Orphan Therapies: Making Best Use of Postmarket Data

Judith C. Maro, Ph.D.[1], Jeffrey S. Brown, Ph.D.[1], Gerald J. Dal Pan, M.D., M.H.S.[2], and Lingling Li, Ph.D.[1]

[1]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA; [2]Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD, USA

BACKGROUND: Postmarket surveillance of the comparative safety and efficacy of orphan therapeutics is challenging, particularly when multiple therapeutics are licensed for the same orphan indication. To make best use of product-specific registry data collected to fulfill regulatory requirements, we propose the creation of a distributed electronic health data network among registries. Such a network could support sequential statistical analyses designed to detect early warnings of excess risks. We use a simulated example to explore the circumstances under which a distributed network may prove advantageous.

METHODS: We perform sample size calculations for sequential and non-sequential statistical studies aimed at comparing the incidence of hepatotoxicity following initiation of two newly licensed therapies for homozygous familial hypercholesterolemia. We calculate the sample size savings ratio, or the proportion of sample size saved if one conducted a sequential study as compared to a non-sequential study. Then, using models to describe the adoption and utilization of these therapies, we simulate when these sample sizes are attainable in calendar years. We then calculate the analytic calendar time savings ratio, analogous to the sample size savings ratio. We repeat these analyses for numerous scenarios.

KEY RESULTS: Sequential analyses detect effect sizes earlier or at the same time as non-sequential analyses. The most substantial potential savings occur when the market share is more imbalanced (i.e., 90 % for therapy A) and the effect size is closest to the null hypothesis. However, due to low exposure prevalence, these savings are difficult to realize within the 30-year time frame of this simulation for scenarios in which the outcome of interest occurs at or more frequently than one event/100 person-years.

CONCLUSIONS: We illustrate a process to assess whether sequential statistical analyses of registry data performed via distributed networks may prove a worthwhile infrastructure investment for pharmacovigilance.

INTRODUCTION

Developing evidence to assess the safety and effectiveness of orphan therapeutics is challenging.[1,2] Sample sizes are unusually small, and appropriate comparison populations are difficult to identify. Pre-licensure activities are notably limited, necessitating extensive postmarket pharmacovigilance programs. While pharmacovigilance tools like the U.S. Food and Drug Administration (FDA)'s pilot Mini-Sentinel System[3] have data for over 125 million persons, that may be inadequate to study rare diseases when complete capture of every affected patient is desired. A product-specific registry is an alternative postmarket data collection mechanism that features more targeted patient capture and may be more appropriate in the rare-disease context.

However, product-specific registries may still introduce data fragmentation. They can become information silos, delaying the ability to answer important public health questions on the natural history of the disease, class-wide effects, or comparative benefits and risks.[4,5] One remedy is to create a distributed electronic health data network,[6–8] which can increase sample size without centralizing or combining patient-level data.

Distributed networks can be used to study treatments for both rare and common diseases, and particularly, safety and effectiveness outcomes that occur at varying frequency in the treated population.[3,9–11] The Mini-Sentinel System is such a network, comprised of data sources that represent the US commercially insured population. Here, we propose the same infrastructural model—a distributed electronic health data network—with registries as the source databases to optimize postmarket surveillance of special populations, such as those affected by rare diseases.

One advantage of distributed electronic health data networks is their ability to use data from multiple sources without requiring those sources to share data with each other. Previously, pharmacoepidemiologists had to pool data via multi-site database studies[12–14] or perform meta-analyses[15–18] to strengthen support for statistical inference. Another advantageous feature of distributed networks is their ability to support sequential statistical analyses.[19–24] Traditional statistical inferences from a non-sequential study (i.e., those made after full enrollment and complete follow-up) require years. Implementation of sequential statistical analyses can generate early warnings of emerging risks and/or benefits.

Development of the infrastructure to support sequential statistical analyses on a distributed network has substantial front-end cost. However, this investment may be justified if inferences via sequential analyses can be made notably sooner than those via non-sequential and/or meta-analyses. This time saved may translate into better patient care and consequent health or monetary savings, because physicians will gain knowledge of benefit-risk tradeoffs sooner. It is important for decision makers working with limited resources to understand the scenarios that would make sequential surveillance, and by extension, the formation of a distributed network among registries, worthwhile.

We work through a simulated example to explore these scenarios. Specifically, we estimate the time needed to identify a signal of excess risk for competing orphan therapeutics under a variety of assumptions. We compare the time needed for a sequential study and a non-sequential study, and search for circumstances that save time. Then we discuss the costs and benefits of building such a network. We offer a generalized framework for decision makers who are considering investing in a targeted distributed network infrastructure for pharmacovigilance, which may be most appropriate in a rare disease context when complete coverage of the patient population is desired.

BACKGROUND

Distributed Networks. A distributed network is a system that allows secure remote analysis of discrete data sets held by separate institutions. Periodically, each institution generates extracts of its data and stores them in a separate, dedicated, and firewalled location at the institution. These extract databases remain under the institution's direct control, mitigating legal, proprietary, privacy, and security concerns with respect to dealing with privately held, protected, and identifiable patient data.[6–8,25] The extract databases adhere to a common data model, including identical file structures, data fields, and coding systems.

To use a distributed network, authorized users with appropriate credentials first send a standardized executable computer program to the separate institutions (i.e., "query" the data sets). Each institution responds to this query by choosing to either execute the program onsite and report institution-specific summary-level results (i.e., not patient-level data) or to opt out. The initiator collects the institution-specific summary-level results and aggregates them for analysis.

In the context of multiple product-specific registries, the extract databases would be subsets of data from the registries, and therefore the participating institutions would typically be the sponsors of competing orphan products. The data fields to be collected and recorded would be guided by the goals of drug safety surveillance and set during the protocol development process, as these data form the observational database for product-specific studies. Such a protocol could also detail policies and procedures for designing queries and performing analyses.

Sequential Statistical Analyses. In this hypothetical proposal, authorized analysts would perform sequential database surveillance, sometimes referred to as active surveillance. Sequential database surveillance is a near real-time sequential statistical approach to evaluate pre-specified exposure-outcome pairs using data that are frequently updated, often quarterly or biannually.[26,27] The goal is to generate early warnings of some pre-specified effect via interim tests of data as they accumulate. These methods require investigators to set a stopping boundary, or a way to interrupt surveillance through "signaling." The shape of this boundary dictates the likelihood of signaling at various interim tests of the hypothesis and determines some of the tradeoffs between power (i.e., sample size) and the timeliness of signal detection. A statistical "signal" is detected when a stopping boundary is reached, adjusting for the multiple testing inherent in the frequent looks at the data.

Simulated Example—Homozygous Familial Hypercholesterolemia. Recently, the FDA licensed two therapeutics for homozygous familial hypercholesterolemia, a serious condition that leads to early cardiovascular morbidity and mortality.[28] Although the literature describes this condition as occurring in one in one million persons,[29] manufacturers of these therapeutics believe that as many as 3,000 Americans might currently be affected.[30] Pre-licensure clinical trial data suggest various hepatotoxicity events could be associated with both therapeutics as a consequence of their mechanism of action, which increases hepatic fat.

The FDA required both sponsors to create postmarket product-specific exposure registries that function as long-

term prospective observational studies. Each sponsor's registry will collect data only for patients using its own product. The sponsors will enroll patients globally and perform analysis for 10 years after the last patient enrolls. Enrollment is voluntary, i.e., access to the therapy is not conditioned on participation in the registry. Recent experience suggests that distributed electronic health data networks can be used to study class effects or comparative analyses;[31,32] the latter is the study question we simulate. We emphasize that we chose this simulated example to motivate a more general question: how to make best use of postmarket data that accumulate in product-specific registries, which are the preferred data sources when complete patient capture of the population of interest is desired?

METHODS

First, we perform sample size calculations for sequential and non-sequential statistical studies aimed at comparing the incidence of hepatotoxicity following initiation of each therapy. Then, using models to describe the adoption and utilization of two newly licensed therapeutics for homozygous familial hypercholesterolemia, we simulate when these sample sizes are attainable in calendar years. Using these simulations, we calculate the analytic calendar time savings ratio, which describes the proportion of analytic calendar time saved if one conducted a sequential study as compared to a non-sequential study. We repeat these analyses for numerous scenarios.

Conditional Sequential Sampling Procedure. We calculate sample size for sequential and non-sequential statistical analyses using the Conditional Sequential Sampling Procedure (CSSP).[33] The CSSP is a group sequential analysis, meaning that a new hypothesis test is performed whenever a designated "group" of information arrives. We set the group size to the average number of events we would expect to observe quarterly. For example, if we expect 40 events over a 10-year timeframe, then the group size would be one. We assume that new data are available every 6 months, i.e., a biannual update frequency.

We test the null hypothesis that there is no difference in event rates between the two products and use a two-sided overall type I error of 0.05 (i.e., we allow a false positive result to occur 5 % of the time). Because we perform multiple hypothesis tests, we must divide up the potential for a false positive at any given test (i.e., we have to "spread" the 0.05 allowance over multiple tests). We use a quadratic error spending function, and have a smaller allowance for false positives early in the process, when data are sparser.

To allow for direct comparisons of sample size (i.e., events) across different effect sizes, we hold statistical power constant at 90 %. More statistical information is required to achieve the desired statistical power whenever effect sizes are more modest. Given the calculated sample sizes, we then estimate when these sample sizes are attainable using stochastic agent-based models.

Agent-Based Models. Agent-based models describe dynamic person-level activities.[34] We simulate a 3,000-person patient pool, representing the upper estimate of the affected patient population. At each time step, a portion of the patient pool adopts one of the competing therapies. We assume that 25 % of this pool adopts neither therapy. We model the patient's likelihood of adoption using a well-known model in the innovation diffusion literature.[35] Once they adopt, they contribute exposed person-time to surveillance for the duration of their treatment plus a 30-day extension period. This time frame is when we deem them to be at risk of experiencing the outcome of interest in relation to the drug exposure. We assume that 20 % of patients are lost to follow-up for both therapies, and that the losses occur within the first 6 months of adoption. While on either treatment, these patients experience outcomes of interest at pre-determined rates and are then censored. We do not model switching behavior among products. In the base model, the adoption parameters are calibrated to data reported by the sponsors. Specifically, 250–300 new users adopt within the first year that products are available.[30] Full details of the agent-based model are available in the electronic online appendix.

We run the base model and vary the rates of the outcome of interest, the market share of the competing therapeutics, and the effect sizes in the simulation. We perform these simulations with the following incidence rates: one event/100 person-years (i.e., "*common*" per the Council for International Organizations of Medical Sciences[36]), and one event/1000 person-years (i.e., "*infrequent*"). We vary the market share of the more widely adopted therapy from 0.5 to 0.9, assuming therapy A is more widely adopted. Finally, we vary the effect sizes to include incidence rate ratios (IRRs) that range from 10 to 0.1 when comparing therapy A to therapy B. In other words, we model instances when each therapy has elevated levels of the outcome of interest. We run each "setting" 1,000 times, collecting information on exposed time and outcomes for both therapies for a 30-year run of the simulation.

Confounding Adjustment for Heterogeneous Patient Populations. In our base analyses, we model adoption and outcome patterns assuming a homogeneous population for simplicity. We then relax this assumption, allowing the

adoption and outcome patterns to be affected by a binary confounder, and adjust via stratification. Confounding adjustment in distributed networks has been discussed elsewhere.[37–39]

RESULTS

We show a subset of our results in Table 1. The sample size savings ratio—defined as the (non-sequential sample size – sequential sample size)/non-sequential sample size—illustrates the potential sample size advantages of a sequential analysis over a non-sequential analysis. These values depend on the scenario being examined (i.e., IRR, market share, rate of outcome of interest, statistical power, type I error), as well as the chosen group size and error

spending function. The sample size savings ratios are ceilings that can be achieved and are illustrated in the upper panel of Fig. 1. These ratios are calculated independent of the calendar time necessary to achieve them.

Of more practical interest are the median lengths of surveillance in both sequential and non-sequential settings, and the associated analytic calendar time savings ratios, which are illustrated in the lower panel of Fig. 1. If a sample size could not be attained in the 30-year run of the simulation, then the ratio on the Table is listed as "not determined."

In the Table, when one therapy has a markedly higher frequency of the outcome than the other, both the savings ratios and the median length of surveillance are lowest. For more modest differences in risk, the savings ratios increase, as does the median length of surveillance.

In Fig. 1, we illustrate the sample size savings ratios and the analytic calendar time savings ratios for the two

Table 1. Comparison of Analytic Calendar Time Savings Ratio for Multiple Scenarios

Outcome rate of interest (Events/person-years)	IRR* of therapy A to therapy B	Market share for therapy A	Sample size savings ratio†	Median length of sequential surveillance (Calendar years)	Median length of non-sequential surveillance (Calendar years)	Analytic calendar time savings ratio‡
1/100	10	0.9	0.00	6.5	6.5	0.00
		0.8	0.00	4.5	4.5	0.00
		0.7	0.00	4.0	4.0	0.00
		0.6	0.00	3.5	3.5	0.00
		0.5	0.00	3.5	3.5	0.00
	5	0.9	0.23	12.5	16.0	0.22
		0.8	0.06	9.0	9.5	0.05
		0.7	0.00	7.5	7.5	0.00
		0.6	0.20	5.5	6.5	0.15
		0.5	0.14	6.0	6.5	0.08
	2.5	0.9	0.31	>30.0	>30.0	N/D
		0.8	0.33	24.5	>30.0	N/D
		0.7	0.25	20.0	26.0	0.23
		0.6	0.31	16.5	23.0	0.28
		0.5	0.28	16.0	21.5	0.26
	0.4	0.9	0.41	>30.0	>30.0	N/D
		0.8	0.36	21.5	>30.0	N/D
		0.7	0.34	17.5	25.5	0.31
		0.6	0.33	16.0	23.0	0.30
		0.5	0.33	16.0	23.0	0.30
	0.2	0.9	0.36	10.0	14.5	0.31
		0.8	0.05	8.5	9.0	0.06
		0.7	0.26	6.0	7.5	0.20
		0.6	0.23	5.5	7.0	0.21
		0.5	0.17	5.5	6.5	0.15
	0.1	0.9	0.00	6.0	6.0	0.00
		0.8	0.00	4.5	4.5	0.00
		0.7	0.00	4.0	4.0	0.00
		0.6	0.00	3.5	3.5	0.00
		0.5	0.00	3.5	3.5	0.00
1/1000	10	0.9	0.27	>30.0	>30.0	N/D
		0.8	0.26	20.5	27.5	0.25
		0.7	0.13	19.0	21.5	0.12
		0.6	0.19	15.0	17.5	0.14
		0.5	0.08	16.0	17.0	0.06
	0.1	0.9	0.38	>30.0	>30.0	N/D
		0.8	0.30	18.5	25.0	0.26
		0.7	0.10	18.0	19.5	0.08
		0.6	0.18	15.0	17.5	0.14
		0.5	0.08	16.25	17.5	0.07

*IRR, incidence rate ratio
†Sample size savings ratio is calculated as (non-sequential sample size – sequential sample size)/non-sequential sample size. The units for sample size are the median number of events to end surveillance
‡ Analytic calendar time savings ratio is calculated as (median length of non-sequential surveillance – median length of sequential surveillance)/median length of non-sequential surveillance
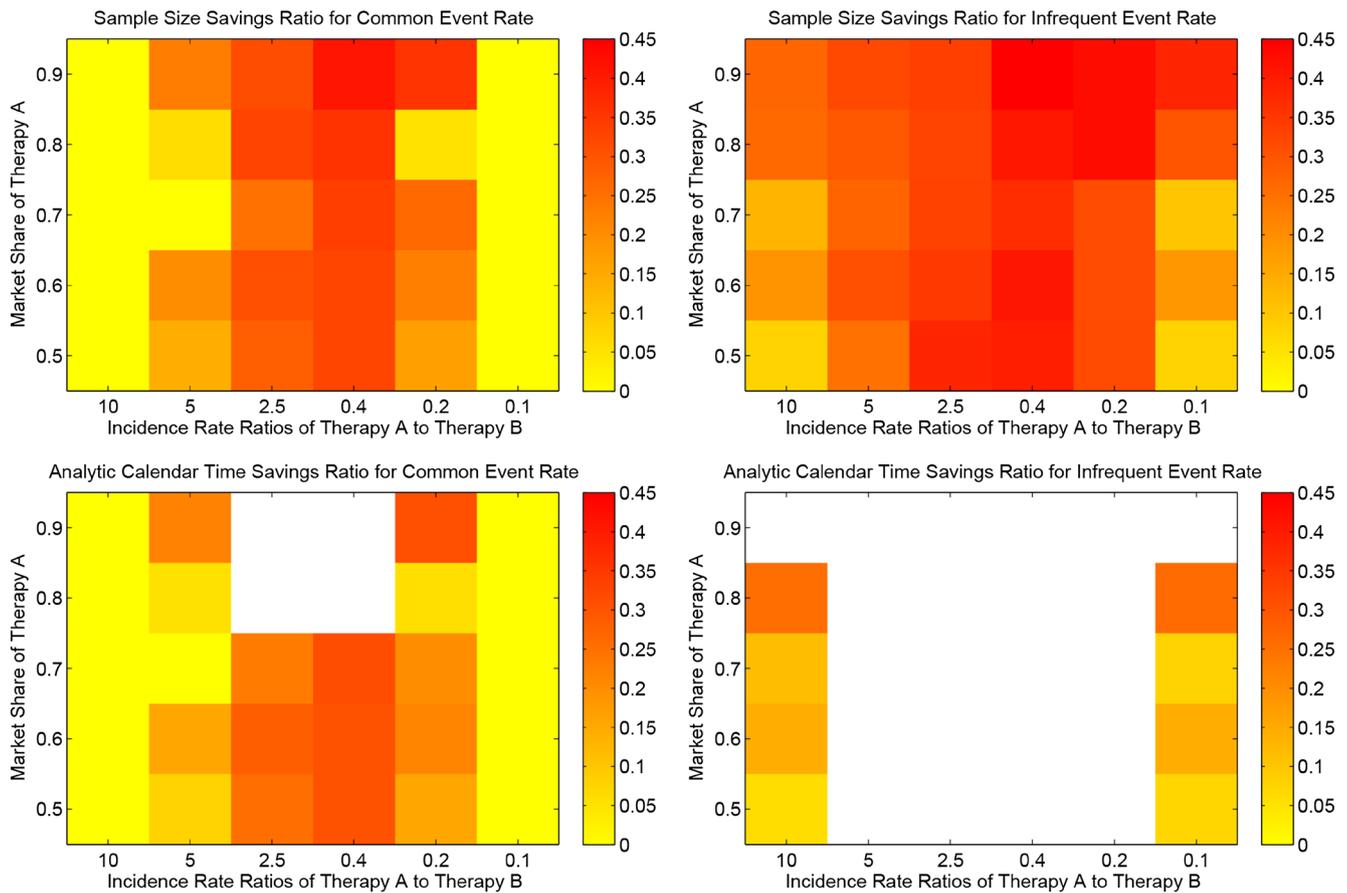
Figure 1. Sample size savings ratios and analytic calendar time savings ratios for common and infrequent outcomes of interest. The bar to the right represents the scale for savings ratios.

outcome rates of interest. The left panel illustrates these ratios for the homogeneous patient population with a *common* outcome of interest. For reference, a common outcome observed in clinical trials was elevated liver enzymes at > 3x the upper limit of normal. The right panel reflects *infrequent* outcomes of interest, which were unobservable in the clinical trial.

With common outcome rates, there were no savings of the sequential model as compared to the non-sequential model for the IRRs of 10 and 0.1. In these cases, it was possible to attain the non-sequential sample size at the very first group sequential test. For the remaining effect sizes, the most substantial savings were possible (i.e., the sample size savings ratios were largest) when the market share was more imbalanced (i.e., 90 % for therapy A) or the effect size was closest to the null hypothesis. However, comparing the upper panel to the lower panel, these savings were often not practically achievable, because the time required to complete surveillance is longer than the 30-year run of our simulation model.

Heterogeneous Patient Population with Confounding Adjustments. We performed the same series of analyses

on a stratified patient population. The sample size savings ratios were not greatly affected.

Sensitivity Analysis for Statistical Power. We relaxed statistical power from 90 to 80 % to determine whether more scenarios could complete surveillance within the 30-year timeframe. We obtained smaller sample size savings ratios from easier-to-achieve power targets. However, many analyses were still not achievable within 30 years.

DISCUSSION

Our intent with this hypothetical example was not to focus on these particular therapies, but instead to illustrate a process to assess whether sequential statistical analyses of registry data performed via distributed networks may prove a worthwhile pharmacovigilance infrastructure investment. While sequential analyses can detect safety signals earlier or at the same time as non-sequential analyses, often these savings cannot be realized because the surveillance time required is intolerably long, underscoring the difficulty in monitoring very low exposures. These difficulties are best

illustrated by comparing the upper-right to the lower-right portion of Fig. 1. With an *infrequent* outcome rate, sequential analysis reduces sample size requirements considerably, but IRRs of 5, 2.5, 0.4, and 0.2 cannot be detected within 30 years in either sequential or non-sequential analyses.

For *common* outcome rates, a smaller group size, enabled by more frequent data updates (i.e., quarterly), could notably improve the relative performance of sequential analyses. Although we do not show it here, we performed our analysis with various group sizes and error spending functions. We chose a group size and a quadratic error spending function that we believed would match the way data would arrive.

Limitations. We make several simplifying assumptions in this simulated example. First, in the absence of historical data, we assume a specific adoption and diffusion function for these therapies that may not reflect real-world adoption patterns. Second, we set the upper estimate of the affected patient population at 3,000. We would not have been able to detect any effect sizes at the more conservative estimate of 300 patients. Third, we assume that the utilization patterns and discontinuation rates observed in the clinical trial are generally representative of this population. Fourth, when patients discontinue one therapy, we do not model switching to the alternative therapy, which may have artificially limited our overall sample size for the two products. Fifth, we do not model competition from newer entrants, which would presumably reduce sample size. Finally, we assume no exposure or disease misclassification because of reliance on primary data.

Disease-Based Registries. To truly eliminate data fragmentation among these specialized populations, a disease-based registry is required.[4,5] However, disease-based registries require substantial financial resources over long periods of time, and it is unclear whether a sustainable funding model exists. We believe the model we propose here is a plausible alternative, because it takes advantage of existing requirements for postmarket registries and uses these data for analysis without requiring manufacturers to share them with one another.

Outcomes of Interest and Prior Information. We do not specify particular hepatotoxic outcomes of interest in these analyses, because our aim was to build a general model. However, to use this simulation study to weigh the advantages of building a distributed network, one must first consider the likelihood that outcomes of interest occur at these rates (e.g., as or more frequently than one event per 100 person-years). If so, one must consider the prior evidence to suggest what comparative effect sizes are

possible. That is, is it possible that therapy A creates ten times as many events as therapy B? Finally, and most importantly, are the potential findings important enough to alter the risk-benefit balance of the therapy, and therefore worth pursuing?

Once decision makers answer these questions, they may then assess the benefits and costs of detecting these effect sizes early in a novel infrastructure model. If these data have already been collected in compliance with postmarket regulatory requirements, then the true costs are software development (the extracting software and the web-based portals for secure communication) and labor costs associated with assessing the results of biannual hypothesis tests. How such costs compare to the benefits of early warnings will determine the value of a targeted distributed network infrastructure for pharmacovigilance.

Corresponding Author: *Judith C. Maro, Ph.D.; Department of Population MedicineHarvard Medical School and Harvard Pilgrim Health Care Institute, 133 Brookline Avenue, 6th Floor, Boston, MA 02215, USA (e-mail: jmaro@mit.edu).*

REFERENCES

1. **Buckley BM.** Clinical trials of orphan medicines. Lancet. 2008;371(9629):2051–2055. doi:10.1016/S0140-6736(08)60876-4.
2. **Haffner ME.** Adopting orphan drugs—two dozen years of treating rare diseases. N Engl J Med. 2006;354(5):445–447. doi:10.1056/NEJMp058317.
3. **Platt R, Carnahan RM, Brown JS, et al.** The U.S. Food and Drug Administration's Mini-Sentinel program: status and direction. Pharmacoepidemiol Drug Saf. 2012;21(Suppl 1):1–8. doi:10.1002/pds.2343.
4. **Lionetti G, Kimura Y, Schanberg LE, et al.** Using registries to identify adverse events in rheumatic diseases. Pediatrics. 2013;132(5):e1384–1394. doi:10.1542/peds.2013-0755.
5. **Hollak CEM, Aerts JMFG, Aymé S, Manuel J.** Limitations of drug registries to evaluate orphan medicinal products for the treatment of lysosomal storage disorders. Orphanet J Rare Dis. 2011;6:16. doi:10.1186/1750-1172-6-16.
6. **Curtis LH, Weiner MG, Boudreau DM, et al.** Design considerations, architecture, and use of the Mini-Sentinel distributed data system. Pharmacoepidemiol Drug Saf. 2012;21(Suppl 1):23–31. doi:10.1002/pds.2336.
7. **Brown JS, Holmes JH, Shah K, Hall K, Lazarus R, Platt R.** Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. Med Care. 2010;48(6 Suppl):S45–51. doi:10.1097/MLR.0b013e3181d9919f.
8. **Maro JC, Platt R, Holmes JH, et al.** Design of a national distributed health data network. Ann Intern Med. 2009;151(5):341–344. doi:10.7326/0003-4819-151-5-200909010-00139.

9. **Baggs J, Gee J, Lewis E, et al.** The Vaccine Safety Datalink: a model for monitoring immunization safety. Pediatrics. 2011;127(Suppl 1):S45–53. doi:10.1542/peds.2010-1722H.

10. **Velentgas P, Bohn RL, Brown JS, et al.** A distributed research network model for post-marketing safety studies: the meningococcal vaccine study. Pharmacoepidemiol Drug Saf. 2008;17(12):1226–1234. doi:10.1002/pds.1675.

11. **Nguyen M, Ball R, Midthun K, Lieu TA.** The Food and Drug Administration's Post-licensure Rapid Immunization Safety Monitoring program: strengthening the federal vaccine safety enterprise. Pharmacoepidemiol Drug Saf. 2012;21(Suppl 1):291–297. doi:10.1002/pds.2323.

12. **Ray WA.** Improving automated database studies. Epidemiol. 2011;22(3):302–304. doi:10.1097/EDE.0b013e31820f31e1.

13. **Strom BL.** Methodologic challenges to studying patient safety and comparative effectiveness. Med Care. 2007;45(10 Supl 2):S13-5. doi:10.1097/MLR.0b013e318041f752.

14. **Schneeweiss S, Avorn J.** A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol. 2005;58(4):323–337. doi:10.1016/j.jclinepi.2004.10.012.

15. **Chowdhury BA, Seymour SM, Levenson MS.** Assessing the safety of adding LABAs to inhaled corticosteroids for treating asthma. N Engl J Med. 2011;364(26):2473–2475. doi:10.1056/NEJMp1104375.

16. **Berlin JA, Cepeda MS, Kim CJ.** The Use of Meta-analysis in Pharmacoepidemiology. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology.* Fifth. Chichester, West Sussex, UK: John Wiley & Sons; 2011:723-756. Available at: http://dx.doi.org/10.1002/9781119959946.ch40.

17. **Berlin JA, Colditz GA.** The role of meta-analysis in the regulatory process for foods, drugs, and devices. JAMA. 1999;281(9):830–834. doi:10.1001/jama.281.9.830.

18. **Temple R.** Meta-analysis and epidemiologic studies in drug development and postmarketing surveillance. JAMA. 1999;281(9):841–844. doi:10.1001/jama.281.9.841.

19. **Avery TR, Kulldorff M, Vilk Y, et al.** Near real-time adverse drug reaction surveillance within population-based health networks: methodology considerations for data accrual. Pharmacoepidemiol Drug Saf. 2013;22(5):488–495. doi:10.1002/pds.3412.

20. **Yih WK, Kulldorff M, Fireman BH, et al.** Active surveillance for adverse events: the experience of the vaccine safety datalink project. Pediatrics. 2011;127(Suppl 1):S54–64. doi:10.1542/peds.2010-1722I.

21. **Greene SK, Kulldorff M, Yin R, et al.** Near real-time vaccine safety surveillance with partially accrued data. Pharmacoepidemiol Drug Saf. 2011;20(6):583–590. doi:10.1002/pds.2133.

22. **Yih WK, Lee GM, Lieu TA, et al.** Surveillance for adverse events following receipt of pandemic 2009 H1N1 vaccine in the Post-licensure Rapid Immunization Safety Monitoring (PRISM) system, 2009-2010. Am J Epidemiol. 2012;175(11):1120–1128. doi:10.1093/aje/kws197.

23. **Tse A, Tseng HF, Greene SK, Vellozzi C, Lee GM.** Signal identification and evaluation for risk of febrile seizures in children following trivalent inactivated influenza vaccine in the Vaccine Safety Datalink project, 2010-2011. Vaccine. 2012;30(11):2024–2031. doi:10.1016/j.vaccine.2012.01.027.

24. **Lee GM, Greene SK, Weintraub ES, et al.** H1N1 and seasonal influenza vaccine safety in the Vaccine Safety Datalink project. Am J Prev Med. 2011;41(2):121–128. doi:10.1016/j.amepre.2011.04.004.

25. **Toh S, Platt R.** Is size the next big thing in epidemiology? Epidemiol. 2013;24(3):349–351. doi:10.1097/EDE.0b013e31828ac65e.

26. **Kulldorff M.** Sequential Statistical Methods for Prospective Postmarketing Safety Surveillance. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology.* Fifth. Chichester, West Sussex, UK: John Wiley & Sons; 2011:852-867. Available at: http://dx.doi.org/10.1002/9781119959946.ch46.

27. **Nelson JC, Cook AJ, Yu O, et al.** Challenges in the design and analysis of sequentially monitored postmarket safety surveillance evaluations using electronic observational health care data. Pharmacoepidemiol Drug Saf. 2012;21(Suppl 1):62–71. doi:10.1002/pds.2324.

28. **Kolansky DM, Cuchel M, Clark BJ, et al.** Longitudinal evaluation and assessment of cardiovascular disease in patients with homozygous familial hypercholesterolemia. Am J Cardiol. 2008;102(11):1438–1443. doi:10.1016/j.amjcard.2008.07.035.

29. **Goldberg AC, Hopkins PN, Toth PP, et al.** Familial hypercholesterolemia: screening, diagnosis and management of pediatric and adult patients: clinical guidance from the national lipid association expert panel on familial hypercholesterolemia. J Clin Lipidol. 2011;5(3 Suppl):S1–8. doi:10.1016/j.jacl.2011.04.003.

30. **Goodman M.** Aegerion Upbeat On Juxtapid Launch, Stares Down PCSK9 Threat. *The Pink Sheet Daily.* http://www.elsevierbi.com/publications/the-pink-sheet-daily/2013/4/30/aegerion-upbeat-on-juxtapid-launch-stares-down-pcsk9-threat. Published April 30, 2013. Accessed March 21, 2014.

31. **Toh S, Reichman ME, Houstoun M, et al.** Comparative risk for angioedema associated with the use of drugs that target the renin-angiotensin-aldosterone system. Arch Intern Med. 2012;172(20):1582–1589. doi:10.1001/2013.jamainternmed.34.

32. **Toh S, Platt R, Steiner JF, Brown JS.** Comparative-effectiveness research in distributed health data networks. Clin Pharmacol Ther. 2011;90(6):883-887. doi:10.1038/clpt.2011.236; 10.1038/clpt.2011.236.

33. **Li L.** A conditional sequential sampling procedure for drug safety surveillance. Stat Med. 2009;28(25):3124–3138. doi:10.1002/sim.3689.

34. **Epstein JM, Axtell R.** Growing artificial societies : social science from the bottom up. Washington: Brookings Institution Press; 1996.

35. **Bass FM.** A new product growth for model consumer durables. Manag Sci. 1969;15(5):215–227.

36. Council for International Organizations of Medical Sciences. Working Group VIII. *Practical aspects of signal detection in pharmacovigilance : report of CIOMS Working Group VIII.* Geneva: CIOMS; 2010.

37. **Toh S, Gagne JJ, Rassen JA, Fireman BH, Kulldorff M, Brown JS.** Confounding adjustment in comparative effectiveness research conducted within distributed research networks. Med Care. 2013;51(8 Suppl 3):S4–10. doi:10.1097/MLR.0b013e31829b1bb1.

38. **Toh S, Reichman ME, Houstoun M, et al.** Multivariable confounding adjustment in distributed data networks without sharing of patient-level data. Pharmacoepidemiol Drug Saf. 2013. doi:10.1002/pds.3483.

39. **Rassen JA, Avorn J, Schneeweiss S.** Multivariate-adjusted pharmacoepidemiologic analyses of confidential information pooled from multiple health care utilization databases. Pharmacoepidemiol Drug Saf. 2010;19(8):848–857. doi:10.1002/pds.1867.