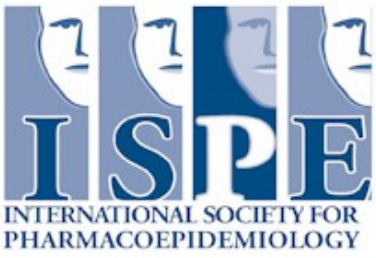

Response to:

US Food & Drug Administration Docket No. FDA 2011-D-0057

Draft Guidance for Industry and FDA Staff on Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets

Submitted by the

International Society of Pharmacoepidemiology

April 15, 2011

FDA "Guidance for Industry and FDA Staff: Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets"

The International Society for Pharmacoepidemiology (ISPE) is very pleased to have the opportunity to offer our perspectives and suggestions, and submits for your consideration the following comments on the FDA "Guidance for Industry and FDA Staff: Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets".

ISPE is an international, nonprofit, professional membership organization dedicated to promoting the health of the public by advancing pharmacoepidemiology, the science that applies epidemiological approaches to studying the use, effectiveness, values and safety of pharmaceuticals. ISPE is firmly committed to providing an unbiased scientific forum to the views of all parties with interests in drug, biologics, and devices development, delivery, use, costs and value, adverse and beneficial effects, and therapeutic risk management.

Moreover, the Society provides an international forum for the open exchange of scientific information among academia, government, and industry and for the development of policy; a provider of education; and an advocate for the fields of pharmacoepidemiology and therapeutic risk management.

The Society's more than 1,000 members represent 45 countries. ISPE members work in academic institutions, the pharmaceutical industry, government agencies, and non-profit and for-profit private organizations. ISPE members are researchers with background and training in epidemiology, biostatistics, medicine, public health, nursing, pharmacology, pharmacy, law, and health economics.

Our comments are based on a careful review of the FDA document by the Database Sepcial Interest Group and the Public Policy Committee as well as by the Society's membership at-large, by ISPE Fellows, Past Presidents, members of the Board of Directors and Executive Committee. Due to the development process of the draft documents in which many ISPE members from academia, research centers and regulatory bodies were involved, some of these comments may have been sent directly to the FDA. We thank the FDA for allowing us the opportunity to comment on this document. ISPE welcomes any future dialogue with the FDA.

Sincerely,

Databse SIG
Public Policy Committee,
Board of Directors,
International Society for Pharmacoepidemiology (ISPE)

The FDA is to be complimented for compiling this document as it provides some useful guidance on an extremely broad topic about which there exists a great deal of opportunity for misunderstanding, namely how to approach, conduct and report pharmacoepidemiologic studies using electronic healthcare databases.
The FDA guidance is focused on what needs to be reported to the FDA and the steps to allow this to happen but there is inconsistency in the depth of general guidance on the use of databases for pharmacoepidemiologic studies. The areas of design, database selection, and analysis are covered to some extent, but not always fully, and other areas are covered in one sentence or not at all (e.g., privacy, data cut). ISPE suggests to either greatly expand the FDA guidance to cover all these areas fully or, alternatively, direct readers to other guidance. In addition to general guidelines, e.g., the ISPE Guidelines on Good Pharmacoepidemiologic Practices[1] and STROBE[2,3] guidance, there are also ISPE guidelines for Quality Conduct in DatabaseResearch[4] and ISPOR guidance on conducting[5] and reviewing database studies[6].

For example, the ISPE Guidance on Database Conduct devotes entire sections and provides clearer guidance to use of multiple resources, data privacy, and selecting a resource. In addition, when drafting a protocol, coding algorithms used to define outcomes or covariates should be included as an appendix. Finally, given the FDA's attempt to not favor a specific data source, study design, or analytic technique, the reviewers felt that the reference section should be more comprehensive and perhaps list publications that cover the broad range of data sources, study designs, analytic techniques and development/validation of coding algorithms in more detail.

Major Comment:

1. p.12, ll. 462-470: As laid out in the Mini Sentinel Taxonomy document of October 2010 and a 2010 PDS paper,[7] we view 2 fundamental design choices based on where variation of exposure is likely to occur (within patients, between patients) and whether confounding by time invariant patient factors is likely strong. Study designs fall into two general categories – one uses "between-person" comparisons (i.e. cohort, case-control, and case-cohort studies) and the other uses "within-person" comparisons (i.e. case-crossover and self-controlled-case series). A "between-person" approach samples comparator experience from other people at the same time, whereas a "within-person" approach samples comparator experience from the same people at a different time. Lines 462-466 list commonly used designs in pharmacoepidemiology. However, the "Comparator Selection" section beginning on line 474 is couched within the context of between-person comparisons only. If within person comparisons are feasible then self-controlled designs, like the case-crossover design or the self-controlled case series design should be applied, both designs that inherently adjust for all time-invariant patient characteristics. If such designs are not feasible or will not provide answers to the specific study questions then between-subject comparisons are the next best choice and in practice the most frequent choice. Between patient comparison will be conducted with cohort designs or sampling strategies embedded in cohorts, i.e. case-control, case-cohort, 2-stage sampling. Particularly in electronic healthcare data sets it might be misleading to portray

case-control studies or case-cohort studies as freestanding designs. They are simply efficient sampling strategies within an underlying cohort. In databases studies we can enumerate this underlying cohort since all the claims data have already been collected. Such efficient sampling strategies are most helpful if additional confounder information will be collected at an extra cost and time commitment. These sampling designs will increase efficiency over the full cohort analysis. Conversely, if no additional data is being collected there is no point in sampling but one would rather prefer to see a full cohort analysis. Computational efficiency is no longer an excuse in epidemiology. A third design choice would be between provider (physician, facility, health plan) or between region comparisons, which in some cases may be analyzed as instrumental variables if several assumptions are fulfilled. We do not advocate making this point explicit in this guidance as it is an infrequent application.

2. Given that FDA does not advocate a particular design approach, the discussion of comparator selection should be more balanced. We suggest that a discussion of the issues surrounding selection of within-person experiences as comparator time be included in the guidelines. For example, the case-crossover approach typically uses a "control" period antecedent to the "case" period, where the case period generally resides just before the abrupt outcome of interest. Other approaches may use person-time both preceding and following the outcome of interest. Bi-directional approaches may be limited by potential issues with immortal person-time. Further, investigators must consider whether exposure status in control periods following the event of interest may be influenced by the event itself. Other issues for consideration when using a self controlled design include bias due to exposure time-trends.[8] These time trends can be particularly problematic in the setting of newly marketed products.

3. P. 15, ll. 624-632: We are glad the new (incident) user design made its way into the document. We see in the reference list Wayne Ray's paper but it is not referenced in this paragraph although it should be. We agree with the technical aspects of defining incident users and that we need sensitivity analyses with varying length of washout periods. What we missed where two points. Firstly a brief explanation why the new user design has so many advantages over a prevalent or mixed user design. Of course Wayne Ray has provided plenty text for that.[9] Secondly, as in PDS 2010[7] we would advocate that researchers should start out with a incident user design when conceiving a new study and should then explain why they did NOT use an incident users design (usually worries about shrinking study size) and discuss the potential biases this decision may cause.

General comments:

1. A glossary should be included in every protocol given that database nomenclature is not always consistent (e.g. look-back period vs. baseline). Perhaps the FDA may suggest standard descriptions.

2. Not all databases are suitable for all research questions. The study question/requirements should be outlined and then the databases selected after checking whether or not they are suitable.

3. Is the protocol developed for mini-sentinel to study saxigliptan and risk of acute myocardial infarction an example of what is expected?

4. For each section, point out where STROBE[2,3] may apply. Perhaps the FDA can provide some example language similar to what STROBE provides.

5. Should there be a section to address reporting of adverse events when the study involves looking in medical charts? For example if – when reviewing EMR records – there is a notation of possible adverse event due to a specific drug

6. No mention of or statement regarding multiple comparisons (without implying that there should be adjustments for multiple comparisons but the issue should probably be raised)

7. There is no discussion of quality control in the data cut/analytic process. This is vital and an area where mistakes are often made – see e.g.,[1]

Specific comments:

1. **Line 47:** Would change "studies" to "evaluations" in keeping with current Mini-Sentinel terminology.

2. **Line 111-114:** The FDA guidance suggests that generally database studies would be particularly useful when other forms of design are infeasible. They don't highlight the advantages such as timeliness (speed), minimization of some biases, and generalizability.

3. **Line 148:** Would re-word to "there is often information missing from published..."

4. **Line 174 and others:** What is the definition of “safety”? We feel this needs an explanation since a safety measure for some drugs can be a benefit measure for others. For example the outcome of Major Acute Coronary Event (MACE) can be a benefit or an adverse effect of a drug. In the context of Comparative Effectiveness, the restriction of the guidance document to drug safety seems out of place.

5. **Line 176:** Taken out of context, this could be taken to mean that FDA wants protocols for all pharmacoepi studies to be submitted. Would it be helpful for FDA to clarify which pharmacoepi studies it would like to see protocols for?

6. **Line 192:** Any recommendation on what is considered concise for a summary (e.g. 1 page)?

7. **Line 198:** Duration of study should be described. For example, if the study is a regulatory commitment, you may want to provide a description that the study will last x number of years or will continue until x number of patients with x number of person-years are captured.

8. **Line 199:** Add inclusion/exclusion criteria to summary

9. **Line 206:** Would change to "Clinical and public health impact"

10. **Line 227:** I would add that the discussion should reflect understanding of the clinical conditions the drugs under study are used for.

11. **Line 249:** How should level of experience be expressed?

12. **Line 261:** Care should made that the document is not advocating post-hoc power analyses. Perhaps reference could be made to this (Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. Ann Intern Med. 1994;121:200-6.)

13. **Line 261:** Consider changing "initial statistical power calculations"to"width of the resulting confidence intervals", as the latter is more informative.

14. **Line 264:** Statement about statistical significance can be easy to achieve in database research should be rephrased since this is often not the case when you are looking at rare outcomes, rare exposures, or within patients with rare conditions/diseases

15. **Line 265:** Would add "and public health" following "clinical".

16. **Line 274:** Would add: "and a discussion of other questions raised."

17. **Line 310:** We agree that it is important that the source data capture exposure and outcomes, but it also seems important that the data source capture relevant covariates.

18. **Line 369-370:** Results of a feasibility assessment should be given

19. **Line 374:** We appreciate the listing of issues to consider when using non-US data sources, but some of these would also apply to US data sources, such as prescribing and utilization practices within a particular health plan.

20. **Line 475:** It seems surprising that the FDA is encouraging multiple comparison groups to enhance validity. Although multiple comparison groups can be useful for the purpose of evaluating different hypotheses, interpretation of the different results obtained can distract from the safety/effectiveness finding. We would think that the focus should be on finding the most suitable comparison group.

21. **Line 485:** It is unclear to us why historical comparators for "cases" are mentioned. Is this assuming a case-control design?

22. **Line 490:** We agree that differences between preventive therapies and treatments may lead to differences in design considerations, but then the selective mention of vaccines and the specific "healthy vaccinee effect" makes this paragraph appear to be about vaccines.

23. **Line 521:** Propensity scores are really not innovative and will become progressively less so. We suggest rewording this to improve the longevity of the document and remove the dichotomy between "traditional" and "innovative" approaches.

24. **Line 529:** Consider adding discussion of other methods of controlling for confounding (e.g., restriction, stratification, multivariable modeling, disease risk scores) in addition to propensity scores.

25. **Line 531:** The word "usually" does not seem correct as used.

26. **Line 536:** Instead of presenting the propensity score model to assess performance and fit (what criteria?) it is better to assess the performance of propensity scores to control for confounding by looking at the balance of important risk factors for the outcome within propensity score strata, propensity score matched, or weighted groups

27. **The paragraph starting with line 540:** We suggest rewording or striking this entire paragraph. As currently written, it does not describe restriction for confounder control, which we agree is a valid and useful technique.[11] Perhaps common techniques for confounder control could be listed and described – restriction, stratification, matching (balancing), modeling, and weighting.

28. **Line 579- 587** (section on sample size and power calculations): The issue of power calculation in its conventional form is closely related to significance testing and p- value thinking that should not be encouraged[1]. Power calculation essentially deals with the numbers required to obtain a "significant result" if there is an association above a certain threshold. With large databases, the situation often is the reverse: The number of subjects (e.g. exposed or cases) is given and the precision of estimates is a function of these numbers. Therefore, researchers should describe what precision in the measure of association they would achieve within a given database, rather than calculate what sample size is required as in a conventional power calculation. This approach also allows researchers to

conclude beforehand that the data source is not large enough, i.e., when the anticipated lack of precision precludes any meaningful interpretation.

29. **Line 579:** Consider adding "/detectable difference" to "statistical power".

30. **Line 626:** We suggest removing “popularity” as a metric for the value of the new (incident) user design. Instead the advantages of the design (addresses time-varying hazards and survivor-based selection among prevalent users)

31. **Paragraph beginning with line 646:** This paragraph should mention OTC medications.

32. **Paragraph beginning with line 655:** This paragraph should mention hospitalization as one mechanism for apparent gaps in drug therapy (Suissa S. Immeasurable time bias in observational studies of drug effects on mortality. Am J Epidemiol 2008;168:329-335.)

33. **Line 679:** Can occur in outpatient data too, but is perhaps “particularly” important when using inpatient data.

34. **Line 726:** The population, database, and timeframe should be discussed with respect to whether any reported validity measure would apply to the study population at hand (i.e., the transportability of measurement error)

35. **Line 769:** Discuss delay in getting mortality records and how this will influence study design (e.g study window for data freeze)

36. **Line 771:** Some EMR databases (e.g.THIN)[10] and claims (e.g. Medicare) have been shown to have good quality death and date of death information - if not cause of death.

37. **Line 802:** We suggest removing the word “both” since this would imply that only unadjusted and adjusted results are available. Often there are unadjusted and a number of adjusted results available (age&sex adjusted, partially adjusted, fully adjusted).

38. **Line 815:** Post-hoc analyses are glossed over without acknowledgement of the controversy in recent years

39. **Paragraph beginning with line 820:** This paragraph seems out of place on its own. We expected to see it combined with some earlier paragraphs.

References:

1. ISPE commentary: Guidelines for good pharmacoepidemiology practices (GPP). Pharmacoepidemiology and Drug Safety 2008;17:200-8.
2. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)statement: guidelines for reporting observational studies. PLoS Med. 2007 Oct 16;4(10):e296. PMID: 17941714
3. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ,Poole C, Schlesselman JJ, Egger M; STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med. 2007 Oct 16;4(10):e297. PMID: 17941715
4. http://www.pharmacoepi.org/resources/Quality_Database_Conduct_2-28-11.pdf
5. Arnold RG, Kotsanos JG, Motheral B, Ramsey S, Crown W, Puder K, Hornbrook M, Wright A, Murray M. Panel 3: Methodological Issues in Conducting Pharmacoeconomic Evaluations—Retrospective and Claims Database Studies. Value Health 1999;2:82-7.
6. Motheral B, Brooks J, Clark MA, et al. (2003) A checklist for retrospective database studies--report of the ISPOR Task Force on Retrospective Databases. Value Health 6: 90-97.
7. Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. Pharmaceopidemiol Drug Safety 2010;19:858-68.
8. Suissa S. The case-time-control design. Epidemiology 1995;6:248-53.
9. Ray, Wayne. Evaluating Medication Effects Outside Clinical Trials: New-User Designs. Am J Epidemiol 2003;158:915-920.
10. Hall GC (2009) Validation of death and suicide recording on the THIN UK primary care database. Pharmacoepidemiol Drug Saf 18: 120-131
11. Schneeweiss S, Patrick AR, Sturmer T, Brookhart MA, Avorn J, Maclure M, Rothman K, Glynn RJ. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. Medical Care 2007;45:S131-42.