Causal Inference and Assessment of Risk in the Health Sciences

Ryan T. Demmer and Panos N. Papapanou

Contents

The quest for causality in the health sciences has been ongoing since time immemorial and causal concepts are known to have been discussed by several philosophers in ancient Greece including Aristotle [1]. In the last two hundred years, many presumed causes of disease have been identified (e.g., smoking as a cause of cancer and cardiovascular disease, LDL-cholesterol and high blood pressure as causes of cardiovascular disease, *Mycobacterium tuberculosis* as a cause of tuberculosis, *Plasmodium falciparum* as a cause of malaria and, in the context of oral diseases, *Streptococcus mutans* as a cause of dental caries). Causal inferences of this nature, albeit often imperfect, have contributed to major advances in the health sciences towards reducing morbidity and extending life expectancy.

R. T. Demmer
Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA

Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA
e-mail: demm0009@umn.edu

P. N. Papapanou (✉)
Division of Periodontics, Section of Oral, Diagnostic and Rehabilitation Sciences, College of Dental Medicine, Columbia University, New York, NY, USA
e-mail: pp192@columbia.edu

I. L. C. Chapple, P. N. Papapanou (eds.), *Risk Assessment in Oral Health*,
https://doi.org/10.1007/978-3-030-38647-4_2

The discipline of Epidemiology has been central to causal inquiry for health outcomes in humans since the beginning of the nineteenth century. Indeed, most definitions of epidemiology are explicit about the importance of understanding determinants of disease as well as describing disease patterns. However, despite numerous historical examples of causal discovery, a number of surprising and/or inconsistent findings, particularly in regard to complex chronic disease aetiology, have weakened confidence in the causal models that helped to vanquish infectious diseases during the early twentieth century. This crisis of confidence in causal inquiry has inspired a level of criticism of epidemiological methodology at large. For example, in 1995, Gary Taubes published an article in Science discussing the challenges and limits of modern epidemiology [2]. Since then, a shift has occurred in the way that the health science professions, as well as the public at large, appreciate epidemiologic inferences as they relate to causal inquiry. Most scientists engaged in human-orientated research studies are well-aware of common misinterpretations: for example, 'correlation does not equal causation' is a commonly cited refrain, which while true, is an oversimplification of a complex thought process. Indeed, although non-causal correlations are abundant, this does not mean that every correlation is non-causal in nature. It is also increasingly common to encounter findings from human studies cited as 'epidemiological', with some scientific journals even cataloguing manuscripts under a specific 'epidemiology' section. Typically, in these situations, 'epidemiological' is an adjective used to specify the descriptive arm of epidemiology or to distinguish observational from interventional etiological epidemiological study designs. As such, this language is either incorrect or redundant.

To appreciate this debate, it is imperative to review the definition of 'a cause' and to understand the underlying logic and models used to identify causal relationships in the health sciences. With respect to the first point, one popular definition of a cause reads as follows: 'any factor without which the disease event would not have occurred, at least not when it did, given that all other conditions are fixed' [3]. To test causal hypotheses and identify causes, epidemiologists utilize a conceptual approach referred to as a 'potential outcomes' or—synonymously—a 'counterfactual framework'. A counterfactual framework observes the disease experience in a group of individuals exposed to a hypothesized cause and then inquires what the disease experience in that same group would have been, had they—counter to fact—*not* been exposed to the hypothesized cause during the same time period, with all other factors kept unchanged. The observations from a theoretical experiment of this nature would then yield a causal effect, which is defined as the proportion of exposed individuals who develop disease during a given time period, divided by the proportion of the same exposed individuals that would have developed disease, had they been unexposed during the same observation period. While this is a valuable thought experiment, it is untenable in reality. Therefore, a cornerstone of etiologic epidemiological designs is the use of group comparisons. All etiologic epidemiological study designs, including observational designs and randomized interventions, have been developed precisely to enable valid group comparisons that can approximate the counterfactual ideal and estimate causal effects.

What Is Risk and How Is It Measured

The concept of risk has served as a fundamental tool for inquiry regarding the occurrence of human health and disease. In the context of a counterfactual (or potential outcomes) framework, risk is a proportion that is numerically equivalent to the probability of disease occurrence defined as follows: the number of people who develop a condition divided by the number of at-risk individuals in the source population under study. In more precise epidemiological terms, risk is often referred to as cumulative incidence (CI); a visual representation of CI and the explicit formula is presented in Figs. 1 and 2. It is worth noting that this definition of risk explicitly requires the passage of time such that disease develops during a follow-up period among a subset of initially disease-free individuals. In contrast to incidence, prevalence reflects the probability of current disease. Prevalence is defined as a ratio of the number of existing cases at a point in time (or during a specific time period) over the total number of individuals in the population under study. For example, if the prevalence of diabetes is 14% in a particular country, this tells us that the probability of any randomly selected inhabitant having diabetes is 0.14 (or ~1 in 7 people). In contrast, if the cumulative incidence (or risk) of diabetes in 2018 is 14%, this tells us that during the 2018 calendar year, the probability of developing diabetes among the initially diabetes-free population is ~1 in 7. Another commonly used measure of disease occurrence is odds, which is defined as the probability of having the disease over the probability of being disease-free (i.e., 1-probability of disease). To state it another way, the cumulative

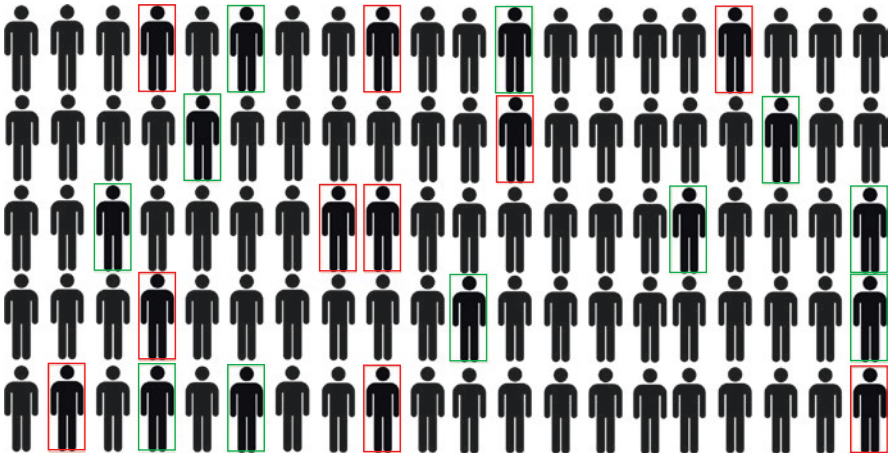
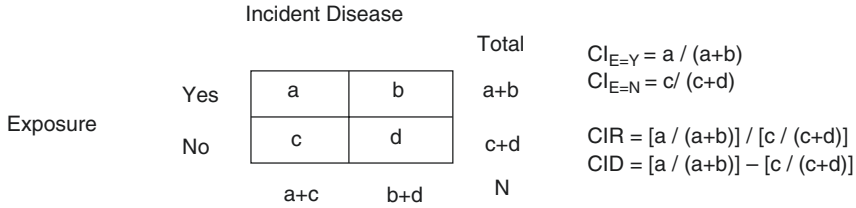


Fig. 1 Measures of disease frequency. n = 100 individuals enrolled into a longitudinal cohort study on January 1st, 2019 and followed for 20 years. Red borders signify disease present at the beginning of the study (baseline), n = 10. Green borders signify disease that developed during follow-up (incident disease), n = 11. Prevalence on January 1st, 2019 = 10/100 = 0.10 or 10%. Cumulative Incidence during the 20 year study period = 11 incident cases/(100 baseline − 10 prevalent cases) = 0.12 or 12%



Epidemiological Measures of Impact

Attributable Risk	$CI_{exposed} - CI_{unexposed}$
Population Attributable Risk	$P_{exposed}(CI_{exposed} - CI_{unexposed})$ or $CI_{all} - CI_{unexposed}$
Attributable Fraction	(CIR –1) / CIR or $(CI_{exposed} - CI_{unexposed}) / CI_{exposed}$
Population Attributable Fraction	$P_{exposed}(CIR -1) / [1 + P_{exposed}(CIR-1)]$ or $(CI - CI_{unexposed}) / CI$

Fig. 2 A 2 × 2 table summarizing the joint distribution of an exposure (i.e., risk factor) and incident disease. Cell a = the number of exposed individuals with disease; b = the number of exposed individuals without disease; c = the number of unexposed individuals with disease; d = the number of unexposed individuals without disease. $CI_{E=Y}$ = cumulative incidence of disease among the exposed. $CI_{E=N}$ = cumulative incidence of disease among the unexposed. CIR = cumulative incidence ratio. CID = cumulative incidence difference

incidence odds (CIO) of disease is simply = [CI/(1−CI)], and prevalence odds would be defined as [prevalence/(1−prevalence)]. Therefore, odds can be calculated and used in the context of both prevalence or incidence. Finally, the concept of incidence rate (or incidence density) is also of central importance to epidemiological inquiry and is closely related to the concept of risk. The incidence rate simply incorporates time explicitly into the denominator as follows: the number of people who develop a condition (incidence) divided by the person time contributed by initially disease-free individuals during the study period. Person time is calculated for each individual as the amount of time that passes between entry into the study and either: (1) the development of disease (or in many study settings disease diagnosis, which often differs from the precise time of disease development); (2) the end of the observation period; or (3) death or loss-to-follow-up. These concepts are demonstrated in Fig. 3.

Risk and Measures of Association

While measures of disease frequency, such as risk (i.e., cumulative incidence), are of value for a number of important reasons, risk is frequently used to assess the evidence for causal associations. This is typically done by comparing risk of disease between two different groups of individuals defined by variation in an ‘exposure’ or hypothesized risk factor. For example, consider the 2 × 2 tables in Figs. 2 and 4 which demonstrate different measures of association derived from risks (or rates) of disease among individuals exposed vs. those unexposed. Figures 2 and 4 define (1)

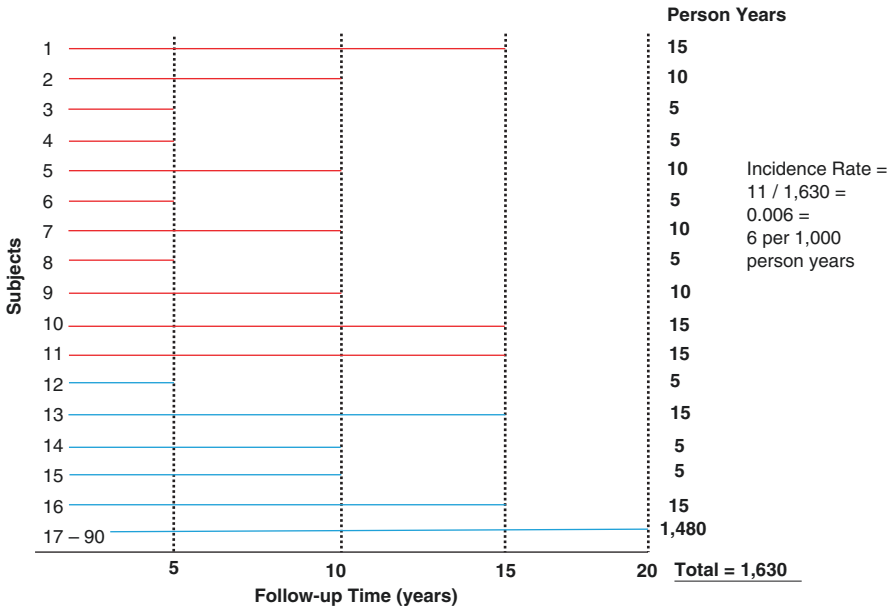


Fig. 3 Visualization of how person time accrues in longitudinal study designs. Red lines signify individuals that develop disease over the observation period, blue lines individuals who remain healthy

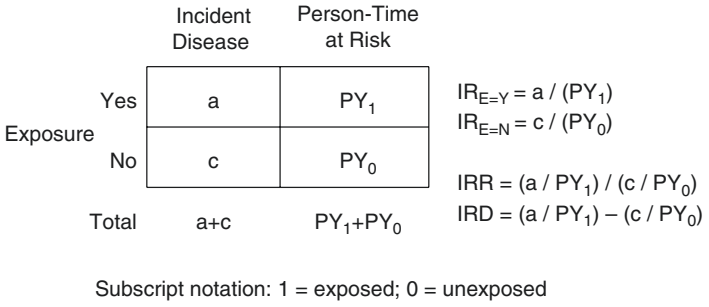


Fig. 4 A 2 × 2 table summarizing the incident disease and person time by exposure (i.e., risk factor) status. Cell a = the number of exposed individuals with disease; PY_1 = the total person time contributed by exposed individuals during the study; c = the number of unexposed individuals with disease; PY_0 = the total person time contributed by unexposed individuals during the study. $IR_{E=Y}$ = Incidence rate among the exposed. $IR_{E=N}$ = Incidence rate among the unexposed. *IRR* incidence rate ratio, *IRD* incidence rate difference

cumulative incidence ratio; (2) cumulative incidence difference; (3) incidence rate; and (4) incidence rate difference.

Based on the aforementioned measures of association, additional measures of impact used in epidemiology can be derived including: (1) attributable risk (AR, synonymous with the cumulative incidence difference—see Fig. 2); (2) population

attributable risk (PAR); (3) attributable fraction (AF); (4) population attributable fraction (PAF). These measures summarize the number of cases of disease that are the result of (i.e., attributable to) the exposure among different populations; the respective populations of interest being the exposed for AR, the total population for PAR, the exposed with disease for AF, and the diseased for PAF, respectively. Formulas for these measures can be found in Fig. 2. Note that the terminology for these measures varies considerably in the literature and one should always take careful note of the underlying formula used when interpreting the meaning of these measures.

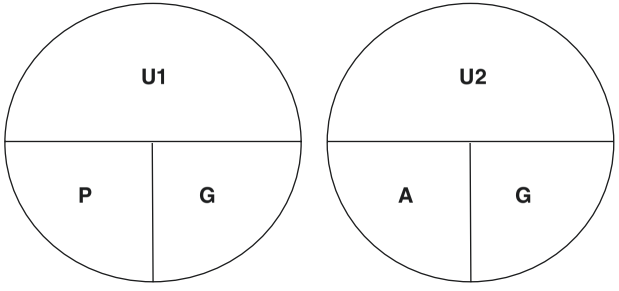
Causal Inference and Causal Models

It is frequently explicitly stated (or intimated) in the literature that observational designs, particularly cross-sectional and case-control designs, cannot be used to infer causality, but can only identify putative causal exposures that require testing in subsequent randomized controlled trials (RCTs) to provide definitive causal estimates. In fact, while experimental designs have the potential to provide less biased and/or confounded causal estimates, observational designs are both capable of and frequently used to inform causal relationships. Some examples follow to demonstrate this point. When exposure status clearly precedes the disease outcome and nature randomizes the exposure, observational designs can be quite powerful. For example, Mendelian randomization embedded in longitudinal observational cohort studies leverages the randomness of the meiotic process to inform whether hypothesized exposures cause disease. Even in a cross-sectional study design, a Mendelian randomization approach could potentially provide strong causal evidence since the genotype clearly precedes the disease outcome (i.e., clearly fulfils the temporality requirement) and the mutation in question was assigned by nature and thus cannot be confounded by events occurring during the life course, such as socio-economic status, access to health care, or health behaviours. Furthermore, in situations, where randomization is unethical, observational designs are generally the only feasible option. The establishment of smoking as a cause of lung cancer and cardiovascular disease using observational designs demonstrates this point. Importantly, a poorly conducted RCT—for example, one in which randomization is not achieved, blinding is not utilized, and/or follow-up rates are low—is prone to all common types of bias and confounding that threaten the validity of observational studies and is, therefore, less likely to enable valid causal inference than a well-conducted observational cohort study. In addition, intervention against a true causative factor in an RCT may still fail to result in lower levels of disease for a variety of reasons, including an immutable or unsuccessfully controlled causative factor, inappropriate timing or intensity of the intervention, and lack of patient compliance. Therefore, the frequently advocated view in recent years that observational designs are of lesser value while a positive RCT outcome is indispensable in the identification of a true causal exposure is misleading and threatens to disregard important scientific progress towards reducing morbidity and mortality in the population.

The use of group comparisons to approximate the ideal counterfactual knowledge under investigation is of critical importance but still fails to provide an explicit causal model linking exposures to disease outcomes. For epidemiological designs to yield meaningful causal inferences, coherent causal models of disease aetiology are necessary, such as the ones employed in studies of infectious disease aetiology (arguably the models that established the discipline of epidemiology). The studies of John Snow on cholera [4] and James Lind on scurvy [1] are classic examples of early epidemiological inquiry. In the context of infectious diseases, causes were generally identified when a microorganism (i.e., the causal factor or 'risk factor') was, or appeared to be, both the necessary and sufficient condition for the disease to occur. In other words, the factor had always to be present in every case of the disease, and the factor alone could produce disease. Accordingly, Koch's postulates were originally developed to provide a framework for establishing a particular microorganism (Mycobacterium) as the cause of tuberculosis. Interestingly, while Koch's postulates were initially quite helpful in elucidating the causal organism of TB, it was realized in retrospect that they were generally less useful in the study of several other infectious diseases. For example, the first postulate posits that a microorganism must be *present in all cases* of disease and *absent in healthy individuals*, a condition which is now known to be false for numerous infectious diseases, including TB. The second postulate states that the microorganism must be isolated from a diseased host and grown in culture, which obviously does not apply to uncultivable microbes or to viruses. The third postulate requires the emergence of disease when a healthy host is inoculated with the causative organism; the existence of asymptomatic carriers for infectious disease (e.g., Typhoid Mary) negates the veracity of this requirement.

As industrialized societies acquired a better understanding of infectious diseases and life expectancy increased during the 1800s and 1900s, the leading causes of death shifted to conditions that are chronic and multifactorial. During this epidemiologic transition, it became apparent that classical causal models were inadequate. Smoking as a cause of lung cancer and cardiovascular disease was a specific and early example of the insufficiency of causal models requiring necessary and sufficient causes of disease. More broadly, causal models that require necessary and sufficient causes are of limited value for all current leading causes of death in the world (e.g., cardiovascular disease, cancer, diabetes, respiratory diseases).

In response to these limitations, a now classic model for causal inference in the context of chronic diseases, that can also be applied to infectious diseases, was proposed by Rothman using a '*sufficient cause*' model of causation [3]. A sufficient cause (SC) is defined as 'a complete causal mechanism that inevitably produces disease'. The SC model visually represents causal hypotheses using causal 'pies' as shown in Figs. 5 and 6. Causal pies are represented as full circles (i.e., sufficient causes) comprised of individual slices termed '*component causes*', each of which is required to assemble in full a sufficient cause and, thus, for disease to occur. According to the main premise of the conceptual model, once all component causes of a sufficient causal pie are in place, disease will inevitably occur. The example in Fig. 5 provides a hypothetical sufficient component causal model for



Sufficient Cause 1 Sufficient Cause 2

Prevalence of U1 and U2 is 100% in population 1 and 2.

Table A. Linking risk factor combinations to periodontitis risk according to sufficient causes 1 and 2

U1	U2	A	P	G	SC	Risk	Population 1	Population 2
1	1	1	1	1	1,2	1	500	500
1	1	1	1	0	None	0	500	500
1	1	1	0	1	2	1	50	350
1	1	1	0	0	None	0	50	350
1	1	0	1	1	1	1	400	100
1	1	0	1	0	None	0	400	100
1	1	0	0	1	None	0	50	50
1	1	0	0	0	None	0	50	50

Table B.
Joint distribution of *P. gingivalis* and periodontitis in population 1

	Periodontitis	No Periodontitis	Total
P. gingivalis present	900	900	1800
P.gingivalis absent	50	150	200

Table C.
Joint distribution of *P. gingivalis* and periodontitis in population 2

	Periodontitis	No Periodontitis	Total
P. gingivalis present	600	600	1200
P. gingivalis absent	350	450	800

Estimates of the causal effect of *P. gingivalis* on periodontitis in two separate populations

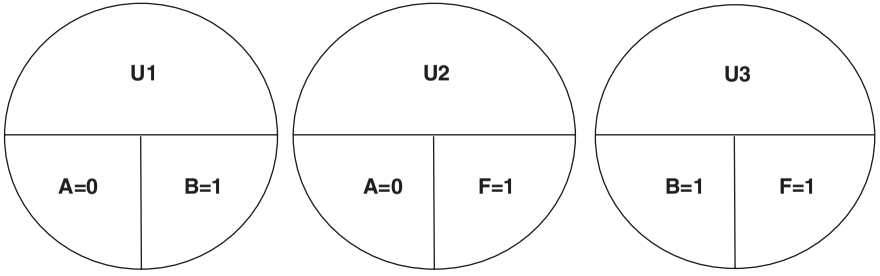
CIR = (900/1800) / (50/200) = 2.0

CID = (900/1800) – (50/200) = 0.25

CIR = (600/1200) / (350/800) = 1.14

CID = (600/1200) – (350/800) = 0.06

Fig. 5 Translating the sufficient component cause model of disease causation into causal estimates of risk in two separate populations, each comprising 2000 individuals. In the upper left, two hypothetical sufficient causes of human periodontitis are described. Sufficient cause (SC) 1 is comprised of the following three component causes: U1 (unknown causal factors assumed to be ubiquitous), P (*Porphyromonas gingivalis* presence), and G (a set of genetic polymorphisms). Sufficient cause (SC) 2 is comprised of the following three component causes: U2 (unknown causal factors assumed to be ubiquitous), A (*Aggregatibacter actinomycetemcomitans* presence), and G (the same set of genetic polymorphisms as in SC 1). The component causes U1, U2, A, P, and G are synonymous with the term 'risk factor' in modern epidemiology. Table A provides all possible risk factor combinations, with 1 = present and 0 = absent, and assuming that U1 and U2 are ubiquitous in both populations. The SC column of Table A indicates which sufficient cause (1 and/or 2 or neither) is completed for each possible risk factor combination. The risk column equals 1 when at least one sufficient cause is completed and 0 if no SC is completed. The population 1 and 2 columns reflect the number of individuals in each population with a given risk factor 1 = present and 0 = absent. Tables B and C reflect the joint distribution of *P. gingivalis* and periodontitis in populations 1 and 2 as derived from Table A. Cumulative incidence ratio (CIR) is defined as the ratio of the proportion of individuals with a certain risk factor that have completed a sufficient cause (i.e., have developed the disease) over the proportion of individuals without the risk factor that have completed a sufficient cause. Cumulative incidence difference (CIDs) is defined as the difference between the above two proportions. Note the difference in the causal effect of *P. gingivalis* on periodontitis in the two populations



Sufficient Cause 1 Sufficient Cause 2 Sufficient Cause 3

Table A. Linking risk factor combinations to diabetes risk according to sufficient causes 1 and 2

A	B	F	SC	Risk	Population 1	Population 2
1	1	1	3	1	900	100
1	1	0	None	0	900	100
1	0	1	None	0	100	900
1	0	0	None	0	100	900
0	1	1	1,2,3	1	100	900
0	1	0	1	1	100	900
0	0	1	2	1	900	100
0	0	0	None	0	900	100

Table B. Joint distribution of *F. nucleatum* and diabetes in population 1

Population 1	Diabetes Present	Diabetes Absent	Total
F. nucleatum present	1900	100	2000
F. Nucleatum absent	100	1900	2000

CIR = (1900/2000) / (100/2000) = 19

CID = (1900/2000) – (100/2000) = 0.9

Table C. Joint distribution of *F. nucleatum* and diabetes in in population 2

Population 2	Diabetes Present	Diabetes Absent	Total
F. nucleatum present	1100	900	2000
F. nucleatum absent	900	1000	2000

CIR = (1100/2000) / (900/2000) = 1.22

CID = (1100/2000) / (900/2000) = 0.1

Fig. 6 Translating the sufficient component cause model of disease causation into causal estimates of risk in two separate populations. In the upper left, three hypothetical sufficient causes of diabetes are described. Sufficient cause (SC) 1 is comprised of the following three component causes: U1 (unknown causal factors assumed to be ubiquitous), A (high dietary fibre consumption), and B (a genetic polymorphism). SC 2 is comprised of the following three component causes: U2 (unknown causal factors assumed to be ubiquitous), A (high dietary fibre consumption), and F (*Fusobacterium nucleatum* colonization in the pancreas). SC 3 is comprised of the following three component causes: U3 (unknown causal factors assumed to be ubiquitous), B (a genetic polymorphism), and F (*Fusobacterium nucleatum* colonization in the pancreas). Table A provides all possible risk factor combinations (assuming U1, U2, and U3 are present in all participants in both populations) in columns A–F (1 = present and 0 = absent). The SC column of Table A indicates which sufficient causes (1, 2, and/or 3 or none) are completed for each possible risk factor combination. The risk column = 1 when at least one sufficient cause is completed and 0 if no SC is completed. The population 1 and 2 columns reflect the number of individuals in each population with a given risk factor distribution. Tables B and C reflect the joint distribution of *F. nucleatum* and diabetes in populations 1 and 2 as derived from Table A. Assuming this causal model is true, the cumulative incidence ratios (CIRs) and cumulative incidence differences (CIDs) have a causal interpretation

the development of human periodontitis in which there are two sufficient causes. In this example, sufficient cause 1 involves the presence of microbial dysbiosis triggered by a particular microorganism (*Porphyromonas gingivalis*) (P), a set of genetic polymorphisms (G) and the additional presence of a number of unknown factors (U1). Sufficient cause 2 is comprised of a different dysbiotic microbial profile, namely dysbiosis triggered by *Aggregatibacter actinomycetemcomitans* (A), the same set of genetic polymorphisms as in SC 1 (G), and another set of unknown factors (U2) which are distinct from U1. In the example visualized in Fig. 5 for periodontitis, G represents a *necessary cause*—i.e., G is a component cause that is present in all sufficient causes of disease and is therefore necessary to be present for periodontitis to occur. However, while G is necessary for the development of periodontitis, G alone is not sufficient to produce periodontitis without the presence of G's causal complements (i.e., P + U1 or A + U2). In contrast, P, A, U1, and U2 represent component causes that are neither sufficient nor necessary to cause periodontitis. If any individual in a hypothetical population completes either SC 1 or SC 2, they will develop periodontitis. A second example (Fig. 6) provides a hypothetical set of sufficient causes positing translocation of *Fusobacterium nucleatum* (F) from the oral cavity to the pancreas as a cause of type 2 diabetes mellitus development. In this example, there are three distinct sufficient causes comprised of six different component causes. This example demonstrates a scenario in which there are no necessary causes.

Two points should be emphasized from the SC model approach presented in Figs. 1 and 2. First, in modern epidemiology, the term 'component cause' is synonymous with the more commonly used term, 'risk factor'. In other words, risk factors are causes of disease that generally work in tandem with other risk factors (i.e., component causes) to produce disease. Note that the term 'risk predictor' is generally used to refer to a variable that predicts risk but for which causality is not assumed (e.g., grey hair is a risk predictor of mortality but not a risk factor). Second, and building on the first point, a somewhat obvious conclusion from the SC model is that there are multiple pathways that lead to the development of a given disease and each pathway involves multiple component causes that work together synergistically. This synergy precisely represents the concept of interaction (or effect measure modification) in statistics and epidemiology. Although we will not discuss interaction in detail here, in the specific context of SC models, when causal factors interact, any one component cause can only cause disease in the presence (or possibly in the absence) of the other component cause(s) in the same SC.

A careful review of the examples in Figs. 5 and 6 demonstrates another important concept that helps us understand why an exposure can cause disease even if the strength of association is weak or varies greatly across different studies (for example, as often observed in a meta-analysis). In the examples presented in Figs. 5 and 6, it is apparent that the cumulative incidence ratio (CIR), i.e., the ratio of the proportion of individuals with a certain risk factor that have completed a sufficient cause (i.e., have developed the disease) over the proportion of individuals without the risk factor that have completed a sufficient cause, and the cumulative incidence

difference (CID), i.e., the difference between the above two proportions, vary across populations in which the distribution of component causes are not equal. This raises a profoundly important point about causal inquiry that is often not appreciated in the health sciences: specifically, the strength of association (using absolute measures) is dependent upon the prevalence of causal complements in the population. The causal complement of a risk factor is defined as the set of all other component causes in all sufficient causes in which a risk factor participates. In the case of Fig. 6, the causal complements of F are A = 0 and U2, or B = 1 and U3. As the prevalence of these causal complements increases, the strength of association between F and diabetes becomes stronger.

So, what are the implications of our causal models for epidemiological research and the ability to identify causes of disease in humans? When we explore risk factors in isolation using reductionist approaches, there can be great variation in the strength of association between a causal factor and a disease outcome across populations. In populations with a low prevalence of causal complements, the strength of association for the main component cause (i.e., risk factor) under investigation will be weak when compared to that in a population with a higher prevalence of causal complements.

In contrast, in disease models where there are multiple sufficient causes in the population, and there is a high prevalence of component causes in sufficient causes where the risk factor of interest does not participate, the observed effect for this particular risk factor will be relatively weak or undetectable. In Fig. 6, note that an increase in the prevalence of individuals with both A = 0 and B = 1 would lead to an increase in the prevalence of individuals susceptible to SC1, yielding weaker associations between F and diabetes because F cannot cause disease in individuals with SC1 already complete (i.e., in individuals that are already 'doomed'). This concept, known as causal redundancy, has been elegantly discussed in a review by Gatto and Campbell [5].

Interestingly, high variability in measures of association across studies conducted in different populations is often taken to suggest lack of evidence for causality. For example, the often-referenced Bradford Hill guidelines for assessing causal evidence [6] include the criteria of 'consistency' and 'strength of association', which imply that inconsistent results across study populations and/or weak associations argue against a causal relationship. While consistently strong associations do increase confidence in a causal hypothesis, lack thereof does not necessarily imply no causality. The examples above clearly demonstrate that under specific causal hypotheses, not dissimilar to the underlying hypotheses of modern chronic disease aetiology, causal effects are *expected* to be inconsistent and at times weak, across different populations, as long as the prevalence of other risk factors varies.

Modern epidemiology is often faced with complex sufficient causal hypotheses, which typically lack a necessary cause. For example, in the cardiovascular disease literature, there has been a long-standing argument as to the usefulness of novel risk factor research [7] because so many risk factors have been identified and nearly all cases of coronary disease have one or several traditional risk factors present [8].

However, while it is evident that almost all individuals with CHD have at least one major risk factor, two facts remain: (1) no one risk factor is always present and (2) a large proportion of CHD-free individuals also tend to have multiple risk factors. Therefore, to date, neither 'necessary' nor 'sufficient' causes of CHD have been identified. Nevertheless, since the classical risk factors are pieces of the causal pie, interventions against one or several of these factors can prevent or delay completion of a causal pie sufficient for disease development in many populations.

There are real-world implications for the aforementioned concepts. It has long been observed that many traditional cardiovascular disease risk factors tend to have weaker associations with clinical outcomes in elderly populations. For example, although the benefits of statin therapy to prevent myocardial infarction have been clearly shown, their use in elderly populations remains debatable. A more recent example involves aspirin use for the primary prevention of cardiovascular events among the elderly, in whom it offers no benefit and may possibly even induce harm, despite long-standing benefits of aspirin use for secondary prevention in younger populations [9].

Using the aforementioned concepts, we can couple a causal model (causal pies) with the counterfactual concept operationalized via real-world study designs and data collection to yield group comparisons that inform causal hypotheses.

Populations Vs. Individuals

The aforementioned examples demonstrating how group comparisons utilized to infer causality inform the long-standing paradox in the health sciences in which research methodologies for identifying causality rely on 'average risk' across the groups being compared (e.g., treatment vs. placebo). From a big picture, public health perspective this concept works well when making policy recommendations regarding prevention and treatment of disease. If a particular intervention reduces disease 'on average', population health improves. However, the paradox arises in clinical situations when treatments, or prevention recommendations, are delivered directly from clinicians to individual patients. This setting is more challenging, given the fact that causality is never certain at the individual level and therefore, one can never know if a particular intervention was successful for a given individual. To paraphrase Jude Pearl [10], risk—along with other measures—can often give profoundly accurate predictions about disease occurrence in populations yet, paradoxically, it has very little accuracy at the individual level. Or stated another way, in most settings, causality can only be determined at the population-level. Reconciling the utility of risk-based measures in populations vs. individuals has been an age-old challenge. While clinical judgement and personalized health care have an important place, the lion's share of treatment decisions with proven efficacy rely on evidence derived from epidemiological designs utilizing group comparisons. The implication of this fact is that treatments will work on average. Underscoring this point is

the often-cited measure of impact in the health sciences, Number Needed to Treat (NNT). The NNT is defined as the inverse of cumulative incidence differenced (Fig. 2) and represents the number of individuals that would need to receive an intervention to prevent one case of disease.

Concluding Remarks: Risk and Causality in the Precision Era

We have entered the era of 'precision-oriented' science, including precision medicine and precision public health [11]. Parenthetically, this popular term seems rather misguided, since the concept of 'precision' relates to measures of reliability (i.e., reproducibility) rather than validity (lack of bias). The term 'personalized medicine' is arguably much better in reflecting a more customized, yet valid, approach to medical care. Irrespective, use of either term may be interpreted to suggest that the pushback against epidemiology is intensifying in favour of other causal models leveraging precision (or more narrowly stated, genomic sciences). However, this is again a misconception. In fact, the 'precision agenda' is merely a refinement of the tools that we use to build sufficient cause models and subsequent study designs. Specifically, the precision era is unlikely to fundamentally change our methods of causal inquiry, although it does offer the potential to clarify causal hypotheses and their underlying sufficient cause models by identifying new, mostly genetic, pieces of the causal pie. In doing so, precision approaches can help to refine inclusion/exclusion criteria for research studies that test hypotheses concerning specific risk factors and/or interventions targeting those risk factors. Refining inclusion/exclusion criteria will enhance future studies by ensuring that 'at-risk' individuals are indeed included in the study samples. Formally speaking, 'at-risk' in the context of an intervention design would be individuals who are about to complete or have completed a sufficient cause that includes the particular risk factor targeted by the intervention. In that setting, the removal of the risk factor would prevent, or in the case of a reversible effect, cure the disease under investigation.

Once new risk factors are identified in more precisely defined populations, it stands to reason that clinical and public health practice will benefit by knowing which treatments to deliver and to whom. If this approach sounds familiar, that is because it is exactly the way causes of disease have been identified and that knowledge translated into evidence-based health policies and treatment guidelines. As such, the current hype surrounding the precision agenda is a refined version of a causal model that has been in place for decades.

In conclusion, causal inference in most contexts relevant to human disease requires group comparison under a counterfactual framework. Because most human disease entities of relevance to public health are complex and likely involve multiple sufficient causes of disease, research designs in humans should be based on well-developed sufficient component cause models to ensure that at-risk individuals are studied in populations where causal redundancy can be minimized.

References

1. Morabia A, Morabia A. Enigmas of health and disease: how epidemiology helps unravel scientific mysteries. New York: Columbia University Press; 2014.
2. Taubes G. Epidemiology faces its limits. Science. 1995;269:164–9.
3. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.
4. Hempel S. The strange case of the Broad Street pump : John Snow and the mystery of cholera. Berkeley: University of California Press; 2007.
5. Gatto NM, Campbell UB. Redundant causation from a sufficient cause perspective. Epidemiol Perspect Innov. 2010;7:5.
6. Hill AB. The environment and disease: association or causation? Proc R Soc Med. 1965;58:295–300.
7. Beaglehole R, Magnus P. The search for new risk factors for coronary heart disease: occupational therapy for epidemiologists? Int J Epidemiol. 2002;31:1117–22.
8. Greenland P, Knoll MD, Stamler J, Neaton JD, Dyer AR, Garside DB, Wilson PW. Major risk factors as antecedents of fatal and nonfatal coronary heart disease events. JAMA. 2003;290:891–7.
9. McNeil JJ, Wolfe R, Woods RL, Tonkin AM, Donnan GA, Nelson MR, Reid CM, Lockery JE, Kirpach B, Storey E, Shah RC, Williamson JD, Margolis KL, Ernst ME, Abhayaratna WP, Stocks N, Fitzgerald SM, Orchard SG, Trevaks RE, Beilin LJ, Johnston CI, Ryan J, Radziszewska B, Jelinek M, Malik M, Eaton CB, Brauer D, Cloud G, Wood EM, Mahady SE, Satterfield S, Grimm R, Murray AM, Group AI. Effect of aspirin on cardiovascular events and bleeding in the healthy elderly. N Engl J Med. 2018;379:1509–18.
10. Pearl J, Mackenzie D. The book of why : the new science of cause and effect. 1st ed. New York: Basic Books; 2018.
11. Chowkwanyun M, Bayer R, Galea S. “Precision” public health - between novelty and hype. N Engl J Med. 2018;379:1398–400.