Comparing Sample Surveys of Health with Official Population Statistics: Some Methodological Issues and Empirical Findings

Brian I. O'Toole[1], Richard P. Marshall[2], Ralph J. Schureck[3], and Matthew Dobson[4]

This article reports an examination of factors that might affect the interpretation of computed relative risks of medical illness obtained when estimates of morbidity derived from a specific subpopulation sample are compared with estimates obtained from official statistical bureau surveys of the parent population. Interpreting a significant or nonsignificant relative risk depends upon how fair the comparison is. Three types of error are considered: (1) those arising from subjects, including coverage, selection, response and comparability; (2) those arising from measurement processes, including data collection instruments, interviewers and data processing; and (3) those arising from confounding, which occurs when a risk factor for outcome is differentially distributed between exposure groups. An example is given from a comparison of two sets of morbidity estimates, one obtained from an epidemiological cohort study of a national random list sample of Australian Vietnam veterans, the other obtained by an Australian Bureau of Statistics national area probability sample. Variables measuring prevalence of 37 recent and chronic illness conditions were derived from both studies and their ratio computed as the relative risk of illness in the veteran sample compared with the Australian population sample. This risk was greater than 1.0 for all except one condition, and 18 of 36 recent conditions and 30 of 37 chronic conditions carried 99% confidence intervals excluding 1.0. Adjustment for variables that were thought to have potentially biased the relative risks gave varied results. The issue of when an adjustment becomes an overadjustment depends upon the meaning of the variable and its place in a putative causal pathway.

Key words: Surveys; health statistics; Vietnam veterans; relative risk; bias adjustment.

1.   Introduction

Epidemiological cohort studies (Liddell 1988) are designed to contrast two groups of people, one of which has been exposed to some particular (usually noxious) agent, the other of which has not and which therefore serves as a ''control'' group. These two groups are then compared to determine differences in disease status that, other things being equal, may be argued to be attributable to the exposure (Hill 1965). Ideal study design selects these two groups at the same time independently from a single well defined target population or ''study base'' (Miettinen 1985), with exposure being assumed to stratify this population into two mutually exclusive strata from which independent selections can be made and subjects followed through time. However, under some circumstances it is not possible to select appropriate unexposed controls at the same time as the exposed subjects. Under these circumstances it may be possible to compare the exposed group with a normative unexposed population, such as might arise from a census or representative population sample survey.

   The inference from such a statistical analysis that the ''exposure'' variable has ''caused'' some particular health ''outcome'' relies critically upon the assumptions made about the population(s) from which the measurements are taken and the assumptions that are made about the parameter(s) that are estimated. Any sources of error affecting these assumptions, for example distinct differences between the exposed and unexposed on some important factor, can bias the relative risk estimates. Subjects need to be selected, sampled and to respond; differences in coverage, sampling methods and response rates between the two groups being compared may degrade the confidence in the comparison. All field surveys suffer some nonresponse rate, and sources of error attributable to nonresponse, particularly differential nonresponse between the subpopulation survey and the population survey, could affect the relative morbidity estimates. Measurement error needs to be minimised, both within each group and between groups, including that arising from interviewing, question design, coding, or other survey conditions (Copeland, Checkoway, McMichael, and Holbrook 1977; de Klerk, English, and Armstrong 1989; Groves 1989). In comparing the prevalence of various conditions between two independent samples, additionally it must be shown that other disturbing (Kish 1985) or confounding variables (Boivin and Wacholder 1985; Grayson 1987) are not operating that may already discriminate between the two groups.

   In 1989–90, the Australian Bureau of Statistics (ABS) conducted a national health survey, selecting an area probability sample and using trained ABS interviewers and standardised survey procedures to derive estimates of population morbidity (ABS 1990). This health survey did not explicitly identify or sample war veterans, whether from World Wars I or II, Korea, Vietnam, the Gulf war, or any other conflict with which Australia has been involved, so that information on the potential effects of war on subsequent health would need to come from specific studies conducted outside of ABS. Prior to the ABS survey, the current authors were planning the Australian Vietnam Veterans Health Study (AVVHS), a national epidemiological cohort study of the physical and psychological health and welfare of Australian Vietnam veterans (O'Toole et al. 1996a, b, c). Because of the difficulty of constructing a similar cohort study for following nonveterans, a cross-sectional comparison was envisaged as used in recent major studies

of American Vietnam veterans (Centres for Disease Control 1988a, b, c, d; Kulka et al. 1990a, b), and the possibility recognised of bridging the gap to the Australian population to compare, cross-sectionally, the current state of health of the veterans with population estimates provided by the ABS survey. By using ABS methods and instrumentation in the AVVHS to derive estimates of morbidity, the comparison of morbidity might have less potential for measurement biases.

When a statistically significant relative risk is observed, indicating a greater health burden on the veterans, then the interpretation of this as a ''real'' difference involves more than inspection of the confidence intervals. Assumptions about the sampling distributions and their effects must also be tempered with assumptions about the nonsampling aspects of measurement and detection of underlying confounding. Some of these have been identified in the circumstances provided by the two samples. Within the framework of subjects, measures and confounding a number of factors are discussed that might affect the relative risks, and an empirical demonstration of their control is presented.

2.   Methods

From October 1989 to September 1990, ABS conducted fieldwork for the 1989–90 National Health Survey. From across Australia an area probability sample of 26,470 households was selected using multi-stage sampling, yielding a total population sample of 54,576 people aged 15 and above. There was some household sample loss, at 16.1%, giving an effective household sample of 83.9% (22,202 households). All members of a household were to be interviewed, although there was again some small loss (3.9%). Refusers and noncontacts were not replaced. All households were initially approached by mail, whereby they were given a brochure about the survey including confidentiality provisions. Interviews took place in-home, and allowed 3–5 calls back for their completion; average interview time was 52 minutes (ABS 1990). ABS does not specifically identify veterans, thus there may be veteran representation in the ABS sample. Australia's population of approximately 18 million, when sampled to a size of 26,470 households, would be unlikely to contain many of the potential 50,000 Vietnam veterans available for study. Whatever veterans' health state, there would be insufficient representation of them in the ABS-HIS to heavily influence the Australian estimates.

The AVVHS began with a simple random sample of 1,000 male Australian Army Vietnam veterans selected from a computer file held by the Australian Army believed to contain the entire complement of men who were posted to Vietnam. The sample was traced to residential addresses using available computerised and manual registers held in government, ex-service and community sectors. Subjects were approached initially by mail, whereby they were asked to confirm their veteran status and Regimental Service Number, were given a brochure about the study and were requested to return a signed card giving consent to participate. Nonrespondents were followed to all known addresses or telephone numbers until all yielded no contact after three attempts and visiting the address where possible (mainly in the major cities) also failed to return a contact.

Of the 1,000 men selected, 50 had died (eight in Vietnam), 259 had no available address, 48 refused initially to participate and a further 13 refused at the time of interview; 35 were not interviewed for logistic reasons, leaving a total of 641 veterans interviewed; this

represents 67.5% of selected eligible subjects and 92.8% of contacted subjects. Subjects were interviewed across Australia by volunteer counsellors from the Vietnam Veterans Counselling Service (VVCS), members of the research team or their assistants and volunteer officers of the Australian Army Psychology Corps. In the AVVHS, the ABS interview was the opening section of a much longer interview, which went on to paper-and-pencil psychology tests, military and civilian experience and combat interviews, and psychiatric diagnostic interviews. Interviews took place mostly in veterans' homes, but also in ex-services clubs, worksites, farms, and prisons and lasted an average of approximately four hours (longest nine hours with a disturbed veteran); the health interview component was of a similar length to the ABS interview.

2.1. Calculation of relative risks

ABS electronic data was available from the Australian Health Survey for comparison with the veteran data. As the age and sex distributions of the veterans and the Australian population sample were different, attention was confined to the ABS data for males within the age range of the veteran sample (38–73) at the time of interview. This gave 11,468 cases in the comparison ABS unit record file. Because of the clear potential for confounding by age (discussed below), standardisation of the ABS data to the age distribution of the veteran sample was computed for all calculations reported here. The prevalence (proportion) of cases of each illness category (e.g., arthritis) in each age stratum among the interviewed veterans was compared with the corresponding expected proportion of cases in that age stratum given by the ABS data. The ratio of the sum across strata of the obtained prevalence in the veterans to the sum across strata of the expected population prevalence was computed as the summary relative risk. Although the ABS morbidity data arises from a sample survey, requiring computation of the relative risk as the ratio of two random variables, here it was assumed that the ABS estimates were population parameters, permitting a simpler form for calculation of confidence intervals (Fleiss 1981). The finite population correction factor was ignored in both the ABS and the AVVHS calculations.

However, significance levels are often set by convention at $p < 0.05$, with a significant relative risk having a 95% confidence interval that does not include 1.0. However, since many confidence intervals were computed it is possible that many may be significant by chance alone. Since most methods of controlling for the multiple test problem involve a change in significance level, and since it is known that varying levels of missing data lead to underestimation of the confidence limits (Rubin 1987), it was decided to compute 99% confidence intervals, permitting readers to judge a less restrictive criterion.

2.2. Nonresponse adjustment

Bias due to nonresponse is well recognised in the survey and epidemiological literature (Groves 1989; Clarke, Aneshensel, Frerichs, and Morgan 1983; Decoufle, Holmgreen, Calle, and Weeks 1991; Eaton, Anthony, Tepler, and Dryden 1992), with the concern that the subjects of most interest (i.e., sickest) will be those less likely to respond. Unfortunately, ABS has published no data in its survey regarding nonrespondents, acknowledging that, in the absence of this information it is ''...not possible to accurately quantify the

nature and extent of the differences between responders and nonresponders..." (ABS 1990, p. 50). The AVVHS, however, was designed as a retrospective cohort study which began at the time of entry into the army, and had access to extensive background data from military records on the entire cohort of veterans. Some of the military items had been shown to be predictive of mortality (O'Toole, Adena, and Jones 1988; O'Toole 1990; O'Toole and Cantor 1995) in earlier studies. They also might be used to compare respondents and nonrespondents in the veteran sample and to adjust statistically for non-response. Military data included pre-enlistment education and employment and the results of Army psychology classification tests, service details (type of enlistment, postings, dates, service milestones, discharge details) and conduct and casualty information.

Adjustment for nonresponse entails applying knowledge about the distribution of a particular variable among the respondents and imputing values for the unknown distribution among the nonrespondents. Regression adjustment was used (Greenland and Finkle 1995; Little and Rubin 1987, pp. 253–254) to impute for nonresponse. The probability P(R) of response ($R = 1$) versus nonresponse ($R = 0$) was computed for response indicator R for each subject in the selected alive cohort ($n = 950$). For the set of k variables $\mathbf{x} = \{x_1, x_2, x_3, \ldots x_k\}$ available from the military records, the probability of response is modelled using the logistic function $P = 1/\{1 + \exp[-(\alpha + \Sigma \beta_i x_i)]\}$. The variables x_i were chosen from the military data available for the whole cohort, selected on the basis of individual statistically significant associations with response. For subselection of data items as candidates for adjustment, instead of setting statistical significance (alpha rate) at the traditional 0.05, the criterion was relaxed to 0.1 in order to permit more sensitivity to detect potentially disturbing variables and thus to evaluate more rather than fewer variables in the analysis that searched for response bias. Thus, of the several hundred military variables available, 29 were significantly different between respondents and nonrespondents; 11 of these had p-values less than 0.05.

These variables were then submitted to prediction (forward stepwise) logistic regression models for each of the 37 recent and chronic illness groupings. Logistic regression was used as illness reports are binary, either absent (coded 0) or present (coded 1). The resulting prediction model for each illness condition was then used as an estimation model among the nonrespondents, to estimate the number of cases that would be expected in the nonrespondents based on the prediction regression model among the respondents. The adjusted prevalence among the total eligible sample of 950 selected veterans was then calculated as the sum of the cases observed in the 641 interviewed veterans plus the cases predicted among the 309 noninterviewed veterans. This was compared with the number of cases that would be expected in the 950 total veteran sample using the Australian population data based on the ABS health survey.

To calculate age adjustments, the five-year intervals used by ABS were used (35–39, 40–44, etc.). To calculate a pseudo ''age-at-interview'' for the noninterviewed subjects the date of the midpoint of fieldwork was used (November 15, 1991) by which time 59.6% of interviews had been conducted. Not all of the 29 military variables were discriminators of illness reports: that is, only some of them were significant in models differentiating those with and those without particular diagnoses. In three conditions the logistic prediction model in the respondents contained no independent variables, thus no regression adjustment for nonresponse was possible.

A single inputed value was used; it is recognised that multiple imputation is preferable in deriving more unbiased estimates of variance (Little and Rubin 1987, chapter 12; Rubin 1987) than ''naive regression'' (Greenland and Finkle 1995). However, for applications where the distributions of covariates are not misspecified (such as misspecifying a log-normal covariate distribution as normal for the purposes of imputation; Greenland and Finkle 1995) single imputation using regression seems to perform as well as multiple imputation procedures.

2.3. Measurement error

Broad differences in measurement that could affect the relative risks arise from the interview schedule (questions, order of administration, visual aids, etc.), interviewer variability in the actual administration of the questionnaire, and data processing (coding, data cleaning, computing). As the ABS interview schedule was used in the AVVHS, the actual questions and their order were the same for both the ABS survey and the AVVHS, with two slight modifications: the demographic section was relocated to the end of the questionnaire instead of at the beginning (as administered by ABS), and the diet and the women's health sections were omitted. The fieldwork, interviewer and office manuals used by ABS were made available to the AVVHS.

Medical conditions were coded into ICD-9 groupings according to procedures in the ABS conditions coding manual. ABS does not publish a coding error rate, but uses the errors detected during quality control procedures to refer to the original questionnaire and feedback is given from office staff to field staff. This was the strategy adopted for the AVVHS. All medical conditions were coded in the AVVHS twice, independently, by different coders and the results compared. All discrepancies were resolved by the chief investigator (BOT). All data were computer-entered twice, independently. Extensive quality control procedures were undertaken to clean the data. While it is not exactly the same type of data cleaning as used by the ABS, the smaller sample size of the AVVHS enabled computer editing and referral to original interview documents procedures to be more intense and repetitive if necessary. While the effects of the differences are not directly assessable, it is unlikely that they would lead to major differences in morbidity estimates.

However, the interviewers in the AVVHS were all academically qualified (most to Master's level and beyond), many were clinical counsellors from psychology and social work disciplines, and many were academic researchers; some were officers in military psychology. Few professional ABS interviewers would have these characteristics. Data on the ABS interview work force qualifications are not published, except for an endorsement of the experience of field staff and a brief description of field and classroom training. Interviewer training for the AVVHS was undertaken by the authors, with classroom and supervised interview sessions. It is therefore possible that differences in interviewer selection and training between the two studies have resulted in differences in interview administration. Standardised interviews tend to be resistant to interviewer variability, however the possibility remains of differential sensitivity to illness reports on the part of clinicians and nonclinicians. To test this hypothesis, the prevalence of illness conditions was simply compared between clinician and nonclinician interviewers.

Other measurement errors are more subtle. It has long been recognised (Cannell, Marquis, and Laurent 1977; Chambers, Spitzer, and Hill 1976; Meltzer and Hochstim 1970) that under-reporting of medical conditions is a potentially greater survey hazard than over-reporting, but both over-reporting and under-reporting have been shown to be serious problems in earlier studies of Australian military personnel (O'Toole, Battistutta, Long, and Crouch 1988). It is possible that the Australian population survey sample tended to under-report illness, while veterans may have been more accurate reporters, being concerned about their own health states. Thus, differential recall between the veterans and the population remains a potential source of bias (Coughlin 1990; Neugebauer and Ng 1991). The veterans participated in interviews that were known in advance to be lengthy and involve other components than the ABS interview. The personal interest taken in the study, the desire to ''tell their story,'' the mistrust of government departments, their expectations of the study, and the lurking compensation issue all render the veteran sample somewhat different from the general Australian population. All of these factors might have influenced the relative risks. Not all such factors are measurable or available for calibration of risks; however, two sources of potential bias that are available for scrutiny are recent doctor visits and perceived overall health state.

It has been long known that recency of medical actions affects recall, which is why the ABS questions are generally confined to the past two weeks. But frequency of attendance might also affect reporting, as the more often one visits the doctor for a particular condition the more likely the doctor is to conduct more intensive investigations, thus making a diagnosis more likely. Although doctors and their patients may not always agree (Kehoe, Wu, Leske, and Chylack 1994), people who visit a doctor more often would be more likely to receive a diagnosis (Sackett 1979) and thus more likely to remember it and to report it. Conversely, those with fewer consultations may be less likely to report diagnoses. Moreover, people with a more optimistic view of their own health generally may be less likely to report illness conditions or health service usage, and vice-versa; thus if veterans are more ''stoic'' than other men in the general population they might be more likely to under-report illness conditions.

To test this view, the relative risks were computed adjusted for the different distributions of doctor visits and perceived health in the AVVHS and ABS data. Adjustment used the strata formed by frequency of doctor visits (six strata) and perceived health (four strata) in each illness condition to compute the ratio of the sum of the age-adjusted expected frequencies across strata (from the ABS data) to the sum of the obtained frequencies in the veterans. This adjusted relative risk was then available for comparison with the unadjusted relative risk.

A major influence on the quality of inference from samples lies in the validity and reliability of the data. Unfortunately, the question of the validity of the excess veteran relative risk cannot be settled in this study. For example, studies of the reliability and validity of subjects' reports for the health survey have not been reported by ABS, nor was assessment of independent clinical records for verification of veterans' medical reports undertaken in the AVVHS. It is known that somatic and psychosomatic complaints are common among former combat soldiers (Solomon and Mikulincer 1987; CDC 1988a, d; Decoufle, Holmgren, Boyle, and Stroup 1992), with the role played by post traumatic stress disorder in the genesis of health complaints not clearly understood. The prevalence

of some illness reports in the AVVHS varied with combat exposure among the veteran sample (O'Toole et al. 1996b), in particular chronic nerves, chronic deafness, chronic rheumatism and back disorders, chronic infectious and parasitic disease, all of which may be expected given the nature of combat in a tropical jungle environment. However, when excess health complaints were observed in one interview study of U.S. Vietnam veterans, few of these were substantiated on medical examination (CDC 1988a, b). Although the interview schema, question wording, and interview details have been adopted following lessons learned over many years and from many official statistical bureau, the tools that are used are known not to be perfect.

2.4. Confounding

Confounding of the relative risks would occur when variables that were associated with an illness condition were differentially distributed in the veteran sample compared with the Australian population (Last 1988). For example, if the prevalence of smoking among veterans was higher than among the population, and smoking was related to certain illnesses, then the relative risks for those illnesses would be confounded by smoking. Age is a clear risk factor for disease that was differentially distributed in the veterans and the Australian population; this was the reason for age-sex adjustment of Australian population estimates before comparing them with the veteran sample. Any differences between the Australian population sample and the veteran sample in the distribution of health risk factors may leave the unadjusted relative risk confounded. Unfortunately, the presence of confounding cannot be assessed using statistical significance (Dales and Ury 1978), since significance is a function of sample size while confounding may be present independently of sample size. However, the size of the change in parameter estimates observed after adjustment is often used in epidemiology as a data-based criterion for assessment of the presence of potential confounding variables (Greenland 1982; Greenberg and Kleinbaum 1985).

To assess the potential for confounding by alcohol and smoking status, the relative risks were computed adjusted for smoking (three strata) and alcohol risk (three strata) using similar procedures as for the adjustment for age, doctor visit and perceived health. In addition, simultaneous adjustment was performed using the joint distribution of all these variables to produce a single estimate adjusted for age, doctor visits, perceived health, smoking and alcohol.

3. Results

Overall, veterans saw their health as poorer; compared with ABS expectations, fewer veterans described their health as ''excellent'' (prevalence, $P = 17.1\%$, relative risk, $RR = 0.58$, 99% confidence interval, $CI = 0.45, 0.71$), that similar number described their health as ''good'' ($P = 50.4\%$, $RR = 0.96$, $CI = 0.86, 1.06$), but more described their health as ''fair'' ($P = 24.3\%$, $RR = 1.70$, $CI = 1.39, 2.00$) or ''poor'' ($P = 8.3\%$, $RR = 2.18$, $CI = 1.45, 2.92$). More had seen a doctor in the past two weeks ($P = 23.7\%$, $RR = 1.49$; $CI = 1.22, 1.77$). There were fewer current smokers ($P = 26.5\%$, $RR = 0.86$, $CI = 0.74, 0.97$) but many more former smokers ($P = 54\%$, $RR = 1.69$; $CI = 1.56, 1.82$) and fewer who had never smoked ($P = 18.1\%$;

$RR = 0.49$; $CI = 0.40$, 0.58). Compared with ABS data there were fewer in the low alcohol risk category ($P = 76.5\%$, $RR = 0.89$; $CI = 0.85$, 0.93), more in the medium risk category ($P = 13.2\%$; $RR = 1.32$; $CI = 1.04$, 1.50) and more in the high risk category ($P = 9.9\%$; $RR = 1.45$; $CI = 1.08$, 1.81). Relative risks for all medical conditions were greater than 1.00 except for recent and chronic asthma; 18 of 36 recent conditions and 30 of 37 chronic conditions carried confidence intervals that excluded 1.0, revealing a general excess of reporting of conditions among the veterans. Veterans reported an average of 2.34 recent conditions (range 0–11, sd $= 1.85$) and 3.79 chronic conditions (range 0–8, sd $= 2.22$), whereas the corresponding means for the Australian population were 1.42 and 1.75 respectively, with similar ranges.

Table 1 shows the relative risks before and after nonresponse adjustment for a selection of the 36 recent (there were no blood diseases) and 37 chronic conditions. Regression adjustment for nonresponse increased the prevalence of 28 recent and chronic conditions, however there was an overall mean increase of only 1.6% for recent conditions and an

Table 1. Relative risks (RR) and 99% confidence intervals (CI) for selected chronic illness reports in Australian Vietnam veterans compared with Australian population estimates both unadjusted and adjusted for nonresponse

	Unadjusted RR (99% CI)	Adjusted RR (99% CI)
Infective and parasitic disease	4.79 (2.47, 10.77)	5.20 (2.79, 7.62)
Endocrine system		
– Cholesterol	2.75 (1.96, 4.85)	2.73 (1.94, 3.52)
– Diabetes	2.72 (1.35, 6.26)	2.71 (1.32, 4.09)
– Gout	3.11 (2.12, 5.70)	3.21 (2.21, 4.22)
– Other endocrine disorders	3.13 (1.88, 6.37)	4.96 (3.42, 6.51)
Mental disorders		
– Nerves, nervousness	8.87 (6.30, 15.73)	9.95 (7.23, 12.67)
– Depression	3.49 (1.12, 9.56)	3.47 (1.11, 5.83)
– Other mental illness	4.15 (2.37, 8.77)	4.69 (2.80, 6.58)
Sensory/Nervous system disorders		
– Disorders of refraction	1.09 (0.98, 1.48)	1.08 (0.96, 1.19)
Circulatory system		
– Hypertension	2.08 (1.63, 3.31)	2.17 (1.71, 2.62)
Respiratory system		
– Hayfever	1.54 (1.21, 2.43)	1.59 (1.27, 1.92)
– Asthma	0.87 (0.43, 2.00)	0.91 (0.46, 1.36)
Digestive system		
– Ulcer	2.83 (1.77, 5.60)	2.72 (1.67, 3.76)
– Hernia	3.28 (2.17, 6.21)	3.18 (2.08, 4.28)
Genito-urinary system	2.31 (1.09, 5.43)	2.34 (1.10, 3.57)
Musculoskeletal disorders		
– Arthritis	1.46 (1.11, 2.42)	1.48 (1.12, 1.84)
– Rheumatism	2.81 (1.02, 7.39)	2.85 (1.04, 4.66)
Injury	4.72 (2.97, 9.29)	4.86 (3.09, 6.63)
Symptoms, signs and ill defined conditions	2.75 (1.86, 5.09)	2.77 (1.88, 3.66)

overall mean decrease of 0.7% for chronic conditions. Moreover, after adjustment for nonresponse the confidence interval changed in only one disease group, recent ulcer, from significant to nonsignificant; all other significant risks remained significant, and nonsignificant risks also remained nonsignificant. Interestingly, the lack of effect of non-response adjustment was also reported in the NVVRS conducted with U.S. veterans (Kulka et al. 1990), using similar military data and regression adjustment techniques.

Differences in prevalence between clinician and nonclinician interviewers in the AVVHS were assessed using Chi-square; only three conditions showed a significant relationship ($p < 0.05$) to interviewer group: on the one hand chronic disorders of refraction, where clinicians recorded about two-thirds of the frequency of nonclinicians (ratio = 0.69), and chronic bronchitis (ratio = 0.56), on the other hand ''symptoms and signs,'' where clinicians recorded a greater frequency than nonclinicians (ratio = 1.73). Only one of these (disorders of refraction) was significant ($p < 0.01$). As there were 36 recent and 37 chronic illnesses, less than one test would be expected to be significant by chance alone at $p < 0.01$, or less than 4 at $p < 0.05$. Thus, little consistent measurement bias might be attributable to the clinical profession of the interviewers, increasing confidence in the standardisation of the ABS interview.

A factor may not be a bias unless it is associated with the exposure and with the outcome. To test the status of the number of recent doctor consultations, perceived health, smoking and alcohol, all were examined for association with each of the illness conditions in the veteran data. Recent doctor consultation was significantly associated with reports of most recent conditions, and with reports of chronic ulcer, hypertension, asthma, other mental conditions, and other endocrine conditions. Perceived health state was related to almost all conditions, in the expected direction. Alcohol risk was associated with recent health actions for rheumatism, injury, other mental disorders, and recent rash, and with chronic conditions of gout, other endocrine conditions, and other mental conditions. Smoking status was associated with recent migraine and hypertension, and with chronic hayfever, asthma, hernia, rheumatism, injury, and disability, but not cancer or heart disease. This may need a further period of observation to accommodate the latency for disease onset, however this confirms the possibility that reporting rates were not independent of medical encounters and perceived health. In addition, alcohol and smoking are well established health risk factors which may need to be controlled when comparing veterans with the Australian population.

Table 2 shows the effects on selected chronic conditions of adjustment for each of the factors discussed: number of doctor consultations, perceived general health, smoking and alcohol. The table shows the ratio of the adjusted relative risk to the unadjusted relative risk for each condition. Asterisks indicate that the 99% confidence interval for the adjusted relative risk does not include 1.00. Adjustment for doctor consultations and perceived health state generally reduced the relative risks. For doctor consultations 33 of 36 recent and 29 of 37 chronic relative risks decreased, with an average of 8.6% for recent and 7.4% for chronic risks. Adjustment for doctor consultation reduced previously significant relative risks for recent hypercholesterolaemia, recent ulcer, recent injury, chronic depression, and chronic hypertension. Apart from these five instances, for the remainder there would be no change to the significance decision after adjustment for doctor consultation.

Adjustment for perceived health state reduced 32 of 36 recent and 34 of 37 chronic

Table 2. Ratio of adjusted to unadjusted relative risks for selected chronic illness conditions reported by Australian Vietnam veterans compared with the age-sex standardised Australian population, with adjustment for nonresponse, recent doctor consultation, perceived health status, smoking and alcohol risk

	Doctor Visits	Perceived Health	Smoking	Alcohol
Infective and parasitic disease	0.99*	0.85*	0.81*	1.01*
Endocrine system				
– Cholesterol	0.92*	0.89*	0.90*	1.03*
– Diabetes	1.00*	0.78	0.81	1.00*
– Gout	0.95*	0.92*	0.90*	0.94*
Mental disorders				
– Nerves, nervousness	0.96*	0.73*	0.80*	0.99*
– Depression	0.89	0.69	0.83	0.98*
– Other mental illness	0.92*	0.73*	0.82*	0.82*
Sensory/Nervous system disorders				
– Disorders of refraction	1.00	0.73	0.82	1.00
Circulatory system				
– Hypertension	0.60	0.84*	0.79*	0.98*
Respiratory system				
– Hayfever	0.99*	0.93*	0.79	0.73
– Asthma	0.90	0.93	0.76	1.02
Digestive system				
– Ulcer	0.94*	0.85*	0.79*	0.96*
– Hernia	0.97*	0.93*	0.81*	0.98*
Genito-urinary system	0.99*	0.83*	0.79	1.02*
Musculoskeletal disorders				
– Arthritis	0.97*	0.84	0.81	1.01*
– Rheumatism	0.99*	0.87	0.83	1.00*
Injury	0.97*	0.98*	0.82*	0.95*
Symptoms, signs and ill defined conditions	0.96*	0.86*	0.80*	1.02*

*99% confidence interval for adjusted relative risk did not include 1.0.

relative risks, with a mean reduction of 17.6% for recent and 8.3% for chronic conditions. Adjustment for perceived health abolished the significance of recent nerves, depression, other mental illness, ulcer, arthritis and injury, and chronic diabetes, arthritis and rheumatism. Adjustment for smoking and alcohol acted differently for recent and chronic conditions. Adjustment for smoking increased 28 of 36 recent relative risks, for a mean increase of 8.6%, but it reduced all chronic risks, averaging 19.2% reduction overall. Adjustment for smoking status instated two previously nonsignificant recent relative risks (for recent gout and other digestive disease) but abolished the significance of 6 chronic conditions risks (for diabetes, depression, hayfever, genitourinary, arthritis and rheumatism). Adjustment for alcohol risk reduced 15 recent relative risks, with a mean increase of 0.2%, and a decrease in 20 chronic relative risks, for a mean decrease of only 1.3%. Adjustment for alcohol changed only one decision, abolishing a previously significant relative risk for chronic hayfever.

To test the possibility that the excess veteran risks are at least partially attributable to the

Table 3.   Relative prevalence of chronic medical conditions reported by Australian Vietnam veterans compared to the age-sex standardised Australian population jointly adjusted for age, nonresponse, number of recent doctor consultations, perceived health state and alcohol and tobacco use

	RR	(99% CI)
Infectious and parasitic disease	4.15	(2.57, 5.74)
Endocrine system		
– Cholesterol	2.53	(1.98, 3.14)
– Diabetes	2.06	(1.19, 2.93)
– Gout	2.83	(2.10, 3.55)
– Other endocrine disorders	2.48	(1.70, 3.26)
Mental disorders		
– Nerves, nervousness	8.45	(6.56, 10.35)
– Depression	2.72	(0.87, 4.57)[a]
– Other mental illness	3.30	(2.21, 4.40)
Sensory/Nervous system disorders		
– Disorders of refraction	1.12	(1.02, 1.22)[b]
Circulatory system		
– Hypertension	1.83	(1.52, 2.15)
Respiratory system		
– Hayfever	1.67	(1.39, 1.96)
– Asthma	0.82	(0.49, 1.15)
Digestive system		
– Ulcer	2.29	(1.57, 3.02)
– Hernia	2.85	(2.04, 3.66)
Genito-urinary system	1.78	(0.84, 2.72)[a]
Musculoskeletal disorders		
– Arthritis	1.31	(1.05, 1.57)
– Rheumatism	1.98	(0.72, 3.24)[a]
Injury	4.53	(3.17, 5.88)
Symptoms, signs and ill defined conditions	2.57	(1.89, 3.25)

[a]Previously significant became nonsignificant.
[b]Previously nonsignificant became significant.

combination of factors that were identified as potential confounders, joint adjustment for all the factors together was computed; that is, separate relative risks for recent and chronic illness groups were computed adjusted jointly for nonresponse, recent doctor consultation, perceived health, and alcohol and tobacco use. The results for selected chronic illness are shown in Table 3. The joint adjustment decreased 24 recent risks for a mean reduction of 6.5% and decreased 30 chronic risks for a mean reduction of 10.3%. After joint adjustment the previously nonsignificant relative risks for recent and chronic disorders of refraction became significant, as did recent injury, and the previously significant risks for chronic depression and chronic genitourinary problems became nonsignificant.

4.   Discussion

The judgment of the credibility of results when comparing statistics estimated from independent surveys depends upon differences in the surveys in terms of subjects,

measurement, and confounding. Subject-related biases arise from coverage errors, sampling errors, and nonresponse. Each of the studies (ABS and AVVHS) had different selection schemes (area probability sampling versus simple random sampling from a frame) and, while coverage using area probability sampling is thought to be very high, the veterans were selected as a simple random sample from a file derived from military records of all known postings to Vietnam, and is thus assumed complete. Response rates were different between the AVVHS and the ABS survey. Nonresponse adjustment was unavailable in the ABS data but regression adjustment for nonresponse in the veteran data changed only a few measures and the significance ($p < 0.01$) of only a few relative risks changed; in almost all cases where the unadjusted 99% confidence interval excluded 1.00 the adjusted confidence interval did so too, and where the unadjusted interval did include 1.00 the unadjusted interval did likewise. Thus, to the extent that the army variables available were adequate predictors of response and morbidity, nonresponse may not have significantly biased the morbidity estimates in the AVVHS, in spite of the lower response rate compared with the ABS survey.

4.1. Measurement

The use of unstandardised or idiosyncratic general health questionnaire instruments has plagued research into the effects of war service on Vietnam veterans (King and King 1990). The AVVHS was designed to overcome this problem, by using the ABS questionnaire, interview procedures, visual aids-show cards, interview and office manuals, coding materials and computerised data. However, interviewer characteristics and training were undoubtedly different although the effect of clinical interviewers compared with lay interviewers, particularly as they may interact with veteran status, seems not to have perturbed the standardisation of the ABS interview schedule.

Two factors that were argued to affect reporting of illness were the number of recent doctor visits and the overall perceived health of the respondent. Adjustment for the number of doctor visits generally reduced the relative risks, although only of the order of 7–8%, but abolished the significance of five of the 48 significant recent and chronic relative risks. Adjustment for perceived health also generally reduced the relative risks 8–18%, and abolished the significance of nine of the 48 significant relative risks. However, in the remainder, Vietnam service was still associated with an excess of reports of many illness conditions.

4.2. Confounding

Confounders that were identified and subject to manipulation were age, smoking and alcohol. Adjustment for alcohol had very little effect on the relative risks, but adjustment for smoking had opposite effects on recent and chronic conditions, increasing the recent relative risks and decreasing the chronic risks. This confirms that potential confounders may operate in both directions, to increase as well as to decrease the relative risk of illness. Adjustment, however, did not generally render the relative risks statistically nonsignificant, indicating that, even after controlling for these potential confounders, Vietnam service seems still to be associated with increased illness reports.

4.3. Interpretation of variables

Whether a potential confounding variable is a real confounding variable or whether a potential bias is a real bias depends upon nonstatistical theory, on what is known about the causal path from exposure to outcome. It is arguable that none of the variables discussed above constitutes a biasing variable in the relation between veteran status and poorer health outcome, except possibly for age. This is because several characteristics of veterans, such as smoking and alcohol status, may be strongly associated with their military experience; for example, very many young men learned to drink and smoke in the army, and they received free cigarettes in their rations and cheap alcohol while in Vietnam. Thus, increased alcohol and tobacco risk may be intrinsically part of the Vietnam experience, and be considered a feature of the ''exposure.'' Therefore, adjusting for smoking and alcohol may in part be adjusting for a variable on the causal pathway between veteran exposure and health outcome, and constitute overadjustment (Day, Byar, and Green 1980).

In addition, it could be the case that veterans indeed have poorer health status than the population, requiring increased health care and accompanied by poorer health outlook. Thus, adjusting for doctor visits and perceived health may also be adjusting for variables on the causal path between exposure and health outcome. For example, in certain of the psychological disorders (nerves/nervousness, depression, other disorders) poor perceived health may be a symptom of the disorder and not a confounder. If this view is persuasive, then the adjusted relative risks are biased; if the view is taken they are legitimate confounders, then the unadjusted relative risks are biased. Statistical theory here is of no assistance, leaving one with the problem of what to make of a statistically significantly increased risk of reports of illness in the case of veterans as compared with the Australian population.

5. References

Australian Bureau of Statistics (1991). Health Interview Survey 1989–90. Canberra: Australian Bureau of Statistics.

Boivin, J.-F. and Wacholder, S. (1985). Conditions for Confounding of the Risk Ratio and of the Odds Ratio. American Journal of Epidemiology, 121, 152–158.

Cannell, C.F., Marquis, K.H., and Laurent, A. (1977). A Summary of Studies of Interviewing Methodology. Vital and Health Statistics, PHS Publication No 1000, Series 2, No 69. Washington DC: U.S. Government Printing Office.

Centers for Disease Control (1988a). Vietnam Experience Study. Health Status of Vietnam Veterans. II. Telephone Interview. Atlanta, GA: Centers for Disease Control.

Centers for Disease Control (1988b). Vietnam Experience Study. Health Status of Vietnam Veterans. III. Medical Examination. Atlanta, GA: Centers for Disease Control.

Centers for Disease Control (1988c). Vietnam Experience Study: Health Status of Vietnam Veterans. I. Psychosocial Characteristics. Journal of the American Medical Association, 259, 2701–2707.

Centers for Disease Control (1988d). Vietnam Experience Study: Health Status of Vietnam Veterans. II. Physical health. Journal of the American Medical Association, 259, 2708–2714.

Chambers, L.W., Spitzer, W.O., and Hill, G.B. (1976). Underreporting of Cancer in Medical Surveys: A Source of Systematic Error in Cancer Research. American Journal of Epidemiology, 104, 141–145.

Clark, V.A., Aneshensel, C.S., Frerichs, R.R., and Morgan, T.M. (1983). Analysis of Non-response in a Prospective Study of Depression in Los Angeles County. International Journal of Epidemiology, 12, 193–198.

Copeland, K.T., Checkoway, H., McMichael, A.J., and Holbrook, R.H. (1977). Bias Due to Misclassification in the Estimation of Relative Risk. American Journal of Epidemiology, 105, 488–495.

Coughlin, S.S. (1990). Recall Bias in Epidemiologic Studies. Journal of Clinical Epidemiology, 43, 87–91.

Dales, L.G. and Ury, H.K. (1979). An Improper Use of Statistical Significance Testing in Studying Covariables. International Journal of Epidemiology, 7, 373–375.

Day, N.E., Byar, D.P., and Green, S.B. (1980). Overadjustment in Case-control Studies. American Journal of Epidemiology, 112, 696–706.

DeClerk, N., English, D., and Armstrong, B.K. (1989). A Review of the Effects of Random Measurement Error on Relative Risk Estimates in Epidemiological Studies. International Journal of Epidemiology, 18, 705–712.

Decoufle, P., Holmgreen, P., Boyle, C., and Stroup, N.E. (1992). Self-reported Health Status of Vietnam Veterans in Relation to Perceived Exposure to Herbicides and Combat. American Journal of Epidemiology, 135, 312–323.

Decoufle, P., Holmgreen, P., Calle, E.E., and Weeks, M.F. (1991). Nonresponse and Intensity of Follow-up in an Epidemiologic Study of Vietnam-era Veterans. American Journal of Epidemiology, 133, 83–95.

Eaton, W.W., Anthony, J.C., Tepper, S., and Dryman, A. (1992). Psychopathology and Attrition in the Epidemiologic Catchment Area Surveys. American Journal of Epidemiology, 135, 1051–1059.

Fleiss, J.L. (1981). Statistical Methods for Rates and Proportions, 2nd ed. New York: Wiley.

Grayson, D.A. (1987). Confounding Confounding. American Journal of Epidemiology, 126, 546–553.

Grayson, D.A., O'Toole, B.I., Marshall, R.P., Schureck, R.J., Dobson, M., Pulvertaft, B., Meldrum, L., and French, M.L. (1996). Interviewer Effects on Epidemiological Diagnoses of Post Traumatic Stress Disorder. American Journal of Epidemiology, 144, 589–597.

Greenberg, R. and Kleinbaum, D.G. (1985). Mathematical Modelling Strategies for the Analysis of Epidemiologic Research. Annual Review of Public Health, 6, 223–245.

Greenland, S. (1982). Interpretation and Estimation of Summary Ratios under Heterogeneity. Statistics of Medicine, 1, 217–227.

Greenland, S. and Finkle, W.D. (1995). A Critical Look at Methods for Handling Missing Covariates in Epidemiologic Regression Analyses. American Journal of Epidemiology, 142, 1255–1264.

Groves, R.M. (1989). Survey Errors and Survey Costs. New York: Wiley.

Hill, A.B. (1965). The Environment and Disease: Association or Causation? Proceedings of the Royal Society of Medicine, 58, 295–300.

Kehoe, R., Wu, S., Leske, M., and Chylack, L.J. (1994). Comparing Self-reported and Physician-reported Medical History. American Journal of Epidemiology, 139, 813–818.

King, D.W. and King, L.A. (1991). Validity Issues in Research on Vietnam Veteran Adjustment. Psychological Bulletin, 1991, 109, 107–124.

Kish, L. (1987). Statistical Design for Research. New York: John Wiley.

Kleinbaum, D.L., Kupper, L.L., and Morgenstern, H. (1982). Selection Bias in Epidemiological Studies. American Journal of Epidemiology, 113, 452–463.

Kleinbaum, D.L., Kupper, L.L., and Morgenstern, H. (1981). Epidemiologic Research: Principles and Quantitative Methods. Belmont, California: Lifetime Learning Publications.

Kulka, R.A., Schlenger, W.E., Fairbank, J.A., Hough, R.L., Jordan, B.K., Marmar, C.R., and Weiss, D.S. (1990a). Trauma and the Vietnam Generation. Report of Findings from the National Vietnam Veterans Readjustment Study. Volume I. New York: Brunner/Mazel.

Kulka, R.A., Schlenger, W.E., Fairbank, J.A., Hough, R.L., Jordan, B.K., Marmar, C.R., and Weiss, D.S. (1990b). The National Vietnam Veterans Readjustment Study, Vol II, Tables of Findings and Technical Appendices. New York: Brunner/Mazel.

Last, J.A. (1988). Dictionary of Epidemiology. 2nd Edition. London: Oxford University Press.

Liddell, F.D.K. (1988). The Development of Cohort Studies in Epidemiology: A Review. Journal of Clinical Epidemiology, 41, 1217–1237.

Little, R.A. and Rubin, D.B. (1987). Statistical Analysis with Missing Data. New York: John Wiley.

Meltzer, J.W. and Hochstim, J.R. (1970). Reliability and Validity of Survey Data on Physical Health. Public Health Reports, 85, 1075–1086.

Miettinen, O.S. (1985). Theoretical Epidemiology. New York: Wiley.

Neugebauer, N. and Ng, S. (1990). Differential Recall as a Source of Bias in Epidemiologic Research. Journal of Clinical Epidemiology, 43, 1337–1341.

O'Toole, B.I. (1990). Intelligence and Motor Vehicle Mortality. Accident Analysis and Prevention, 22, 211–221.

O'Toole, B.I., Adena, M.A., and Jones, M.P. (1988). Risk Factors for Mortality in Australian Vietnam-era National Servicemen: A Case-control Study. Community Health Studies, XII, 408–417.

O'Toole, B.I., Battistutta, D., Long, A.A., and Crouch, K. (1986). A Comparison of Costs and Data Quality of Three Health Survey Methods: Mail, Telephone and Personal Home Interview. American Journal of Epidemiology, 124, 317–328.

O'Toole, B.I. and Cantor, C. (1995). Risk Factors for Suicide in Vietnam-era Australian National Servicemen. Suicide and Life Threatening Behavior, 24, 474–487.

O'Toole, B.I., Marshall, R.P., Grayson, D.A., Schureck, R.J., Dobson, M., Ffrench, M.L., Pulvertaft, B., Meldrum, L., Bolton, J., and Vennard, J. (1996a). The Australian Vietnam Veterans Health Study, I. Study Design and Response Bias. International Journal of Epidemiology, 25, 307–318.

O'Toole, B.I., Marshall, R.P., Grayson, D.A., Schureck, R.J., Dobson, M., Ffrench, M.L., Pulvertaft, B., Meldrum, L., Bolton, J., and Vennard, J. (1996b). The Australian

   Vietnam Veterans Health Study. II. Self-reported Health of Veterans Compared to the Australian Population. International Journal of Epidemiology, 25, 319–330.

O'Toole, B.I., Marshall, R.P., Grayson, D.A., Schureck, R.J., Dobson, M., Ffrench, M.L., Pulvertaft, B., Meldrum, L., Bolton, J., and Vennard, J. (1996c). The Australian Vietnam Veterans Health Study. III. Psychosocial Health Status and Relation to Combat. International Journal of Epidemiology, 25, 331–340.

Rubin, D.B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: John Wiley.

Solomon, Z. and Mikulincer, M. (1987). Combat Stress Reactions, Post Traumatic Stress Disorder and Somatic Complaints Among Israeli Soldiers. Journal of Psychosomatic Research, 31, 131–137.