

Can Current Electronic Systems Meet Drug Safety and Effectiveness Requirements?

Anne Holbrook, MD, PharmD, MSc, FRCPC^{1,2,3,4}, Paul Grootendorst, PhD^{1,3}, Don Willison, ScD^{3,4}, Charles Goldsmith, PhD^{3,4}, Rolf Sebaldt, MD, FRCPC^{2,3,4}, Karim Keshavjee, MD, MBA³. ¹ Faculty of Pharmaceutical Sciences, University of Toronto, ² Division of Clinical Pharmacology & Therapeutics, Department of Medicine, ³ Centre for Evaluation of Medicines, St Joseph's Healthcare, ⁴ Department of Clinical Epidemiology & Biostatistics, all at McMaster University.

Abstract

Background: Every health policy jurisdiction is endeavoring to enhance its ability to evaluate drug effectiveness, safety and cost in the real world (pharmacovigilance).

Methods: A nominal group consensus conference of stakeholders finalized data items deemed necessary for pharmacovigilance. Large administrative datasets (LADs), electronic health records (EHRs) and electronic patient registries (PRs), were investigated as sources of this information and for their vulnerability to methodologic bias. Health data privacy legislation and research guidelines were systematically reviewed for their constraint to linked data resource analyses.

Results: More than 129 data items were strongly recommended for routine pharmacovigilance. LADs had very complete information, but restricted to a small number of required data items. EHRs, especially with e-pharmacy links, offer by far the most complete set of health information domains but data entry completeness is highly variable. Adjustment methods for channeling bias are inadequate to mimic randomized trials. Anonymized, linked data held within a secure academic research environment, poses the least privacy concerns. **Conclusions:** Notwithstanding major technical, methodologic and privacy challenges, individual-level linkage of health data resources poses the best option for pharmacovigilance today. In future, drug regulators and reimbursement agencies should consider mandatory post-marketing randomized trials.

Introduction

Drugs are the most rapidly growing sector in healthcare in many countries but very little research addresses benefit: harm ratios for drugs in real patients in usual care and even fewer studies examine cost-effectiveness. The current Canadian system, which is similar to that in the United States, Australia and many parts of Europe, derives estimates of

benefit from (generally) small, well controlled pre-marketing trials. In contrast, evidence about harm comes largely from low quality post-marketing studies. Widely publicized withdrawals of commonly used drugs, such as cisapride, nefazodone, cerivastatin and rofecoxib because of fatal or near-fatal adverse reactions which were not publicly known based on the pre-marketing trials, have highlighted the urgent need for better post-marketing monitoring of drugs. To complicate matters, the federal jurisdictions which control market access in most countries, do not examine drug cost-effectiveness. A drug product which is not cost-effective particularly if alternatives exist, wastes resources which could be better utilized to fund other health interventions. While the most scientifically sound solution might be to mandate large, simple randomized trials that collect, analyze and report harms and costs as rigorously as benefit in typical patients, this occurs rarely.(1,2,3)

For many years, Canada and other developed countries have been contemplating drug regulation that would provide an initial conditional license to market for a finite period by the end of which time sufficient data on effectiveness, safety and cost-effectiveness could be gathered and analyzed to determine eventual listing and re-imbursement. Aside from large, simple randomized trials in usual care, more sophisticated analyses of observational data will be required. We use the term "pharmacovigilance" to mean the regular monitoring of medications for benefits, harms and costs in real clinical practice.

Before any system of pharmacovigilance can be established, several key questions need answers. These became the focus of this study, with four main research questions as follows:

1. Which data items are important to collect for pharmacovigilance?
2. Do current health data resources hold these items and can retrieve them?
3. Are certain biases inherent in observational data "fatal flaws" that cannot be surmounted?

4. Will privacy legislation prohibit the sharing of anonymized health information for pharmaco-surveillance purposes?

Methods

1. Which Data Items Are Important?

A formal nominal group consensus process was used to vet a list of information items eligible to be considered for optimal pharmaco-surveillance regulatory decisions.⁽⁴⁾ Two types of lists had been previously developed by the investigators in collaboration with other clinical pharmacology and health data experts - one for routine pharmaco-surveillance and one for “special situations”. The routine pharmaco-surveillance list contained those information items thought necessary to be regularly available on very short notice to inform all types of post-marketing benefit, harm and cost decisions (the “gold standard” information set). Special situations were defined as the less common clinical situation (e.g., outcomes related to a rare disease). The consensus group was comprised of experts from across the country representing primary care, clinical pharmacology, internal medicine, pharmacy, drug regulation, pharmaceutical industry, consumer concerns, health data management and analysis, epidemiology/methodology, health economics, drug formulary decisions. Inclusion of an item required endorsement by at least 66% of the consensus panel on the final round of voting. Included items formed the “gold standard information set”.

2. Data Availability

Once the gold standard information set was developed, we tested whether the information was available within our three leading health data resources. We examined linked large administrative datasets (LADs), electronic health records (EHRs) and a patient registry (PR). The LAD is a provincial resource commonly used in pharmacoepidemiologic research, the EHR database is a community primary care EHR network, and the PR was for a recent study of musculoskeletal prescribing. Each data resource was first characterized in terms of database type, size, data sources, coverage and availability to researchers and regulators. Then each resource was searched for data fields corresponding to each of the information items in the gold standard set (data field availability). Then the corresponding fields were each examined to see if any data existed within the field (data completeness).

3. Bias in Observational Studies

For this portion of the study, we focused on allocation (or channeling) bias, that is the prescribing of certain drugs for certain patients for reasons based on the physician-patient-system interaction but which are often invisible in health databases. This bias has previously been identified as a cause of misleading conclusions of even the highest quality observational studies.⁽⁵⁾ We investigated the ability of three primary methods - propensity score matching, linear regression and instrumental variables, to correct for these biases in a large observational database. The association between hormone replacement therapy (HRT) and cardiovascular events was used as the clinical example. Two landmark randomized trials (RCTs), the HERS study ⁽⁶⁾ and the WHI study ⁽⁷⁾, provided the “truth”, which markedly contrasted with a previous series of published observational studies. We used a large cohort of linked administrative data on patients with cardiovascular disease to form our observational dataset.

4. Health Information Privacy

Finally, health information privacy legislation (federal, 12 provinces and territories) and relevant national research guidelines were systematically reviewed for their regulation of the use of anonymized, individual-level linked health data that might be used for pharmaco-surveillance research and policy. Although large administrative datasets are already a compilation of linked data resources, we examined the implications of further linkages between, for example, LADs, EHRs, and registries across jurisdiction boundaries.

Results

1. Which Data Items Are Important?

After a round of voting in isolation, facilitated discussion of controversial items followed by a final round of individual voting again, the nominal consensus group finalized the “gold standard” information set. It contained 138 of the original 193 information items, ranging from 0 items of family history to 64 items related to current and past medication use. Unanimously endorsed were patient gender, birth year, weight, current diagnoses, hospital admission and discharge dates, discharge diagnoses, inpatient case weight, drug generic name, route, duration and reason prescribed, concurrent drugs, suspected current adverse reaction with severity and medication allergies including drugs and allergy type. In contrast, items such as voice biometric, fingerprint, next of kin, consent for organ donation, advance directives regarding life support and prior criminal convictions, were universally rejected.⁽⁸⁾

2. Data Availability

The linked LAD contained only a portion of the data fields required for optimal pharmacosurveillance with 54 of 138 (39%) of the recommended items but scored highly for data completeness (close to 100% compliance with data entry in the field) particularly for core demographics, hospitalization data and drug dispensing data. Linked LADs in general are limited by their variable coverage of populations, their lack of timeliness (for example, hospitalization data may be delayed for months), and their lack of coverage of diagnostic information, patient self-reported information, allergy and adverse reaction details, drugs prescribed (as opposed to dispensed) and non-drug cost information. The EHR database had more data fields with 90/138 (65%) data items accessible in structured data fields but suffered 3 main limitations. Depending on the data structure of the particular EHR, the required data could be in a structured field or in text, the latter being difficult to search systematically. The database design and organization of EHRs tends to be highly variable, often arcane and poorly designed for data extraction. Finally data completeness, except for common fields such as basic demographics, chief complaint, current medications, can be very poor. This reflects the primary purpose of the EHR which is as a charting technology used by hurried physicians for their own clinical review, rather than external data extraction. The patient registry database contained only 19 of the 138 desired fields (14%) and the restricted nature of the field contents (focused on a single disease area) rendered the data only suitable for same disease "special situation" pharmacosurveillance.

3. Bias in Observational Studies

Initial analysis of our observational dataset (N = 8716 women taking HRT) without adjustment, produced similar (misleading) results to previously published observational studies. Women taking HRT appeared to have a significantly lower risk of coronary events. Each of the three methods used to adjust for allocation bias did transform the association between HRT and cardiovascular events towards less benefit or towards harm but none allowed us to completely reproduce the results of the RCTs. Instrumental variables was the most promising method as it may adjust for unknown as well as known confounders.(9)

4. Health Information Privacy

All jurisdictions had freedom of information and personal privacy protection legislation in place that covers activities of the public sector. New federal legislation, the Personal Information Protection and Electronic Documents Act (PIPEDA), covers commercial activities across jurisdictions by

regulating the collection, use and disclosure of personal information including health information.(10) However, PIPEDA is not specific enough to deal with many of the special circumstances of health information, for example: a) personal health information is viewed as more sensitive than other types of information, b) multiple providers need access to confidential information to provide care, c) electronic communication is required for timeliness and patient safety, d) there is ambiguity about public versus private distinctions, e) personal information is held in multiple locations. For this reason, several provinces have enacted or plan to enact specific health privacy legislation.(11) This legislation, where deemed to supercede PIPEDA, makes express provision for consent requirements, for specific circumstances where consent can be waived, for a health data oversight body and for a formal complaints procedure. For example, in Ontario's recently approved Personal Health Information Privacy Act, procedures for accessing personal health information for research purposes discuss the need for a written proposal with scientific and community merit and consent, approval by a research ethics board, scrutiny by a health information custodian, ensuring compliance of the researcher. (12)

CIHR (Canadian Institutes of Health Research) have also updated their draft privacy guidelines for research. (13) These discuss the need to justify the data request, keep the collection of personal data to a minimum, ensuring a fair and informed consent, security of data and accountability for proper conduct of the research.(13) For purposes of secondary uses of large datasets for research, the collection of individual consent is problematic and highly likely lead to skewed and misleading results. (14,15,16) For this reason, such research is allowed if a research ethics board or similar body can be convinced that consent would not be feasible to obtain, prohibitively expensive or would seriously bias the results. In these circumstances, the approving body must agree that the research is in the public interest, is not harmful to the individuals involved, is of sound quality and the research group has adequate provisions for data security and confidentiality. The latter include a code of conduct, usually adapted from the Canadian Standards Association Model Code for the Protection of Personal Information, leadership by health professionals and provisions for duty of confidentiality on the part of researchers.

Discussion

There are several implications and limitations of this study. Firstly, a data resource suitable for routine,

rapid investigation of pharmacosurveillance issues would likely require linkage across LADs and EHRs. However, a comprehensive EHR which incorporates an e-prescribing link with inclusion of dispensing details, would cover virtually all required information. Although not directly studied in this project, our knowledge of each data resource would indicate that the technical feasibility of linkages across data resources, so-called meta-linkages, will be a major problem. Data standards (in some cases) and detailed data integration standards (in all cases) for drugs, diagnoses, diagnostic testing, demographics, etc. need further refinement. Secondly, we did not perform a comprehensive national or international search for databases potentially relevant to pharmacosurveillance. Before any country might embark on pharmacosurveillance, they would be wise to develop an inventory of health data resource availability, technical standards and content and to ensure that the data are properly managed, supported and protected. The co-ordination and support for many data resources, including registries, research-quality electronic health records and linked databases, frequently is grant-dependent and erratic at best. To avoid unethical collection, use and analyses of personal information as well as to allow the greatest public benefit of these resources, we have recommended that registration of all databases be required, as is currently unfolding for clinical trials. (17,18)

Third, further work is required to adjust for the effect of biases inherent in observational data. Without this, interpretations of cause and effect are flawed and may be fatal, and the databases will only be useful for retrospective evaluation of practices or for generating hypotheses for further rigorous, randomized controlled trial (RCT) testing. While we concentrated on the issue of allocation or channeling bias, missing data in these practice-based databases is a major problem as well. For the foreseeable future, RCTs will remain the "gold standard" for understanding benefit and harm.

Fourth, information privacy guidelines stipulate that very strong data security, data confidentiality and research competence be maintained by data custodian researchers. Health information privacy legislation and guidelines are evolving swiftly. Confusion still exists over the level of anonymization of information required to supercede the need for informed consent to analyze the data for research purposes. This confusion has significantly delayed progress on data integration issues. We have found in previous studies that anonymization of data does not equate with lack of privacy concerns on the part of patients or physicians.(19) Both groups are most comfortable with anonymized health information being held by

university researchers rather than private insurers, government or industry.

Conclusion

Most of the information required to monitor benefit, harm and cost-effectiveness of therapies does exist in currently available health database resources. Unfortunately it does not reside entirely in one database, so the difficult task of database integration would generally be required for adequate pharmacosurveillance. Biases inherent in observational data suggest that they should not be relied upon alone for major pharmacosurveillance decisions. Health information privacy legislation and practice still is evolving with respect to the use of anonymized, individual-level linked data for pharmacosurveillance purposes but already indicates that academic research organizations should take the lead in managing and analyzing such sensitive data resources.

Funding Source

This project was funded by a grant from the Health Policy Research Program of Health Canada. Project Number 3795-15-2001/4410013.

References

1. Wright JM. Why don't we initiate more large simple randomized controlled trials? *CMAJ* 2003; 169(11):1170-1171.
2. Sandercock P. What question can large, simple trials answer? *Cerebrovascular Diseases* 1998; 8(Suppl 2):30-36.
3. Hotopf M, Churchill R, Lewis G. Pragmatic randomised controlled trials in psychiatry. *British Journal of Psychiatry* 1999; 175:217-223.
4. Delbecq AL, Van de Ven AH, Gustafson DH. *Group Techniques for Program Planning: a guide to nominal group and delphi processes*. Green Briar Press, Wisconsin 1986.
5. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. *Stat Med* 1991; 10(4):577-581.
6. Hulley S, Grady D, Bush T et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *JAMA* 1998;280:605-13.
7. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML *et al*. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA* 2002;288:321-33.

8. Holbrook A, Goldsmith C, Grootendorst P, Willison D, Levine M, Gaebel K et al. What information is required for optimal pharmacovigilance? *Pharmacoepidemiol. Drug Saf* 12(Suppl 1), S138-S139. 2003.
9. Grootendorst P, Baggar J, Holbrook A, Levine M, Thabane L, Wilson A. Estimating treatment effectiveness using observational data: The case of HRT and coronary heart disease. *Can J Clin Pharmacol* 10[1], 26. 2003.
10. Privacy Commissioner of Canada. The Personal Information Protection and Electronic Documents Act. Available at: http://www.privcom.gc.ca/legislation/02_06_01_01_e.asp . 2000.
11. Cavoukian A. Health Information Protection Act: A major step in healthcare privacy. Available at: <http://www.ipc.on.ca/docs/051704ac.ppt> . May. 2004.
12. Legislative Assembly of Ontario. Bill 31 2003. An Act to enact and amend various Acts with respect to the protection of health information. Available at: http://www.ontla.on.ca/documents/Bills/38_Parliament/Session1/b031_e.htm. 2004.
13. Canadian Institutes of Health Research and Canadian Institute for Health Information. Personal information protection and electronic documents act. Questions and answers for health researchers. 2001. Ottawa, Public Works and Government Services Canada.
14. Tu JV, Willison DJ, Silver FL, Fang J, Richards JA, Laupacis A et al. Impracticability of informed consent in the Registry of the Canadian Stroke Network. *N Engl J Med* 2004; 350(14):1414-1421.
15. Wilson P. Commentary: Legal issues of data anonymisation in research. *BMJ* 2004; 328(7451): 1300-1301.
16. Peto J, Fletcher O, Gilham C. Data protection, informed consent, and research. *BMJ* 2004; 328(7447):1029-1030.
17. Tonks A. A clinical trials register for Europe. *BMJ* 2002; 325(7376):1314-1315.
18. McCray AT. Better access to information about clinical trials. *Ann Intern Med* 2000;133(8):609-614.
19. Holbrook A, Schmidt T, Thabane L. Patient views on electronic health information privacy. *Can J Clin Pharmacol* 11(1), e55. 2004.