Commentary

Measuring effectiveness of drugs in observational databanks: promises and perils

Eswar Krishnan and James F Fries

Division of Immunology, Department of Medicine, Stanford University, Palo Alto, CA, USA

Corresponding author: Eswar Krishnan (e-mail: eswar_krishnan@hotmail.com)

Received: 11 Dec 2003 Accepted: 20 Jan 2004 Published: 5 Feb 2004

Arthritis Res Ther 2004, **6**:41-44 (DOI 10.1186/ar1151)
© 2004 BioMed Central Ltd (Print ISSN 1478-6354; Online ISSN 1478-6362)

Abstract

Observational databanks have inherent strengths and shortcomings. As in randomized controlled trials, poor design of these databanks can either exaggerate or reduce estimates of drug effectiveness and can limit generalizability. This commentary highlights selected aspects of study design, data collection and statistical analysis that can help overcome many of these inadequacies. An international *meta*Register and a formal mechanism for standardizing and sharing drug data could help improve the utility of databanks. Medical journals have a vital role in enforcing a quality checklist that improves reporting.

Keywords: bias, cohort study, confounding, data banks, randomized controlled trial, rheumatoid arthritis.

Introduction

The decision to terminate the Women's Health Initiative (WHI) study, a randomized controlled trial (RCT) of hormone replacement therapy, and the public anxiety caused by the subsequent media publicity have put the hierarchy of evidence in epidemiology in the spotlight. Clinical medicine including rheumatology has also sometimes witnessed similar contradictions between the results of RCTs and observational studies. For example, RCTs indicated an efficacy for auranofin greatly exceeding that observed in observational studies or in clinical practice [1–3]. A meta-analysis of RCTs in 1990 [4] concluded that the efficacy of injectable gold salts, penicillamine and sulfasalazine did not differ from that of methotrexate in patients with rheumatoid arthritis. By contrast, and more in line with clinical experience, observational research reports indicated that courses of methotrexate were continued for much longer time than other agents, suggesting a better experience with this drug. Currently penicillamine and auranofin are almost never used for treating rheumatoid arthritis. Thus, some prominent clinical trials published in well-respected journals reached conclusions that were not validated in clinical practice.

The tools of observational epidemiology become critical 'when the perfectionist demands of clinical trials crash against the shoals of real-world conditions' [5]. There can never be an RCT for every single clinical question. Many important observations over the past two decades in rheumatology would not have been possible without observational research. Recognition of outcomes such as work disability, functional disability, and increased mortality rates in rheumatoid arthritis required long-term observational studies. More recently, the success of 'inverted pyramid' strategies for patients with rheumatoid arthritis has been documented [6]. The problem of gastrointestinal bleeding, ulcers, and obstruction associated with non-steroidal anti-inflammatory drugs was not apparent from RCTs but rather from long-term observational databases. Furthermore, the wide differences in toxicity between the non-steroidal anti-inflammatory drugs themselves were not demonstrated by the multiple RCTs.

Agreement between observational studies and RCTs increases our confidence that the effect of a drug is real [7]. The problems arise when there is discordance. Here we attempt to suggest reasons that results from RCTs

CONSORT = Consolidated Standards of Reporting Trials; HAQ = Health Assessment Questionaire; OMERACT = Outcome Measures in Rheumatology; RCT = randomized controlled trial; SF-36 = short form 36.

Table 1

Some limitations of randomized controlled trials
Patient selection limited by inclusion and exclusion criteria
Short time frame, as long-term clinical trials are ethically or logistically not possible
Differential drop-out patterns between arms of the trial
Statistically significant results might not necessarily be clinically significant, and vice versa
Surrogate markers such as joint tenderness might be suboptimal indicators of prevention of severe long-term outcomes such as radiographic destruction and work disability
Chance (bad luck) can lead to unbalanced groups
Inflexible dosage schedules
'Dose creep' from trial to clinic, rendering trial obsolete
Inability to identify rare adverse events
Hawthorne effect: patients in a study alter their behavior when they are told to be in the study
Design bias: randomized controlled trials might be designed to maximize the probability of a particular outcome, namely the superiority of the new drug

might sometimes differ from clinical practice and observational studies. The scientific rigor of the process of experimentation, the unflinching focus on the question 'Is drug A performing better than the comparator?' comes with a price, often poor generalizability. Results are not necessarily similar over the long term, in less selected populations or after 'dose creeps' have moved the doses used in clinical practice far from those of the RCT. The seldom-enumerated limitations of RCTs (Table 1) are such that short-term efficacy data from clinical trials must be supplemented with analyses of long-term effectiveness using observational research databases.

The Food and Drug Administration of the USA has introduced a requirement for post-marketing surveillance of newer drugs including biological agents; these are now being pursued by pharmaceutical industry, which has set up several surveillance databanks. In addition to monitoring for safety, these databanks collect information that has potential business applications. Such information includes drug dosage and drug switching patterns of the manufacturer's drugs as well as those of their competitors. It is not known to what extent these data are put to use for drug marketing. In addition, many of these databanks might not adhere to recommended standards for longitudinal studies [8,9].

Limitations of observational studies

One of the biggest criticisms of observational databanks results from potential bias in assignment of treatment by a physician. 'Confounding by indication' means that certain treatments are preferentially given to sicker patients and certain treatments preferentially to healthier patients. Thus, it is not uncommon for aspirin to be associated with increased risk for acute myocardial infarction in observational studies, because it is prescribed to those with a higher risk for coronary events. Many studies use statistical methods such as propensity scores that purportedly adjust for such bias. In this method of adjustment, the probability (propensity) of each patient's receiving a treatment is calculated on the basis of the collected information such as age, gender, and education. This propensity score can then be used for 'adjusting' for the effect of confounders by matching, by stratification, and by regression models. However, propensity scores might not adjust for unobserved covariates [10], especially if such covariates are not correlated with observed covariates. Furthermore, once data are collected, there is no fully satisfactory means to determine whether the adjustment is proportionate to the magnitude of the underlying confounding effect.

The second set of potential limitations results from patient self-selection. Very few databank studies report the number and characteristics of patients who were invited to be a part of the study but who eventually declined, whereas a lack of similar information in a report of an RCT might be considered unacceptable. Selection might also occur if patients or physicians receive financial incentives to complete questionnaires or enroll in studies (such as those studies sponsored by pharmaceutical industry). Another major issue is attrition or subject drop-out. Nonrandom drop-outs from studies are inevitable, and selective attrition of subjects can result in biased (often exaggerated) estimates of drug effectiveness. Very few databanks have formally reported the issue of attrition among their subject population.

The third set of limitations involves measurement of outcomes. Although questionnaire-based self-reports of outcomes might be considered to be as informative as

physician-based measures [11], the practicalities of measurement, analysis, and interpretation raise several issues. Longitudinal observational studies typically measure outcomes in specified intervals of 3, 6, or 12 months. Because the start and end of a drug course do not necessarily correspond to the measurement dates, difficulties can arise in correlating outcomes with drug courses. Thus, patient outcomes from drug courses shorter than the interval between measurements tend to be selectively lost. Because early termination of drug courses might indicate failure due to toxicity or inefficacy, the loss of information from these drug courses has the potential to bias the effectiveness estimates upwards. Besides, the absence of a 'washout period' in observational studies makes it difficult to disentangle the effects of current therapies from the residual effects of past therapies, particularly when the clinical half-life is varied and long [12].

Strengthening observational databanks

Observational studies need to be protocol-driven, with prospective data collection including the Health Assessment Questionnaire (HAQ) or its variants, short form 36 (SF-36), or a similar instrument at regular intervals [8,9]. Where drop-outs occur, careful documentation of the details (change in address, refusal, worsening health, and so on) of such losses is required. Rigor in data collection in observational databanks can and should be equivalent to that of RCTs.

We believe the criticism of unobserved bias has been overused. It should not be applied uncritically unless a specific, plausible unmeasured confounder is specified. Such potential confounders need to meet both of the two criteria of confounding, namely (1) association with outcome and (2) no association with the observed variables used for statistical adjustment. We agree with Moses [13] that it is important for the treating physician to record why the patient is being given the therapy selected. This information should be a powerful adjustment variable; 'arranging to collect it will call for imaginative thinking, experimentation, and patience, but it is an idea deserving much effort' [13].

Several steps could be taken within the existing framework for clinical research that can go a long way in improving the use of databanks. Many of the problems with observational studies can be minimized with careful planning in advance of the study. Ideally the subjects in longitudinal databanks should be truly representative of the population. Short of that, a databank should include all consecutive patients observed at the databank center.

We propose an international online registry for observational databanks similar to the *meta*Register of Controlled Trials (*m*RCT; http://www.controlled-trials.com/mrct/, accessed 10 January 2004). All the databanks in such a register should meet certain minimum methodological standards such as those proposed by the Outcome Measures in Rheumatology (OMERACT). This register could collate the data collection protocols and list of publications from each member databank and serve as a convenient reference for publications. This register would also help the users to be certain that they are aware of all the observational evidence relevant to a particular question, avoid duplication of effort, and encourage collaboration.

Patients who participate in databanks do so primarily on the basis of altruism. Patients trust their physicians to use their information for the greatest good of all others with the same disease. Although researchers who obtain funding and collect data deserve to have credit in terms of primacy and publications, data more than, say, 5 years old could very well be shared. Currently such informal data sharing exists through academic networking but the potential is probably not fully used. Research organizations such as the National Institutes of Health and the Centers for Disease Control have placed large amounts of data online, ready to be downloaded. There is little reason why similar sharing of data from rheumatic disease databanks for non-commercial purposes could not be phased in over time.

Medical journals have a key role in enforcing quality standards on reporting observational studies. Unfortunately, journals do not explicitly insist on the guidelines such as those by OMERACT. Providing checklists of reporting requirements similar to the CONSORT (Consolidated Standards of Reporting Trials) checklist for RCTs [14] would streamline the reporting of drug effectiveness data from observational studies.

Patient databanks are here to stay. Our plea here is for methodologically sound observational studies to raise the bar in the performance of clinical research.

Competing interests

None declared.

Acknowledgements

This work was supported by grant AR43584 from the National Institutes of Health to the Arthritis, Rheumatism and Aging Medical Information Systems (ARAMIS).

References

1. Menard HA, Beaudet F, Davis P, Harth M, Percy JS, Russell AS, Thompson JM: **Gold therapy in rheumatoid arthritis. Interim report of the Canadian multicenter prospective trial comparing sodium aurothiomalate and auranofin.** *J Rheumatol Suppl* 1982, **8:**179-183.
2. Bombardier C, Ware J, Russell IJ, Larson M, Chalmers A, Read JL: **Auranofin therapy and quality of life in patients with rheumatoid arthritis. Results of a multicenter trial.** *Am J Med* 1986, **81:** 565-578.
3. Pincus T: **Limitations of randomized clinical trials to recognize possible advantages of combination therapies in rheumatic diseases.** *Semin Arthritis Rheum* 1993, **23**(2 Suppl 1):2-10.

4. Felson DT, Anderson JJ, Meenan RF: **The comparative efficacy and toxicity of second-line drugs in rheumatoid arthritis. Results of two metaanalyses.** *Arthritis Rheum* 1990, **33:**1449-1461.
5. Anon: **Epidemiology and randomized clinical trials [editorial].** *Epidemiology* 2003, **14:**2.
6. Krishnan E, Fries JF: **Reduction in long-term functional disability in rheumatoid arthritis from 1977 to 1998: a longitudinal study of 3035 patients.** *Am J Med* 2003, **115:**371-376.
7. Hill A: **The environment and disease: association or causation.** *Proc R Soc Med* 1965, **58:**295-300.
8. Wolfe F, Lassere M, van der Heijde D, Stucki G, Suarez-Almazor M, Pincus T, Eberhardt K, Kvien TK, Symmons D, Silman A, van Riel P, Tugwell P, Boers M: **Preliminary core set of domains and reporting requirements for longitudinal observational studies in rheumatology.** *J Rheumatol* 1999, **26:**484-489.
9. Silman A, Symmons D: **Reporting requirements for longitudinal observational studies in rheumatology.** *J Rheumatol* 1999, **26:** 481-483.
10. Joffe MM, Rosenbaum PR: **Invited commentary: propensity scores.** *Am J Epidemiol* 1999, **150:**327-333.
11. Wolfe F, Pincus T: **Listening to the patient: a practical guide to self-report questionnaires in clinical care.** *Arthritis Rheum* 1999, **42:**1797-1808.
12. Fries JF, Williams CA, Singh G, Ramey DR: **Response to therapy in rheumatoid arthritis is influenced by immediately prior therapy.** *J Rheumatol* 1997, **24:**838-844.
13. Moses LE: **Measuring effects without randomized trials? Options, problems, challenges**. *Med Care* 1995, **33**(4 Suppl):AS8-AS14.
14. Rennie D: **How to report randomized controlled trials. The CONSORT statement [editorial]**. *JAMA* 1996, **276:**649.

Correspondence

Eswar Krishnan MD, 1000 Welch Road, Suite 203, Palo Alto, CA 94304, USA. Tel: +1 650 776 6484; fax: +1 610 375 6210; e-mail: eswar_krishnan@hotmail.com