ABSTRACT: The article presents some considerations about causality in Pharmacoepidemiology and Pharmacovigilance. To begin, we provide a brief introduction about the importance of the issue, noting that the understanding of causal relationships is considered one of science’s greatest achievements and has been, over time, a continuous and central concern of philosophers and epidemiologists. Next, we describe definitions and types of causes, demonstrating their influences on pharmacoepidemiological thought. After that, we present Rothman’s multi-causal model as one of the founding explanations of multiple causality and the issue of causality assessment. We conclude with some comments and reflections on causality from the perspective of health surveillance, particularly with regard to regulations on pharmacovigilance.

INTRODUCTION

Drug use is related to the occurrence of adverse events (adverse drug events - ADE), which, by definition, include health problems resulting from the exposure to medications\(^1\) and are a frequent cause of hospitalization and death\(^2,3\). One of the objectives of pharmacoepidemiology and pharmacovigilance is to identify and gather consistent evidence on the associations between drug use and the occurrence of adverse events. Such evidence grounds decision-making processes with regard to health surveillance.

The use of consistent evidence presupposes that causality inferences can be designed by examining the etiological link between drug exposure and an adverse event\(^4\). Determining causality presupposes an association between at least two phenomena and, in a simplified way, consists of answering the following question: Is (or would) the factor F a cause the adverse event E? This question assumes a temporal relation in which exposure precedes occurrence of the event\(^5\), and the supposed causal relation is reinforced by the frequency of such observation\(^6\).

The understanding of causality is regarded as one of science’s greatest achievements\(^6\) and has been, over time, a continuing concern of both philosophers and epidemiologists. The former have devoted themselves to studying the fundamental meaning of the notion of cause and the general principles of causality, while epidemiologists are interested in the identification of causes, in the quantification and characterization of effects, in the design causal models, and in examples of cause and effect relationships\(^7-9\). For Bhopal (2002, p.98)\(^6\), the purpose of studying causality in epidemiology is to produce knowledge on the prevention, cure, treatment and control of diseases and other health problems.

This article aimed to present considerations on one of the main conceptual and methodological challenges in pharmacoepidemiology and pharmacovigilance: the causal relationships between drugs and adverse events. It emphasizes the definition of causes and their types, Rothman’s multi-causal model\(^10\), the determination of causality, and the causal relationship from the perspective of sanitary surveillance. It does not intend to exhaust the
subject nor reach a consensus on the presented content, but rather to contribute to an initial sketch of causality in pharmacoepidemiology and pharmacovigilance, while trying to connect it to the practices of sanitary surveillance.

THE DEFINITION OF CAUSE

Hume, in the eighteenth century, wrote that causes are invariably followed by their effects. In contemporary epidemiological thought, Rothman defines cause as an act, event or a state of nature that initiates or permits, alone or in conjunction with other causes, a sequence of events that result in an effect. Susser resorts to Hume to propose essential properties in the recognition of cause that are fundamental to infer causality:

1. there must be an association (cause and effect occur together);
2. temporal order (cause precedes effect);
3. connection or direction (links between cause and effect can be predicted).

Rothman and Greenland incorporate the dimension of temporality into the definition of cause: a cause of a disease is an event, condition or characteristic that preceded the illness and without which the disease would not have occurred at all. For Bhopal, cause is comprised of something that alters the frequency of the disease, health status or associated factors in the population.

No definition of the term “cause” was found in didactic materials on pharmacoepidemiology. The omission may be related to the difficulty in defining a cause in the context of determining the causality between medication and an adverse event, since, in general, it may involve a mixture of causes, and it is therefore preferable not to speak of “causes”, but rather of “risk factors” or “association factors”. Both terms do not necessarily imply a causal relationship. Another possible explanation is that causality, not the elements that compose it (cause and effect), is a topic studied and prioritized in the chapters on methods used in pharmacoepidemiology and pharmacovigilance, although the definition of what is a cause has been the source of discussion by epidemiologists.

TYPES OF CAUSES

Parascandola and Weed found five different definitions of the types of causes in the area of epidemiology: production, necessary, component-sufficient, probabilistic, and counterfactual causes.

The definition of cause production presupposes that a cause “creates” or “produces” effects, giving rise to an “ontological distinction” between causal and non-causal associations, even though this characterization is somewhat vague in the author’s own view. The necessary cause depicts a condition without which the effect will not occur. On the other hand, its presence does not result unequivocally in the occurrence of the event. The view that
causes must be necessary for the occurrence of their effects is traditionally associated with germ theory, in which there is the assumption that the disease is motivated by at least one specific infectious agent. Such a situation is rare within the drug-adverse event causal relationship, and examples such as gray baby syndrome following the use of chloramphenicol are difficult to find in clinical practice.

Some causes are necessary - though not sufficient - for the occurrence of a disease. One explanation for this is that the name of some syndromes or diseases is defined by the triggering exposure, that is, the causal agent. Thus, berylliosis cannot occur in the absence of exposure to beryllium, and the framing of a problem such as ADE requires that a drug has been used. For example, in various diagnostic codes of the International Classification of Diseases (ICD-10), the drug is identified in the description of the clinical problem.

One of the limitations experienced by the process of causal inference in pharmacoepidemiology and pharmacovigilance is that some diseases have causal factors without a defined component as a necessary cause. In a series of cases of fulminant hepatitis, depending on the epidemiological situation, it cannot be stated that all occurrences are related to a drug, since other agents such as hepatitis A or B viruses and cytomegalovirus are also triggers of this health problem. In most situations the “drug” cause for a given adverse event is competing or perhaps interacting with several other possible causes.

The definition of a component-sufficient cause was proposed by Rothman and expands the view of the necessary cause, although it retains a similar deterministic scientific language. Component-sufficient causes are composed of a number of component causes that, alone, are not sufficient for the occurrence of the adverse event. A set of component causes constitutes a sufficient cause, ensuring the occurrence of the event in question.

The main characteristic of the probabilistic causes is that they increase the possibility of the occurrence of their effects, that is, C causes E if, and only if, C increases the probability of the occurrence of E. The probabilistic cause does not need to be necessary or sufficient, nor does it exclude the necessary and sufficient causes. Its definition is broader than that of a component-sufficient cause.

Finally, a counterfactual cause is the difference in the result - or the probability of the result - when it is present compared to when it is absent, that is, it is the result of T minus the absence of T. It compares the frequency of the occurrence of observed outcomes among exposed individuals compared to those that were not exposed, ensuring comparability. Counterfactuality is explored from experimental and observational studies, and in the latter, the presence of residual confounding does not always allow for unambiguous inferences of cause and effect to be established.

The definition of probabilistic cause combined with the counterfactual condition provides better substrate in pharmacoepidemiology. That is, C causes E if the probability of E - in the presence of C - is greater than that of E in the absence of C, under conditions that ensure comparability. An example of the probabilistic-counterfactual combination is the tetanus vaccine (TT), which causes Guillain-Barré syndrome (GBS) with limb paralysis if the probability P (GBS | TT). As such, the probability of GBS in the presence of TT is greater than that of GBS in the absence of TT – this, in ceteris paribus conditions.
These requirements simplify the causal structure, giving opportunities for the acceptance of the probabilistic-counterfactual combination\(^{23}\).

In the perspective of pharmacovigilance, Edwards\(^5\) also defines two types of causes to explain the possible ways in which they can act:

1. Contributory causes: actively involved in the adding of an effect, such as a relative overdose or drug interactions;
2. Contingent causes: essential for the occurrence of the effect, but which have no causal effect per se and which may be unknown, such as a particular cytochromometabolic enzyme phenotype that makes certain patients more susceptible to the effects of a drug.

The former are possibly intrinsic and influence other causes, but are not necessary, like the long half-life of a drug. They can be modified, influencing the causal relation, while the contingents cannot be altered.

They are also rarely routinely investigated in pharmacovigilance practices, as the emphasis is often on the drug as a possible cause rather than on other contributory causes, as is the case with medication errors\(^5\).

There is still another typology of causes in the epidemiological literature that is based on the processes of health determination, which presuppose causal determinants, characterizing them as distal (structural or socioeconomic), intermediary (behavioral characteristics, for example) and proximal (biological characteristics, for example) causes. In the accidental intoxication of children, the drug would be classified as a proximal cause, while the omission of sanitary legislation on safety systems in drug packaging would be framed as distal.

Bégaud’s\(^4\) pharmacoepidemiological dictionary defines two types of cause: necessary and sufficient. Laporte and Tognoni’s\(^15\) book, a reference in the area of pharmacoepidemiology and pharmacovigilance, makes no mention of the types of causes addressed in this article. This finding was also verified in other books on pharmacoepidemiology\(^16\)\(^-\)\(^19\).

The definitions and types of causes have influenced pharmacoepidemiological thinking on causality\(^6\), promoting the design of causal models, such as the multicausal approach proposed by Rothman\(^10\). The mapping of the different causes, including typifying them in the context of a causal structure under sanitary investigation, can facilitate the identification of the causes that are essential in the maintenance of the problem. Identification makes the planning and prioritization of regulatory corrective actions in pharmacovigilance more efficient, since as the causal structure is broken, the event can be reduced, eliminated or prevented\(^24\).

**ROTHMAN’S MULTI-CAUSAL MODEL**

Epidemiologists have sought to construct causal models in attempts to understand and explain how events occur from certain causes\(^10\)\(^-\)\(^24\). None of the pharmacoepidemiology books surveyed mention any kind of causal model as a form of representation of the causal relationship,\(^4\)\(^\)\(^-\)\(^19\), although most of these references cite the expression “multifactorial causality”\(^4\)\(^\)\(^\)\(^-\)\(^15\)\(^\)\(^-\)\(^18\).
The years 1957 and 1960 showed the first mention of multi causality models (“causation web”) in the epidemiological literature. According to Krieger, this model was not designed to provide explanations for causal relationships, but to increase the capacity of epidemiologists to describe and study the complex interrelationships between risk factors and diseases. Based on his work, important inferences about prevention and research have been made, and they remain to this day as part of epidemiological thinking: “to carry out preventive measures, it is not necessary to understand causal mechanisms in their totality” and “even the knowledge of a small component may allow some degree of prevention.”

In an article published in 1976, Rothman proposed a conceptual model of multi-causality called component-sufficient causes. It illustrates several relevant principles about causes. Perhaps most importantly, it shows that a given disease can be motivated by more than one causal mechanism and that each involves the joint action of various causes. In this model, the causal agent may be composed of a constellation of causes (three or more, for example) that are also considered to be sufficient for the occurrence of an adverse event. That which makes up a sufficient cause is called a component, and there may be a minimum number of components necessary for an adverse event to occur. In the composition of a constellation of causes, there is almost always a genetic and environmental origin.

In suspected ADE, the drug is referred to as a component cause, even though other causes are needed. For example, there are people who, by virtue of their genetic makeup or environmental experience, are susceptible to anaphylactic reactions caused by medications, while others are not. These susceptibility factors are component causes of complete causal mechanisms through which the drug causes this type of reaction. A complete causal mechanism forms a sufficient cause, and there may or may not be a sharing of common component causes between different complete causal mechanisms.

Rothman’s multicausal model further postulates that several causal components act together to produce an effect. This does not necessarily imply that component causes must act at the same time. Patients’ hypersensitivity reactions to drugs justify this assertion, because during the patient’s first contact with a drug, this type of adverse reaction does not usually manifest itself, but rather what occurs is the production of antibodies that will be a component cause for the manifestation of the side effect in the patient’s second exposure to the same drug or to another drug that is part of a similar therapeutic class. This whole process occurs at different time intervals.

Drug intoxication, a type of ADE, can be seen in Rothman’s multi-causal model as the result of a number of different sufficient causes, each of which presents distinct doses of the drug as a component cause. Smaller doses probably require a more complex set of component causes for intoxication to occur, when compared to exposure in higher doses. It should be remembered that dosage is directly related to increase of risk and inversely to time, and alterations in the composition of these variables give rise to different intensities of the effect of the drug with regard to intoxication. In addition, the individual’s susceptibility is a component cause to be considered in the manifestation of this type of adverse event, although it seems probable that there are similarities in the components of sufficient causes of different individuals.
According to Rothman and Greenland\textsuperscript{25}, it can be assumed that no cause is self-sufficient to result in the occurrence of an injury. In order to explain why an adverse event, such as Reye’s syndrome, occurred in a patient who ingested acetylsalicylic acid, other components of a complex causal model can be identified, such as underlying disease, the patient’s age, genetic predisposition, and nutritional state. In this case, such a model, which may involve different causal mechanisms, was sufficient for the onset of this syndrome\textsuperscript{22}.

In practice, surveillance and control of causative components of a causal model that was designed in response to an injury can be understood in terms of time and geographic space. This argument corroborates the existence of a different type of decision-making, which saves time (or not) when adopting corrective actions, as well as increases sanitary authorities’ perception of risk in countries facing an imminent risk to society. An example of this was the delay (time) by health authorities in Brazil and other countries (geographical area) in the prohibition of the use of thalidomide for nausea in pregnancy, which resulted in a greater number of children born with congenital malformations\textsuperscript{26}. In this public health tragedy, a key component that allowed for the continuation of the phocomelia outbreak was the delay in defining sanitary legislation, which would prohibit or restrict the indication of the use of this drug.

**DETERMINING A CAUSAL RELATIONSHIP**

Establishing causal inferences between medications and adverse events is a complex and difficult process\textsuperscript{15,27}. It is worth noting that causal models in pharmacoepidemiology and pharmacovigilance do not always refer to the typology of causes and their meanings mentioned above. In some situations, causality may be direct and the cause and effect can be easily perceived and defined. Fire is a direct cause of a burn\textsuperscript{1}. In other circumstances, establishing a causal relationship is quite challenging. For example, it took 20 years to associate the use of aspirin with the increased incidence of gastric ulcer bleeding,\textsuperscript{28} and it is even more difficult and complex to identify and associate the cause with the effect in chronic drug intoxications\textsuperscript{29}. One of the difficulties in determining a causal relationship is the presence of alternative explanations, biases (systematic errors) and confounding factors.

Faced with complexity, technical-scientific thinking about the causality of ADE has evolved in two broad areas of public health: pharmacoepidemiology and pharmacovigilance\textsuperscript{30}. This evolution is demonstrated through the prominent mention of alternative explanations for the determination of causality in textbooks on pharmacoepidemiology\textsuperscript{4,15-18}. Strom, for example, devotes a chapter exclusively to bias and confusion in this science,\textsuperscript{17} while Yang and West-Strum cite Sir Austin Bradford Hill’s causal aspects for its determination\textsuperscript{18}.

Pharmacoepidemiology is a branch of epidemiology that includes the use of epidemiological concepts and methods in studies looking at the uses and effects of drugs in large populations\textsuperscript{31}. The first mention of this science in the scientific literature occurred in the early 1980s.\textsuperscript{32} The American academic view encompasses pharmacovigilance and epidemiological studies, that is, it is not restricted to drug safety alone\textsuperscript{32}.
Much of the consistent evidence regarding the causal relationship of drug-adverse events comes from pharmacoepidemiological studies. Nevertheless, a study on its own, however well designed, is not enough to refute a causal relationship in an individual. Other relationships, including causal ones, are still possible, although unlikely, when they involve unexpected and unrecognized alternative explanations that may lead to an adverse event of a lesser incidence than the power of the study could demonstrate.

A constant challenge in pharmacovigilance – the science and activity related to the detection, evaluation, understanding and prevention of adverse reactions or any other possible problems related to drugs - is that complete data that allows for the evaluation of causality using different methods, such as the Naranjo algorithm used to establish cause-and-effect relationships, is not always available. And this is a critical aspect for pharmacovigilance systems. Equally important is the presence of uncertainties regarding the nature of the evidence of causation, which arises from spontaneous reports of adverse events, characterized mainly by high underreporting. Another limitation concerns the binomial suspension-reintroduction of the drug — considered as a reference standard for the establishment of cause and effect relationships in pharmacovigilance — which is not very effective, for reasons of safety and ethics.

In order to bypass its limitations, health regulations in Europe have incorporated different methods that contribute to the surveillance and control of ADE, including epidemiological methods, such as observational studies. Unlike pharmacoepidemiology, pharmacovigilance is exclusively concerned with drug safety.

Although there are advantages to the inferential process of causality from prospective pharmacoepidemiological studies in comparison to retrospective ones, for both types of study, determining that a causal relationship exists requires a large number of exposed and non-exposed individuals that can be monitored for long periods of time. Thus, these studies are not the best option for short and medium term decision-making in public health emergency situations, particularly in sanitary surveillance. Investigations in the field of pharmacoepidemiology may be useful with regard to this limitation.

In addition to analytic studies, Hill’s views, published in 1965, have been suggested as important aspects to be considered when inferring causality between exposure and outcome from noncommunicable diseases. They include: strength of association, consistency, specificity, temporality, biological gradient, plausibility, coherence, experimental evidence and analogy. These points of view are presented and discussed in textbooks on pharmacoepidemiology. Hill’s aspects were important in order to affirm the counterfactual approach in the development of epidemiological studies, but they are still of little empirical weight.

Shakir and Layton cite that Hill’s points of view can be applied when characterizing causality inferences in pharmacovigilance as long as the specificities of the available data, and factors such as underreporting, misclassification and poor quality of information are considered.

Any causal consideration based only on pharmacoepidemiological studies is not logically defensible and certainly not socially acceptable. To give an example, the safety signal has been convincingly initiated because of a series of spontaneous reports of cases of
severe arrhythmias following the use of cisapride, including evidence following the re-introduction of the drug. No cardiac arrhythmia had been observed in these studies and the safety signal was not accepted until a causal mechanism was discovered — a prolongation of the QT interval — on the echocardiogram. This serious effect should have led to timely corrective actions, such as warnings about the risk of its use in patients prone to heart rhythm irregularities.

Another important point in determining causality is the difference between association (correlation) and causality. A substantial part of the available scientific evidence in the literature is based on the association of variables, which are insufficient to demonstrate causality. The assumption that A causes B, simply because A is associated with B, is an error, and in ecological studies is called an ecological fallacy. However, sometimes a reverse fallacy occurs, and is represented by the discarding of certain associations, as if they were not sufficient to demonstrate causality. It is emphasized that a single association is not sufficient to reach a conclusion on causality, but multiple associations may conclude that A causes B, in conjunction with biological plausibility.

In pharmacovigilance, when investigating an individual case of an adverse event, whose drug is considered suspect, the probability of having a causal relationship is not enough to ensure definitive conclusions. Lack of an experimental design limits the validity of causal inference in pharmacovigilance. This does not reduce the relevance of the research process in this science. An analysis of at least five aspects should be considered when determining causality in pharmacovigilance:

1. temporal relationship;
2. data on suspension-reintroduction of the drug;
3. relationship between the event and the underlying disease;
4. presence of a more probable cause;
5. information on biological plausibility.

FINAL CONSIDERATIONS FROM THE HEALTH SURVEILLANCE PERSPECTIVE

One of the possible questions that can be asked in the area of health surveillance is how to use the knowledge about causality produced by pharmacoepidemiology and pharmacovigilance to avoid damage or improve population health by generating stable evidence. This is a complex task because certain observational studies in the literature have suggested causal relationships that were later refuted when tested in subsequent studies.

Causality is of great practical relevance in sanitary surveillance in that it signals the need or at least the possibility for basing regulatory decisions that reduce exposure to damaging drugs or increase exposure when the drug is beneficial. Decisions range from the inclusion of new warnings and label information to the drug being withdrawn from the market. This gradient of decisions implicitly and sometimes explicitly incorporates the recognition of the causal relationship, as well as considers the frequency and severity of the event in
question. In other emergency situations in health surveillance, although the causal enigma has not been satisfactorily elucidated, decisions are made, including that of maintaining the status quo or making use of the precautionary principle\textsuperscript{42,43}.

Despite recognizing that there are limitations on the causality evidence from pharmacoepidemiology and pharmacovigilance, it is still possible, in some contexts, to use them for health regulatory actions that promote and protect population health. The main challenge is to decide whether the knowledge generated about causality is reproducible and stable, that is, if a finding will not be quickly countered by subsequent scientific research. A practical way of experiencing this is to demonstrate that the causal relationship in question cannot be easily mistaken, given the best current scientific evidence available\textsuperscript{9}.

As such, the role of pharmacovigilance systems is to identify results that have a good chance of being true and consistent. It should be pointed out that such results can be in clear opposition to strong economic interests, which are capable of generating new research often with a high risk of bias\textsuperscript{9}. Tobacco companies did this by funding a large amount of research that contradicted the claim that there was a causal relationship between smoking and certain diseases. However, the plausibility in the causal nexus proved sufficient to support the development of widely recognized policies that control tobacco smoking in several countries\textsuperscript{9}.

However, many decisions in sanitary surveillance are not adopted merely by the scientific and technical evidence of causality. There is always an “extra” component in addition to the evidence. For example, the government may choose not to directly ban a drug because of civil liberties, and propose legislation that only restricts its use,\textsuperscript{9} because evidence of causation alone cannot tell whether a total ban on a drug corresponds to the most appropriate political measure.

An illustration of the “extra” component is the controversy surrounding the resolution of the Brazilian Health Regulatory Agency (Agência Nacional de Vigilância Sanitária – ANVISA), which in 2011 banned the marketing of three appetite suppressants and imposed restrictions on the sale of sibutramine in Brazil. The decision was based on causal evidence pointing to how the risks caused by the drugs outweigh the benefits in the treatment of obesity. After almost three years, the sanitary standard was repealed by Legislative Decree No. 273 on September 5, 2014. One of the justifications of the decree was that ANVISA’s resolution caused “great dissatisfaction among the medical profession, constituting a setback to the treatment of obese people in the country”\textsuperscript{45}.

Considering different scenarios arising from academic and regulatory debate, determining causality presupposes the integration of theoretical and methodological references, such as pharmacoepidemiology and pharmacovigilance. It continues to be matter of judgment, rarely based on only one study\textsuperscript{46} and it is fundamental in the evaluation of drug safety\textsuperscript{30,47}. The adoption of these principles contributes to the improvement, the strengthening and the opportunity of sanitary surveillance actions with respect to the safety of the patient and the population. New analytical approaches using large data volumes, including data and text mining strategies and so-called “machine learning”, are already a reality in pharmacoepidemiology and pharmacovigilance,\textsuperscript{48} and are requiring new and dynamic approaches in causality\textsuperscript{49}. 


REFERENCES