Association and causation

A guide for policymakers on the science of determining the causes of diseases and other health outcomes

VICTORIA State Government | Health and Human Services

To receive this publication in an accessible format phone 9096 2237 using the National Relay Service 13 36 77 if required, or email Health Intelligence unit <health.intelligence@dhhs.vic.gov.au>.

Authorised and published by the Victorian Government, 1 Treasury Place, Melbourne.

ISBN 978-1-76069-402-9 (pdf/online/MS word)

Available at Health status of Victorians <https://www2.health.vic.gov.au/public-health/population-health-systems/health-status-of-victorians>

Contents

Introduction **3**

Intended audience 3
Evidence-based policymaking 3
Determining causation starts with identifying an association 2

A case study – smoking and lung cancer **3**

Background 3
The evidence trail 3
Summary and conclusions 3

Definitions of forms of evidence **4**

1 Strength of the association 4
2 Consistency of the association 4
3 Temporal relationship 4
4 Dose–response relationship 4
5 Alternative explanations 5
6 Biological plausibility of the association 5

References **7**

Introduction

Without evidence, policy makers must fall back on intuition, ideology, or conventional wisdom—or, at best, theory alone. And many policy decisions have indeed been made in those ways. But the resulting policies can go seriously astray, given the complexities and interdependencies in our society and economy, and the unpredictability of people's reactions to change (Banks 2009).

Intended audience

The purpose of this paper is to describe to policymakers and others working with evidence or data the difference between **association** and **causation**. These are terms used by epidemiologists to study and analyse population health, and they help describe how and when a health risk factor can be deemed to *cause* a disease or other health outcome. While this paper uses a case study that relates specifically to health, the same principles described in this paper can be applied to areas beyond health, such as human services.

Just because two things are associated, does not necessarily mean that one thing causes the other.

Several studies found an association between drinking coffee and lung cancer (Alicandro et al. 2017). However, drinking coffee does not cause lung cancer. The association exists because coffee drinkers are more likely to smoke.

Often, the media, the general public, and indeed policymakers working with evidence and data jump to a conclusion of cause-and-effect when two things are found to be associated with each other. However, just because two things are associated, does not necessarily mean that one causes the other. Understanding this distinction is crucial for governments to design sound evidence-based policies for better health and human services outcomes.

Evidence-based policymaking

Evidence-based policymaking has seen significant progress in recent years at the federal, state and local levels. Evidence-based policymaking is critical for governments to ensure that they address the most important contemporary challenges while ensuring that public funds are used as effectively and efficiently as possible (Evidence-based Policymaking Collaborative 2018).

The first step in evidence-based policymaking is to identify the problem that needs to be addressed. For health, it is usually a disease or health risk factor. Understanding the cause of a disease or health risk factor is crucial to determining the best course of action to treat or prevent current or future cases and improve the overall health of the population.

This paper will talk through the steps that determine whether an exposure causes an outcome. This will support policymakers and others to think critically about different forms of evidence.

Exposure refers to any factor that may be associated with an **outcome** of interest. For example, smoking is the exposure and lung cancer is the outcome.

Determining causation starts with identifying an association

How do we determine if an exposure causes an outcome? The first step to identifying a potential causal relationship between an exposure and an outcome is to investigate whether there is an 'association' between the exposure and the outcome.

An **association** is defined as an exposure and outcome occurring together more or less often than would be expected by chance. This means there is a statistical relationship between the exposure and the outcome.

If the probability of the outcome is significantly greater in the presence of the exposure than in the absence of the exposure, then it can be deemed that there is an association between the exposure and the outcome. For example, the occurrence of lung cancer among people who smoke is far greater than the occurrence of lung cancer among people who do not smoke.

The difference between association and correlation

In everyday language, 'association' and 'correlation' tend to be used interchangeably. Technically, 'correlation' has a statistical meaning to do with the strength of the relationship between two things; they both increase and decrease together, or as one increases the other decreases or vice versa. For example, the more hours you spend in direct sunlight the more severe your sunburn. 'Association' is a broader concept around whether two things are related in some way. Correlation can help define this relationship. Importantly, as this paper discusses, just because two things are related (either by association or correlation), does not mean that one causes the other.

Once an association between an exposure and an outcome has been demonstrated, the following questions need to be considered before any claim that the exposure caused the outcome can be made:

1. What is the strength of the association?
2. Is the association between an exposure and an outcome a consistent finding?
3. Is there a temporal relationship between the exposure and the outcome?
4. Is there a dose–response relationship between the exposure and outcome?
5. Is there something else that might explain the association between the exposure and the outcome?
6. Is it biologically plausible that the exposure could cause the outcome?

These questions are considered throughout the following case study. More detailed descriptions of each question follow the case study.

A case study – smoking and lung cancer

Background

Prior to the 20th century the type of lung cancer that we now know to be caused by smoking was such a rare disease that only 140 cases were documented in the published medical literature in 1900 (Proctor 2012). In the late 1800s, cigarette-making machines were invented and enabled the rapid expansion of the tobacco industry supported by mass marketing of their product.

Gradually the scientific community began to notice a dramatic increase in cases of lung cancer, and by the 1940–50s it became apparent that we were in the midst of an unprecedented epidemic of lung cancer. The fact that this coincided with a rapid increase in the uptake of cigarette smoking presented the first clue.

In 1950 the epidemiologists Richard Doll and Bradford Hill published their seminal paper in the *British Medical Journal* that presented the evidence to suggest that smoking was the cause of the lung cancer epidemic (Doll and Hill 1950).

However, it took until 1964 for the US Surgeon General to formally announce that smoking caused lung cancer. Thereafter, the obvious policy response was to try to reduce and ultimately eliminate smoking.

The evidence trail

Determining causation starts with identifying an association.

As health practitioners began to notice the rapid increase in cases of lung cancer, they also noticed that this rise coincided with a rise in cigarette consumption. In 1939 the first study was published that compared 86 people who had been diagnosed with lung cancer with people who did not have lung cancer. The study found that those with lung cancer were more likely to smoke than those without lung cancer. This was the first scientific demonstration of an *association* between smoking and lung cancer.

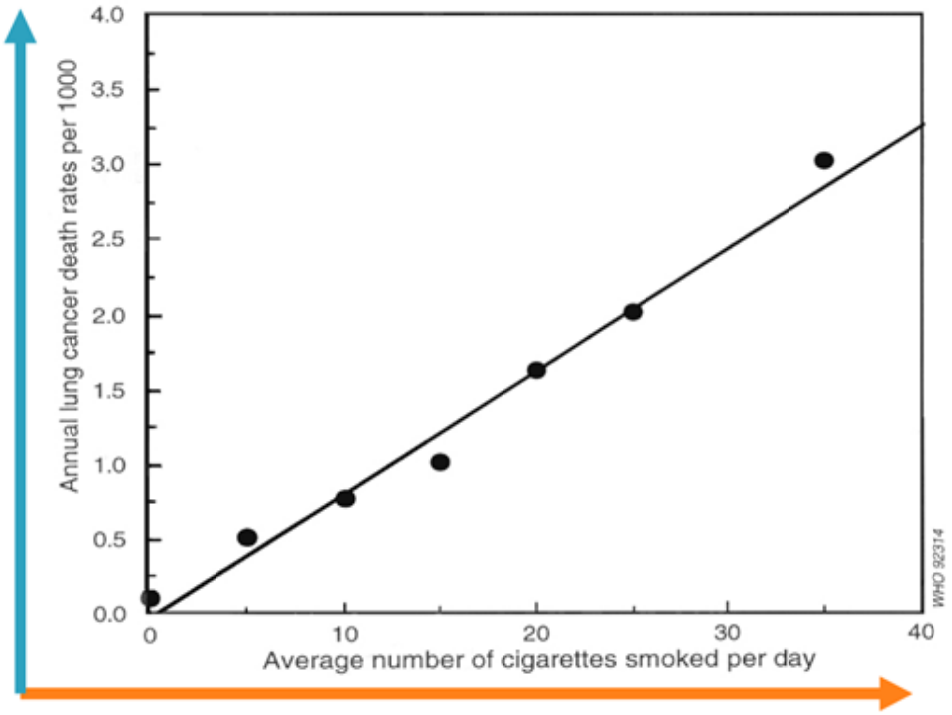
1. The *strength of the association* was high: smokers who smoked 35 cigarettes or more per day were 40 times more likely to develop lung cancer than non-smokers (Proctor 2012).

2. Many more studies in the United Kingdom and the United States confirmed the original finding of an association between smoking and lung cancer, demonstrating the *consistency of the association*.

3. Then came several studies that sought to determine if smoking preceded lung cancer: smokers and non-smokers were followed over time to see who developed lung cancer. The results were unequivocal; smoking preceded lung cancer, demonstrating a *temporal* relationship.

4. These studies also showed a *dose–response relationship* where the greater the number of cigarettes smoked, the more likely a person was to die from lung cancer (Figure 1).

Figure 1: Death rates from lung cancer (per 1,000 people), by number of cigarettes smoked, British doctors, 1951–1961 (Doll and Hill 1964)



5

While early epidemiologists lacked the statistical tools that are available to epidemiologists today to test for potential confounders, they were able to rule out *other explanations* for the rise in lung cancer, such as asphalt dust from newly tarred roads and exposure to poison gas in World War I (Proctor 2012).

A **confounder** is a third factor or variable that has a hidden effect on both an exposure and the outcome of interest. The term 'confound' comes from the Latin *confundere* meaning to pour together or mix. The English word 'confuse' arises from the same Latin root (Babyak 2009).

6

During this period, parallel evidence from other investigative fields began to accrue and provided the evidence base for the *biological plausibility* of smoking causing lung cancer (Proctor 2012):

- Experiments in animals revealed that cancerous tumours could be induced on the skin by painting the skin with cigarette smoke tars.
- Pathologists began seeing damage to the cilia of deceased people who had died of lung cancer as well as other precancerous changes. The cilia are the hair-like structures that line the upper respiratory tract and help to clear the airway passages of mucus and dirt. They speculated that this might cause the particulate matter in cigarette smoke to become trapped in the lungs.
- Other scientists began to discover the presence of a number of cancer-causing chemicals in cigarette smoke.

The confluence of these diverse forms of evidence, combined with diminishing evidence of alternative explanations, resulted in the American Cancer Society's National Board of Directors publicly declaring in 1954 that there may be a causal link between smoking cigarettes and lung cancer. Figure 2 summarises the evidence.

Figure 2: A summary of the evidence that smoking causes lung cancer

Does cigarette smoking cause lung cancer?	
Start with an association	**Yes, there is an association**
① **Strength**	Very strong
② **Consistency**	Highly replicable
③ **Temporality**	Smoking precedes lung cancer
④ **Dose response**	Yes – the more cigarettes smoked, the more likely a person is to develop lung cancer
⑤ **Alternative explanations**	Ruled out asphalt, poison gas from World War I, etc.
⑥ **Biological plausibility**	Animal experiments, clinical observation, discovery of cancer-causing chemicals in cigarette smoke

As noted earlier, it was not until 1964 that the US Surgeon General formally announced that smoking causes lung cancer. The work of Doll and Hill is a fascinating case study of how scientific evidence alone is rarely enough to sway wider society about the causes of disease and ill-health. The powerful tobacco industry successfully delayed the attempts of health professionals to convince the public that smoking causes lung cancer.

Much can be learnt from the tobacco story for today's pressing health concerns such as obesity. For further reading on this subject, we recommend an article by Kelly Brownell and Kenneth Warner from the Rudd Center for Food Policy and Obesity at Yale University (Brownell and Warner 2009) .

Summary and conclusions

Robust and publicly available evidence serves as an important counterweight to the influence of sectional interests. However, evidence building takes time, and when there is lack of sufficiently good evidence, policymaking cannot come to a standstill. Therefore policies are developed while the evidence base continues to build. For example, policymaking around halting and reversing the obesity epidemic continues in the face of a lack of scientific consensus about its cause. As the evidence base grows, some policy directions will need to be strengthened while others revised or abandoned. Understanding the distinction between association and causation and keeping up to date with the literature will inform and maximise good policymaking.

Definitions of forms of evidence

1 Strength of the association

There can be strong or weak associations between an exposure and an outcome of interest. A strong association is more likely to be causal than a weak association. A weaker association is more likely to be explained by measurement error and other study design biases. However, the presence of a weaker association does not rule out causation.

Case study

The strength of association between smoking and lung cancer is very strong. Various studies have found that smokers are between four and 20 times more likely to have lung cancer than non-smokers (Proctor 2012).

2 Consistency of the association

A single study is insufficient evidence of a causal link between an exposure and an outcome. We need to make repeated observations of an association between an exposure and an outcome in different studies, under different circumstances, at different times, by different investigators and in different populations.

When we observe a consistent association between an exposure and an outcome, this strengthens the argument that an exposure may cause an outcome. This does not mean that every single study must find an association between the exposure and the outcome. Consistency is achieved if the majority of studies find an association.

3 Temporal relationship

The exposure must come before the outcome. If the exposure did not precede the outcome then logically it cannot have caused it. Therefore a lack of temporality **rules out** causation – hence this is one of the most important criteria for determining causation.

A common error is to use cross-sectional data as evidence that an exposure causes an outcome. The most common example of cross-sectional data is data that is obtained from a population health survey. For example, while survey data may identify an association between obesity and depression, the data cannot be used to determine causality or its direction because there is no information to indicate whether an exposure preceded or followed an outcome.

Policymakers use population health surveys because they are a relatively inexpensive tool that monitors the health of a population and identifies and quantifies associations between exposures and outcomes. These insights can generate hypotheses about potential causal pathways that can be tested in future epidemiological studies.

4 Dose–response relationship

A dose–response relationship exists when the level or ‘dose’ of an exposure increases and the frequency of the outcome of interest also increases.

Case study

As the number of cigarettes smoked per day increases, so does the risk of lung cancer. The presence of a dose–response relationship is strong evidence that a causal relationship exists.

If a dose–relationship is found to exist, this strengthens the argument for causality. However, the absence of a dose–response relationship does not necessarily rule out a causal relationship. In some cases there is a threshold-level of exposure where the outcome will not occur until the threshold level has been surpassed. For example, a study investigating the risk of miscarriage among pregnant women exposed to trihalomethanes (a contaminant of chlorinated water supplies) found no increased risk of miscarriage until the contaminant level reached 75 mg per litre (Waller et al. 1998).

5 Alternative explanations

This is arguably the most important criterion to determine whether there is a causal relationship between an exposure and an outcome. Even if an association is observed between an exposure and an outcome and that association is strong and consistent and the exposure precedes the outcome, it does not mean that the exposure caused the outcome.

Case study

Coffee drinking does not cause lung cancer. However, coffee drinking is associated with lung cancer. This is because coffee drinkers are more likely to be smokers. The true cause of lung cancer is a third factor (smoking) that, when taken into account, eliminates the association between drinking coffee and lung cancer. Smoking is the real causal exposure, and the coffee drinking was merely an indicator of the presence of a third factor – smoking (Alicandro et al. 2017).

Ruling out alternative explanations is crucial to proving causation. To rule out alternative explanations, one must be able to identify potential alternative explanations. This is not always possible and depends on the current state of knowledge, which may change with time as further studies accumulate. This is the primary reason why causation is very hard to prove and why there is no consensus on what constitutes a sufficient body of evidence to demonstrate causation.

However, we cannot sit around waiting for scientists to prove beyond the shadow of a doubt that an exposure of interest causes an outcome before acting, particularly where there may be a serious danger to health. Nevertheless, when the evidence does accrue to disprove a causal link that might have been assumed and formed the basis for a policy response, policymakers will need to review and update policy responses accordingly. Understanding the distinction between association and causation and keeping up to date with the literature will inform and maximise good policymaking.

Case study

A study showed that people who smoked a pipe were more likely to die from cancer than people who smoked cigarettes: 36 per cent of pipe smokers died compared with 21 per cent of cigarette smokers. However, people who smoked a pipe were more likely to be older than people who smoked cigarettes. When the age difference between pipe and cigarette smokers was accounted for, the association linking the type of smoker (pipe or cigarette) and death not only vanished but reversed – the age-adjusted death rate was 30 per cent among cigarette smokers and 20 per cent among pipe smokers. Age acted as a third factor or ‘confounder’ that misled the investigators into wrongly concluding that pipe smoking was more harmful than cigarettes, when in fact it was the other way around (Babyak 2009).

6 Biological plausibility of the association

Biological plausibility refers to the scientific reasoning used to determine whether an exposure causes an outcome. It essentially answers the question: ‘Is there a logical and plausible biological mechanism to explain the relationship?’ To answer this the investigator looks at the current body of knowledge on

possible mechanisms by which an exposure could cause an outcome. If a causal relationship between the exposure and outcome is consistent with the body of knowledge then the criterion of biological plausibility is met. However, it is not essential that this criterion be met because a lack of biological plausibility may just reflect a current lack of sufficient knowledge about the pathogenesis of a disease.

The benefit of considering biological plausibility is that it is an objective use of logic and attention to the wider knowledge base. The alternative is the more subjective approach of relying on prior beliefs.

Case study

It is not biologically plausible that coffee would cause lung cancer given that people do not inhale coffee and that coffee is not associated with overall cancer risk (Alicandro et al. 2017).

References

Babyak, M. A. (2009). "Understanding confounding and mediation." Evid Based Ment Health **12**(3): 68-71.

Banks, G. (2009). Evidence-based policy making: What is it? How do we get it? (ANU Public Lecture Series, presented by ANZSOG, 4 Fenruary), Productivity Commission, Canberra. https://www.pc.gov.au/news-media/speeches/cs20090204/20090204-evidence-based-policy.pdf.

Brownell, K. D. and K. E. Warner (2009). "The perils of ignoring history: Big Tobacco played dirty and millions died. How similar is Big Food?" Milbank Q **87**(1): 259-294.

Doll, R. and A. B. Hill (1950). "Smoking and carcinoma of the lung; preliminary report." Br Med J **2**(4682): 739-748.

Doll, R. and A. B. Hill (1964). "Mortality in Relation to Smoking: Ten Years' Observations of British Doctors." Br Med J **1**(5396): 1460-1467 CONCL.

Evidence-based Policymaking Collaborative (2018). Principles of evidence-based policymaking. http://www.evidencecollaborative.org/principles-evidence-based-policymaking.

Proctor, R. N. (2012). "The history of the discovery of the cigarette-lung cancer link: evidentiary traditions, corporate denial, global toll." Tob Control **21**(2): 87-91.

Waller, K., S. H. Swan, G. DeLorenze and B. Hopkins (1998). "Trihalomethanes in drinking water and spontaneous abortion." Epidemiology **9**(2): 134-140.