Adverse Event Detection in Drug Development: Recommendations and Obligations Beyond Phase 3

Premarketing studies of drugs, although large enough to demonstrate efficacy and detect common adverse events, cannot reliably detect an increased incidence of rare adverse events or events with significant latency. For most drugs, only about 500 to 3000 participants are studied, for relatively short durations, before a drug is marketed. Systems for assessment of postmarketing adverse events include spontaneous reports, computerized claims or medical record databases, and formal postmarketing studies.

We briefly review the strengths and limitations of each. Postmarketing surveillance is essential for developing a full understanding of the balance between benefits and adverse effects. More work is needed in analysis of data from spontaneous reports of adverse effects and automated databases, design of ad hoc studies, and design of economically feasible large randomized studies. (*Am J Public Health.* 2008; 98:1366–1371. doi:10.2105/AJPH.2007.124537)

Jesse A. Berlin, ScD, Susan C. Glasser, PhD, and Susan S. Ellenberg, PhD

REPORTS OF DEVASTATING adverse events suffered by patients create public doubt about whether drugs are safe. Developing "safe" drugs presents a high hurdle, because every drug carries potential for harm ("risk"). Drug safety cannot be considered an absolute; it can only be assessed relative to the drug's benefits. At the time of marketing, however, the amount of information on benefits and risks, especially long term, is relatively small, and often based on highly selected populations with respect to age, comorbidities, use of concomitant medications, and other factors.

We discuss drug development and assessment of adverse events and offer recommendations for continued evaluation of benefits and harms after a medicinal product becomes marketed.

DRUG DEVELOPMENT PROCESS

The drug development process, from discovery to market, is long and costly.[1,2] Rigorous processes are in place during clinical trials that protect the safety of study participants and also ensure that collection of adverse event data is complete. This completeness, coupled with the randomized design, also helps develop an understanding of the benefits and side effects of a new medicine by strengthening the validity of the comparisons between the new drug and the comparator, which could be a placebo or an active therapy for the condition under study.

Preclinical Testing

Prior to being studied in humans, a drug candidate undergoes an extensive series of laboratory and animal tests to study possible therapeutic and adverse effects. Preclinical studies are also used to characterize the pharmacokinetics and pharmacodynamics of the drug, including absorption, distribution, metabolism, excretion, and persistence of pharmacological effects.

A preclinical evaluation of safety includes in vitro and in vivo studies in animals to search for unintended pharmacological and toxic effects at the whole-animal level and on specific organs and tissues. In addition, carcinogenicity and mutagenicity studies are conducted, along with specific tests of effects on cardiac rhythms. If results suggest the product can be used safely and may produce the desired beneficial effects, the stage is set for testing in humans. There is generally a low threshold for rejecting drugs for safety reasons; the assumption is that unfavorable preclinical results are predictive of human safety problems (although the validity of this assumption may be questionable). Most drug candidates, whether for safety concerns or insufficient potential for efficacy, will never complete the development process; only 1 of every 5000 to 10 000 compounds that enter preclinical testing will become approved for marketing.[3]

Application for Study and Clinical Testing

US law requires manufacturers to petition the US Food and Drug Administration (FDA) to allow the study of investigational drugs or of new indications or dosages for approved drugs. This process ensures that the FDA can make sponsors and investigators aware of potentially unsafe uses of drugs before studies in humans are initiated.

Traditionally, clinical testing of investigational drugs proceeds in a phased fashion. We describe the research program and refer to human participants in a clinical trial as *participants* and users of medicines outside of a research setting as *patients*.

Phase-1 studies evaluate the safety and pharmacology of a compound to determine a range of tolerable doses; preliminary pharmacodynamic data, involving small numbers of participants (20–100), are sometimes obtained. Phase 2 looks for initial indications of efficacy and more data on safety among somewhat larger numbers of participants (typically 100–500), as well as optimal dosage and method of drug delivery. Phase-3 studies are the final step in obtaining the primary evidence of efficacy and safety prior to seeking drug approval. These studies range widely in size (dozens to thousands of participants), depending on the prevalence of the condition being treated and the rate of the event of primary interest, and usually involve random assignment of participants to new treatment or "control" treatment in a blinded manner, to allow an unbiased comparison of both the efficacy outcome and the adverse event profiles in the treatment and control groups. These studies may test one or several doses of the compound.

Approval to market a drug often involves commitments by the sponsor to perform additional studies. These may include randomized or cohort studies that examine the benefits and potential harms of the new drug in a different population or under somewhat different conditions from those originally studied, or special monitoring in a high-risk population—often by establishing a registry of such patients. These studies are sometimes designed to define more carefully an identified signal not adequately quantitated in the premarketing setting.

One such example involves a postmarketing commitment to conduct a prospective, multicenter registry of 4000 adult patients with psoriasis treated with infliximab in the United States. In some of the clinical trials of drugs in this class, including infliximab, increased incidences of some cancers and serious infections have been observed in other populations, such as individuals with rheumatoid arthritis. Given these prior observations, a registry was initiated to characterize and assess the incidence of malignancies and serious infections as well as other adverse events of interest in individuals treated for psoriasis.[4,5] In addition to these measures, health authorities often require the submission of a risk management plan with the marketing application to improve detection or to mitigate potential harms for new medicinal products.

LIMITATIONS ON DETECTING RISK

In phase-3 programs that enroll 3000 participants or more, even for adverse events occurring at a frequency of 1 in 1000, at least one such event will probably be observed. However, observing an adverse event is not equivalent to identifying that event as an adverse reaction to the drug. To do the latter, one needs to show that the rate of the event in those treated with the drug is greater than the rate in the control group.

An elementary principle in the design of studies is that the number of participants needed to detect an increased rate of an adverse event depends on how confident one wants to be of identifying a risk of a given magnitude (i.e., the desired statistical power). For example, with 1000 participants, we have a greater than 80% chance of detecting a true doubling in the rate of an adverse event from 5% to 10%, but we have far less confidence (only a 17% chance) in detecting a doubling from 1% to 2%. We would need to study at least 50 000 participants to achieve 80% power of detecting a doubling of a 0.1% event rate (Table 1). Although such an event rate seems very small, if the treatment is used by millions of individuals, the number of excess adverse events resulting from an increase to 0.2% will be substantial.

Thus, premarketing studies of new pharmaceuticals cannot reliably detect rare, but potentially important, adverse events. Moreover, events that take time to be observed (i.e., that have a latency period) may not be seen in trials of relatively short duration, which are typical in a development program. Drugs are therefore, *as a rule*, made available for public use before rare but potentially serious reactions have been identified and their probability quantified. Moreover, the adverse event profile of the drug, which is usually well defined in relatively small, carefully controlled, premarketing studies, may not adequately reflect the profile that will emerge with widespread use after approval, for several reasons: (1) study participants may represent a somewhat healthier and select subset of all participants, (2) they may receive better care than "real-life" patients, (3) study drugs will (of necessity) be given for shorter durations in studies than in postmarketing use, and (4) neither concomitant medications administered in clinical trials nor comorbidities of study participants will represent all those possible outside the trial setting.

ADVERSE EVENTS AFTER MARKETING

Data on adverse events after marketing of the drug include

TABLE 1—Statistical Power (%) to Detect a Doubling of Adverse Event Rates in Clinical Studies of Drugs, by Sample Size

Sample Size	From 5% to 10%, %	From 1% to 2%, %	From 0.1% to 0.2%, %
1 000	82	17	5
5 000	>99	80	7
10 000	>99	>98	17
50 000	>99	>99	79

Source. Ellenberg.[6]
Note. A statistical power of 80% or higher is generally considered acceptable.

spontaneous case reports, computerized claims or medical record databases, and data collected in prospective postmarketing studies.[7] Such systems have been extensively reviewed.[8–16] Clearly, a comprehensive drug safety program also includes evaluation of other relevant clinical findings (e.g., laboratory test results, vital signs, cardiac or other specialized testing) that we do not address.

Spontaneous Reporting

Spontaneous reports refer to unsolicited reports of clinical observations originating outside of a formal clinical study that are submitted to drug manufacturers or regulatory agencies.[16] Some of the events will represent true adverse effects of treatment; many will be symptoms of the disease being treated, or coincidental events that are unrelated to disease or treatment.

The most important reports are either new (i.e., not included in the product label), rare, serious events associated with the drug's use, or recognized adverse events occurring at a higher than anticipated rate. Other reports may reflect medical errors, inappropriate dosing or other misuse of the drug, and product defects.[17] Spontaneous reporting systems can "signal" emerging problems and thereby have the potential for uncovering previously unknown adverse reactions. Because these reports are written by health care professionals whose clinical judgment is valued, the companies spend a great deal of time analyzing individual reports and any patterns underlying these reports.

The limitations of spontaneous reports include substantial and unquantifiable underreporting (thus, such systems do not produce accurate estimates of incidence for a given adverse event) as well as lack of verification of important clinical details.

Adverse events may be spontaneously reported at disproportionately high rates at various times in the drug's marketing life cycle. For example, health care professionals and patients are more inclined to report adverse reactions when a drug is newly introduced (Weber effect),[18] when the events are medically very significant, when the event occurs very close in time to the administration of treatment, or when negative publicity emerges, such as the increased number of cases of rotavirus vaccine–associated intussusception reported after the Centers for Disease Control and Prevention recommended suspension of the rotavirus vaccination program.[19]

Sophisticated statistical approaches to formalize the "signal generation" aspect of spontaneous reports, aimed at determining when a particular type of adverse event is reported disproportionately relative to other adverse events associated with a given drug, have been developed. Such systems, often using Bayesian statistical methods, are used and evaluated by safety reviewers employed by regulatory authorities and pharmaceutical companies, as well as by an increasing number of academic investigators.[20,21] These methods may be useful as automated searching tools, especially as the number of spontaneous reports increases.

Once a signal is detected, the correct course of action is usually not obvious. If the signal is compelling—that is, if the increase in risk seems very large and is consistent with the known mechanism of action of the drug, the attribution to the drug is certain, and the event in question is clinically significant—there may be a need to initiate action immediately. (For very severe events, immediate action may be warranted even if they are not obviously related to the drug's mechanism of action.) If the signal is suggestive but too preliminary for immediate action, a more rigorous follow-up investigation may be needed to support taking any action. The dilemma is that if action comes too late, more people may suffer harm; if too early, people may stop taking beneficial medications unnecessarily.

In some cases, preliminary analysis of adverse event rates may suggest an apparent risk, although ultimately that risk could be dismissed as artifactual. For example, preliminary review of data from the Vaccine Adverse Event Report System showed that higher rates of serious adverse events were reported for children who received a specific brand of recombinant hepatitis B vaccine; subsequent analysis from a retrospective cohort study showed that there was no difference in rates of adverse events for the 2 vaccines.[22] Signals from spontaneous reporting therefore need to be interpreted very cautiously.

Electronic Databases

Electronic databases are broadly classified as based on claims or on medical records. Claims-based databases, such as Medicaid–Medicare in the United States, are usually set up by health maintenance or other health insurance organizations, or by government programs, and contain useful information on reimbursable expenses (prescriptions, hospitalizations), often with diagnoses from a definable population of patients. Their use is limited by lack of clinical detail, necessitating the access of other sources of information, such as hospital charts, to obtain further information.[23] Such databases often provide little information on outpatient events, including deaths. Studies using these databases can take substantial time to complete if they involve obtaining information from the supplemental sources described, and therefore cannot usually provide a rapid "check" on a worrisome (but very possibly false) signal.[8]

Medical practice databases contain patient medical records and prospectively recorded information on medical events, such as prescriptions, previous history, diagnoses, and test results. One widely used database for pharmacoepidemiological research is the General Practices Research Database in the United Kingdom.[11] This database is a unique resource because it includes very detailed medical information, symptoms, and signs in a well-defined, representative, and stable population, and it is also validated (i.e., information on diagnoses and on prescriptions has been found to agree with that recorded on paper charts or provided by physicians). It is, however, limited with respect to exposures to recently marketed drugs and may be therefore better suited to studying older, well-established drugs or drug classes. Another limitation is the duration of patient follow-up, which tends to be only a few years.[23]

Medical practice data sets are not designed to collect specific safety information with clear definitions. For example, differences among physicians in how a particular clinical presentation is coded in an insurance claim can make it difficult to identify all cases of a particular end point

without also identifying spurious cases. In many databases, only the fact that a prescription was written and filled may be recorded. Whether the patient was actually taking the medication in the clinically relevant time period preceding the event cannot always be reliably determined; thus, assumptions about timing and adherence generally need to be made to infer a given patient's exposure status.

The recently enacted Food and Drug Administration Amendment Act (or FDAAA; Public Law 110-85, signed into law September 27, 2007) calls for the establishment, under the auspices of an independent foundation, of a database of health insurance claims data for 100 million people by July 2012. This database is to be used to generate signals for further investigation.

Ultimately, a comprehensive health care database, including all interactions between an individual and the health care system (outpatient and inpatient visits, laboratory and other diagnostic results, and prescriptions), could provide the necessary breadth and depth for assessing safety and effectiveness in actual clinical practice. Conceivably, specific additional measures of effectiveness (e.g., patient- or physician-reported measures of symptoms) might be included as supplemental items in the electronic data capture systems that would give rise to a truly comprehensive database.

Observational Studies

Observational studies (typically, cohort or case–control designs), which do not include an intervention, can provide considerable information on the probability of specific adverse events.

Cohort studies. Cohort studies evaluate individuals who have a certain condition (e.g., epilepsy) or receive a particular treatment (e.g., anti-epileptic drug) over time. Their experience may be compared with that of others who are not affected by the condition under investigation, or are exposed to medications other than the one of interest. *Registry* is another name for a cohort study, which may be disease based (e.g., epilepsy) or product based (e.g., individuals with any condition who are prescribed a particular anti-epileptic medication).

Cohort studies may or may not include a comparison group. For example, one could estimate the incidence of cancer in participants with rheumatoid arthritis, with or without comparing that incidence to the rate of cancer among those without rheumatoid arthritis. Similarly, one could estimate the risk of cancer in a cohort of participants with rheumatoid arthritis who are exposed to a particular rheumatoid arthritis medication, with or without studying participants who are exposed to other rheumatoid arthritis medications. Omitting a comparison group, however, can make it difficult to interpret rates observed in a single group, because there is no set reference rate determined by using similar methods of data collection for different groups.

Even when there is a comparison group, cohort studies are not as reliable for making comparisons as randomized controlled studies, because the 2 groups of participants in a cohort study may differ in ways other than in the variable under study. For example, if more severely ill rheumatoid arthritis participants are selectively prescribed a particular medication, their cancer risk may differ because of the medication or some other factor related to the severity of the underlying illness. Statistical "adjustment" for those differences may not be sufficient, because there may be important unknown or unmeasured selection factors that contribute to the observed outcomes. When adverse events, such as cancer, occur infrequently or develop slowly, large numbers of participants, long-term follow-up, or both are required.

Case–control studies. In case–control studies, information is collected retrospectively from "cases" (participants who already have a certain condition) and "controls" (those who do not have the condition). This design can assess whether certain characteristics are associated with the specific condition or adverse event being studied.

Case–control studies are most appropriate when the event rate is low,[23] because the number of participants required is far smaller than would be needed for a cohort study. This efficiency stems from identifying the cases after they occur and needing to study only a relatively small sample of noncases, rather than having to follow a large number of participants to observe development of cases in "real time." The most important limitation of case–control studies is that a statistical relationship between an exposure and an outcome does not necessarily mean that the exposure caused the event. For example, participants with gastrointestinal bleeding may tend to have received H2-receptor blocker drugs more commonly than controls without gastrointestinal bleeding. Because these drugs are used for the prevention of such bleeding, an explanation could be that participants exhibiting early, subtle signs of bleeding are preferentially prescribed these medications. Other types of biases in these studies can also arise.[24] Still, case–control studies are often the only feasible way to study the relationship between an exposure and a rare adverse outcome, and have been used to establish several important relationships, including those between use of diethylstilbestrol in mothers and vaginal cancer in their daughters, use of aspirin and Reye's syndrome, and use of oral contraceptives and thromboembolic events.[8]

Large Simple Safety Trials

Large simple safety trials are conducted in larger numbers of participants than registration studies (typically, many thousands). They are considered when a product is to be used widely and when it is important to ensure that the risks of severe adverse events (usually not assessable in typically sized registration trials) are sufficiently low. Study conduct is facilitated by use of broad eligibility criteria consistent with the expected target population. Streamlined trial entry procedures and minimal data collection are essential to permit large studies without placing unacceptable burdens on investigators.[12]

A trial that compared ibuprofen with acetaminophen for treatment of fever in children is an interesting example of how a large trial can be done simply. The study was conducted when ibuprofen, a nonsteroidal anti-inflammatory drug, was proposed for over-the-counter use in children; this was at a time when there was relatively little pediatric experience with ibuprofen and concerns about the safety of

antipyretics had been heightened by the association of aspirin with Reye's syndrome. Primary outcomes were hospitalization for events known to be associated with use of nonsteroidal anti-inflammatory drugs in adults (e.g., acute gastrointestinal bleeding, kidney failure, anaphylaxis) as well as Reye's syndrome. Results from this large controlled trial of more than 84 000 children showed no differences in the outcomes of interest[25] and provided sufficient reassurance to support the move to over-the-counter status.

RECOMMENDATIONS

The solutions to the challenges of postmarketing evaluation of drug safety are complex and will require highly collaborative interactions among the FDA, other health authorities, and industry regarding new product labels and the development of epidemiological and statistical methodology for detecting and interpreting adverse event signals.[26–30] Methods and resources for rapid collection of adverse event data and further study when necessary should be considered and specified before marketing.

To attain this goal, we recommend a variety of approaches that may need to be tailored to specific situations. For example, although all new drugs require careful monitoring, a drug that is the first to be used for a particular therapeutic purpose, that has a unique mechanism of action, or about which a safety question was raised during the development stage may need closer scrutiny. One approach involves more-systematic ascertainment of adverse events, which may include prospective follow-up of a defined cohort—either disease based or product based—such as is used in an observational study registry. This improved ascertainment could be complemented by use of claims databases, but that use should be accompanied by refinement of methodology for monitoring data and detecting signals and enhancement of availability of electronic data to facilitate rapid study, with careful attention to ethical and legal considerations. For products that are widely used in fundamentally healthy populations, even low rates of serious adverse events may affect large numbers of people. In such cases, if the adverse events are clinically important, expanded simple trials, preferably involving randomization, might be considered; all adverse events, or a specific set of adverse events of concern, would then be the primary or coprimary outcome measures.

All of the aforementioned measures will come at a price. The additional financial and staff resources at the FDA that the Institute of Medicine has recommended[31] will be difficult to achieve in the current government fiscal environment. The Food and Drug Administration Amendment Act provides additional funding to the FDA, but the resources proposed—for example, for the creation of the very large claims database—are likely to be inadequate and may require public–private partnerships. Such partnerships should fully engage all stakeholders—especially industry, academic, and regulatory scientists—in methodological discussions.

Greater drug company expenditures for more and larger studies may mean higher drug prices or possibly a declining rate of new drug development. Lastly, greater concern about the balance of a drug's benefits and potential harms may mean longer delays until new drugs become available, unless there is public acceptance of more-limited information at the time of approval coupled with commitments to expanding information and understanding about the drug's safety profile throughout its life cycle.

Balancing a drug's benefits against its potential harms is a complex task. Improved statistical methods are needed for automated signal detection and the ability to rapidly perform follow-up studies to confirm or refute signals. Such improvements will require expansion—and expanded availability—of databases containing information on exposures and outcomes of large numbers of individuals. More work is needed in (1) the analysis of data from spontaneous reports of adverse effects and claims databases; (2) the design of ad hoc studies to assess favorable and unfavorable drug effects in actual practice in an unbiased manner, with appropriate measures to reduce false positive findings; and (3) the design of economically feasible, large, randomized studies to identify small but serious risks that may have public health significance. Appropriate interpretation and communication of findings are also important. As the amount of information available increases in the media and on the Internet, the average person—even the average physician—may need help in understanding the practical implications of that information.

In conclusion, quantification of the potential for harm is a critical goal before and after marketing. The objective of drug development and subsequent postmarketing evaluation must be to provide information that allows physicians and patients to make educated decisions about the potential benefits and harms of a drug. It is also important to identify products whose benefits are outweighed by their harms; they should be used only for specific indications or in populations likely to benefit from them, or they should be removed from the market. It is also important to identify products with an unfavorable overall balance between benefits and harms: such products should be used only for specific indications or populations in which the benefits do not outweigh the harms or they should be removed from the market. ■

About the Authors

Jesse A. Berlin and Susan C. Glasser are with Johnson & Johnson Pharmaceutical Research & Development, LLC, Raritan, NJ. Susan S. Ellenberg is with the University of Pennsylvania School of Medicine Center for Clinical Epidemiology and Biostatistics, Philadelphia, PA.

Requests for reprints should be sent to Jesse A. Berlin, ScD, Johnson & Johnson Pharmaceutical Research & Development, 1125 Trenton-Harbourton Rd, PO Box 200, Mail Stop K-304, Titusville, NJ 08560 (e-mail: jberlin@prdus.jnj.com).

This essay was accepted December 21, 2007.

Contributors

J.A. Berlin and S.C. Glasser prepared the initial draft, with major substantive additions and revisions provided by all authors. All authors contributed equally to the conceptual development of the essay before any drafts were prepared.

Acknowledgments

We thank Patricia Molino and Audrey Phillips of Johnson & Johnson for their advice and counsel and Wendy Battisti of Johnson & Johnson Pharmaceutical Research & Development, LLC, for assistance in the preparation of this essay.

References

1. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. *J Health Econ.* 2003;22:151–185.

2. DiMasi JA. New drug development in the United States from 1963 to 1999. *Clin Pharmacol Ther.* 2001;69:286–296.

3. *Pharmaceutical Industry Profile 2006*. Washington, DC: Pharmaceutical Research and Manufacturers of America; March 2006.

4. Remicade [package insert]. Malvern, PA: Centocor Inc; 2007.

5. Department of Health and Human Services, Food and Drug Administration Remicade (infliximab) psoriasis approval letter 2006, BL 103772/5129. Available at: http://www.fda.gov/cder/foi/appletter/2007/103772s5129LTR.pdf. Accessed June 28, 2007.

6. Ellenberg SS. Safety considerations for new vaccine development. *Pharmacoepidemiol Drug Saf.* 2001;10:411–415.

7. Strom BL. How the US drug safety system should be changed. *JAMA*. 2006;295:2072–2075.

8. Strom B, ed. *Pharmacoepidemiology*. 4th ed. Chichester, England: John Wiley & Sons Ltd; 2005.

9. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med*. 1984;3:409–422.

10. Hasford J. Drug risk assessment: a case for large trials with lean protocols. *Pharmacoepidemiol Drug Saf.* 1994;3:321–327.

11. Rodriguez L, Gutthann S. Use of the UK General Practice Research Database for pharmacoepidemiology. *Br J Clin Pharmacol*. 1998;45:415–425.

12. Ellenberg SS, Foulkes MA, Midthun K, Goldenthal KL. Evaluating the safety of new vaccines: summary of a workshop. *Am J Public Health*. 2005;95:800–807.

13. Mitchell AA, Lesko SM. When a randomised controlled trial is needed to assess drug safety. The case of paediatric ibuprofen. *Drug Saf.* 1995;13:15–24.

14. Wood L, Coulson R. Revitalizing the General Practice Research Database: plans, challenges, and opportunities. *Pharmacoepidemiol Drug Saf.* 2001;10:379–383.

15. Jick SS, Kaye JA, Vasilakis-Scaramozza C, et al. Validity of the general practice research database. *Pharmacotherapy*. 2003;23:686–689.

16. Ahmad SR. Adverse drug event monitoring at the Food and Drug Administration. *J Gen Intern Med*. 2003;18:57–60.

17. Edwards R, Faich G, Tilson H. Points to consider: the roles of surveillance and epidemiology in advancing drug safety. *Pharmacoepidemiol Drug Saf.* 2005;14:665–667.

18. Weber JCP. *Epidemiology of Adverse Reactions to Nonsteroidal Antiinflammatory Drugs*. New York, NY: Raven Press; 1984:1–7. Rainsford KD, Velo GP, eds. *Advances in Inflammation Research*; vol 6.

19. Niu MT, Erwin DE, Braun MM. Data mining in the US Vaccine Adverse Event Reporting System (VAERS): early detection of intussusception and other events after rotavirus vaccination. *Vaccine*. 2001;19:4627–4634.

20. Almenoff JS, DuMouchel W, Kindman LA, Yang X, Fram D. Disproportionality analysis using empirical Bayes data mining: a tool for the evaluation of drug interactions in the post-marketing setting. *Pharmacoepidemiol Drug Saf.* 2003;12:517–521.

21. DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA Spontaneous Reporting System. *Am Stat*. 1999;53:177–202.

22. Niu MT, Rhodes P, Salive M, et al. Comparative safety of two recombinant hepatitis B vaccines in children: data from the Vaccine Adverse Event Reporting System (VAERS) and Vaccine Safety Datalink (VSD). *J Clin Epidemiol*. 1998;51:503–510.

23. Abenhaim L, Moore N, Begaud B. The role of pharmacoepidemiology in pharmacovigilance: a conference at the 6th ESOP Meeting, Budapest, 28 September 1998. *Pharmacoepidemiol Drug Saf.* 1999;8:S1–S7.

24. Csizmadi I, Collet J-P, Boivin J-F. Bias and confounding in pharmacoepidemiology. In: Strom B, ed. *Pharmacoepidemiology*. 4th ed. Chichester, England: John Wiley & Sons Ltd; 2005:791–810.

25. Lesko SM, Mitchell AA. An assessment of the safety of pediatric ibuprofen. A practitioner-based randomized clinical trial. *JAMA*. 1995;273:929–933.

26. Wood AJ, Stein CM, Woosley R. Making medicines safer—the need for an independent drug safety board. *N Engl J Med*. 1998;339:1851–1854.

27. Ray WA, Stein CM. Reform of drug regulation—beyond an independent drug-safety board. *N Engl J Med*. 2006;354:194–201.

28. Taylor MR. Protecting FDA's ability to protect public health. *Food Drug Law J*. 2006;61:805–808.

29. Curfman GD, Morrissey S, Drazen JM. Blueprint for a stronger Food and Drug Administration. *N Engl J Med*. 2006;355:1821.

30. Furberg CD, Levin AA, Gross PA, Shapiro RS, Strom BL. The FDA and drug safety: a proposal for sweeping changes. *Arch Intern Med*. 2006;166:1938–1942.

31. Committee on the Assessment of the US Drug Safety System, Baciu A, Stratton K, Burke SP, eds. *The Future of Drug Safety: Promoting and Protecting the Health of the Public*. Washington, DC: National Academies Press; 2006. Available at: http://www.iom.edu/CMS/3793/26341/37329.aspx. Accessed June 26, 2007.