Assessing the Safety of New Arthritis Drugs: Are We There Yet?

SAMY SUISSA

J Rheumatol 2008;35;2295-2297
http://www.jrheum.org/content/35/12/2295

1. Sign up for TOCs and other alerts
http://www.jrheum.org/alerts

2. Information on Subscriptions
http://jrheum.com/faq

3. Information on permissions/orders of reprints
http://jrheum.com/reprints_permissions

The Journal of Rheumatology is a monthly international serial edited by Earl D. Silverman featuring research articles on clinical subjects from scientists working in rheumatology and related fields.
Assessing the Safety of New Arthritis Drugs: Are We There Yet?

The management of patients in rheumatology involves several classes of drugs, many of which have only recently been added to the physician’s armamentarium. Despite the proven effectiveness of these drugs, their safety profile remains largely unknown at the time of marketing, so that the balance between benefit and risk is often uncertain. Rheumatology has been at the center of concerns over drug safety, including the recent spotlight on the cardiovascular risks of cyclooxygenase-2 inhibitors, numerous analgesics, and disease modifying antirheumatic drugs (DMARD). With today’s newest advances in information technology and the influx of new, costly drugs in rheumatology, it is pertinent to reexamine the current approach to drug safety evaluation and to ask whether risks can be identified and quantified earlier.

The randomized controlled trial is the fundamental scientific approach to evaluate a drug’s effectiveness and to obtain its approval for use. While such trials provide some safety information, their primary objective, data analyses, and power calculations all focus on measures of effectiveness and benefit, so that safety assessments are problematic. A typical large trial on the effectiveness of a new DMARD in rheumatoid arthritis (RA) will report the most frequent adverse events, usually around 40 different ones, so that several uncertainties will necessarily arise regarding the safety of the new drug. First, the large number of different adverse events introduces statistical uncertainty from multiple testing, as it is impossible to establish whether an adverse event found to be “statistically significant” at the p = 0.05 level emerged simply by chance, from the large number of statistical tests conducted. Statistical techniques developed to account for such multiple testing will tend to correct the “significant” p value to a “nonsignificant” one because of the large number of statistical tests, thereby possibly concealing a real safety concern. Second, the study size is problematic, since an adverse event that occurs in 1/1000 patients will only be detected with a seldom-seen trial size of 3000 patients per arm. Further, latent adverse events that only occur after longterm exposure may not even be observed within the typical one-year followup. Thus, the information that randomized trials can provide in terms of safety is very limited, by their restricted size and short followup, particularly for infrequent and latent adverse events. An investigation of the withdrawal of 5 drugs from the US market for serious adverse drug reactions found a major gap between the number of patients exposed in the clinical trials and the number exposed subsequently. For example, while the nonsteroidal antiinflammatory drug bromfenac was evaluated in 2400 patients prior to marketing, the associated serious hepatotoxic effects that led to its withdrawal from market were observed after 2.5 million people had used the drug.

Once the drug is marketed, spontaneous reporting systems are effective tools for detecting rare, unexpected adverse effects; this is a relatively inexpensive strategy, covering the entire patient population. This approach, however, has several limitations, including difficulty in identifying reactions with long latency such as carcinogenicity; insufficient information to rule out other causes; factors associated with reporting such as media bias, market share, and time on the market; and most important, the serious underreporting that can vary between physicians, regions, and over time, and can be differential across drugs. The latter, compounded with the absence of “denominator” data on the population exposed to the drug, do not permit valid calculation of incidence rates. Rather, reporting rates based on sales data are sometimes computed as an alternative approach, and the proportional reporting ratio has been proposed as a statistical approach for safety signal generation.

To quantify the risk of a drug, measures such as the incidence rate and the rate ratio can be estimated using observational studies that involve no intervention besides observing the natural course of disease longitudinally in a large patient population. Observational studies have become an important scientific tool to assess safety signals, and at times also to estimate drug benefit. These studies permit this evaluation in situations of real-life use, providing information on actual patterns of drug use, and allow the assessment of rare outcomes or latent effects that may not have been observable in the randomized trials. Two strategies to generate data for observational studies are currently in place in rheumatology: computerized health databases and registries.
Computerized health databases, constituted from administrative data routinely collected for insurance management purposes or clinical data collected by general practitioners, have revolutionized the conduct of observational research in drug safety. Because the data are already collected, the use of such computerized health databases allows the relatively quick and inexpensive conduct of observational studies. These databases offer several important advantages, including their large size, that make it possible to study rare, drug-related adverse events. Also, they can allow the study of entire populations, thus avoiding selection bias, and often provide complete and accurate prescription medication information, with major outcome data from physician billing, hospitalization, and mortality records. However, they also have limitations, such as the questionable validity of diagnostic information, incomplete data on risk factors and confounders, no information on over-the-counter drug use and in-hospital medication, and limited followup and often high patient turnover, among others. Despite these limitations, such databases can be used reliably to assess many drug safety issues. However, studies using such databases have often lacked the methodological rigor essential in the proper conduct of observational studies, leading to inconsistencies created by inadequate study design or inappropriate statistical analysis, rather than by database limitations. For example, soon after the signal of an elevated cardiac risk with rofecoxib appeared from the VIGOR randomized trial, 2 observational studies performed to assess the veracity of this risk presented discordant findings, one reporting an excess risk, the other no. As the databases were similar in structure, these discordancesses were likely due to methodological flaws in study design and data analysis. For example, soon after the signal of an elevated cardiac risk with rofecoxib appeared from the VIGOR randomized trial, 2 observational studies performed to assess the veracity of this risk presented discordant findings, one reporting an excess risk, the other no. As the databases were similar in structure, these discordancesses were likely due to methodological flaws in study design and data analysis. For example, soon after the signal of an elevated cardiac risk with rofecoxib appeared from the VIGOR randomized trial, 2 observational studies performed to assess the veracity of this risk presented discordant findings, one reporting an excess risk, the other no. As the databases were similar in structure, these discordancesses were likely due to methodological flaws in study design and data analysis. For example, soon after the signal of an elevated cardiac risk with rofecoxib appeared from the VIGOR randomized trial, 2 observational studies performed to assess the veracity of this risk presented discordant findings, one reporting an excess risk, the other no. As the databases were similar in structure, these discordancesses were likely due to methodological flaws in study design and data analysis. For example, soon after the signal of an elevated cardiac risk with rofecoxib appeared from the VIGOR randomized trial, 2 observational studies performed to assess the veracity of this risk presented discordant findings, one reporting an excess risk, the other no. As the databases were similar in structure, these discordancesses were likely due to methodological flaws in study design and data analysis. For example, soon after the signal of an elevated cardiac risk with rofecoxib appeared from the VIGOR randomized trial, 2 observational studies performed to assess the veracity of this risk presented discordant findings, one reporting an excess risk, the other no. As the databases were similar in structure, these discordancesses were likely due to methodological flaws in study design and data analysis. For example, soon after the signal of an elevated cardiac risk with rofecoxib appeared from the VIGOR randomized trial, 2 observational studies performed to assess the veracity of this risk presented discordant findings, one reporting an excess risk, the other no. As the databases were similar in structure, these discordancesses were likely due to methodological flaws in study design and data analysis. For example, soon after the signal of an elevated cardiac risk with rofecoxib appeared from the VIGOR randomized trial, 2 observational studies performed to assess the veracity of this risk presented discordant findings, one reporting an excess risk, the other no. As the databases were similar in structure, these discordancesses were likely due to methodological flaws in study design and data analysis. For example, soon after the signal of an elevated cardiac risk with rofecoxib appeared from the VIGOR randomized trial, 2 observational studies performed to assess the veracity of this risk presented discordant findings, one reporting an excess risk, the other no. As the databases were similar in structure, these discordancesses were likely due to methodological flaws in study design and data analysis.

Registries of patients with RA have been developed specifically around the issues of safety of new biologic agents. Thus, registries such as The National Data Bank (USA; n = 13,600), ARTIS (Sweden; n = 7354), BSRBR (UK; n = 8973), and RABBIT (Germany; n = 1529) have formed cohorts of patients with RA by accumulating longitudinal clinical data. While these registries include very accurate clinical information, their size remains somewhat limited to assess rare adverse events. For example, the ARTIS registry reported 2 relative risks of lymphoma with tumor necrosis factor (TNF) antagonists, namely 5.0 (95% CI 0.9—27.0) and 1.1 (95% CI 0.6—2.1), using different versions of the registry and comparison cohorts, with low statistical precision being a possible explanation for the discrepancy. Recently, the 3 European registries reported somewhat discrepant estimates of the relative risk of serious infection with TNF antagonists, again possibly because of random variation. Nevertheless, these registries have the potential to produce important information on several safety outcomes associated with DMARD, although their limited size will remain a major obstacle.

Where does this leave us? First, while it is important to continue monitoring drug safety in randomized trials, we cannot be deceived by their reputation: a gold standard in clinical research methodology that can confer a false sense of security — “if we don’t see it, it’s probably not there.” Second, despite their numerous limitations, spontaneous reporting systems remain the first line of defence in identifying drug risks for newly marketed drugs, and recent developments in statistical tools such as the proportional reporting ratio will help to provide early warnings of safety concerns after marketing. Third, observational studies using large computerized databases must be recognized as a powerful and efficient tool to assess drug safety in the real-world setting. However, while the quality of the databases has been improving with more powerful computer technologies, more careful attention is to be given to methodological aspects of the epidemiological designs and statistical analyses of these studies to avoid awkward discrepancies and potentially misleading results. Finally, the numerous registries of arthritis patients, quite unique in medicine and clearly a jewel for rheumatology, must be encouraged, here again with special care for epidemiological and statistical methodologies.
REFERENCES