ORIGINAL REPORT

Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics[†]

Sebastian Schneeweiss MD, ScD*

Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

SUMMARY

Background   Large health care utilization databases are frequently used to analyze unintended effects of prescription drugs and biologics. Confounders that require detailed information on clinical parameters, lifestyle, or over-the-counter medications are often not measured in such datasets, causing residual confounding bias.
Objective   This paper provides a systematic approach to sensitivity analyses to investigate the impact of residual confounding in pharmacoepidemiologic studies that use health care utilization databases.
Methods   Four basic approaches to sensitivity analysis were identified: (1) sensitivity analyses based on an array of informed assumptions; (2) analyses to identify the strength of residual confounding that would be necessary to explain an observed drug-outcome association; (3) external adjustment of a drug-outcome association given additional information on single binary confounders from survey data using algebraic solutions; (4) external adjustment considering the joint distribution of multiple confounders of any distribution from external sources of information using propensity score calibration.
Conclusion   Sensitivity analyses and external adjustments can improve our understanding of the effects of drugs and biologics in epidemiologic database studies. With the availability of easy-to-apply techniques, sensitivity analyses should be used more frequently, substituting qualitative discussions of residual confounding. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS — confounding; bias; sensitivity analysis; claims databases; epidemiologic methods; pharmacoepidemiology

BACKGROUND

Results from pharmacoepidemiologic research often have immediate and far-reaching clinical, regulatory, and economic implications. Like others in similar non-experimental research, pharmacoepidemiologists need to carefully evaluate the causality of an association between a prescription drug and a health outcome. Although a variety of systematic errors may bias non-experimental research,[1] confounding bias is of particular concern in epidemiologic studies of drug effects.[2]

Large health care utilization data sets are often the best sources of data to analyze the relation between prescription drugs or biologic use and unintended and infrequent health events. A major advantage of health care utilization data is that they reflect routine practice for large and representative populations, in contrast to the much smaller and often healthier patient

* Correspondence to: Dr Sebastian Schneeweiss, Department of Medicine, Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital and Harvard Medical School, 1620 Tremont Street, Suite 3030, Boston, MA 02120, USA. E-mail: schneeweiss@post.harvard.edu
[†]No conflict of interest was declared.

Table 1. Clinical, behavioral, and socioeconomic factors often not measured in pharmacoepidemiologic database studies and that may cause residual confounding

Potential confounders often unmeasured in pharmacoepidemiologic database studies	Examples of drug—disease outcome associations possibly affected by residual confounding in epidemiologic database studies			
	Anti-TNFα therapy and lymphoma in patients with rheumatoid arthritis	Statins and fractures	Cox-2 inhibitors and myocardial infarction	NSAIDs and short-term mortality
Body mass index		X	X	X
Over-the-counter aspirin use,			X	
Smoking		X	X	X
Frailty		X	X	X
Functional impairment		X		
Cognitive impairment		X		
Educational attainment		X	X	X
Income status		X	X	X
Laboratory values, for example, EBV antibody titer, lipid level, CRP level	X		X	
Results of invasive and non-invasive exams, for example, bone mineral density measure (DXA), ECG,		X	X	
Disease-specific severity markers	X			

populations in clinical trials.[3] They are large enough to assess the frequency and etiology of rare drug effects and avoid the delays common in the collection of primary data.

Despite their importance, pharmacoepidemiologic claims data studies have been criticized for the incompleteness of their information on potential confounders such as the use of over-the-counter medications (e.g., aspirin in studies of NSAIDs), markers of clinical disease severity, body mass index, smoking status, functional status, laboratory values, among others (Table 1). Such factors may lead to selective prescribing of drugs, which may result in biased estimates of the association between drugs and health outcomes.[4] All too often research studies discuss the potential for residual confounding only qualitatively without any quantitative assessment of the magnitude of such bias. Sensitivity analyses were described as 'the last line of defense against biases after every effort has been made to eliminate, reduce, or control them in study design, data collection, and data analysis.'[5] The basic concept of sensitivity analyses is to make informed assumptions about potential residual confounding and quantify its effect on the relative risk estimate of the drug-outcome association. If suitable data sources can be identified, these assumptions can be substituted by empirical estimates and then be used for external adjustment of the Drug-Disease Outcome Association. Figure 1 shows how sensitivity analyses and external adjustment fit into the methods tool kit of pharmacoepidemiologists to better understand and possibly control confounding.

Existing sensitivity analyses include the production of a grid of estimates as a function of several assumptions with limited knowledge of the true parameter constellation. Several epidemiologic studies on occupational safety using employment records with limited information on workers' health status used this approach.[6,7] Recent studies have explored how strong unmeasured confounding must be to explain the elevated relative risks observed in studies of drug effects using health care utilization databases.[8–11] If additional information is available through surveys, external adjustment can be attempted with increasing methodological complexity.

Among the several recent examples of observational database studies on drug effects that struggle with the potential for residual confounding bias are those of the associations between newer sedative hypnotics and hip fractures, statin use and cancer, selective COX-2 inhibitors and cardiovascular events, and anti-TNFα therapy and lymphatic malignancies.

This paper demonstrates a framework of techniques for sensitivity analysis and external adjustment for residual confounding using several examples. First, it explains simple sensitivity analyses in the absence of external information. Next, it demonstrates the use of external information for external adjustment of effect estimates for single binary covariates and follows with an examination of techniques for externally adjusting
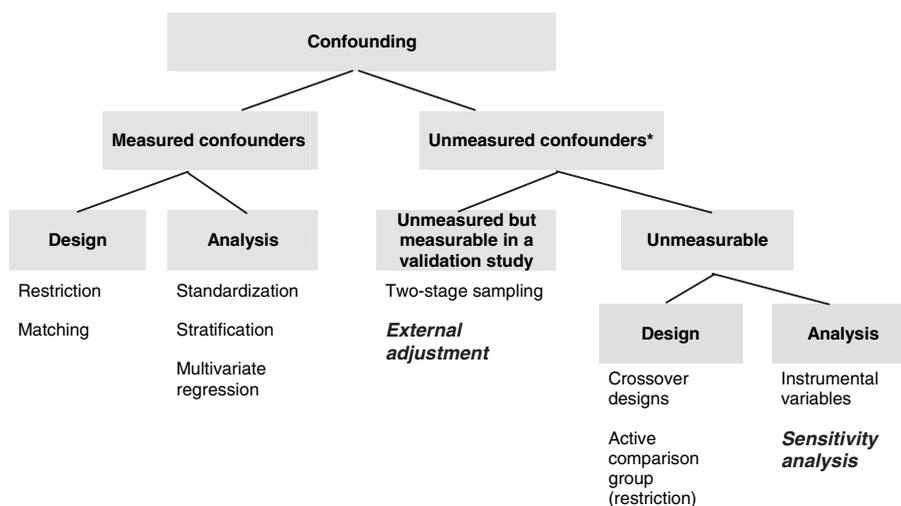
* These strategies generally also adjust for measured confounders but come with additional assumptions or restrictions to generalizability

Adapted from Schneeweiss S: Confounding. In: Hartzema A, Chan A, Porta M, Tilson H: Pharmacoepidemiology, 4th edition in press.

Figure 1. Strategies to control for unmeasured confounders in pharmacoepidemiology

multiple confounders of various distributions. For most analyses, spread-sheets with example data can be downloaded at http://www.drugepi.org.

Residual confounding and the basics of sensitivity analyses

Physicians prescribe drugs in light of diagnostic and prognostic information available at the time of prescribing. The factors influencing this decision vary by physician and over time[12] and frequently involve clinical, functional, or behavioral characteristics of patients. If these factors are also independent predictors of the study outcome, failing to control for such factors can lead to confounding bias. The confounding, thus, results from an informed selection or channeling of patients into drug-exposure groups based on indications and contraindications,[13] and is, therefore, widely referred to as confounding by indication.[4]

A typical example would be the prescribing of non-steroidal anti-inflammatory drugs (NSAIDs) for pain and their effect on gastrointestinal (GI) hemorrhage. Non-selective NSAIDs are known for their potential to cause gastric and duodenal ulcer, erosive gastritis, and GI hemorrhage. Physicians who act rationally and follow treatment guidelines will prescribe COX-2-selective NSAIDs to patients with a history of GI irritation or hemorrhage;[14] this subgroup of NSAIDs

has demonstrated reduced gastric side effects in randomized clinical trials (RCT).[15,16] Because these patients are at higher risk for the development of a GI hemorrhage independent of drug use, epidemiologic studies may show an apparent association between selective COX-2 inhibitor use and GI bleeding.[17] A related example of confounding bias in studies of the intended effect of drugs using observational data is one on the efficacy of gastroprotective drugs among NSAID users. The study found an apparent 10-fold increase in risk of gastric bleeding or perforation among users of gastroprotective drugs that is likely due entirely to confounding.[18]

Causal graphs (Figure 2) are helpful for illustrating confounding.[19] A factor can be a confounder only if that factor is associated with drug exposure ($OR_{EC} \neq 1$, see notation in Table 2) and is also an independent risk factor of the disease outcome ($RR_{CD} \neq 1$). Note that the association OR_{EC} can be causal or incidental. If either association is non-existent, there is no confounding. Factors that are not independent predictors of the study outcome ($RR_{CD} = 1$) cannot be confounders even if they are not balanced among drug exposure groups. Likewise, if a risk factor is not associated with the drug exposure—for example, through random assignment of drug exposure—this factor is not a confounder.

Ideally, we would be able to fully assess the history of GI irritation and group patients into strata of similar baseline risk of GI hemorrhage. Comparing users of

(a) Measured confounding
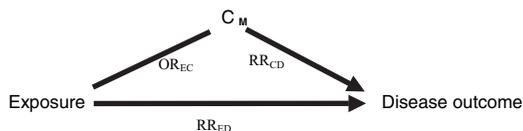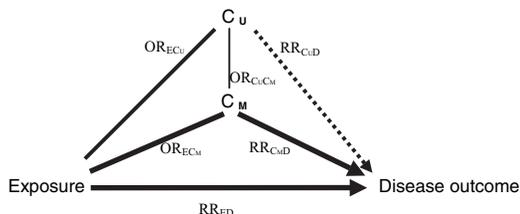

(b) Measured and unmeasured confounding

Figure 2. A graphical concept of measured and unmeasured confounding.

selective COX-2 inhibitors with users of non-selective NSAIDs within these strata would then find, similar to the findings of randomized trials, a reduced risk of GI hemorrhage in users of selective COX-2 inhibitors.[14,20] However, often physicians consider subtle risk factors for GI hemorrhage that are not recorded and are, therefore, unmeasured confounders (C_U in Figure 2b) when they prescribe selective COX-2 inhibitors. This is less likely to happen in studies with primary data collection than in studies using medical records or claims data, in which the choice of covariates is limited by the data source. Most non-experimental studies using claims data with limited patient information to compare selective with non-selective NSAIDs will not be able to fully measure and adjust such confounders and will therefore be unable to show a gastroprotective effect of COX-2 inhibitors due to residual confounding.

Table 2. Notation

RR	'True' or fully adjusted exposure relative risk
ARR	Apparent exposure relative risk
RR_{CD}	Association between confounder and disease outcome
P_C	Prevalence of confounder
P_{C1}	Prevalence of confounder in the exposed
P_{C0}	Prevalence of confounder in the unexposed
P_E	Prevalence of drug exposure
OR_{EC}	Association between drug use category and confounder
RD	'True' or fully adjusted exposure risk difference
ARD	Apparent exposure risk difference
C_M	Measured confounder
C_U	Unmeasured confounder

Array approach

The confounded relative risk (RR), which we call apparent RR (ARR), can be expressed as the 'true' or fully adjusted RR times Bias ($ARR = RR \times Bias_M$), which is an expression of the imbalance of a binary confounding factor among exposed (P_{C1}) and unexposed (P_{C0}, using the notation in Table 2):[21]

$$ARR = RR \times \frac{P_{C1}(RR_{CD} - 1) + 1}{P_{C0}(RR_{CD} - 1) + 1} \qquad (1a)$$

Similar to this multiplicative model of confounding, an additive model for risk differences (RD) can be derived: [11]

$$ARD = RD + (P_{C1} - P_{C0})RD_{CD} \qquad (1b)$$

In basic sensitivity analyses on residual confounding, we try to understand how the strength of an unmeasured confounder and imbalance among drug exposure categories affects the observed or apparent RR. By solving Equation (1a) for RR

$$RR = \frac{ARR}{\left[\frac{P_{C1}(RR_{CD}-1)+1}{P_{C0}(RR_{CD}-1)+1}\right]} \qquad (2)$$

and plugging in a range of values for P_{C1}, P_{C0}, and RR_{CD} for a given ARR, one can calculate the 'true' or more fully adjusted value of RR under these circumstances. The difference between ARR and RR is the absolute bias on the relative risk scale, while bias is also expressed as the ratio of ARR/RR[7] or the proportion of bias of the true RR: percent bias $= [(ARR - RR)/(RR - 1)] \times 100$. [22] In case of an ARR of 2.0 and a true RR of 1.5, the percent of the effect negated by bias would be 100% accordingly. The advantage of this metric is that it recognizes that the null value of a relative risk measure is 1.0. Percent bias was also proposed as $[(ARR - RR)/RR] \times 100$ which can be rewritten as $[(ARR/RR) - 1] \times 100$. In case of an ARR of 2.0 and a true RR of 1.5, the percent bias would then be 33% because the metric is assuming 0 as the null value and not 1. Neither way of quantifying the amount of confounding in a single number has any clear advantage and inevitably information is lost when two numbers are combined into one. Most basic sensitivity analyses consider only the value of the point estimate and not the precision of its estimation. More complex sensitivity analyses based on re-sampling can quantify random error.[23]

Following our example of COX-2 inhibitor use and the risk of GI hemorrhage, we wish to examine the role of residual confounding in a hypothetical observational study that found a relative risk of ARR = 1.5. The prevalence of a history of GI bleeding or peptic ulcer disease in the control group of users of non-selective NSAIDs is $P_{C0} = 0.1$. These two factors will not be altered in the following sensitivity analysis. In Figure 3, two factors were varied: the strength of the confounder-disease association (1.0 to 5.5) and the prevalence of the confounder in the coxib group (0.0 to 0.5). With an increasing imbalance of the confounder, that is, patients receiving coxibs were more likely to have a history of GI problems, the 'true' or fully adjusted RR moves closer to the results observed in randomized clinical trials of about 0.7. The back wall of Figure 3 represents the situation when $RR_{CD} = 1$, which means that the potential confounder is assumed to not be associated with the outcome. This scenario would result in unbiased estimates, as would that of

perfectly balanced distributions of the confounder among drug groups, that is, $P_{C0} = P_{C1} = 0.1$. This is an example of fairly strong confounding, as can be expected when one studies anticipated drug effects that is known to prescribers.

In realistic settings, ARR is already adjusted for a set of measured covariates and the interest is in assessing the residual confounding by additional covariates not measured in the main study (Figure 2b). To the extent that measured (and adjusted) confounders are correlated with unmeasured characteristics, residual confounding caused by the unmeasured factors will be reduced or partially adjusted. People who take vitamin supplements are also more likely to have a healthy lifestyle with more physical activity and careful selection of nutrients. While it is easy to measure intake of vitamins, the construct of health-seeking lifestyle is much harder to quantify. Most techniques for basic sensitivity analysis assume that measured and unmeasured covariates are independent given

ARR = 1.5
$P_{C0} = 0.10$

Unbiased result if $P_{C0} = P_{C1}$

RR (fully adjusted)

RR_{CD}

P_{C1}
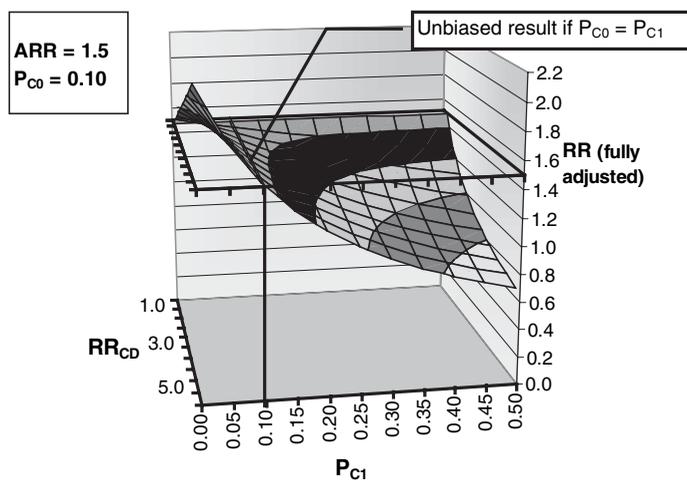
* The curved surface represents the "true" or fully adjusted RR assuming an apparent relative risk of ARR=1.5 and a prevalence of an unmeasured confounder of $P_{C0} = 0.1$ in the unexposed group. The prevalence of the unmeasured confounder in the exposed group (P_{C1}) is varied between 0.0 to 0.5 on the x-axis. The strength of the confounder-disease association (RR_{CD}) is varied between 1.0 and 5.5 on the z-axis. With an increasing imbalance of the confounder (on the x-axis) the difference between the "true" or fully adjusted RR and the apparent RR increases. The back wall represents the scenario of $RR_{CD} = 1$, i.e. the potential confounder is assumed to be not associated with the outcome resulting in no bias. The vertical line represents perfect balance of the potential confounder among drug exposure groups, i.e. $P_{C0} = P_{C1} = 0.1$, which results in an unbiased ARR.

Figure 3. Sensitivity analysis of residual confounding: array approach

exposure ($OR_{CuCM} = 1$ in Figure 2b)[24] and may, therefore, overestimate the amount of residual confounding in practical settings. The Appendix gives an

$$P_{C1} = \frac{P_E - P_E^2 - (P_E P_C RR_{CD} ARR) + (P_E P_C ARR) + (P_E^2 ARR) - (P_E ARR)}{(P_E ARR) - (P_E RR_{CD} ARR) - RR_{CD} + 1 + (P_E RR_{CD}) - P_E} \tag{5}$$

additional example on the association between TNFα-blocking agents and the risk of lymphatic malignancies.

Rule-out approach

The array approach is helpful for exploring the effect of residual confounding over a wide range of parameter constellations. However, researchers may want to tailor a sensitivity analysis to their specific study findings and assess the extent of confounding necessary to fully explain the observed findings, that is, the observed point estimate would move to the null (ARR = 1). The hope is that a number of unmeasured possible confounders can then be ruled out because they cannot possibly be strong enough confounders to explain the observed association. This approach was also called target-adjustment sensitivity analysis.[25]

To answer this question, we wanted to find all combinations of OR_{EC} and RR_{CD} (the left and right sides of the confounding triangle in Figure 2a) necessary to move the observed point estimate of RR to 1. More formally, we wanted to plot the relationship between OR_{EC} and RR_{CD} for a given ARR, RR_{CD}, P_C, and P_E.

Assuming a two-by-two table of a dichotomous exposure and a dichotomous confounder, the association between the confounder and exposure can then be measured by the confounder-exposure odds ratio or OR_{EC}, which is a function of the prevalence of the confounder among exposed (P_{C1}) and the marginal probabilities of exposure P_E and confounder P_C:

$$OR_{EC} = \frac{P_{C1}[1 - P_C - P_E + P_{C1}]}{[P_C - P_{C1}][P_E - P_{C1}]} \tag{3}$$

Assuming no underlying true exposure-disease association or $RR_{ED} = 1$, Walker[26] showed that the apparent relative risk (ARR) is a function of P_{C1}, the marginal probabilities P_E and P_C, and the confounder-disease association RR_{CD}:

$$ARR = \frac{P_{C1}[RR_{CD} - 1] + P_E}{[P_C - P_{C1}][RR_{CD} - 1] - P_E + 1} \frac{1 - P_E}{P_E} \tag{4}$$

If the primary interest is to explore the relationship between OR_{EC} and RR_{CD} for a given ARR, RR_{CD}, P_C, and P_E, we need to solve Equation (4) for P_{C1}:

and substitute the derived term for P_{C1} in Equation (3).

An example is provided by Psaty et al.[10] The authors found an increased risk (ARR = 1.57) of acute myocardial infarction (MI) in hypertensive patients who were using calcium channel blockers as compared with those using beta-blockers.[27] Their findings were criticized for being caused by residual confounding by unmeasured factors channeling the prescribing of calcium channel blockers to patients at higher risk of MI.[28] His group produced a graph (reproduced in Figure 4a) that demonstrated that very strong risk factors of cardiovascular events must be unmeasured and uncontrolled to explain the observed association. In Figure 4a, all parameter combinations of OR_{EC} and RR_{CD} above and to the right of the curve representing the ARR = 1.57 line would move the point estimate of the association to 1. It becomes clear that strong risk factors that are fairly imbalanced among exposure groups must be unmeasured and uncontrolled. Psaty et al. noted that most known, strong, independent risk factors of MI, including diabetes, coronary heart disease (CHD), or smoking, were already adjusted and that any unmeasured confounder of the required strength would also have to be independent of the adjusted confounders, that is, correlated confounders such as partial occlusions of coronary arteries are to some extent adjusted by factors like preexisting CHD. They repeated their sensitivity analysis for the value of the observed lower 95% confidence limit (ARR = 1.3) to determine the constellations in which the 95% confidence interval would cross the null. For this sensitivity analysis the prevalence of the unmeasured confounder was fixed at $P_C = 0.2$. Figure 4b shows the same sensitivity analysis but now allows P_C to vary from 0.1 to 0.5. This type of sensitivity analysis is insightful but under-utilized, although it is easy to perform using an Excel spread sheet.

Despite the conclusion of Psaty et al. that there is substantial evidence of a relation between dehydropyridine calcium channel blocker use and the risk of acute MI, randomized trials have now shown that there is no increased risk of acute MI.[29,30] This example is a reminder that sensitivity analyses are based on assumptions—implicit or explicit. An

(a) P_E is constant at 0.01 and P_C is constant at 0.2:*
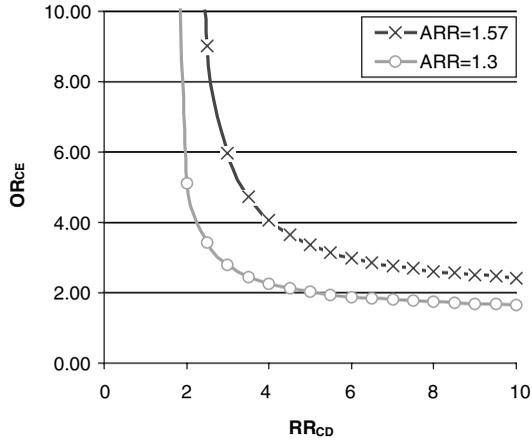
* Each line splits the area into two: the upper right area represents all parameter combinations of OR_{EC} and RR_{CD} that would create confounding by an unmeasured factor strong enough to move the point estimate of the apparent RR (ARR = 1.57) to the null (ARR=1) or even lower, i.e. make the association go away. Conversely, the area to the lower left represents all parameter combinations that would not be able to move the ARR to the null. This example by Psaty et al.[10] assumed a prevalence of the confounder (P_C) of 0.2 and a prevalence of the exposure (P_E) of 0.01.

(b) P_E is constant at 0.01 but P_C can vary from 0.1 to 0.5**
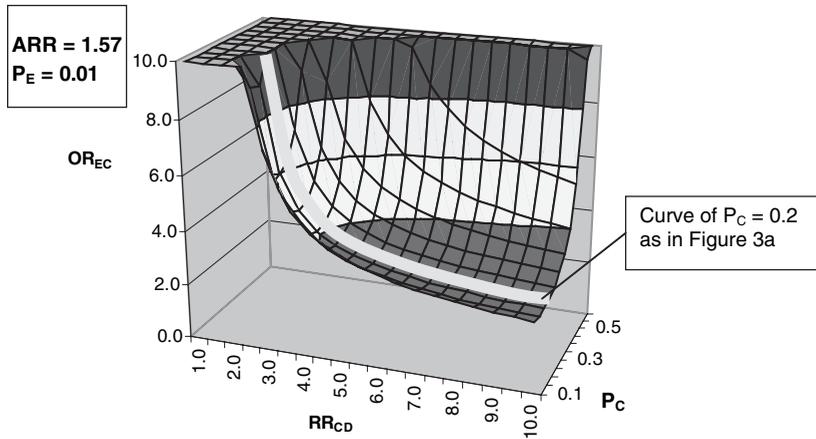
** In addition to Figure 4a, P_C is now varied between 0.1 and 0.5. The corresponding curve of Figure 4a fixed at P_C=0.2 is highlighted.

Figure 4. Sensitivity analysis of residual confounding: Rule-out approach

298 S. SCHNEEWEISS

implicit assumption in Psaty's case was that it would be very unlikely to miss such a strong single confounder. It is conceivable that several weaker confounders may have acted together and explained the apparent RR. Alternatively, other biases may have led to an increased ARR in this case-control study.[31]

Limitations of these methods are that they are constrained to one binary confounder, which is sufficient for illustrative purposes but may be not helpful if several confounders are unmeasured and the joint effect of such confounders is unknown, and that these methods, unlike external adjustment, do not provide an assessment of the magnitude of the existing residual confounding in a specific study, or the choice of reference group.

External adjustment: algebraic solution for single binary confounders

If additional information is available, for example, a detailed survey in a sample of the main database study, such univariate sensitivity analyses can be used to correct for confounders unmeasured in the main study. [32] If internal validation studies are not feasible or are too costly, external data sources can be used under certain assumptions. For example, the Medicare Current Beneficiary Survey (MCBS) studies a representative sample of about 12,000 Medicare beneficiaries to measure a wide variety of characteristics not captured in Medicare claims data, such as limitations in activities of daily living,[33] cognitive impairment, and physical impairments.[34] Since the study outcomes of many pharmacoepidemiologic studies on the safety of drugs are rare moderately sized validation studies can be used to assess the imbalance of confounders among drug exposure groups (OR_{EC}) but can rarely be used to assess the confounder-outcome association (RR_{CuD}). Therefore, researchers extract the independent effects of the individual confounders on the study outcome from the literature. Given estimates of OR_{EC} and RR_{CuD} for each of the unmeasured confounders, an assessment of confounding bias by unmeasured factors and therefore adjustment for these factors can be achieved.[32] Thus the MCBS can be used for external adjustment of unmeasured confounders in a variety of drug studies using Medicare claims data.[35,36]

An example using medicare claims data and MCBS survey information

To illustrate how to correct effect estimates for unmeasured confounding using external information,

we used the association between selective COX-2 inhibitor use and the incidence of MI.[32] Several epidemiologic studies using health care utilization databases reported such an association. In this example, external adjustment was applied to a recent study of Medicare beneficiaries who had complete drug coverage through a pharmacy assistance program that fully covered all selective COX-2 inhibitors and nonselective NSAIDs.[37] In this main study, a number of potential confounders were adjusted using multivariate analyses ($C_{Measured}$ in Figure 2b).

We used the MCBS to estimate the associations between predefined drug exposure categories and selected confounders not measured in Medicare data ($C_{Unmeasured}$ in Figure 2b) and estimates of the confounder-disease associations abstracted from the medical literature. The MCBS is a representative sample of current Medicare beneficiaries living in the community or in institutions.[34] Data are obtained from face-to-face interviews by trained interviewers in the beneficiaries' homes or facilities. In these surveys, the response rate is generally high (between 85% and 95%) and data are very complete.[38,39] Specifically, five patient characteristics not measured in Medicare utilization data that might act as confounders were identified as: body-mass index (BMI), over-the-counter aspirin use, current smoking status, economic status, and educational attainment.

The prevalence of exposure, P_E, the prevalence of potential confounders, P_C, and the association between exposure and confounder, OR_{EC}, were estimated from the MCBS study population. Logistic regression was used to calculate the corresponding age-sex-adjusted OR_{EC}, which was used for all subsequent analyses. For initial bias estimates, we assumed the null hypothesis of no association between selective COX-2 inhibitor exposure and the incidence of MI (RR = 1). Bias estimates were later applied to the effect estimates observed in the main study that may be different from 1. Estimates of the confounder-disease associations that are unmeasured in the main study, RR_{CuD}, were derived from the literature. Literature estimates were derived from large cohort studies or randomized trials after an intensive literature search and expert consultations. If several valid literature estimates were identified, the most extreme (farthest away from the null) was chosen for the base-case analysis, which would lead to more extreme estimates of bias. Alternatively, averaged RR_{CuD} values could be used.

If the primary interest is to estimate ARR as a function of OR_{EC}, RR_{CD}, and the marginal probabilities P_E and P_C, we need to solve Equation (3) for P_{C1}:

Copyright © 2006 John Wiley & Sons, Ltd. *Pharmacoepidemiology and Drug Safety*, 2006; **15**: 291–303

$$\underbrace{P_{C1}^2(OR_{EC} - 1)}_{a}$$
$$+ \underbrace{P_{C1}[-P_C OR_{EC} - P_E OR_{EC} + P_E + P_C - 1]}_{b}$$
$$+ \underbrace{P_C OR_{EC} P_E}_{c} = 0$$

(6)

and P_{C1} can be found as the solution of a quadratic equation of the form

$$P_{C1} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

(7)

which will then be substituted for P_{C1} in Equation (4).

The joint distribution of unmeasured confounders can rarely be assessed with this approach because stable literature estimates are usually not available for several confounder combinations. A practical solution is to sum bias estimates of all confounders weighted by the prevalence of each confounder in the validation sample. Table 3 shows the calculation of such bias estimates. More details on how values for this table were derived are provided in an earlier publication.[32] A maximum range of bias by summing all negative biases to yield a realistic lower-bound estimate and all positive confounders to yield a realistic upper-bound estimate. It is theoretically possible that the effects of individual biases are multiplicative. In such less likely scenarios the additive bounds described here would be falsely narrow.
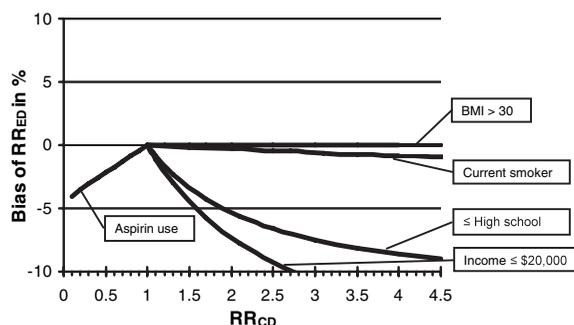

Figure 5. Sensitivity of bias estimates as a function of misspecification of the confounder–disease association (RR_{CD}) from the literature

This external adjustment depends on the correct specification of RR_{CD} from outside sources. Since the literature estimate of RR_{CD} may not be generalizable to a specific study setting, graphic exploration of the sensitivity of the external adjustment procedure to reasonable changes in RR_{CD} is recommended (Figure 5).

Based on this sensitivity analysis suggesting minimal confounding by five unmeasured risk factors for MI we felt more comfortable to publish results from our database main study describing an association between rofecoxib and an increased risk of MI.[37] A year later a large randomized controlled trial of rofecoxib confirmed the elevated risk in a healthier and younger population.[40]

The additional estimation error of OR_{EC} from the validation study can be incorporated in the overall error term. If $RR = ARR \times Bias_M$ and $Var(ARR)$ is the

Table 3. COX2 external adjustment: quantitative assessment of confounding bias in relative risk estimates of the association between the selective COX-2 inhibitor rofecoxib and myocardial infarction compared with naproxen

Potential confounder Data source:	RR_{CD} literature	P_C MCBS	OR_{EC}* MCBS	RR assumed	P_E MCBS	ARR[†]	Percent bias[‡]
Obesity (BMI $\geq$30 kg/m² versus <30 kg/m²)	1.7	0.20	1.00	1.00	0.51	1.000	0.01
Aspirin use (use versus non-use)	0.7	0.10	1.60	1.00	0.51	0.987	−1.28
Smoking (current versus former/never)	3.1	0.07	0.95	1.00	0.51	0.994	−0.61
Educational Attainment ($\leq$high school versus >high school)	2.1	0.70	0.64	1.00	0.51	0.944	−5.61
Income status ($\leq$$20,000 versus >$20,000)	2.1	0.51	0.63	1.00	0.51	0.922	−7.78
Net confounding							
Sum of all negative biases:							−15.3
Weighted average:							−5.10
Sum of all positive biases:							0.01

*Age- and sex-adjusted.
[†]Apparent relative risk between exposure (COX-2 use) and MI outcome if the potential confounder was not controlled, under the assumption that the fully adjusted relative risk RR equals 1.0.
[‡]Bias = [(ARR − RR)/RR] × 100.

variance estimate from the main study and $Var(Bias_M)$ is the variance from the validation substudy, then

$$Var(RR) = Var(ARR \times Bias_M) = ARR^2 \\ \times Var(Bias_M) + Bias_M^2 \times Var(ARR) \quad (8)$$

This easy-to-use approach to assessing the direction and magnitude of unmeasured confounding makes several simplifying assumptions that limits its use. Most importantly, confounder and outcome can only be coded as dichotomous variables, which may over-simplify the relation between some confounders and outcomes. Another important limitation is that this approach cannot consider the joint distribution of unmeasured confounders. Instead, bias estimates are summed over all confounders, weighted by the prevalence of each confounder as a pragmatic approximation of the net bias. Alternatively, the range between the most extreme bias combinations, assuming additive or multiplicative biases, could be considered, providing a most conservative and least informative interpretation of the data. Also, this approach still makes the simplifying assumption that the unmeasured confounders are independent of the measured confounders conditional on exposure status.[24] An association between measured and unmeasured confounders can lead to an overestimation of the magnitude of bias.

Of course the validity of external adjustment also depends on the accuracy of the confounder assessment in the validation substudy and may be limited by the number of important confounders assessed in such studies.

Despite this list of limitations, this approach is valuable for describing or ruling out meaningful confounding that can be assessed in validation substudies but not in claims data main studies.

External adjustment: multiple confounders of various distributions

External adjustment methods were recently extended to a multivariate adjustment for unmeasured confounders that uses a new technique of propensity score calibration (PSC), which can be applied when external information is available that does not contain outcome information.[41] In a validation study for each subject, the full database record is available along with detailed survey information. The goal is to compute within the validation population an error-prone exposure propensity score using only database information, as well as an improved exposure propensity score that also includes survey information for each subject (Table 4). The error component in the database propensity score in the validation study is then quantified and can be used to correct the propensity score in the database main study, using established regression calibration techniques.[42] PSC implicitly takes into account the joint effect of unmeasured confounders that are measured only in the validation study, as well as the relation between measured and unmeasured confounders ($OR_{C_UC_M}$ in Figure 2b). PSC can,

Table 4. Illustration of the data structure used for the PSC approach

Data items	Main study	Validation study	
		Error prone measurement	'Gold standard' measurement
		MCBS survey	
	Medicare Claims Data	MCBS claims	MCBS claims plus survey
Drug exposure	X[‡]	X	X
Demographics	X	X	X
Diagnoses	X	X	X
Procedures	X	X	X
Visits	X	X	X
Smoking			X
Aspirin			X
Body-mass index			X
Functional status			X
Cognitive status			X
Education			X

The shaded headings represent the notation used in the measurement-error literature, while the other headings represent data from our example using Medicare data and the MCBS survey.
[‡] Pharmacy dispensing information linked from pharmacy assistance programs.

therefore, elegantly overcome major limitations of the algebraic approach to external adjustment described above, although it may not perform well in situations were the surrogacy assumption of regression calibration is violated.[42,43]

In contrast to two-stage sampling designs[44–46] or multiple imputation,[47] PSC does not require that outcome information in the validation study performs well, which is an important advantage in drug safety studies of rare adverse effects.[48]

Simulation-based sensitivity analyses

For completeness, simulation-based sensitivity analyses should be mentioned, although they require more technical understanding and programming skills than the above approaches. Such analyses, also called Monte Carlo sensitivity analysis (MCSA), sample bias parameters and then invert the bias model to provide a distribution of bias-corrected estimates. Although MCSA still relies on additional empirical information or structural assumptions, these techniques can provide measures of variation in addition to bias-corrected point estimates. Applications can be found in a number of publications.[23,49–51] Greenland provides a comprehensive overview of such techniques[52] and discusses how they can be expanded beyond confounding bias to sequentially assess the impact of a cascade of systematic errors possible in epidemiology[1] including misclassification and sampling bias.

CONCLUSION

The absence of information on potential confounders is a common criticism of non-experimental studies based on health care utilization data. Easy-to-apply sensitivity analyses and external adjustments using validation study data should be applied more frequently to quantitatively assess confounding bias in pharmacoepidemiologic studies that use claims data. Despite such quantitative approaches to residual confounding the interpretation of what constitutes an analysis that is insensitive to unmeasured confounders remains a matter of judgment and any external adjustment relies on the quality of validation data.

ACKNOWLEDGEMENTS

KEY POINTS

(1) Pharmacoepidemiologic studies using administrative data are often criticized for their limited ability to measure clinically important confounders.

(2) Quantitative sensitivity analyses are easy to perform using spreadsheets (www.drugepi.org) and can much improve our understanding of residual confounding.

(3) If validation study data are available for a subset then external adjustment is possible using simple algebraic formulas.

(4) Propensity Score Calibration is a very promising tool for external adjustment of multiple confounders simultaneously.

(5) Sensitivity analyses should be done more frequently although their results must be interpreted cautiously.

APPENDIX 1

Alternative example of the array approach: an association between biologic anti-TNFα therapy and the risk of lymphatic malignancies (LM)

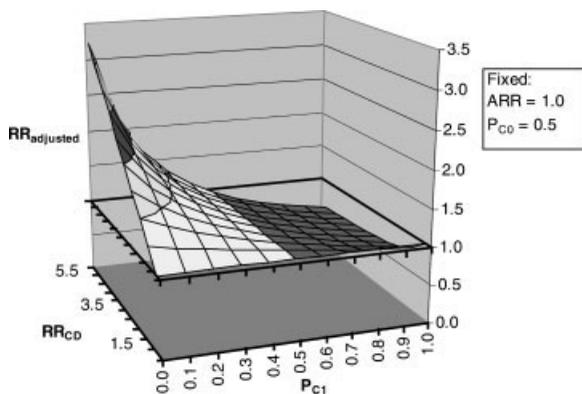1. ARR: Several recent cohort studies showed an increased risk of LM in patients with rheumatoid

Figure A1

arthritis (RA) who used biologics, often in combination with methotrexate (MTX). Wolfe found a threefold increased risk for lymphoma among patients in an RA registry treated with anti-TNFα as compared with SEER cancer registry data from the general public.[53] A Swedish registry study found an 11-fold increased risk for lymphoma in anti-TNFα users as compared with the general public.[54]

2. RR_{CD}: There is some evidence from observational studies that patients with more severe RA are more likely to develop LM. [55]

3. OR_{EC}: It was suspected that patients with more severe RA were more likely to receive biologic anti-TNFα therapy.

In this simplified scenario, we have now established that RA severity may act as a confounder of the biologics—LM association. However, it is not clear which exact components of severity of RA, including acute inflammation, pain, functional impairment, long-term disability, are associated with the decision to prescribe biologics.

Since current evidence on RR_{ED} and OR_{EC} is insufficient, calculating the sensitivity of ARR as a function of RR_{CD}, P_{C0}, and P_{C1} may provide a better understanding of the uncertainties involved in making claims about the safety of biologics in RA patients. Figure A1 demonstrates such a sensitivity analysis, analogous to Figure 3 in the text.

REFERENCES

1. Maclure M, Schneeweiss S. Causation of bias: the episcope. *Epidemiology* 2001; **12**: 114–122.
2. MacMahon S, Collins R. Reliable assessment of the effects of treatment on mortality and major morbidity, II: observational studies. *Lancet* 2001; **357**: 455–462.
3. Schneeweiss S, Avorn J. Using health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005; **58**: 323–337.
4. Walker AM. Confounding by indication. *Epidemiology* 1996; **7**: 335–336.
5. West SL, Strom BL, Poole C. Validity of pharmacoepidemiology drug and diagnosis data. In *Pharmacoepidemiology* (3rd edn), Strom BL (ed.). Wiley: Chichester, 2000; 668.
6. Greenland S, Salvan A, Wegman DH, Hallock MF, Smith TJ. A case-control study of cancer mortality at a transformer-assembly facility. *Int Arch Occup Env Health* 1994; **66**: 49–54.
7. Schlesselman JJ. Assessing effects of confounding variables. *Am J Epidemiol* 1978; **108**: 3–8.
8. Walker AM. Newer oral contraceptives and the risk of venous thromboembolism. *Contraception* 1998; **57**: 169–181.
9. Wang PS, Bohn RL, Glynn RJ, Mogun H, Avorn J. Zolpidem use and hip fractures in older people. *J Am Geriatr Soc* 2001; **49**: 1685–1690.
10. Psaty BM, Koepsell TD, Lin D, *et al.* Assessment and control for confounding by indication in observational studies. *J Am Geriatr Soc* 1999; **47**: 749–754.
11. Gail MH, Wacholder S, Lubin JH. Indirect corrections for confounding under multiplicative and additive risk models. *Am J Ind Med* 1988; **13**: 119–130.
12. Schneeweiss S, Glynn RJ, Avorn J, Solomon DH. A Medicare database review found that physician preferences increasingly outweighed patient characteristics as determinants of first-time prescriptions for cox-2 inhibitors. *J Clin Epidemiol* 2005; **58**: 98–102.
13. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. *Stat Med* 1991; **10**: 577–581.
14. MacDonald TM, Morant SV, Goldstein JL, Burke TA, Pettitt D. Channelling bias and the incidence of gastrointestinal haemorrhage in users of meloxicam, coxibs, and older, non-specific non-steroidal anti-inflammatory drugs. *Gut* 2003; **52**: 1265–1270.
15. Bombardier C, Laine L, Reicin A, *et al.* Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. *N Engl J Med* 2000; **343**: 1520–1528.
16. Silverstein FE, Faich G, Goldstein JL, *et al.* Gastrointestinal toxicity with celecoxib vs nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis: the CLASS study: a randomized controlled trial. Celecoxib Long-term Arthritis Safety Study. *JAMA* 2000; **284**: 1247–1255.
17. Laporte JR, Ibanez L, Vidal X, Vendrell L, Leone R. Upper gastrointestinal bleeding associated with the use of NSAIDs: newer versus older agents. *Drug Saf* 2004; **27**(6): 411–420.
18. McMahon AD. Observation and experiment with the efficacy of drugs: a warning example from a cohort of nonsteroidal anti-inflammatory and ulcer-healing drug users. *Am J Epidemiol* 2001; **154**: 557–562.
19. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999; **10**: 37–48.
20. Wolfe F, Flowers N, Burke TA, Arguelles LM, Pettitt D. Increase in lifetime adverse drug reactions, service utilization, and disease severity among patients who will start COX-2 specific inhibitors: quantitative assessment of channeling bias and confounding by indication in 6689 patients with rheumatoid arthritis and osteoarthritis. *J Rheumatol* 2002; **29**: 1015–1022.
21. Bross IDJ. Spurious effects from an extraneous variable. *J Chronic Dis* 1966; **19**: 637–647.
22. Maldonado G, Greenland S. Simulation study of confounder-selection strategies. *Am J Epidemiol* 1993; **138**: 923–936.
23. Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology* 2003; **14**: 451–458.
24. Hernan MA, Robins JM. Letter to the editor: assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics* 1999; **55**: 1316–1317.
25. Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology* 2003; **14**: 459–466.
26. Walker AM. *Observation an Inference*. Epidemiology Resources Inc.: Newton, 1991; 120–124.
27. Psaty BM, Heckbert SR, Koepsell TD, *et al.* The risk of myocardial infarction associated with antihypertensive drug therapies. *JAMA* 1995; **274**: 620–625.
28. Messerli FH. Case-control study, meta-analysis, and bouillabaisse: putting the calcium antagonist scare into context. *Ann Intern Med* 1995; **123**: 888–889.

29. ALLHAT Working Group. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic. The antihypertensive and lipid-lowering treatment to prevent heart attack trial (ALLHAT). *JAMA* 2002; **288**: 2981–2997.

30. Blood Pressure Lowering Treatment Trialists' Collaboration. Effects of different blood-pressure-lowering regimens on major cardiovascular events: results of prospectively designed overviews of randomized trials. *Lancet* 2003; **362**: 1527–1535.

31. Wacholder S. Design issues in case-control studies. *Stat Methods Med Res* 1995; **4**: 293–309.

32. Schneeweiss S, Glynn RJ, Tsai EH, Avorn J, Solomon DH. Assessment of bias by unmeasured confounders in pharmacoepidemiologic claims data studies using external data. *Epidemiology* 2005; **16**: 17–24.

33. Saliba D, Orlando M, Wenger NS, Hays RD, Rubenstein LZ. Identifying a short functional disability screen for older persons. *J Gerontol* 2000; **55**: 750–756.

34. Eppig FJ, Chulis GS. Matching MCBS and medicare data: the best of both worlds. *Health Care Financ Rev* 1997; **18**: 211–229.

35. Schneeweiss S, Wang P. Association between SSRI use and hip fractures and the effect of residual confounding bias in claims database studies. *J Clin Psychopharm* 2004; **24**: 632–638.

36. Schneeweiss S, Wang P. Claims data studies of sedative-hypnotics and hip fractures in the elderly: exploring residual confounding using survey information. *J Am Geriat Soc* 2005; **53**: 948–954.

37. Solomon DH, Schneeweiss S, Glynn RJ, Kiyota Y, Levin R, Avorn J. The relationship between selective COX-2 inhibitors and acute myocardial infarction. *Circulation* 2004; **109**: 2068–2073.

38. Adler GS. A profile of the Medicare Current Beneficiary Survey. *Health Care Financ Rev* 1994; **15**: 153–163.

39. Adler GS. Medicare beneficiaries rate their medical care. *Health Care Financ Rev* 1995; **16**: 175–187.

40. Bresalier RS, Sandler RS, Quan H, *et al.* Adenomatous polyp prevention on vioxx (APPROVe) trial investigators. Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. *N Engl J Med* 2005; **352**: 1092–1102.

41. Sturmer T, Schneeweiss S, Avorn J, Glynn RJ. Correcting effect estimates for unmeasured confounding in cohort studies with validation studies using propensity score calibration. *Am J Epidemiol* 2005; **162**: 279–289.

42. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *Am J Epidemiol* 1990; **132**: 734–745.

43. Sturmer T, Schneeweiss S, Glynn RJ. Performance of propensity score calibration (PSC). *Am J Epidemiol* 2005; **161**(suppl): S75.

44. Walker AM. Anamorphic analysis. sampling and estimation for covariate effects when both exposure and disease are known. *Biometrics* 1982; **38**: 1025–1032.

45. Cain KC, Breslow NE. Logistic regression analysis and efficient design for two-stage studies. *Am J Epidemiol* 1988; **128**: 1198–1206.

46. Flanders WD, Greenland S. Analytic methods for two-stage case-control studies and other stratified designs. *Stat Med* 1991; **10**: 739–747.

47. Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. *Stat Med* 1991; **10**: 585–598.

48. Strom BL. Sample size considerations for pharmacoepidemiology studies. In *Pharmacoepidemiology* (3rd edn), Strom BL (ed.). Wiley: Chichester; 2000.

49. Steenland K, Greenland S. Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. *Am J Epidemiol* 2004; **160**(4): 384–392.

50. Lash TL, Silliman RA. A sensitivity analysis to separate bias due to confounding from bias due to predicting misclassification by a variable that does both. *Epidemiology* 2000; **11**: 544–549.

51. Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology* 2003; **14**: 459–466.

52. Greenland S. Multiple-bias modelling for analysis of observational data. *J R Statist Soc A* 2005; **168**: 267–306.

53. Wolfe F, Michaud K. Lymphoma in rheumatoid arthritis: the effect of methotrexate and anti-tumor necrosis factor therapy in 18,572 patients. *Arthritis Rheum* 2004; **50**: 1740–1751.

54. Geborek P, Bladstrom A, Turesson C, *et al.* Tumour necrosis factor blockers do not increase overall tumour risk in patients with rheumatoid arthritis, but may be associated with an increased risk of lymphomas. *Ann Rheum Dis* 2005; **64**: 699–703.

55. Baecklund E, Ekbom A, Sparen P, Feltelius N, Klareskog L. Disease activity and risk of lymphoma in patients with rheumatoid arthritis: nested case-control study. *BMJ* 1998; **317**: 180–181.