

PERSPECTIVES AND DEBATES

New approaches to disease causation research based on the sufficient-component cause model

Abdul Hakeem Alrawahi

Head of Research, Department of Research and Studies, Oman Medical Specialty Board, Muscat, Oman

Abstract

Up to date, the sufficient-component cause model seems to be a theoretical framework for disease causation in epidemiology, and its implications in epidemiological research methods is currently still limited. Recently, pitfalls in current epidemiological research methods were addressed based on the sufficient-component cause model; hence, new research approaches are needed as alternatives. Therefore, this paper aims to review and suggest new epidemiological methods used to assess disease causation. A new approach was discussed to identify potential mechanisms of disease occurrence which may be useful for risk prediction and disease prevention. In addition, a novel “exposed case-control” design was introduced to identify potential component causes. Furthermore, this paper suggested a new approach of conducting a systematic review/meta-analysis related to causation studies.

Introduction

In the 1970s, Rothman introduced one of the most discussed causal models in epidemiology. According to this model, sufficient cause for an outcome is determined by a set of minimal conditions and events that inevitably produce the stated outcome.¹ This implies that all of the minimal conditions or events are necessary for the outcome to occur.^{1,2} Each component cause is therefore a necessary part of the causal mechanism it contributes to; in other words, no one factor is stronger than any of the others. A specific component cause may play a role in one, two or more sufficient causes. A component which is present in all sufficient causes is known as a necessary cause. For example, if A in Figure 1 represents smoking in causing lung cancer, then the factors acting with it in various causal mechanisms are the complementary factors for smoking in producing lung cancer. In disease causation, most identified component causes are neither necessary nor sufficient to cause a disease by themselves.²

Nevertheless, diseases may be caused by either a few or many sufficient causes, while each sufficient cause may also include either a few or many individual components. Different component factors may act at different times, and some may not be known or cannot be expected, making it difficult to anticipate a specific causal mechanism. However, these are not limitations to the suffi-

cient-component cause model, but merely indicate the overall complexity of disease causation.³ Fortunately, outcomes can be controlled by regulating common components in shared sufficient causes.² Despite certain limitations,⁴ the sufficient-component cause model seems to successfully explain some of the real-world questions that could not be answered by previously held theories.³

Classical research methods used to assess disease causation

Different study designs are often utilised in the scientific literature to assess disease associations and causation, including cross-sectional, case-control, cohort and experimental designs.^{5,6} Among these, prospective cohort and experimental studies are thought to be the optimal designs in causation research, due to their ability to establish temporality and yield high quality data.^{5,6} However, experimental study designs are usually restricted by costs and ethics, particularly when it comes to assessing disease causation.² In such designs, Relative Risk (RR) or Odds Ratio (OR) as an approximation of RR, as well as the risk difference are the most important measures of effect that are usually reported.

Despite the wide utilization of the classical causation research methods in the literature, pitfalls in those methods were addressed in a recently published article.³ The addressed pitfalls were all based on the sufficient-component cause model, suggesting the need for alternative research approaches.³ At present, the sufficient-component cause model is still considered as a theoretical framework in the field of epidemiology.^{3,7} In other words, incorporation of the implications of the sufficient-component cause model in current epidemiological research is rare. In this regard, only one article could be retrieved in the literature that used a new research approach based on this model.⁸ In addition, up-to-date, no epidemiological research methods have been introduced to study potential component causes that may act together in common sufficient causes. Studying such entities is very important in disease prevention, risk prediction and personalized medicine. Therefore, the current paper aims to review and suggest alternative research methods in the field of disease causation that are based on this model. These suggestions are open to constructive criticisms and may hopefully serve as the basis for further ideas in this area.

Significance for public health

This paper presented several novel epidemiological research approaches based on the sufficient-component cause model. The new approaches are of great importance in disease prediction and prevention, and precision medicine. Therefore, the future use of the suggested new epidemiological methods in the area of medicine and public health is promising.

Causation research approaches based on the sufficient-component cause model

Although the sufficient-component cause model was introduced since long time, literature review revealed only one article used a new research approach that was based on the model. In this regard, Reiber *et al.* attempted to determine potential sufficient causes for lower-extremity ulcers among diabetic patients.⁸ In this approach, Reiber *et al.* predefined specific potential mechanisms (sufficient causes) that can lead to lower-extremity ulcers, and then in a group of cases with ulcers they studied which mechanism was the most likely cause in each case. Although the approach used in Reiber *et al.*'s study is subject to certain limitations, it was successful in inspiring new ideas to advance methods in disease causation research. However, disease prevention can be achieved by controlling common factors that play a shared role in sufficient causes of a disease; as such there is no urgent need to identify potential sufficient causes.²

As mentioned previously, it is very difficult to anticipate or suggest a likely sufficient cause in individual patients without a reference comparison because there may be hundreds of sufficient causes for an outcome, of which each may include many different component causes which may in turn be unknown or unexpected or act at different times. Although sufficient causes can be grouped into overall classes (*e.g.* all sufficient causes that are related to smoking are included within a smoking class),⁷ it is still difficult to determine if an outcome is due to a specific class as the actual disease process is unknown and may be due to a sufficient cause that is unexpected and far removed from the expected mechanism.

Consider an example scenario of a 70-year-old patient with lung cancer who has smoked 10 cigarettes/day for 5 years, along with other factors thought to fit the disease mechanism. In this case, the lung cancer could surprisingly be due to exposure to an environmental substance during a one-week holiday the patient had taken five years previously, along with other component causes that fit a separate sufficient cause. Hence, the patient's smoking status and age may have nothing to do with the disease. In fact, assuming that smoking class sufficient cause is the more likely mechanism for lung cancer in this patient without a reference comparison may be due to a preoccupation with the role of smoking in current lung cancer research.

Novel causation research approaches based on the sufficient-component cause model

Modified Reiber *et al.*'s approach

In fact, Reiber *et al.*'s approach can be considered as the beginning for innovating new epidemiological methods. In this article, we suggest a modified version of Reiber *et al.*'s approach. The suggestion is to involve a comparable control group who do not have the outcome along with the case group. In exactly the same way as for the case subjects, each individual in the control group should be assessed for the presence of any possible combination of component causes that may fit a likely sufficient cause. Assessors should be as blinded as possible to the outcome; this can be easily achieved in the previous example of lower-extremity ulcers by applying the same wound dressing to the feet of all individuals in both the case and control groups. However, in this type of study design, it would not be fair to compare the proportion of individuals with a specific suggested sufficient cause/mechanism among both cases and controls, as per the usual method. This is because the sufficient-component cause model states that the sufficient cause inevitably produces the outcome; therefore, the presence of a suggested sufficient cause in a few individuals in the control group is sufficient to exclude the mechanism. Accordingly, it will be very challenging to come to valid conclusions. However, this can be dramatically overcome by using big data studies and Machine-Learning analytic programs.^{9,10} Another way to improve this approach is to let Machine Learning computer program to build algorithms (potential mechanisms) instead of assessing predefined specific mechanisms for the outcome. In addition, the Modified Reiber *et al.* approach is considered as a case-control design, in which data quality is threatened. However, this can be improved by using a prospective study design. Hopefully, the future of big data held by Machine-Learning software to build algorithms for predicting disease occurrence, using the modified Reiber *et al.* approach, is promising.

Exposed case-control study design

Instead, a new alternative approach suggested in this paper, is to identify likely common complementary component causes that act in common sufficient causes of a specific factor in producing a

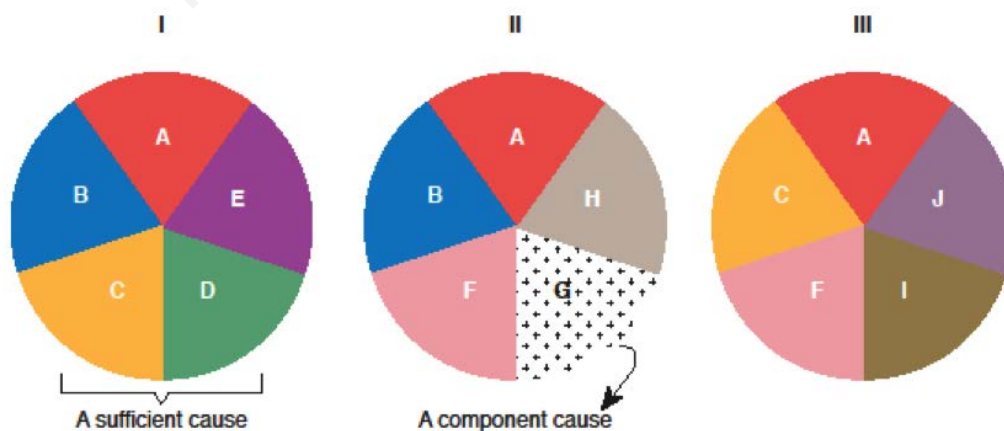


Figure 1. Three sufficient causes of an outcome. Each pie (a sufficient cause) consists of specific factors. Each of the letters A, B, C, etc. represents a component cause (factor) and each may participate in one or more sufficient causes.

specific disease/outcome. For example, if we can identify common complementary factors for smoking in causing lung cancer, then it would be of a great benefit. This may be useful in planning alternative preventative measures for individuals who are exposed to a certain factor which cannot easily be controlled at either the population or individual levels. This may also be applicable for other causation studies including therapeutic studies where a certain agent (e.g. drug) is tested. In this regard, if we can identify complementary factors that act with the studied agent in causing cure or any other favourable outcome, then this would be of a great importance in precision medicine.

For this purpose, cohort studies or experimental studies (if feasible) which yield large RR or OR values are needed. In this regard, in the studies that identify larger RR values for a factor (or intervention), the exposed groups have a larger proportion of outcome cases attributable to that factor's sufficient causes, compared to study samples with lower RR values, provided that the exposed and unexposed groups are comparable.³ In other words, in samples with a higher RR value, the availability of complementary factors is greater. Therefore, we may label study samples which determine a large RR for a specific factor as "golden samples", and exposed groups in such studies deemed to be "golden exposed groups". While there is no exact cut-off RR value for golden samples, we consider a RR value of >2.0 as a reasonable limit, at which more than 50% of outcome cases in the exposed group are attributable to the studied factor. However, studies with larger RR values are preferable and their findings should be given more weight and attention. It is important to note that exposed and unexposed groups in a study should be comparable in order to deem the sample to be golden. Subsequently, potential common component causes in golden samples can be identified at two levels: the study sample level and stratification subgroup level.

At the study sample level, cases in a golden exposed group (*i.e.* golden cases) should be given special attention to further identify

potential complementary characteristics that might act together with the factor of interest. Golden cases in which the outcome manifested should be compared to the exposed subjects who did not develop the outcome in the same golden group. Simply put, the characteristics of exposed outcome cases should be studied and compared to those of exposed control subjects who remained free of the outcome. The revealed unique characteristics should be considered as potential complementary factors for the factor under study. This new design may be called an "exposed case-control study design". Figure 2 shows the basic framework of this new design considering smoking and lung cancer as examples of exposure and outcome respectively.

In such a design, the exposed cases and exposed controls can be nested from a cohort study (as discussed above), however, exposed cases and exposed controls can be selected from general populations or from hospital settings, as the case in the classical case-control studies. However, exposed cases and controls nested from cohort studies with large RR are better due to higher chance that the cases are attributed to the studied exposure.

Beforehand, if possible, "filtering" of the outcome cases in the golden group should be attempted, so that if we can remove at least some cases (from the golden cases) that are likely to be caused by other sufficient causes (that do not belong to the factor of interest), then we may end with golden cases that are more likely to be caused by sufficient causes that belong to the studied factor. This is because there is a baseline proportion of outcome cases in the golden exposed group that are thought to be caused by sufficient causes related to other factors.³ The filtering of golden cases can be performed using a "best-matching" method. In this approach, outcome cases that arise in the unexposed group (baseline cases) are matched according to their characteristics and risk factor profile as closely as possible to golden cases, either by means of statistical software or manually. The assumption here is that a case in the golden group that appears very similar (in terms of its potential

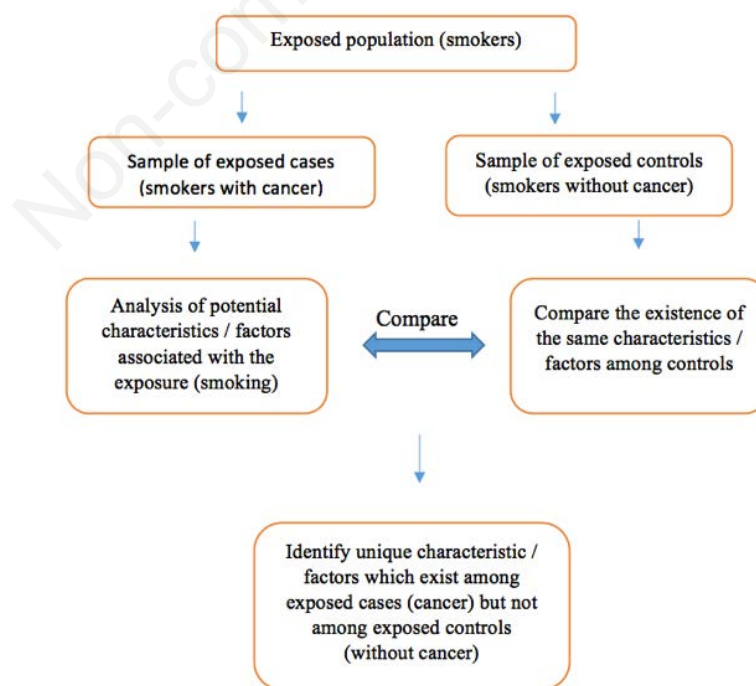


Figure 2. Basic framework for the exposed case-control design used to identify common complementary factors acting with a specific exposure/intervention in producing specific outcome.

risk factors and characteristics) to one or more cases in the unexposed group is likely to be caused by sufficient causes of other factors and not the factor of interest, thus representing baseline cases. While it is not necessary to find matching cases for all suspected baseline cases, removing a few from among the golden cases is worthwhile. Once filtered, golden cases should then be compared to exposed subjects who did not develop the outcome, as an attempt to identify some of the potential complementary factors acting with the studied factor. Results revealed by such studies may then be compared or compiled later on during a systematic review/meta-analysis, as discussed later in this paper.

At the second level, the stratification level, the RR value observed for a factor in a golden sample is further studied in subgroups (strata) related to other stratifying factors in a similar method to that classically used to identify confounding factors. Stratifying factors should be selected with caution and may include risk factors as well as potential confounders. This method allows for the analysis of likely confounding, intermediate and complementary factors. If the exposed and unexposed subgroups in each stratum are comparable, then the following possibilities should be considered. If the RR value is insignificant in all strata, then the stratifying factor may be a confounder or an intermediate factor. If all strata show significant RR values, then the factor being studied is independent of the stratifying factor in producing the outcome. However, in this case, sufficient causes that share both factors may still exist. If the RR value is significant for some strata only, then it is likely that the factor of interest and the risky level of the stratifying factor which reveal a significant RR value are components in common sufficient causes. In the latter two situations, the exposed subgroups (strata) that yield larger significant RR values for the studied factor have a higher availability of complementary factors acting together with the factor of interest. Similar to golden samples, such are deemed golden subgroups because a larger proportion of the outcome cases can be attributed to the studied factor's sufficient causes compared to subgroups with smaller RR values. Therefore, golden cases in these exposed subgroups at the stratification level are known as golden cases. Moreover, the outcome cases arising in such subgroups should undergo further analysis for characteristics thought to act together with the factor of interest.

A golden subgroup should only be labelled as "golden" if it is comparable to its control subgroup in the stratification analysis. As for the study sample level, filtering golden outcome cases in the golden subgroup at the stratification level would likewise be worthwhile, in the same way and for the same reason as previously explained. To this end, outcome cases that arise in the unexposed subgroup in the same strata should be matched according to their key characteristics with outcome cases in the golden case subgroup. Thereafter, a golden case that appears very similar (in terms of its potential risk factors and characteristics) to one or more cases in the unexposed subgroup should be excluded. Once again, the removal of even a few suspected cases is advantageous.

As before, the remaining golden cases should then be compared to other subjects within the exposed subgroup, who did not develop the outcome, so as to identify unique characteristics of golden outcome cases. This process should then be repeated for other subgroups according to other stratifying factors. Subsequently, results revealed by different studies would then be more useful for subsequent systematic reviews/meta-analyses. However, if the strata subgroups are not comparable, then the revealed potential characteristics may be used to form a working hypothesis which can then be confirmed or disproved in subsequent studies.

For example, if the RR value of current smoking status in pro-

ducing a lung cancer outcome in a golden cohort sample is observed to be highest among obese patients, and the smoker obese subgroup is comparable to its control subgroup (*i.e.* non-smoker non-obese individuals), then the smoker obese subgroup can be considered a golden subgroup and the outcome cases as golden cases. After filtering is attempted, the characteristics of golden cases can be compared to those of obese smokers who did not develop lung cancer in order to identify their unique potential characteristics. While the smoker obese subgroup cannot be called a golden subgroup if it is not comparable to its control subgroup, studying the unique characteristics of non-comparable subgroup cases can nevertheless be performed in a similar way; however, as mentioned previously, findings from such comparisons cannot be used to draw firm conclusions but should be used as the basis of a hypothesis to be investigated in further studies. Furthermore, the same analytic process should be undertaken using other stratifying factors.

It is important to note that this new method is subject to certain limitations, as it is built on assumptions and conditions which are often not that easy to assess or meet. For instance, it is relatively uncommon to find a sample with a RR value of 3-4 or higher, however, samples with smaller RR are still valuable. In addition, ensuring comparability between groups can be challenging especially in stratification subgroup analysis. However, studies with big data is the key to get comparable stratification subgroups for such analysis. Moreover, the concept of an additional filtering step can be a 'double-edged sword'. However, most current research methods and statistical approaches are also based on similar assumptions, with each method subject to specific advantages and disadvantages. Simply, we have no other perfect way when dealing with the unseen.

New approach to systematic review/meta-analysis

Another methodological issue is related to systematic reviews/meta-analyses. It seems that the use of the classical approach of conducting systematic review/meta-analysis is threatened.^{3,11} It seems that mixing studies that address specific exposure and showing different results in different populations in order to measure the net effect is unreasonable.³ There is no point in letting different studies in different populations averaging the effect of each other since the prevalence of complementary factors for the studied exposure differ in different populations.³ Therefore, the observed RRs (or OR) 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. An example of such a classical meta-analysis is the one conducted by Aune *et al.* to investigate the association between total meat intake and type 2 diabetes risk.¹² In this classical meta-analysis, the overall effect shown in the net RR suggests no association. However, taking in consideration that the availability of complementary factors of meat consumption differ in different populations, meat consumption seems to be a significant risk factor in some populations and seems a protective factor in others, as it is obvious in the forest plot. However, this variability in the effect of this factor has disappeared by analyzing the net effect that shows no difference, which is a misleading result. The classical net effect that showed no association may lead to ignore this factor in the affected populations.

Therefore, novel approaches related to methods of conducting systematic reviews/meta-analyses related to causation studies are discussed here. First, this research design may be used to assess the likely existence of a factor's sufficient causes - and hence to determine if there is an overall relationship between the factor (or agent) of interest and an outcome - after the factor has been studied in different samples. For this purpose, high-quality studies among dif-

ferent populations whose primary purpose is to study the factor of interest should be included. Most importantly, included studies should ensure the comparability of the exposed and unexposed groups. Conversely, the clinical heterogeneity (variability in the study setting and participants across studies), statistical heterogeneity (variability in the results) and representativeness of the samples are not important criteria for the inclusion of the studies in such a systematic review/meta-analysis. In this regard, without analysing the classical net effect, if some high-quality studies have indicated the likely existence of a factor's sufficient causes (in the form of a significant RR or OR value or risk difference), then it can be concluded that there is a likely association between the factor of interest and the outcome. Due to the possibility of bias and poor academic integrity in which poor-quality data are falsely reported to be of good quality, such conclusions cannot be based on a single study. However, if several high-quality studies among diverse populations have indicated no association, with no conflicting results, then it is equally possible that i) there is no association between the factor and the outcome or ii) that the prevalence of the factor's sufficient causes in the studied samples/populations is rare.³ In such situations, it is preferable to interpret such results as indicative of no observed association to date. Second, a systematic review/meta-analysis can be used to assess the contribution of a factor of interest in causing the outcome of interest within a specific population. This can be assessed by estimating the likely prevalence of a factor's sufficient causes within the exposed group (risk difference) or the related population (population-attributable risk).³ Different studies of the same population may reveal different findings due to variations in the representativeness of the chosen samples, study designs and quality of data, as well as differences in the characteristics of each sample. Therefore, the population must be as specific as possible so that samples share at least certain key characteristics such as, but not limited to, ethnicity, lifestyle and culture. Hence, high-quality studies conducted among the same population should be included within such analysis, and the representativeness should be one of the most important criteria utilised to include/exclude various studies. Studies in which the exposed or unexposed group is not representative of the chosen population (apart from the factor of interest) must be excluded because in this case the risk difference does not represent the prevalence of the factor's sufficient causes in the exposed population, and the calculated population attributable risk does not represent the prevalence of the factor's sufficient causes within the total population.³ Notably, the clinical homogeneity of included studies is important, while statistical heterogeneity is not. As such, the prevalence of the factor's sufficient causes among the exposed population (risk difference) is averaged based on sample size and hence the likely population attributable risk (*i.e.* the prevalence of the factor's sufficient causes within the population) can be calculated by multiplying the averaged risk difference in the prevalence of that factor within the total population.³

More importantly, systematic reviews/meta-analyses should be used to study the unique characteristics (*i.e.* the potential complementary factors acting together with the factor of interest) observed for golden cases in different studies, at both the study sample and stratification levels (as discussed in a former part of this article). However, studies with higher RR or OR values should be given more weight, while the clinical and statistical homogeneity are not important criteria for the inclusion of studies. In such analysis, studies conducted in different settings and showing different results are required to reach valid conclusions. As a result of critically comparing the results of included studies, the analysis may yield important conclusions related to common complementary factors acting with a specific factor. Later on, different meta-analysis conducted to assess

complementary factors for different classes of sufficient causes may be included in one big analysis to identify common complimentary factors acting in different classes of sufficient causes. For example, after studying complementary factors for smoking and complementary factors for obesity (each in a separate meta-analysis), then we can identify the common complementary factors acting with both exposures. This will be of a great value in epidemiology and preventive medicine. Additionally, systematic reviews/meta-analyses may be used to compare the distribution of different classes of sufficient causes in various populations. While this is similar to the classical method of approaching a systematic review/meta-analysis, the aim here is not to study the net averaged effect (*e.g.* net RR or OR) which is thought to be meaningless,^{3,11} but to compare the prevalence of a factor's sufficient causes within different exposed populations (risk differences) and their prevalence in the total populations (population-attributable risk),³ revealed by different studies in diverse populations. However, the usefulness of such a review is questionable since different populations are independent of each other.

Conclusions

For the first time in literature, this paper sought to suggest several promising new research approaches based on the sufficient-component cause model. A new approach was discussed to identify potential mechanisms of disease occurrence which may be useful for risk prediction purposes. In addition, a novel suggestion was made to identify potential component causes in common sufficient causes by focusing on "golden" cases using a novel "exposed case-control" study design. Furthermore, due to the threatened classical systematic review/meta-analysis approach, an alternative approach of conducting a systematic review/meta-analysis was introduced. Future research may be conducted to verify and utilize the newly suggested approaches.

Correspondence: Abdul Hakeem Alrawahi. Head of Research, Department of Research and Studies, Oman Medical Specialty Board, Muscat, Oman.
Tel. +968.99337891.
E-mail: suhaibfather@gmail.com

Key words: Sufficient-component cause model; causation; cohort design; meta-analysis; new research methods; exposed case-control design.

Conflict of interest: The author declares no conflict of interest.

Acknowledgements: The author would like to thank Dr Murtadha Alkhabori, a Senior Consultant at Sultan Qaboos University Hospital, for his suggestions related to this paper work.

Received for publication: 28 December 2019.
Accepted for publication: 26 August 2020.

©Copyright: the Author(s), 2020

Licensee PAGEPress, Italy

Journal of Public Health Research 2020;9:1726

doi:10.4081/jphr.2020.1726

This work is licensed under a Creative Commons Attribution NonCommercial 4.0 License (CC BY-NC 4.0).

References

1. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health* 2005;95:S144-50.
2. Rothman KJ, Greenland S, Poole C, Lash TL. Basic concepts, Section I. In: *Modern epidemiology*. 3rd edition. Philadelphia: Lippincot Williams & Wilkins; 2008.
3. AlRawahi AHH. Classical causation research practices and sufficient-component cause model – Appraisal and pitfalls. *Epidemiol Biostat Public Health* 2017;14. doi: 10.2427/12576.
4. Parascandola M, Weed DL. Causation in epidemiology. *J Epidemiol Community Health* 2001;55:905-12.
5. Carlson MD, Morrison RS. Study design, precision, and validity in observational studies. *J Palliat Med* 2009;12:77-82.
6. 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;1:145-69.
7. Hoffmann K, Heidemann C, Weikert C, et al. Estimating the Proportion of Disease due to Classes of Sufficient Causes. *Am J Epidemiol* 2006;163:76-83.
8. Reiber GE, Vileikyte L, Boyko EJ, et al. Causal pathways for incident lower-extremity ulcers in patients with diabetes from two settings. *Diabetes Care* 1999;22:157-62.
9. Inza I, Calvo B, Armañanzas R, et al. Machine learning: an indispensable tool in bioinformatics. *Methods Mol Biol Clifton NJ* 2010;593:25-48.
10. Obermeyer Z, Emanuel EJ. Predicting the future — Big data, machine learning, and clinical medicine. *N Engl J Med* 2016;375:1216–9.
11. Olsen J. What characterises a useful concept of causation in epidemiology? *J Epidemiol Community Health* 2003;57:86-8.
12. Aune D, Ursin G, Veierød MB. Meat consumption and the risk of type 2 diabetes: a systematic review and meta-analysis of cohort studies. *Diabetologia* 2009;52:2277-87.