Observation and Experiment with the Efficacy of Drugs: A Warning Example from a Cohort of Nonsteroidal Anti-inflammatory and Ulcer-healing Drug Users

Alex D. McMahon

Observational data are well suited for many types of medical research, especially when randomized controlled trials are inappropriate. However, some researchers have attempted to justify routine use of observational data in situations in which randomized controlled trials are normally conducted. Literature searches cannot be used to directly compare the results of the two types of research, because invalid observational studies normally are not publishable in the journal literature. The author created a study (1989–1994) to determine the efficacy of one exposure (ulcer-healing drugs) in preventing the serious upper gastrointestinal toxicity associated with another exposure (nonsteroidal anti-inflammatory drugs (NSAIDs)). A cohort of subjects from Tayside, Scotland, receiving both NSAIDs and ulcer-healing drugs appeared to experience a large rise in their risk of gastric bleeding and perforation (e.g., the rate ratio was 10.00 (95% confidence interval: 6.68, 14.97) when this cohort was compared with one receiving NSAIDs alone). This increased risk was due to confounding. Thus, use of a “restricted cohort design” was not able to eliminate uncontrollable bias. It is possible that if many different studies were carried out, then observational research would be found to be only occasionally useful for studying drug efficacy. Am J Epidemiol 2001;154:557–62.

anti-inflammatory agents, non-steroidal; cohort studies; drug evaluation; observation; pharmacoepidemiology; ulcer

It has been traditional to place observational research lower down the quality ranks of clinical evidence than research based on experiment. Cohort studies and case-control studies are usually considered inferior to randomized controlled trials (1). At worst, observational studies are reckoned to be useful only in generating hypotheses that subsequently can be tested in randomized studies. However, it is well known that randomized controlled trials cannot answer every question and that observational data are useful in many different situations (2, 3). The regulatory requirement to carry out randomized controlled trials applies to only the small percentage of clinical studies conducted to licence newly discovered drugs for their targeted indication. Observational data are particularly suited to studying the unintended effects of interventions, such as unexpected drug toxicity, which may be rare and/or have a long delay of onset (4). One researcher made a reasonable defense of observational studies in 1996 and requested more studies of direct comparisons between observational and randomized studies (2).

When the two types of studies are compared directly, observational studies are often criticized (5). However, recently there has been somewhat of a backlash against this view. An effort is under way to move observational studies further up the rankings of “strength of evidence” (6). The debate has been going on for some time, and every now and then a paper is published attempting to justify observational research on the efficacy of therapies (7). A paper by Britton et al. in Health Technology Assessment was devoted to this issue (8), and the research was later summarized in BMJ (9). The conclusion of this substantial piece of research was that neither method of study tended to produce larger estimates of the effect of treatment. Two recent articles in The New England Journal of Medicine have added more fuel to the controversy (6, 10). One set of authors explicitly stated that their results challenge the consensus on the hierarchy of strength of evidence (6). The implied suggestion is that observational studies may be just as acceptable in informing clinical decisions as their experimental counterparts.

These developments published in the journal literature are a matter of some concern for researchers of pharmaceutical interventions. These papers tend not to reference what could be considered the key papers studying the effectiveness of drugs by using epidemiologic methods (11–14), although one paper (8) fleetingly refers to the study by Miettinen (11). There are indeed some situations in which good observational studies may be performed that investigate drug efficacy. Examples are well known, such as in the study of vaccines or insulin use in diabetics (2, 15). The only known adverse reaction to the recent papers is limited to several

Received for publication August 29, 2000, and accepted for publication April 16, 2001.

Abbreviation: NSAID, nonsteroidal anti-inflammatory drug.

From the Robertson Centre for Biostatistics, Boyd Orr Building, University of Glasgow, Glasgow, G12 8QQ, Scotland (e-mail: alexm@stats.gla.ac.uk). (Reprint requests to Dr. Alex D. McMahon at this address).
The main thrust of these papers is that the literature on a variety of topics has been searched for direct comparisons of the two study methodologies. No systematic differences between the two methodologies have been noted. It has been pointed out that authors who have found differences (5) have unfairly included observational studies that made use of historical controls (7, 17, 19). One suggested reason for the similar results with the two methods is that observational studies have been improving and now are conducted by adhering to a higher standard (10, 17). This line of thought exposes a flaw in the logic of those who would replace randomized controlled trials with more inexpensive observational research (6, 19). The fact that the results of controlled trials and observational studies are similar may simply be a result of experienced epidemiologists knowing when a study is invalid because of uncontrollable bias. Invalid studies tend not to be carried out anymore, so this source of conflict with the randomized controlled trial will not manifest itself as published research. Some authors have gone even further and have criticized randomized controlled trials for producing contradictory results and being prone to selection bias (6). It also has been said that the paucity of good comparative papers is due to observational studies not being “trustworthy,” which turns the argument on its head.

The purpose of the current study was to conduct an examination of drug efficacy that is difficult to carry out by using observational research. It is hoped that this study will serve as both a reminder and a warning that dangerous sources of bias such as confounding by indication are very real (12, 20, 21) and have not been eliminated by using modern epidemiologic techniques. The effects of confounding by indication are still a research topic (22, 23). The lesson about the confounded association between warfarin and thrombosis seems to have been forgotten (11). Some of the serious sources of bias associated with the study of drugs are described by Salas et al. (24). Confounding by indication occurs when the indication for a treatment is a confounder, and confounding by severity (a related issue) arises when the severity of the disease acts as a confounding variable. Protopathic bias occurs when a treatment is given for early symptoms of the outcome being studied. One recent author has conceded that observational studies of efficacy cannot be used when a treatment is routinely given to the sickest patients (10). This phenomenon is very common, even when drugs are prescribed for the same illness (23). At one time it was hoped that statistical adjustment of confounding variables would enable observational study of drug efficacy so that databases could replace randomized studies (7), but this was a forlorn hope (25). A research team has even implied that observational data are particularly useful for the study of drug efficacy (6).

It is not disputed that some observational studies of drug efficacy are indeed possible. Careful use of subject inclusion criteria in the manner of a clinical trial has been advocated, and this type of study has been dubbed the “restricted cohort design” (6, 19, 29). The idea is to exclude subjects with risk factors that are strong indications for or contraindications to a given treatment, similar to the screening rules in clinical trials. Thus, it may be possible to create groups that are “similar for prognostically important clinical severity” (19).

As an aside, it has been argued that the clinical trial ultimately fails as a paradigm for observational research (27, 30). Attempts to introduce a “time zero” into the restricted cohort design (19) (to “approximate the point of randomization”) may be thwarted because an observational study may not have a true baseline time point. If no intervention has been applied to the study subjects, then the start of exposure may be an artificial concept, especially when routine medical records are used. There may be no strongly identifiable anchor of time upon which to base the study other than when the outcome occurs at the end of the study.

The current study attempted to remove all possible bias by using very strict subject selection criteria. The data were taken from a previous study of gastric bleeding and perforation associated with use of nonsteroidal anti-inflammatory drugs (NSAIDs) (31). Unlike the previous study, the current one was designed to specifically examine the protective effects of ulcer-healing drugs when used in combination with NSAIDs. This study looked at the efficacy of one exposure (ulcer-healing therapy) in preventing the toxicity of another (NSAIDs). The aim was to use observational data to determine whether confounding could be combated and to demonstrate that ulcer-healing drugs can protect NSAID users from serious upper gastrointestinal toxicity.

MATERIALS AND METHODS

A study population of all identifiable residents of Tayside, Scotland, was created. The MEMO record-linkage database was used to provide details on prescriptions for NSAIDs and ulcer-healing drugs from 1989 to 1994 (32). The year 1989 was used as a screening period so that bias could be removed by applying the “sacrifice of early data” principle (33). Therefore, subjects who had received prescriptions for NSAIDs or ulcer-healing drugs during 1989 were excluded from the study base. Subjects for whom the database start date occurred after January 1, 1989, were also excluded. A summary of all inclusion and exclusion criteria is given in table 1.

In addition to providing a set of new users of the study drugs, the screening period was also required so that the order in which the two types of drugs were prescribed could be determined. This order could be known only after a prescription-free period. The primary outcome was a subject’s first emergency hospitalization at a Tayside hospital for a serious upper gastrointestinal diagnosis (i.e., bleeding peptic ulcer or perforated peptic ulcer). The study start date was the date on which the subject received the first prescription (either NSAID, ulcer-healing drug, or both on the same day) after the screening period, that is, from January 1, 1990, onward.

Am J Epidemiol Vol. 154, No. 6, 2001
Exposure

The first period of NSAID exposure was defined as the chain of all overlapping or consecutive NSAID use beginning on the study start date. A consecutive prescription was defined as one that started the day after the previous one ended. Similarly, the second, third, and following periods of NSAID exposure were estimated. A similar definition was used for periods of exposure to ulcer-healing drugs.

It was anticipated that it would be difficult to realistically define “continuous” periods of exposure exactly given the dates that the pharmacies dispensed the drugs. Short gaps between prescriptions were possible, especially if a patient was not required to take the treatment every day (e.g., the regimen was “as required”). Therefore, before exposure periods were calculated, a period of 7 days was added to the duration of each prescription. Two study groups were created.

The combination therapy group. This group consisted of those subjects whose first period of ulcer-healing drug exposure overlapped a period of NSAID use. Thus, the combined effects of NSAIDs and ulcer-healing drugs could be examined. The “end of exposure” was defined as the day before the second period of ulcer-healing drug exposure began (when there was more than one period of exposure). A subject’s exposure was truncated at this point, because the outcome rates for further periods of ulcer-healing drug exposure would have been confounded by the earlier ones.

The NSAID group. Subjects in this group were experiencing their first period of NSAID use and were not part of the combination therapy group. The group included those subjects who had either never received ulcer-healing drugs or received their ulcer-healing drugs when they were not using NSAIDs. Only the first exposure period was considered, because only one type of person-years from this group could be used as the referent. The “end of exposure” was defined as the last day of the first period of NSAID exposure.

<table>
<thead>
<tr>
<th>Purpose of screening rule</th>
<th>Rule*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ensure a minimum of 6 months of follow-up</td>
<td>Include only those subjects whose exposure to the drugs started more than 6 months before the end of the study.</td>
</tr>
<tr>
<td>Exclude prior events and related risk factors</td>
<td>Exclude subjects if they had any hospitalizations for a gastrointestinal diagnosis prior to the start of exposure. Exclude subjects if they had any endoscopies prior to the start of exposure.</td>
</tr>
<tr>
<td>Ensure new users of the study drugs</td>
<td>Subjects must be in the study population during the entire screening period. Exclude users of NSAIDs† during the screening period. Exclude users of ulcer-healing drugs during the screening period.</td>
</tr>
<tr>
<td>Ensure that the NSAID was used first</td>
<td>Exclude subjects if an ulcer-healing drug was taken first.</td>
</tr>
</tbody>
</table>

* Subjects were required to meet all screening rules. † NSAID, nonsteroidal anti-inflammatory drug.

After these exclusions were made, a further set of study exclusions was applied. To ensure that all subjects had at least 6 months of follow-up, all subjects whose study start date was on or after June 1, 1994, were excluded. Subjects who had any hospitalizations for a gastrointestinal diagnosis (not just one of the study outcomes) prior to the study start date also were excluded (data were available from 1980 onward). In addition, all subjects with an endoscopy prior to the study start date were excluded (data were available from 1980 onward). If the first prescription (after the study start date) was for an ulcer-healing drug, then these subjects also were excluded.

There were two reasons for these last three exclusions. Firstly, to minimize confounding problems, subjects who had had gastrointestinal events and endoscopies were excluded. Thus, subjects thought to be at high risk of gastrointestinal problems would be prescribed different patterns of drugs from those prescribed to subjects thought to be at low risk. In a previous study, the risk was found to be higher for subjects with prior events; however, the authors did not find any excess NSAIDs-associated toxicity in these subjects (31). If subjects who received ulcer-healing drugs in addition to NSAIDs were compared with those who did not, confounding would be particularly problematic.

Secondly, if subjects had received an ulcer-healing drug before an NSAID, then clearly they already were under suspicion (at least) of having peptic disease and probably had some gastrointestinal symptoms. Therefore, this occurrence would again result in confounding, and these subjects were excluded. This study was designed to determine whether ulcer-healing drugs might protect against NSAID toxicity, which might suggest that prior use of ulcer-healing drugs would be of interest. However, because subjects were not randomly allocated to this prior exposure, which would have been prescribed for existing gastrointestinal disease, they were excluded. Note that if the first ulcer-healing drug was prescribed on the same day as the first NSAID, then the subject was allowed into the study.

Am J Epidemiol Vol. 154, No. 6, 2001
Study end date

The study end date was defined as December 31, 1994, the date of death, the date of the end of exposure, or the date of a study outcome, whichever was the earliest.

Exposure classification scheme

For each subject in the combination therapy group, person-years were partitioned into the following types of exposure or nonexposure (figure 1): 1) prior exposure to NSAIDs (possibly several periods), 2) prior nonexposure (possibly several periods), 3) simultaneous exposure to both NSAIDs and ulcer-healing drugs, 4) exposure to ulcer-healing drugs only (if the prescription lasted longer than the one for NSAIDs), 5) “after” exposure to NSAIDs (if the NSAID prescription lasted longer than the one for ulcer-healing drugs, and possibly several more exposure periods had occurred), and 6) “after” nonexposure (possibly several periods). Subjects in the NSAID group simply had “NSAID exposure,” which was collected from their first period of NSAID use.

Analysis

Each type of exposure for subjects in the combination therapy group was compared with the NSAID group experience (i.e., the referent category of risk). The number of days of exposure was totaled separately for each type of exposure. Incidence rates were calculated by using the number of events per 1,000 person-years of exposure. Rate ratios and 95 percent confidence intervals were calculated.

The main contrast of interest was “combination therapy” versus the NSAID group. Also of interest was the contrast between “prior exposure to NSAIDs” and the NSAID group. This contrast was intended to give some indication of how successful the attempts were to control for confounding. If there was no differential prescribing based on any perceived risk of peptic disease, then the incidence rates in these two groups would be similar.

RESULTS

In the NSAID group, the incidence rate for bleeding and perforation was 6.30 events per 1,000 person-years of exposure (table 2). In the combination therapy group, the incidence rate was 4.70 during the periods of nonexposure before the combination of NSAIDs and ulcer-healing drugs started. This rate is slightly but not significantly lower than the rate for the NSAID group. For prior NSAID use, the rate rose to 56.40, which was significantly higher than the rate for the NSAID group; the rate ratio was 8.95 (95 percent confidence interval: 6.63, 12.08). During combination therapy with ulcer-healing drugs, the incidence rate rose even higher to 63.00, and the rate ratio was 10.00 (95 percent confidence interval: 6.68, 14.97).

<table>
<thead>
<tr>
<th>Study group and type of exposure</th>
<th>No. of events</th>
<th>1,000 person-years</th>
<th>Incidence rate</th>
<th>Rate ratio</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSAID* group (n = 65,542)</td>
<td>47</td>
<td>7.46</td>
<td>6.30</td>
<td>Referent</td>
<td></td>
</tr>
<tr>
<td>Combination drug therapy group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noexposure before combination</td>
<td>10</td>
<td>2.13</td>
<td>4.70</td>
<td>0.76</td>
<td>0.38, 1.47</td>
</tr>
<tr>
<td>(n = 1,769)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSAID exposure before combination</td>
<td>76</td>
<td>1.35</td>
<td>56.40</td>
<td>8.95</td>
<td>6.63, 12.08</td>
</tr>
<tr>
<td>(n = 2,311)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combination exposure (n = 2,815)</td>
<td>23</td>
<td>0.37</td>
<td>63.00</td>
<td>10.00</td>
<td>6.68, 14.97</td>
</tr>
<tr>
<td>UHD* exposure after combination</td>
<td>1</td>
<td>0.12</td>
<td>8.51</td>
<td>1.35</td>
<td>0.19, 9.72</td>
</tr>
<tr>
<td>(n = 1,414)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSAID exposure after combination</td>
<td>4</td>
<td>0.56</td>
<td>7.12</td>
<td>1.13</td>
<td>0.41, 3.14</td>
</tr>
<tr>
<td>(n = 1,547)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonexposure after combination</td>
<td>11</td>
<td>2.42</td>
<td>4.55</td>
<td>0.72</td>
<td>0.38, 1.39</td>
</tr>
<tr>
<td>(n = 2,274)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* NSAID, nonsteroidal anti-inflammatory drug; UHD, ulcer-healing drug.
DISCUSSION

The combination therapy group had a high incidence of toxicity. This finding was probably due to confounding, despite the study design. Ulcer-healing drugs did not reduce the rate either, because the drugs were consequently prescribed to subjects at an increased risk. The fact that the risk was very high in the combination therapy group, even before treatment with ulcer-healing drugs, demonstrates that the two groups of patients were not comparable with regard to gastrointestinal risk. It has been shown that some types of NSAIDs are channeled toward patients at a higher gastrointestinal risk than others are (23). It is reasonable to assume that this effect would be even more pronounced when comparing those who have and have not been exposed to ulcer-healing drugs when they are also using prescribed NSAIDs. Some evidence suggests that ulcer-healing drugs can prevent gastric toxicity associated with exposure to NSAIDs, although this evidence is by no means conclusive (especially for H2-antagonists) (34–36). Even if these drugs are not effective in preventing this type of NSAID toxicity, they certainly are not responsible for a 10-fold increase in toxicity.

In this instance, observational data were found to be unsuitable for detecting the intended effects of a pharmaceutical intervention. This study failed to be internally valid, which is obviously a prerequisite to being externally valid (37). In the United States, the government body called the Agency for Healthcare Research and Quality (AHRQ) accepts applications for funding of observational studies of effectiveness, provided they are methodologically rigorous and both internally and externally valid. The current study provides an example of a treatment and disease pair that was not suitable for study without introduction of random allocation of treatments. When the study was designed, every effort was made to exclude every source of measurable bias. Every subject in the population who had ever received an endoscopy, or even an ulcer-healing drug, was excluded. Therefore, the principles of the “restricted cohort design” were followed.

Unfortunately, diagnostic notes from general practitioners were not accessible, because that type of information was not available in the particular record-linkage database used in the study (31, 32). It may be possible to access further information regarding symptoms and severity of disease in a more detailed observational study, for example, in a prospective study that uses patient interviews. That type of study would have some advantages over a “database study.” However, using that type of information probably would not have changed the conclusions drawn from the present study, even if the rate ratios had decreased slightly. One advantage of the database used in the study is that every prescription dispensed for the entire population was available. Everything that could be done was done. Note that analytical techniques such as propensity scores (38) are only useful for reducing confounding that has been measured.

What can be deduced from the results of the current study? If proper refutational logic (39, 40) is followed, observational and experimental studies do not always produce equal estimates of a treatment effect. However, other authors similarly have been unable to prove that observational studies are always as good as randomized controlled trials (a claim that very few people would make anyway). The implicit message is that observational studies of drug efficacy may often or even usually produce valid estimates. I suspect that observational studies of drug efficacy will usually not produce accurate results. This idea is testable in principle but not in practice. The hypothesis that both techniques are similar could be refuted by carrying out a large number of head-to-head tests for various classes of pharmaceuticals. However, this research would be very unsatisfying for researchers to conduct and would never be funded, and publication would no doubt be problematic. Nevertheless, the current study should serve as a timely caution for researchers thinking of applying data from observational databases to research into the efficacy (or effectiveness) of pharmaceutical treatments.

ACKNOWLEDGMENTS

The author thanks the research team of the MEMO unit in Dundee, Scotland, especially Professor Tom MacDonald, Dr. Josie Evans, Dr. Mark McGilchrist, Gary White, and Douglas Boyle.

REFERENCES