Pfizer Inc
235 East 42nd Street
New York, New York
10017

06 June 2008

Division of Dockets Management (HFA–305)
Food and Drug Administration
5630 Fishers Lane, Room 1061
Rockville, Maryland 20852

RE: Docket No. FDA–2008–N–0234; Developing Guidance on Conducting Scientifically Sound Pharmacoepidemiologic Safety Studies Using Large Electronic Healthcare Data Sets; Public Workshop; Request for Comments (73 Federal Register 21963; 23 April 2008)

Submitted electronically: *http://www.regulations.gov*

Dear Sir or Madam:

Pfizer Inc appreciates the opportunity to submit comments to the Agency for consideration as the Agency develops Draft Guidance on Conducting Scientifically Sound Pharmacoepidemiologic Safety Studies Using Large Electronic Healthcare Data Sets. Pfizer is the world's largest private research-based pharmaceutical company dedicated to the discovery and development of novel medicines and treatments to improve the quality of life of people around the world. Our mission is to meet patients' needs by providing innovative medicines and health management services; these are enabled by advancing the quality and safety of healthcare through research.

Patient safety is a priority for Pfizer. We have a long-standing and vital interest in adverse event collection, reporting, follow-up, and evaluation and in analyzing the benefits of our products relative to the known risks. Pharmacoepidemiology, including the use of large electronic healthcare data sets, is a key component of our approach to patient safety and we engage in this and related activities on a daily basis to protect patients.

Pfizer strongly supports the Agency's stated goals of: (a) Providing guidance to Industry on designing and conducting scientifically-sound pharmacoepidemiologic safety studies using large electronic healthcare data sets and (b) Providing criteria for FDA to use in conducting consistent reviews when evaluating protocols and study reports in this important discipline. We also appreciate the Agency's demonstrated desire to seek information and views from a broad range of stakeholders on best practices and principles for the design and evaluation of such studies. Along with many other stakeholders, Pfizer participated in the public workshop held on May 7, 2008, in Silver Spring, Maryland (73 Federal Register 21963). We reference those proceedings and we are pleased to expand on comments we made at that workshop. In this letter, we provide general comments as well as specific comments on the questions raised in the Federal Register notice and at the public workshop.

Pfizer's approach to the use of large electronic healthcare data sets is guided by these overarching principles and we suggest that these principles have general applicability:

- **Patient First:** Patients' interests must drive the evolution of the use of large electronic healthcare data sets. Essential patient interests include safety, quality, individualized care, privacy, and access to care. These impact the public health. Use of large healthcare data sets in scientifically-sound pharmacoepidemiologic studies is part of this continuum.

- **Primacy of Clinical Judgment:** Large electronic data sets should provide greater access to health information and provide support that ultimately enables providers to make better healthcare decisions in collaboration with the patient.

In light of these principles, when preparing the Draft Guidance, we urge the Agency to consider the following seven points that derive from the above principles:

- **Set Forth Guiding Principles:** The Draft Guidance should not be prescriptive but, rather, it should embody a set of guiding principles, i.e., a framework that describes minimum and recommended elements for studies;

- **Harmonize with Other Regions:** The Draft Guidance should be harmonized and developed collaboratively with The European Medicines Agency (EMEA), especially since the EMEA and other European regulators have similar efforts underway through the European Union Pharmacovigilance and Pharmacoepidemiology Network;

- **Consider Other Methods:** The Draft Guidance should provide advice that Sponsors will continue to need to consider the option of proposing *de novo* randomized and cohort/registry methods to investigate drug safety questions, since large electronic data sets are not sufficient to address all safety concerns;

- **Utilize Existing Guidelines:** The Draft Guidance should incorporate or reference relevant guidelines developed by professional organizations, such as the International Society for Pharmacoepideimology's (ISPE) Good Pharmacoepidemiology Practice (GPP) Guidelines;

- **Underscore Comprehensive Rigor:** Guiding principles in the Draft Guidance should describe and underscore the importance of validation studies and quality programs as essential components of any research program that uses healthcare databases for drug safety research;

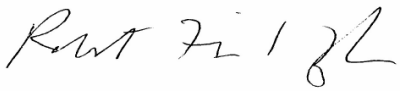
- **Develop Consensus Outcomes Definitions:** There is a need for standardized definitions for outcomes in epidemiology database studies and such definitions should be developed by the stakeholder community via a consensus process; and

- **Document Rationale:** Guiding principles in the Draft Guidance should outline the importance of performing and documenting feasibility assessments, as well as the scientific rationale for selecting a particular data, and should include recognition that data sources outside of the United States are a critical resource.

We have specific comments for each of the questions posed by the Agency in the April 23, 2008, Federal Register Notice and discussed at the May 7, 2008, public workshop. These comments are organized as presented in the Notice and are provided as Annex I. When relevant, references from the published scientific literature are included with commentary for the respective question.

Pfizer commends the Agency for stimulating dialogue on this important aspect of patient safety and we thank the Agency for the opportunities to provide our perspectives and to participate in the public workshop. We look forward to working with the Agency and other stakeholders as the Draft Guidance is developed and made available for consultation under the Agency's Good Guidance Practice guidelines. We would like to continue to provide input on this important topic and we would be glad to meet with the Agency to explain our comments or respond to any questions.

Respectfully submitted,
PFIZER INC



Robert F. Reynolds, Sc.D.
Executive Director, Epidemiology

Attachment: Annex I

Annex I: **Comments of Pfizer Inc.**

Developing Guidance on Conducting Scientifically Sound Pharmacoepidemiologic Safety Studies Using Large Electronic Healthcare Data Sets; Public Workshop; Request for Comments

06 June 2008

Docket No. FDA–2008–N–0234

73 Federal Register 21963 (23 April 2008)

Specific Comments to Questions Posed in the Federal Register Notice of 23 April 2008 and Discussed at the Public Workshop on 07 May 2008:

Panel 1 Questions

1. What information and what level of detail are needed for FDA to ensure the appropriateness of the data source to address the product safety questions being asked? How does this differ by type of data source (electronic medical records (EMR) vs. claims)?

To ensure the scientific integrity of pharmacoepidemiologic drug safety research, automated healthcare data sets should meet minimum standards of data quality. Automated healthcare databases used in pharmacoepidemiologic research vary greatly with respect to their reliability and validity. It is important to understand the differences between different types of data sources that impact on their suitability for drug safety research.

Points to be considered when conducting or selecting appropriate data sources to conduct pharmacoepidemiologic safety studies:

i) One of the most important criteria in the conduct of pharmacoepidemiologic research using automated healthcare databases is the ability to validate the automated data with original medical records. The most desirable data sources are those in which it is possible to accurately measure the frequency or rate of occurrence of a health outcome without the need for extensive data validation. In North America, examples of data sources that have already undergone extensive validation include Kaiser Permanente, Group Health of Puget Sound, and Saskatchewan Health. In Europe, databases such as GPRD and THIN in the UK, and population-based healthcare registries in Denmark and Sweden have also been extensively validated, or have the ability to access the original medical records for validation purposes. Therefore, it may be appropriate to use either US or non-US data sets or both to address a specific safety question;

ii) Data sources which do not allow for validation present challenges to the pharmacoepidemiologist. Given the healthcare system in the US, some insurance plans may increase the incentive for upcoding on billing claims. In

such instances, claims-based automated data may be unreliable. In this situation, it is important to validate at least a sample of the outcomes;

iii) The ability to conduct record linkage of exposure and outcome data (endpoints and important confounding variables) is important;

iv) The ability to conduct linkage with vital statistics data is important;

v) Given differences in healthcare delivery systems, in general, membership of subjects enrolled in non-US databases tend to be more stable than in US databases;

vi) It is important for identified populations to be characterized on the basis of enrollment for the purposes of measuring person-time. Data sources that do not allow for the measurement of person-time present an important limitation;

vii) Quality assurance, including routine quality control measures, should be accounted for in any plan to use database(s) in pharmacoepidemiologic research;

viii) It may be useful for FDA to conduct similar drug safety studies in different databases to compare and contrast results, and to identify differences that may suggest the occurrence of important biases based on the source data used;

ix) Representativeness of the study population, i.e., whether the targeted treated patient population is well represented in the database, is important. For example, certain populations, such as the elderly, are usually underrepresented in HMO data and, thus, it would not be appropriate to evaluate a medication that treats a medical condition primarily occurring in the elderly, such as age-related macular degeneration, using HMO data;

x) Population size of the data source must be sufficient for the study, especially when rare events are being evaluated;

xi) Are there any formulary restrictions in the health care system in which the data are derived? This is particularly relevant when designing a study to evaluate the effect of medication in the same class or in different classes to treat the same medical condition;

xii) Misclassification of drug exposure should be considered, medication prescription or dispensing may not reflect the actual use of medication. It is also possible that patients receive medications outside of the data system;

xiii) The event (outcome) of interest should be identifiable using diagnostic code(s), such as ICD-9 codes; and

xiv) It should be recognized that data on important risk factors/confounders, such tobacco use, alcohol intake, etc., may not be available in some of the databases.

2. What are the challenges of using enrollment data for defining study populations in claims databases? Describe effective strategies for addressing the absence of formal enrollment in some EMR systems.

Automated healthcare databases in which there is high turnover of membership due to the transient nature of the population, i.e., changes in employment, present significant challenges in terms of defining cohorts, measuring outcomes with sufficient follow-up, etc. Also, it is difficult to ascertain whether the individual has left due to an event, which raises validity concerns. It is less of a concern when evaluating the acute effects of medications. Studies conducted in such data sources may not necessarily be representative of the rest of the population. In contrast, healthcare databases that are population-based or represent a specific geographic region where there is low turnover are better suited for pharmacoepidemiologic studies.

It is important to define what is meant by the Electronic Medical Record (EMR). In general, EMRs that are intended as stand-alone data sources in which to conduct pharmacoepidemiologic studies should meet the same criteria established for other types of automated databases as described earlier. Enrollment data are lacking in most EMR systems in the US, which presents an important challenge for defining the study population and in measuring exposure. *It is not apparent that there are any effective strategies for addressing the absence of enrollment data in EMRs.*

3. Under what circumstances should FDA consider studies using non-U.S. electronic data sources in its assessment of product safety questions?

Non-US data sources should be considered under the following circumstances:

i) When the target patient population is not well represented in the US data sources, such as the elderly;
ii) When a medication has been marketed in non-US countries for a period of time prior to the approval of the same medication in the US, retrospective evaluation of non-US data can address the safety concern in a timely manner; and
iii) When special cohorts, such as birth registries, or patient cohorts, such as cancer registries, exist in non-US countries but not in the US.

Different electronic databases may have been devised with different users in mind and may be specifically suited to answer certain questions. The needs generated by the question(s) should be the guiding criteria when selecting a specific database to use for a given pharmacoepidemiologic study. The population included should be representative of the population from which it is drawn, and ideally, one would want to use the most extensive database, to include in- and outpatient data, laboratory tests, diagnostic and prognostic tests, use of prescribed and OTC medications, etc. Information on race and ethnicity, together with other possible confounders, e.g., tobacco use, alcohol intake, etc., are also desirable.

With the proviso that representative and complete data are available from a fully-described population, the use of non-US data sources should always be considered for a pharmacoepidemiologic study.

The characteristics of healthcare systems need to be considered, but these may vary within the US as well as in non-US countries. The quality of care may vary by many factors, including access to health coverage, and this can have an impact on the generalizability of the results, for example, on management of chronic illnesses. Recent work has shown that the outcomes for the management of diabetes in the UK match those for the insured population in the US and that both of these groups have considerably better outcomes than the uninsured US population [1].

Reference

1. Mainous AG, Diaz VA, Saxena S, Baker R, Everett CJ, Koopman RJ, Majeed A. Diabetes management in the USA and England: comparative analysis of national surveys. *J Roy Soc Med*. 2006; 99:463-469.

Panel 2 Questions

1. How can FDA assure that the study design accurately captures the clinical events, exposures of interest, and confounding factors needed to answer the product safety question under investigation?

Similar elements that are fundamental to the study design of any epidemiologic study apply to the study design of a pharmacoepidemiologic database study [1]. An appropriate study design is driven by a research question of interest, which requires clear operational definitions of outcomes (clinical events), exposures and confounders, as well as the selection criteria and comparison groups [2]. An operational definition is defined as one that can be implemented independently using the data available in the proposed study [2].

Given that healthcare databases were not created for research purposes, numeric coding such as international Classification of Diseases version 9 (ICD-9) or Common Procedural Terminology version 4 (CPT-4) is used to define clinical events. The use of discharge diagnoses for the identification of cases can potentially result in several sources of error including variation in coding procedures, coding errors, incomplete coding, lack of specificity in available codes and error in the clinical diagnoses [3, 4]. In order to overcome these potential problems in a study, validation studies are frequently required [5, 6]. Timing of a clinical event needs to be well defined in order to ensure that an exposure of interest preceded an event [7]. Exposure definitions need to be well-defined in terms of timing, dose and duration of use and relevant to a causal inference [3, 7]. Depending on the latency period associated with a clinical event of interest, not all databases may be adequately used to answer questions about long term drug effects [8]. For both exposure and outcome definitions, it is critical to consider based on a research question at hand whether the study population should be limited to the incident exposure/cases or prevalent exposure/case or both incident and prevalent exposure/case [7, 9].

Information on confounders such as disease severity, tobacco use, occupation, alcohol intake or laboratory values is often limited or lacking in healthcare databases. In the case of disease severity, several proxies have been used including information on the intensity of therapy [10, 11]. At the same time there can be a large number of measures for each construct of a confounding factor [12]. When no prior knowledge is available of which measure is optimal, it is possible to have a large number of covariates. Several approaches that involve data reduction have been developed to address this problem

including exposure propensity scores [13-15] and disease risk scores [14, 16, 17].

The choice of adequate comparison group is critical to the valid design of a study. However, choosing a comparison group is frequently complex [12]. Ideally, the comparison group should comprise patients with identical distributions of measured and unmeasured risk factors of the study outcome [12]. Depending on the research question, a comparison group can be comprised of patients who use drugs with the same perceived medical indication. Alternatively, in those cases where there is either no comparator drug with a similar indication to the study drug or a suspected class effect, a comparison group can be comprised of nonusers, i.e., subjects who did not use any drug of this class [12]. The comparability of the treatment/exposed group and the comparison group, such as patient characteristics and severity of the disease, should be evaluated when interpreting study findings.

References

1. Strom, B.L., *Study Designs Available for Pharmacoepidemiology Studies*, in *Pharmacoepidemiology*, B.L. Strom, Editor. 2005, John Wiles & Sons Ltd: West Sussex, p. 17-28.
2. *Guidelines for good pharmacoepidemiology practices (GPP).* Pharmacoepidemiol Drug Saf, 2008; **17**(2):200-8.
3. Sorensen, H.T., Johnsen, S.P., Norgard, B. , *Methodological issues in using prescription and other databases in pharmacoepidemiology.* Norwegian Journal of Epidemiology, 2001; **11**(1):13-18.
4. Steinberg, E.P., J. Whittle, and G.F. Anderson, *Impact of claims data research on clinical practice.* Int J Technol Assess Health Care, 1990; **6**(2):282-7.
5. West, S.L., Strom, B.L. and Poole, C., *Study Designs Available for Pharmacoepidemiology Studies*, in *Pharmacoepidemiology*, B.L. Strom, Editor. 2005, John Wiles & Sons Ltd: West Sussex, p. 709-766.
6. Strom, B.L., et al., *Using a claims database to investigate drug-induced Stevens-Johnson syndrome.* Stat Med, 1991; **10**(4):565-76.
7. Jick, H., L.A. Garcia Rodriguez, and S. Perez-Gutthann, *Principles of epidemiological research on adverse and beneficial drug effects.* Lancet, 1998; **352**(9142):1767-70.
8. Strom, B.L., *Data validity issues in using claims data.* Pharmacoepidemiol Drug Saf, 2001; **10**(5):389-92.
9. Ray, W.A., *Evaluating medication effects outside of clinical trials: new-user designs.* Am J Epidemiol, 2003; **158**(9):915-20.
10. Spitzer, W.O., et al., *The use of beta-agonists and the risk of death and near death from asthma.* N Engl J Med, 1992; **326**(8):501-6.
11. Suissa, S., L. Blais, and P. Ernst, *Patterns of increasing beta-agonist use and the risk of fatal or near-fatal asthma.* Eur Respir J, 1994; **7**(9):1602-9.
12. Schneeweiss, S. and J. Avorn, *A review of uses of health care utilization databases for epidemiologic research on therapeutics.* J Clin Epidemiol, 2005; **58**(4):323-37.
13. Rubin, D.B., *Estimating causal effects from large data sets using propensity scores.* Ann Intern Med, 1997; **127**(8 Pt 2):757-63.
14. Glynn, R.J., S. Schneeweiss, and T. Sturmer, *Indications for propensity scores and review of their use in pharmacoepidemiology.* Basic Clin Pharmacol Toxicol, 2006; **98**(3):253-9.
15. Sturmer, T., et al., *Analytic strategies to adjust confounding using exposure propensity scores and disease risk scores: nonsteroidal antiinflammatory drugs and*

short-term mortality in the elderly. Am J Epidemiol, 2005; **161**(9):891-8.
16. Feltelius, N., et al., *Results from a nationwide postmarketing cohort study of patients in Sweden treated with etanercept.* Ann Rheum Dis, 2005; **64**(2):246-52.
17. Schneeweiss, S., et al., *Outcomes of reference pricing for angiotensin-converting-enzyme inhibitors.* N Engl J Med, 2002; **346**(11):822-9.

2. What are effective strategies to address confounding by indication and the effect of measured and unmeasured confounders?

One common limitation of observational studies of drug safety is the presence of confounding by indication [1-5]. Confounding by indication is present when characteristics that differ between the two groups are frequently associated with the outcome under investigation. This can confound the association between drug use and the outcome, in a similar fashion as the selection bias in traditional observational studies. For example, patients who are prescribed insulin to treat diabetes are at high risk of cardiovascular events due to the presence of diabetes. Therefore, one might expect that given a population where insulin treatment was not randomized, insulin users would be at a higher risk of cardiovascular events than non-users. Without accounting for the indication for the use of insulin, one might conclude that insulin therapy increases the risk of cardiovascular events even though it may actually decrease risk among those individuals who share the same diabetes risk profile. In certain situations, these factors can be well characterized and this confounding can be controlled for using standard methods such as multivariable-adjustment of regression models [2]. However, when the indication cannot be characterized in detail, traditional methods will not sufficiently account for the confounding [2, 4-6]. Newer statistical methods including the use of group treatment variables, propensity score adjustment [7-15] and marginal structural models [6, 16, 17] can be used to adjust for confounding by indication.

Well-measured confounders can be controlled for using a variety of statistical techniques including stratified analysis and multivariable regression approaches [18, 19], while the difficulty in characterizing the indication for therapy in observational studies is linked to a more general problem of observational studies in general: that of unmeasured confounders [2-4, 20]. This is particularly problematic in large database studies where several strong confounding factors may not be available from the data source, including medical history and lifestyle factors such as smoking [2-5]. An initial approach to dealing with unmeasured confounders might be to construct causal diagrams, such as directed acyclic graphs [21-26], to determine whether controlling for measured covariates may be sufficient to control for the effects of these unmeasured covariates. Frequently, however, it is not possible to find a sufficient adjustment set that does not include at least one of these unmeasured confounders. Therefore, approaches such as propensity score methods [7-9, 12, 27-30] or bounding the range of possible associations using sensitivity analysis that use standard statistical analyses [31-35] can be used to help control for confounding by these unmeasured covariates.

Despite these advances, controlling for confounding by indication is not easily or completely done in practice. It may not be possible to reliably measure the severity of the underlying illness or there may be unknown factors that influence the selection of a medicine, factors, which by definition, will be unmeasured. In these instances, studies using baseline randomization and observational follow-up, i.e., large simplified trials, may be more appropriate for addressing the safety question.

References

1. Strom BL. Methodologic challenges to studying patient safety and comparative effectiveness. Med Care, 2007; **45**:S13-5.
2. Strom BL, Miettinen OS, Melmon KL. Postmarketing studies of drug efficacy: when must they be randomized? Clin Pharmacol Ther, 1983; **34**:1-7.
3. Sorensen HT, Lash TL, Rothman KJ. Beyond randomized controlled trials: a critical comparison of trials with nonrandomized studies. Hepatology, 2006; **44**:1075-82.
4. Strom BL. Pharmacoepidemiology. Chichester ; Hoboken, NJ: J. Wiley, 2005.
5. Schneeweiss S. Developments in post-marketing comparative effectiveness research. Clin Pharmacol Ther, 2007; **82**:143-56.
6. Cole SR, Hernan MA, Robins JM, et al. Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models. Am J Epidemiol, 2003; **158**:687-94.
7. Fernandez M, Calvo-Alen J, Bertoli AM, et al. Systemic lupus erythematosus in a multiethnic US cohort (LUMINA L II): relationship between vascular events and the use of hormone replacement therapy in postmenopausal women. J Clin Rheumatol, 2007; **13**:261-5.
8. Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. Basic Clin Pharmacol Toxicol, 2006; **98**:253-9.
9. Mapel DW, Nelson LS, Lydick E, Soriano J, Yood MU, Davis KJ. Survival among COPD patients using fluticasone/salmeterol in combination versus other inhaled steroids and bronchodilators alone. COPD, 2007; **4**:127-34.
10. Schmoor C, Caputo A, Schumacher M. Evidence from nonrandomized studies: a case study on the estimation of causal effects. Am J Epidemiol, 2008; **167**:1120-9.
11. Seeger JD, Kurth T, Walker AM. Use of propensity score technique to account for exposure-related covariates: an example and lesson. Med Care, 2007; **45**:S143-8.
12. Sturmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. J Clin Epidemiol, 2006; **59**:437-47.
13. D'Agostino RB, Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med, 1998;**17**:2265-81.
14. Kurth T, Walker AM, Glynn RJ, et al. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. Am J Epidemiol, 2006; **163**:262-70.
15. Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika, 1983; **70**:41-55.
16. Cole SR, Hernan MA, Margolick JB, Cohen MH, Robins JM. Marginal structural models for estimating the effect of highly active antiretroviral therapy initiation on CD4 cell count. Am J Epidemiol, 2005;**162**:471-8.
17. Brotman RM, Klebanoff MA, Nansel TR, et al. A Longitudinal Study of Vaginal Douching and Bacterial Vaginosis--A Marginal Structural Modeling Analysis. Am J Epidemiol, 2008; doi:10.1093/aje/kwn103.
18. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008.
19. Selvin S. Statistical analysis of epidemiologic data. Oxford, New York: Oxford University Press, 2004.
20. Eg Hansen AB, Gerstoft J, Kirk O, et al. Unmeasured confounding caused slightly

better response to HAART within than outside a randomized controlled trial. J Clin Epidemiol, 2008; **61**:87-94.
21. Greenland S, Brumback B. An overview of relations among causal modelling methods. Int J Epidemiol, 2002; **31**:1030-7.
22. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology, 1999; **10**:37-48.
23. Hernan MA, Hernandez-Diaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. Am J Epidemiol, 2002; **155**:176-84.
24. Oakes JM, Kaufman JS. Methods in social epidemiology. San Francisco, CA: Jossey-Bass, 2006.
25. Pearl J. Causality: models, reasoning, and inference. Cambridge, U.K.; New York: Cambridge University Press, 2000.
26. Robins JM. Data, design, and background knowledge in etiologic inference. Epidemiology, 2001; **12**:313-20.
27. Oakes JM, Church TR. Invited commentary: advancing propensity score methods in epidemiology. Am J Epidemiol, 2007; **165**:1119-21; discussion 1122-3.
28. Sturmer T, Glynn RJ, Rothman KJ, Avorn J, Schneeweiss S. Adjustments for unmeasured confounders in pharmacoepidemiologic database studies using external information. Med Care, 2007; **45**:S158-65.
29. Sturmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. Am J Epidemiol, 2005; **162**:279-89.
30. Sturmer T, Schneeweiss S, Rothman KJ, Avorn J, Glynn RJ. Performance of propensity score calibration--a simulation study. Am J Epidemiol, 2007; **165**:1110-8.
31. Cole JA, Rothman KJ, Cabral HJ, Zhang Y, Farraye FA. Migraine, fibromyalgia, and depression among people with IBS: a prevalence study. BMC Gastroenterol, 2006; **6**:26, doi:10.1186/1471-230X-6-26.
32. 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**:384-92.
33. Greenland S. Sensitivity analysis, Monte Carlo risk analysis, and Bayesian uncertainty assessment. Risk Anal, 2001; **21**:579-83.
34. Greenland S. Basic methods for sensitivity analysis of biases. Int J Epidemiol, 1996; **25**:1107-16.
35. MacLehose RF, Kaufman S, Kaufman JS, Poole C. Bounding causal effects under uncontrolled confounding using counterfactuals. Epidemiology, 2005; **16**:548-55.

3. What are other challenges to internal and external validity in studies using EMR and claims databases? What are the best practices for addressing them?

In considering the term “other” as posed in Question 3, we preface our response with the understanding that “other” challenges refer to challenges in addition to those outlined in Questions 1 & 2, i.e., beyond exposure classification and beyond confounding by indication. Thus, “other” would include claims data set/EMR and system-related challenges.

Reliance on secondary data poses a significant challenge to internal validity and external validity. Administrative databases collect information only when there is an encounter with the physician that is accompanied by a diagnosis, medical procedures, or prescribing of

medications, while electronic medical records (EMR) typically are comprised of detailed clinical information, i.e., patient and family history, laboratory results [1]. Although the consequences of study design and study conduct issues were previously addressed in Questions 1 and 2, along with reference to the extant literature, other challenges inherent in claims data and EMR utilization that affect internal data validity include:

i) Misfiling, i.e., false ranking of primary diagnosis [1];
ii) Miscoding, i.e., drugs and doses at pharmacies, procedures [1];
iii) Over-reporting, i.e., depiction of more diagnostic codes unintentionally or based on an incentive system [2];
iv) Under-reporting, i.e., secondary diagnoses, failure to file claims [1];
v) Changes in hardware/ software/ coding practices/ health care company mergers [1]; and
vi) Referral bias [2].

Although the EMR provides a comprehensive method of capturing medical information, time lag between patient visit and input of data by the physician, as well as incomplete capture of informational exchange, may also occur [3]. A best practice for addressing drug exposure misclassification is to rely on electronic pharmacy dispensing records, which are reimbursed by insurers based on detailed filed claims and are viewed as the gold standard of drug exposure information compared to self-reported information [4]. To determine the integrity of linked claims data in the event of claims database mergers, descriptive analyses of the population over time can be conducted [5]. Medical record validation of the patients' primary records help ensures high specificity of the diagnosis, thus minimizing misclassification bias [1, 6, 7].

External validity in claims data is challenged by the underlying populations' insurance status, income level, geographic location, regional practice patterns, and reimbursement/cost differences [2]. External validity must be addressed by ascertaining to whom the study results are most generalizable. Stratified or sensitivity analyses according to clinically relevant subgroups, i.e., comparison group, and subjects with contraindications, very low adherence levels, and high/low risk, may assist with evaluation of treatment effects across groups and generalizability to the broader population [8]. The conduct of small-area variation studies ascertains differences in regional practice patterns such as lengths of stay, hospital admission, and rates of surgical procedures/physician visit costs [2]. To account for cost differences which vary across time and place, assessments should be normalized to adjust for geographic and longitudinal variations [2].

Researchers suggest conduct of both pre-post and database cohort comparisons to allow for more robust scientific inferences, rather than using either approach alone [9-11]. This approach aims to improve internal and external validity by assessing:

i) Whether the cohorts differed before event onset;
ii) Whether the cohorts differed after event onset; and
iii) Whether the patterns of change (pre and post event) differed across cohorts [2].

Furthermore, researchers must be cognizant of the similarities and differences between the database study sample and study population when making inferences based on the study results.

References

1. Schneeweiss, S. and J. Avorn, *A review of uses of health care utilization databases for epidemiologic research on therapeutics.* J Clin Epidemiol, 2005; **58**(4):323-37.
2. Motheral, B.R. and K.A. Fairman, *The use of claims databases for outcomes research: rationale, challenges, and strategies.* Clin Ther, 1997; **19**(2):346-66.
3. Porter, S.C. and K.D. Mandl, *Data quality and the electronic medical record: a role for direct parental data entry.* Proc AMIA Symp, 1999: p. 354-8.
4. West, S.L., et al., *Recall accuracy for prescription medications: self-report compared with database information.* Am J Epidemiol, 1995; **142**(10):1103-12.
5. Hennessy, S., et al., *Descriptive analyses of the integrity of a US Medicaid claims database.* Pharmacoepidemiol Drug Saf, 2003; **12**(2):103-11.
6. Evans, J.M. and T.M. MacDonald, *Misclassification and selection bias in case-control studies using an automated database.* Pharmacoepidemiol Drug Saf, 1997; **6**(5):313-8.
7. Strom, B.L., *Methodologic challenges to studying patient safety and comparative effectiveness.* Med Care, 2007; **45**(10 Supl 2):S13-5.
8. Schneeweiss, S., et al., *Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results.* Med Care, 2007; **45**(10 Supl 2):S131-42.
9. Sclar, D.A., et al., *Antidepressant pharmacotherapy: economic evaluation of fluoxetine, paroxetine and sertraline in a health maintenance organization.* J Int Med Res, 1995; **23**(6):395-412.
10. Sclar, D.A., et al., *Antidepressant pharmacotherapy: economic outcomes in a health maintenance organization.* Clin Ther, 1994; **16**(4):715-30; discussion 74.
11. Soumerai, S.B., et al., *Effects of a limit on Medicaid drug-reimbursement benefits on the use of psychotropic agents and acute mental health services by patients with schizophrenia.* N Engl J Med, 1994; **331**(10):650-5.