analysis. *Mult Scler* 2005; **11**: 459–63.
36. **Weatherby SJ, Hawkins CP.** Does trauma trigger multiple sclerosis? 1: A controversy. *Hosp Med* 2003; **64**: 581–84.
37. **Ackerman KD, Heyman R, Rabin BS, Anderson BP, Houck PR, Frank E** *et al*. Stressful life events precede exacerbations of multiple sclerosis. *Psychosom Med* 2002; **64**: 916–20.
38. **Mortensen JT, Bronnum-Hansen H, Rasmussen K.** Multiple sclerosis and organic solvents. *Epidemiology* 1998; **9**: 168–71.
39. **Perneger TV.** What's wrong with Bonferroni adjustments. *Br Med J* 1998; **316**: 1236–38.
40. **Lucas RM, McMichael AJ.** Association or causation: evaluating links between "environment and disease". *Bull World Health Organ* 2005; **83**: 792–95.
41. **Rothman KJ, Greenland S.** *Modern epidemiology* (2nd edn). Lippincott-Raven Publishers, 1998: 1–711.
42. **Last J,** ed. *A dictionary of epidemiology.* Oxford University Press, 2001.
43. **Sackett D.** Bias in analytic research. *J Chronic Dis* 1979; **32**: 51–63.
44. **Kricker A, Vajdic CM, Armstrong BK.** Reliability and validity of a telephone questionnaire for estimating lifetime personal sun exposure in epidemiologic studies. *Cancer Epidemiol Biomarkers Prev* 2005; **14**: 2427–32.
45. **English DR, Armstrong BK, Kricker A.** Reproducibility of reported measurements of sun exposure in a case-control study. *Cancer Epidemiol Biomarkers Prev* 1998; **7**: 857–63.
46. **Lo CW, Paris PW, Holick MF.** Indian and Pakistani immigrants have the same capacity as Caucasians to produce vitamin D in response to ultraviolet irradiation. *Am J Clin Nutr* 1986; **44**: 683–85.
47. **Lucas RM, Ponsonby AL.** Ultraviolet radiation and health: friend and foe. *Med J Aust* 2002; **177**: 594–98.
48. **Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L** *et al*. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". *Ann Neurol* 2005; **58**: 840–46.
49. **Dwyer T, Blizzard L, Gies PH, Ashbolt R, Roy C.** Assessment of habitual sun exposure in adolescents via questionnaire – a comparison with objective measurement using polysulphone badges. *Melanoma Res* 1996; **6:** 231–39.
50. **Hypponen E, Laara E, Reunanen A, Jarvelin MR, Virtanen SM.** Intake of vitamin D and risk of type 1 diabetes: a birth-cohort study. *Lancet* 2001; **358**: 1500–503.
51. **Thomas SL, Wheeler JG, Hall AJ.** Contacts with varicella or with children and protection against herpes zoster in adults: a case-control study. *Lancet* 2002; **360**: 678–82.

52. **Freedman DM, Dosemeci M, Alavanja MC.** 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**: 418–21.
53. **Balandraud N, Meynard JB, Auger I, Sovran H, Mugnier B, Reviron D** *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**: 1223–28.
54. **Hennekens C, Buring J.** *Epidemiology in medicine*, (1st edition). Little, Brown and Company, 1987.