Classical causation research practices and sufficient-component cause model - Appraisal and pitfalls

Abdul Hakeem Alrawahi[1]

(1) School of Medicine, Griffith University, Queensland, Australia. E-mail: suhaibfather@gmail.com

CORRESPONDING AUTHOR: *Abdul Hakeem Alrawahi, School Medicine, Griffith University, Gold Coast Campus, Parklands Drive, Southport, Qld 4222. Phone No. +61411632695, E-mail: suhaibfather@gmail.com*

DOI: ***10.2427/12576***
Accepted on June 12, 2017

ABSTRACT

The sufficient-component cause model is one of the most discussed recent theories in disease causation. Despite some limitations, it seems one of the best theories to explain real world phenomena. It seems that this model has many implications on the current classical research methods related to disease causation. However, these implications have not been sufficiently explored. Therefore, based on the sufficient-component cause model, this paper aims to review, identify and rectify pitfalls in the classical causation research practices, especially those related to cohort study design.

Key words: sufficient component cause model; implications; classical research; cohort design; meta-analysis; pitfalls; new approaches.

INTRODUCTION

The sufficient-component cause model introduced by Rothman in 1970s is one of the most discussed causal models in epidemiology. According to this model, a "sufficient cause" can be defined as a set of minimal conditions and events that inevitably produce an outcome [1]. The word "minimal" implies that all of the conditions or events are necessary to the occurrence of the outcome. In disease etiology, the completion of a sufficient cause may be considered equivalent to the onset of the disease [1,2].

Figure 1 presents a diagram of three sufficient causes for an outcome, in the form of pie diagrams, in a hypothetical individual. Each combination of component causes (risk factors) represented in Figure 1 is sufficient to produce the outcome. In other words, there is no extraneous or redundant component cause [2]. In addition, each component cause is a necessary part of that specific causal mechanism; that is, no factor is stronger as a superhero. A specific component cause may play a role in one, two, or more of the causal mechanisms. A component cause which presents in all sufficient causes, is called a "necessary cause". In disease causation, most identified factors are neither necessary nor sufficient alone to produce disease [2].

In fact, it is not that simple, and the complexity of disease causation mechanism can be seen in the following points. First, each sufficient cause may include elements as simple as two or three component causes and may include as many as hundreds of component causes. Also, sufficient causes are likely to include factors that are not yet known, or are not expected to be involved in the disease causation

FIGURE 1. Rothman diagram illustrating three sufficient causes of an outcome.

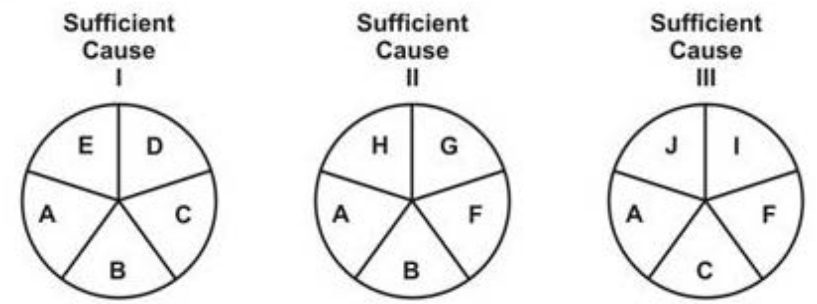

Figure caption: Each pie consisting of a specific combination of component causes represents a sufficient cause for the outcome of interest. Each of the letters A, B, C, etc. represents a component cause (risk factor) and each may participate in one or more sufficient causes

mechanism. In addition, different component factors may act at different times, making it further complicated to anticipate a specific sufficient cause. Moreover, disease outcome may be caused by one or two sufficient causes, but also, it can be caused by as many as hundreds of sufficient causes. Although similar sufficient causes can be grouped into classes or categories (e.g. in lung cancer, smoking sufficient causes means all sufficient causes that include smoking), it is still difficult to anticipate a likely factor sufficient cause in a specific patient. This is because the real process that produces the disease is unknown, and it can be due to a sufficient cause that is not expected at all.

However, all of above are not considered to be limitations for the model, but the complexity of disease causation nature. As for preventive actions, the control of an outcome can be achieved by controlling one of the components in each "pie", and if there is a factor common to all pies, the disease would be eliminated by removing that alone [2]. Otherwise, controlling common component causes in common sufficient causes is the clinical way for disease prevention.

Despite some inherent limitations [3], this model seems to be closer to the truth in explaining the real world questions that cannot be answered by the previous model, "the web of causation" [4,5]. For example, why some individuals develop an outcome and others don't, while the later may have higher levels of potential known risk factors?! Simply, because no pies have been formed yet. Another question is why do some good quality studies reveal conflicting results?! While a study concludes a significant association between a factor and an outcome, another concludes no association, and both are of the same design and quality!! In fact, the sufficient cause theory is the easiest way to answer such questions as appears in the following discussions.

However this model has faced criticism [6], such criticism seems to be related to the complexity of disease causation and limitation of human knowledge, rather than limitations for the model. Negative criticism for the sake of rejecting or under-estimating other's work is a road to nowhere. In fact, what we need is to positively criticise each other's work and if possible build on it to achieve advancement in science.

Classical research methods used to assess causation

In this article, we mean by "causation research" any research directed to assess a cause and effect relationship between a factor, state or a characteristic and an outcome. Causation research is not limited to risk factors and diseases as in epidemiology, but it is also related to other fields of medicine and other sciences. Pharmacologists aim to find effectiveness of treatment in disease cure and the adverse effects of drugs, dieticians try to assess relationship between vitamins and health outcomes, and education scholars assess the effectiveness of new educational strategies in students' achievement.

All types of observational study designs have been utilized in the literature to assess associations and causations. These include cross-sectional, case control and cohort studies [7,8]. However, prospective cohort studies are considered the optimal among those, due to its ability to establish temporality, and can yield better quality of data [7,8]. On the other hand, experimental study designs, especially in disease causation research, are restricted by ethics and costs [2]. Therefore, this paper discussion concentrates on the cohort study design and its measures of effect; the relative risk (RR) (or the odds ratio (OR) as an approximation for RR) and the risk difference.

However, to the best of our knowledge, the sufficient-component cause model seems to be a theoretical framework for epidemiology, and its implications on the current classical causation research practices are still not well explored [9]. Therefore, based on the sufficient-component cause model, this paper aims to review, identify and rectify the pitfalls in the classical causation research practices, especially those related to cohort study

designs. As a limitation of human mind, the raised pitfalls need to be critically reviewed; however, these may be the base for positive criticism and other new suggestions.

Basic concepts

According to *Rothman*, the RR (or OR) of a factor, observed in a cohort study, is not related to an inherent strength of the association between the outcome and that factor. Rather, it is more related to the prevalence (availability) of the complementary component causes that act together with the factor of interest in the same sufficient causes [10]. In a study sample, higher availability of component causes that are complementary for a factor, makes the sample environment more favorable for that factor to produce the disease compared to a sample with low availability of the same complementary factors, and hence larger RR is observed. This means that the RR of a factor depends on the prevalence of the factor sufficient causes (i.e. all sufficient causes related to the factor of interest) in the studied sample. However, the RR doesn't depend on that only, but also it depends on the prevalence of the other factors sufficient causes (which do not belong to the factor of interest), that compete with the factor of interest to produce the outcome [2]. When viewed in this manner, the RR is really not related to an inherent strength of the studied factor.

To further clarify this argument, we emphasize the following assumptions and then discuss some theoretical examples. First, provided that the control group (unexposed group in cohort studies) is comparable to the exposed group and both from same population, we assume that the outcome incidence in the unexposed group represents the baseline incidence of the outcome, so that the exposed group would have similar baseline incidence of the outcome if its individuals are not exposed to the factor of interest. Second, the sufficient causes inevitably produce the outcome, so that the prevalence/incidence of all sufficient causes in a sample is nearly equal to the incidence of the outcome observed in that sample, assuming that no or minimal cases are cured or die after being produced by the sufficient causes. In other words, the prevalence of a factor sufficient causes equals the incidence of the outcome caused by that factor sufficient causes. However the classical calculation of the RR relies on same assumptions [11]; the only difference is that we are talking about the factor sufficient causes instead of the factor alone.

Now, the following theoretical examples can be discussed for further clarifications. In a cohort study, it was found that the incidence of lung cancer increased from 2% among non-smokers to 6% among smokers, with RR of 3.0. In a similar second study, asbestos exposure increased the incidence from 4% among non-exposed to 10% among exposed, with RR of 2.5. Considering that the two groups are comparable and from same population in each study, the 2% incidence among non-smokers represents the baseline outcome incidence, which reflects the prevalence of the other sufficient causes (which do not belong to smoking) that are usually present in both groups. Therefore, the prevalence of smoking sufficient causes in the smokers group equals to 6% - 2% = 4%. However, 4% is only the risk difference. Similarly, the risk difference for asbestos in the second study is 6% which reflect the prevalence of asbestos sufficient causes among the exposed group, while the prevalence of other factors sufficient causes here is 4% (the baseline). However the prevalence of smoking sufficient causes among the exposed group is less than that of the asbestos, the RR of smoking is larger. So, RR doesn't depend on the prevalence of the factor sufficient causes alone, but also on the prevalence of other factors sufficient causes that compete with the factor of interest to produce the disease. This means that when the sample characteristics is more favorable for a factor of interest (i.e. higher availability of the complementary factors acting with the factor of interest) to produce the outcome and much less favorable for other factors sufficient causes (due to low availability of their complementary factors), this leads to higher RR for the factor of interest.

Simply, a third study with a good quality as the first and second studies, conducted in a different population with different characteristics and environment that make the sample highly favorable for asbestos to produce disease, and unfavorable for other competing sufficient causes, may reveal more obvious effect for asbestos, with RR as large as 5.0 compared to smoking in the first study, as well as, compared to asbestos in the second study. However, the three studies are of a good quality and their results are likely to be true for their samples. So, smoking has a larger RR than asbestos in the first comparison not because smoking itself is a stronger factor inherently, but because the baseline incidence was low and the sample environment was favorable for smoking to produce a good number of cases relative to the baseline number of cases, which in total, make the difference more obvious for smoking.

Pitfalls in current classical causation research practices

Based on the above assumptions and arguments, the following issues are raised. First, the classical interpretation of the *P* value, the RR (or the OR) and the risk difference in cohort studies should be reviewed. In fact, significant association results means that in the studied sample, it is likely that there are sufficient causes that exist for the factor under the study. On the other hand, insignificant results means that it is unlikely that the study sample includes participants with sufficient causes related to the studied factor. On the other hand, the RR (along with the *P* value), seems to be linked to assess if there is an obvious

association rather than to assess the strength or magnitude of association. Simply, and to make it easy, larger significant RR (significant RR means RR with confidence interval not crossing the 1 or RR with *P* value < 0.05) means more obvious association or a likely obvious existence of sufficient causes related to the factor under study, assuming that the exposed and unexposed groups are comparable and from the same population. Interpreting P value and RR results in the context of a factor's sufficient causes and the availability of its complementary factors – instead of the classical perspective that "there is a significant / insignificant association between the studied factor (alone) and the outcome"– may inspire researchers to broaden their thinking and facilitate criticism of classical research methods in relation to causation research.

Second, in the light of the sufficient cause theory, it is not only that various risk factors (component causes) in a specific sufficient cause are equally important [10], but this seems also true even if the two factors belong to different sufficient causes. In the previous examples, asbestos showed smaller RR in one instance and larger RR in another instance compared to smoking, depending on the sample characteristics. Tending to say smoking is inherently stronger than any other factor in causing lung cancer because most of related studies show larger RR for smoking, seems to be unfair.

Third, if there is something called "strength" or "magnitude" of association, then it should be the risk difference. As mentioned earlier, the risk difference reflects the prevalence of a factor's sufficient causes in the exposed group, provided that the two groups are comparable and from the same population. However, this strength is not an inherent characteristic of the factor, and it varies across populations and samples. Classically comparing the RR (or the risk difference) of two factors, revealed by two or more studies that involved completely different samples, is an unfair comparison. In the previous examples, it is usual that the smokers group and the asbestos group are completely different, since the smoker group may represent smokers in a community while the asbestos group may represent a smaller population working in a factory, despite the fact that the two samples may originate from one source population. Therefore, it is unfair to judge whether smoking or asbestos has greater a role in lung cancer, unless the two study samples are similar in some key characteristics and both represent the same population or subpopulation.

However, if two factors are studied in two (or more) study samples that represent the same population (or subpopulation), then the role of these factors in that population can be compared by the population attributable risk (i.e. the prevalence of each factor's sufficient causes in the total population) that can be calculated from the risk difference (the prevalence of each factor's sufficient causes in the exposed population). This can be calculated in the same manner the classical population attributable risk is usually calculated. Therefore, the population prevalence of a factor's sufficient causes equals the prevalence of the same among the exposed group (risk difference) multiplied by the prevalence of the factor in the total population, provided that the studied exposed group is comparable to the unexposed group and it is representative for that population apart from the factor of interest.

For example, if a cohort study with comparable groups is directed to assess the role of hypertension in developing coronary heart disease in a sample that represents type 2 diabetic patients attending a specific clinic, then this study can be compared to another similar study directed to assess the role of high serum cholesterol in developing the same outcome in the same diabetic population attending the same (or similar) clinic with a similar setting, provided that all patients belong to the same source population. For example, if the first study revealed a risk difference (prevalence) of 10% for hypertension sufficient causes in the hypertensive group and other data showed that the prevalence of hypertension in that population is around 30%, then the prevalence of hypertension sufficient causes in that population equals 10 x 30/100 = 3%. Similarly the prevalence of high cholesterol sufficient causes can be calculated for the same population and then compared to hypertension sufficient causes since both factors are common in that population and the exposed samples studied are representative for the same population (apart from the exposure of interest).

Fourth, most researchers are worried about the representativeness of study samples. As claimed by *Rothman*, this is not always required, especially in analytic studies [12]. Where the researcher is concerned about the comparability of the exposed and non-exposed groups, taking the two groups from the same population may balance unknown factors; however, again, representativeness seems not to be important. However, taking a sample that represents the population is still required where the magnitude of the contribution of a factor's sufficient causes is studied, or where the roles of two factors are to be compared in the same population, for example in order to plan priority preventive measures, then representativeness is important. The previous examples assessing and comparing the roles of hypertension and high cholesterol are good examples where representativeness is important. If the hypertensive group is not representative for the meant population, then the risk difference does not represent the prevalence of hypertension sufficient cause in the hypertensive subpopulation, and therefore the calculated population prevalence of the same sufficient causes will be biased.

Fifth, explaining conflicting results in different good quality studies should not be limited to the classical context of "controversy", "possible bias" and "confounding". Rather, observing positive significant association between a factor and an outcome in some good quality studies is enough to conclude the association. On the other hand, insignificant results does not mean no association, but

should be interpreted within the context of the studied population. In other words, "rare or un-availability of the complementary factors" may be a better explanation. Therefore, ruling out associations based on insignificant RR results observed for a factor, even in too many good quality studies where bias is minimized and confounders are well controlled, is not true, since the positive association can be established in different populations or in different times where/when the sample environment becomes more favorable for the action of that factor.

Sixth, the argument above means that risk factor studies are population specific and can be even sample specific to some extent. Even if extensive studies revealed insignificant association between a factor and an outcome, it is still a good rationale and a good objective to study it in different populations and in different samples from the same population and/or in different times. This is more worthy in case of insignificant results revealed by previous studies. On the other hand, the classical practice of re-studying a factor which has already showed positive association in some good quality studies, in order to further assess/verify the association is not a good objective and it should be viewed as cost-ineffective. However, repeating such studies in different samples from the same population may be useful for a subsequent systematic review/meta-analysis, to assess the likely prevalence of a factor's sufficient causes in an exposed specific population (risk difference) or in a specific total population (population attributable risk).

Seventh, this point is related to systematic reviews and meta-analysis. Based on the earlier arguments, mixing studies that showed different results in different populations in order to measure the net effect seems unreasonable. There is no point in letting different studies in different populations averaging the effect of each other, since the matter is not about an inherent strength of association, and different sample characteristics can yield different results. The observed RRs in different good quality studies that involve different populations, are independent of each other and therefore, all should be assumed to be true for their samples. In fact, the net effect (e.g. net averaged RR or OR) revealed by the classical meta-analysis which includes studies from different populations, seems to be related to a hypothetical population that consists of a mixture of various studies' populations included in the meta-analysis. In other words, a net RR of 2.0 for a factor means that in a hypothetical unreal population, which is a mixture of the samples involved in the meta-analysis, the prevalence of all sufficient causes in the exposed group is twice that of the unexposed group. On the other hand, a net RR of 1.0 means that in such a hypothetical population, the association is not observable. The same applies for the net risk difference. However, knowing whether there is association and knowing the magnitude of net effect in a hypothetical unreal population is useless.

If systematic review/meta-analysis has to be conducted, then it should not be directed to the classical assessment of the net effect, but it may be used to assess the existence of the general association between a factor of interest and an outcome, to estimate the likely prevalence of a class of sufficient causes in a specific exposed population or in a specific total population, and to identify common component causes acting together in common sufficient causes. All of these suggestions in addition to other new methodological approaches are to be elaborated in another paper.

CONCLUSION

In this paper, some of the classical causation research methods were reviewed and appraised based on the sufficient-component cause model. RR along with the P value seems to be related to whether the association is obvious or not, and its interpretation should be reviewed. In addition, it seems that no factor is stronger than another in producing an outcome, and the role of different factors can only be compared under specific circumstances. Moreover, representativeness of study samples is not always a concern; however it seems necessary in some situations. Furthermore, the current classical systematic review/meta-analysis approach seems to be threatened and needs to be critically reviewed and rectified.

Conflicts of interest

No conflict of interests was involved in this work

Funding

No funds were received for this work

References

1. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. Am J Public Health. 2005;95 Suppl 1:S144-150. URL: https://www.ncbi.nlm.nih.gov/pubmed/16030331
2. Rothman KJ, Greenland S, Poole C, L. Lash T. Basic Concept. In: Modern Epidemiology. Third edition. Philadelphia, PA: Lippincot Williams & Wilkins; 2008.
3. Parascandola M, Weed DL. Causation in epidemiology. J Epidemiol Community Health. 2001 Dec 1;55(12):905–12. URL: http://jech.bmj.com/content/55/12/905
4. Krieger N. Epidemiology and the web of causation: has anyone seen the spider? Soc Sci Med 1982. 1994;39(7):887–903. URL: https://www.ncbi.nlm.nih.gov/pubmed/7992123
5. Vineis P, Kriebel D. Causal models in epidemiology: past inheritance and genetic future. Environ Health. 2006; 21;5:21. URL: https://www.ncbi.nlm.nih.gov/pubmed/16859537
6. Karhausen E by LR. Causation in Epidemiology: a Socratic

dialogue: Plato. Int J Epidemiol. 2001;1;30(4):704–6. URL: https://academic.oup.com/ije/article/30/4/704/705896/Causation-in-Epidemiology-a-Socratic-dialogue

7. Carlson MDA, Morrison RS. Study Design, Precision, and Validity in Observational Studies. J Palliat Med. 2009;12(1):77–82. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2920077/
8. Zapf D, Dormann C, Frese M. Longitudinal studies in organizational stress research: a review of the literature with reference to methodological issues. J Occup Health Psychol. 1996;1(2):145–69. URL: https://www.ncbi.nlm.nih.gov/pubmed/9547043
9. Miettinen OS. Proportion of Disease Caused or Prevented by a Given Exposure, Trait or Intervention. Am J Epidemiol. 1974;99(5):325–32. URL: https://www.ncbi.nlm.nih.gov/pubmed/4825599
10. Rothman KJ. Causes. Am J Epidemiol. 1995;141(2):90–5. URL: https://academic.oup.com/aje/article-abstract/104/6/587/139202/CAUSES. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2853157/
11. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. Int J Epidemiol. 2013;42(4):1012–4. URL: https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/dys223

*