

Jick H (2008). Learning how to control biases in studies to identify adverse effects of drugs: a brief personal history.



© Hershel Jick, Boston Collaborative Drug Surveillance Program, Boston University School of Medicine, Lexington, MA 02421, USA. Email: hjick@bu.edu

Prior to 1966, almost all available information on drug safety for marketed medicines was strictly qualitative. It was predominantly based on anecdotal reports and clinical opinions derived primarily from the known pharmacology of particular drugs. While the medical environment for identifying and quantifying drug toxicity was generally informal and undirected, the use of anecdotal reports to detect serious drug toxicity was sufficient for identifying the toxicity associated with thalidomide ([McBride 1961](#); Brynner and Stephens 2001). But the mechanism of discovery was unacceptably inefficient since an estimated 9,000 babies were born with thalidomide-caused phocomelia before the toxic effect was recognized and described in 1960.

Cite as:

Jick H (2008). Learning how to control biases in studies to identify adverse effects of drugs: a brief personal history. JLL Bulletin: Commentaries on the history of treatment evaluation (www.jameslindlibrary.org). [Personal reflection]

As a result of the thalidomide experience, government agencies in North America and Europe were established to collect spontaneous reports of purported adverse reactions to drugs, primarily from practising physicians. In addition, pharmaceutical companies began to establish departments to collect spontaneous reports of purported adverse reactions to their medicines. This system of collecting information on drug safety has proved useful over the years ([Venning 1982](#)) and several medicines have been withdrawn from the market based primarily on spontaneous anecdotal reports.

The early 1960s were a critical time for the development of more formal techniques for identifying and quantifying drug toxicity for the many medicines on the market. In addition to the thalidomide experience, which provided striking evidence that devastating toxicity could be present and unrecognized by doctors, the 1960s saw the development and marketing of a large number of new, highly potent medicines. These included oral contraceptives, postmenopausal hormones, antihypertensive agents, major tranquilizers, and anti-tumour agents. As a result of these developments, it became clear that large-scale formal techniques were required to quantify the known toxicity of medicines and to identify unsuspected drug toxicity. Formal methods differ from spontaneous reports in that they imply that studies are performed on defined populations of people with standardized collection of information on exposure to medicines and the illnesses that develop subsequently.

I first became aware of the issue of medicines safety in the spring of 1966. At that time, Tom Chalmers, well known for his interest in randomized trials, came to see me. He was at that time the chief of medicine at Lemuel Shattuck Hospital in Boston, where I was doing research in renal physiology and pharmacology (between 1961 and 1969). Chalmers had just come back from a meeting with other senior public health officials in Washington DC, where it was concluded that there was an urgent need to develop formal, structured research in the field of drug safety. He asked me to look into the research possibilities, and I agreed to do so.

A review of the medical literature yielded no articles that had either described the nature or the extent of the problem or how to approach the issue in a scientifically valid way. On reflection, it seemed that there was only one conceptual and feasible method to obtain useful and credible quantitative information on the adverse effects of the hundreds of marketed medicines. I concluded that, in its simplest form, a planned drug safety study would begin by identifying a group of people who start to use a medicine. This 'cohort' of users would then be followed up to see which newly diagnosed illnesses developed. For short-term drug use, the follow-up time would normally need only a few weeks or less; for medicines used over a long time, it might be one or more years. The ultimate task was to distinguish which, if any, of the newly diagnosed illnesses were caused by the medicine under study from illnesses that had occurred spontaneously, independently of the medicine under investigation.

A feasible option was to identify users of medicines in the hospital setting, and follow them forward. We started the first continuous large-scale, multipurpose, formal study of drug effects in 1966; the first six months of the design was tested at the Shattuck Hospital (Jick 1968). The technique was restricted to the study of hospitalized patients where drug exposure could be fully recorded and for whom careful follow-up was feasible. Specially trained nurse monitors were assigned to particular hospital wards to record relevant clinical information on a regular basis, using standardized

data forms. The information recorded and put on computer included patient demographics, a history of medicines taken prior to admission, and a complete record of the medicines prescribed and all medical events occurring during hospitalization.

Over the years, an identical study design was introduced in some 40 hospitals in seven countries (Jick et al. 1970), and, by 1982, the information encompassed about 70,000 patients. This research enterprise, known as the Boston Collaborative Drug Surveillance Program, provided a broad range of quantitative information on the acute toxicity of the medicines used in hospitalized patients during the study period. The primary objectives of the Boston Collaborative Drug Surveillance Program, from the start, were to identify previously unrecognized adverse (or beneficial) drug effects; to quantify known effects; and to identify factors which modify drug effects, such as gender, age, body mass index and even possibly cigarette smoking. More than 150 papers describing findings using this study design have been published (Cohen and Weaver 1992).

In 1974, in a talk at the Harvard University School of Public Health, I reviewed the first eight years of data on some 11,500 patients who had received more than 100,000 courses of drug therapy. The frequency of serious adverse drug reactions was surprisingly low given the high use of drugs in the hospital setting, averaging more than nine drugs per patient. Franz Ingelfinger, then editor of the *New England Journal of Medicine*, was in the audience, and invited me to write a paper on this topic (Jick 1974).

The complexity of observational drug safety research became clear as we continued to explore the multitude of specific safety issues. Drug safety research differs from more classic epidemiological research in the range and complexity of the exposure variable. Every course of drug therapy is unique to a particular individual and drug use tends to vary over time. Importantly, drug therapy is always related to a clinical condition which initiates its use. In addition, drug use varies by age, sex, calendar time, and geography, variables that must be controlled correctly in the study design and analysis. I published another paper in the *New England Journal of Medicine* in 1977, entitled *The discovery of drug-induced illness*, which attempted to organize and describe basic guidelines, related to the magnitude of a drug effect and the baseline risk of the disease, in the choice of an individual study design (Jick 1977). I noted adverse drug effects may be rare or somewhat common. They may be relatively benign (transient nausea, for example), serious (acute liver disease, for example), or even life threatening (stroke, for example). They may occur soon after a drug is started or after months or even years of use. A very important consideration is the background spontaneous frequency of the illness caused by the drug. If the illness is common (for example, myocardial infarction in an elderly population) and a drug causes the illness only rarely, this effect will not be readily identifiable in an observational study, or indeed in a randomized clinical trial unless the latter is very large. On the other hand, if an illness occurs rarely (phocomelia, for example) and a drug causes the illness frequently, this effect may be recognized without properly designed observational studies (McBride 1961; Venning 1982).

The first phase of activity of the Boston Collaborative Drug Surveillance Program emphasized acute drug related events during hospitalization (Slone et al. 1966). In the early 1970s, the Program enlarged its scope. Using the history of drug use data item included on the standardized data forms and the in-hospital monitoring data, it was also possible to evaluate the risk of hospitalization for certain illnesses caused by medicines used prior to hospitalization. We used the case-control design for this, which compares patients hospitalized for a specific disease with patients hospitalized for other reasons and relates the two groups to medicines used prior to hospital admission (Slone et al. 1977). Published studies include the negative association between aspirin use and myocardial infarction (Boston Collaborative Drug Surveillance Program 1974a; Elwood 2004), and the positive association between estrogen use and gallbladder disease (Boston Collaborative Drug Surveillance Program 1974b). Both findings have been confirmed in clinical trials. A brief history of the early phases of the Boston Collaborative Drug Surveillance Program was written by Lawson (1980).

In 1968, Martin Vessey and Richard Doll published the results of a case-control study on the relation between use of oral contraceptives and venous thromboembolism and, for the first time, identified and described the fundamental principles of the case-control design in drug safety research (Vessey and Doll 1968). The Boston Collaborative Drug Surveillance Program used the identical case-control design to conduct and publish a paper on oral contraceptives and venous thromboembolism in 1973, which yielded virtually the same results as those found earlier by Vessey and Doll (Boston Collaborative Drug Surveillance Program 1973). Martin Vessey and I became colleagues and friends in the late 1960s (Vessey 2006) and we published our first article together in 1969 (Jick et al. 1969). It demonstrated a strong association (present in the US, Britain, and Sweden) between ABO blood type and oral contraceptive-related venous thromboembolism.

Subsequently, in 1978, Martin Vessey and I also published a paper detailing the principles and methods of the case control design as applied to drug safety research (Jick and Vessey 1978; Vessey 2006). In the case-control design, people who are hospitalized because of an illness of interest (for example, venous thromboembolism - cases) are compared to persons hospitalized for other conditions (controls). The controls must be closely similar to the cases in age, gender, and general health status. This concept is critical in the conduct of epidemiologic studies of drugs. The risk of an outcome illness has to be the same for the cases and the controls prior to disease onset. To achieve this,

the case and controls should be as homogenous as is feasible. When this condition is met, if an association is present, one may consider the possibility of a causal interpretation.

The in-hospital monitoring design had a number of important limitations. It was restricted to hospitalized patients: by definition, they were suffering from a current illness which itself was often associated with a number of symptoms. Thus the attribution of drug causation for newly developed medical symptoms is more complex than it is for patients whose medical conditions are stable. Also, the study design was extremely expensive. It required hiring and training a large number of nurse monitors. In addition, staff were required to review and validate the incoming paper forms and to transfer the information onto computer. After 15 years, the cost of additional data collection was such that continuing the study was no longer regarded as cost-effective.

For those of us who were engaged full-time in the conduct of drug-safety studies in the mid-1970s, it was clear that far greater efficiency was required to conduct the necessary research for thousands of marketed medications ([Colombo et al. 1977](#)). The increasing availability of computer-recorded medical information offered, at least in principle, the opportunity to achieve a major advance in the efficiency of obtaining the required information. So it was that, in 1978, the Boston Collaborative Drug Surveillance Program developed a co-operative agreement with Group Health Co-operative of Puget Sound, a health maintenance organization in Seattle, Washington, which had a membership of about 300,000 people. In 1972 it had begun to computerize detailed information about the hospitalizations of all its members. In addition, all of the local pharmacies had become fully computerized by 1976. Since medicines were provided free of charge or at a reduced cost, it seemed likely that virtually all medicines prescribed by the participating physicians would be recorded on computer. The pharmacy record included complete, accurate, and detailed information on drug use. These attributes allowed us to define drug exposure in a way that strengthened the possibility of valid identification of drug effects in terms of timing, dose, and duration of use. Finally, because all original clinical records were located in a central record department, we had rapid access to records needed to validate diagnoses. Review of the original clinical records and interview of patients confirmed the high quality and completeness of the computer-recorded information on medicines dispensed and hospital diagnoses (Jick et al. 1979).

In view of the expense and administrative tediousness of the previous means of conducting drug safety studies, it was immediately evident that the availability of this computerized data resource represented a major advance in the ability to conduct drug safety studies. One could now identify large cohorts of users of prescribed medicines directly from a computer-recorded resource and link the cohorts of users with subsequent illnesses requiring hospitalization. The Group Health Co-operative resource did have a number of limitations, however, mainly because the size of the population encompassed was only 300,000 people, the formulary was limited (so that a number of important medicines could not be studied), and because we were unable to study directly outcomes that occurred in outpatients.

The first paper using Group Health Co-operative data showed that postmenopausal estrogen use increases more than 8-fold the risk of endometrial cancer in women with a uterus, and that this risk manifests itself only after four years of estrogen use, and is completely reversed when estrogen therapy is stopped (Jick et al. 1979a). We showed subsequently that the adverse estrogen effect is preventable by concomitant use of progesterone (Jick et al. 1993). In total, we published over 50 studies based on Group Health Co-operative data.

In 1979 we published a paper entitled *Postmarketing follow-up* (Jick et al. 1979b). This included a review of the available experience and methods, as well as identifying potential sources of data for drug safety research. We encouraged the proper use of valid, comprehensive, computerized information as the way forward for identifying and organizing large quantities of data for analysis by scientists trained in this field.

While the initial use of computerized medical information had provided new, highly efficient research output about drug safety, the available data sources limited the quantity and types of studies that could be conducted. In June 1984, at the request of a number of drug company medical directors, the Boston Collaborative Drug Surveillance Program organized a workshop in Minster Lovell, England, on the principles and methods of drug safety research. These workshops have continued through 2007.

The UK provides a unique medical environment for drug safety research because general practitioners are the repositories for virtually all of the relevant medical data and prescribing a record on each patient. Achieving a database to capture this information seemed a pipe dream at the time. However, in 1988, a small private company (Vamp Health), which was producing office computers for general practitioners, used computer software designed to record comprehensive medical information on individual patients, including patient demographics, all prescriptions, and all clinical diagnoses, together with considerable additional information on patient medical care.

We worked together with Vamp Health to validate the data required for research. The database, now known as the General Practice Research Database, was organized so that the information from hundreds of individual general practitioners was merged into a single file organized into a form which was efficient and designed for conducting drug safety studies. The drug use information is a complete record of prescriptions dispensed (the prescription is actually produced by the computer after entry by the general practitioner), and this information has been confirmed to be of

high quality. After excluding general practices that provided unsatisfactory data, we were left with practices generating good quality data on approximately 3 million patients. After validation of the information on prescriptions and outcomes, we concluded that the database could be relied upon to provide efficient access to clinical information suitable for drug safety studies, and that it was possible to identify large cohorts of drug users. The first paper based on the General Practice Research Database was published in the *BMJ* in 1991 (Jick et al. 1991).

In the mid-to-late 1990s, numerous papers purporting to describe adverse and beneficial effects of medications based on observational data were published in the medical literature. They provided conflicting and controversial results, which were reported uncritically in the lay media as well as in medical journals. This unacceptable state of affairs called for an update of the principles of observational drug safety research using automated databases (Jick et al. 1998). Together with two colleagues I wrote "...these published studies have been reported in the media and led to substantial public confusion and unnecessary and sometimes heated controversy among the investigators. Our view is that these controversies arise from misunderstandings about epidemiological principles by investigators and journal reviewers."

Studies conducted using the General Practice Research Database have demonstrated the extraordinary utility of this resource. More than 20 groups, including the Food and Drug Administration and Medicines and Healthcare Products Regulatory Agency, have now signed on to receive part or all of the General Practice Research Database data. The Boston Collaborative Drug Surveillance Program continues to receive annual updates and has now published some 200 papers based on the General Practice Research Database, which now contains 18 years of information, encompassing some 50 million person-years of follow-up.

I have found the last 41 years to be personally rewarding and productive. The papers we have published have included more than 200 authors from five continents, many of whom have become friends. The availability of large computerized medical databases that contain comprehensive and accurate medical information has now been proven to be extraordinarily valuable for drug safety research. However, the science of drug safety research is highly complex and subtle and, therefore, it is important that those who use the automated resources for drug safety research are fully trained and experienced in this field.

References

Boston Collaborative Drug Surveillance Program (1973). Oral contraceptives and venous thromboembolic disease, surgically confirmed gallbladder disease, and breast tumors. *Lancet* 1:1399-1404.

Boston Collaborative Drug Surveillance Program (1974a). Regular aspirin intake and acute myocardial infarction. *BMJ* 1:440-443.

Boston Collaborative Drug Surveillance Program (1974b). Surgically confirmed gallbladder disease, venous thromboembolism, and breast tumors in relation to postmenopausal estrogen therapy. *New England Journal of Medicine* 290:15-19.

Brynner R, Stephens TD (2001). *Dark Remedy: the impact of thalidomide and its revival as a vital medicine*. New York: Perseus Books.

Cohen MR, Weaver J (1992). A compilation of abstracts and an index of articles published by the Boston Collaborative Drug Surveillance Program 1966-1991. *Hosp Pharm* 4S:3-55.

Elwood P (2004). The first randomized trial of aspirin for heart attack and the advent of systematic overviews of trials. *The James Lind Library* (www.jameslindlibrary.org).

Jick H (1968). Drug surveillance program. *Medical Science* 18:41.

Jick H (1974). Drugs-remarkably nontoxic. *New England Journal of Medicine* 291:824-828.

Jick H (1977). The discovery of drug induced illness. *New England Journal of Medicine* 296:481-485.

Jick H, Vessey MP (1978). Case-control studies in the evaluation of drug-induced illness. *Am J Epidemiology* 107:1-7.

Jick H, Vessey M, Westerholm B, Inman WH, Vessey MP, Shapiro S, Lewis GP, Worcester J (1969). Venous thromboembolic disease and ABO blood type: a cooperative study. *Lancet* 1:539-542.

Jick H, Miettinen OS, Shapiro S, Lewis GP, Siskind V, Slone D (1970). Comprehensive drug surveillance. *JAMA* 213:1455-1460.

Jick H, Watkins RN, Hunter JR, Dinan BJ, Madsen S, Rothman KJ, Walker AM (1979a). Replacement estrogens and endometrial cancer. *New England Journal of Medicine* 300:218-222.

Jick H, Walker AM, Spriet-Pourra C (1979b). Postmarketing follow-up. *JAMA* 242:2310-2314.

Jick H, Jick SS, Derby LE (1991). Validation of the information recorded on general practitioner computerized data resource in the United Kingdom. *BMJ* 302:766-768.

Jick SS, Walker AM, Jick H (1993). Estrogens, progesterone, and endometrial cancer. *Epidemiology* 13:212-217.

Jick H, Garcia Rodriguez LA, Pérez Gutthann S (1998). Principles of epidemiologic research on adverse and beneficial drug effects. *Lancet* 352:1767-1770.

Lawson D (1980). Intensive monitoring in hospitals – I: Boston Collaborative Drug Surveillance Program. In: Inman WHM, ed. *Monitoring for Drug Safety*. Philadelphia: Lippincot, p 225-273.

McBride WG (1961). Thalidomide and congenital abnormalities. *Lancet* 2:1358.

Slone D, Jick H, Borda I, Chalmers TC, Feinleib M, Muench H, Lipworth L, Bellotti C, Gilman B (1966). Drug surveillance utilizing nurse monitors. An epidemiological approach. *Lancet* 2:901-3.

Slone D, Shapiro S, Miettinen OS (1977). Case-control surveillance of serious illness attributable to ambulatory drug use. In: Colombo et al., eds. *Epidemiological evaluation of drug use*. Amsterdam: Elsevier, p 57-70.

Venning GR (1982). Validity of anecdotal reports of suspected adverse drug reactions: the problem of false alarms. *BMJ* 284:249-254.

Vessey MP (2006). Learning how to control biases in studies to identify adverse effects of drugs. The James Lind Library (www.jameslindlibrary.org).

Vessey MP, Doll R (1968). Investigation of relation between use of oral contraceptives and thromboembolic disease. *BMJ* 2:199-205.

[Home](#)

[Contents](#)