Evaluating differences in disease occurrence between two groups to determine the effect size of an observed factor based on a multi-variate model

Hui Liu (liuhui60@dlmedu.edu.cn)
Dalian Medical University

Research article

Keywords: risk evaluation, analysis model, cohort study, relative risk, absolute risk transformation, outcomes

Posted Date: November 26th, 2019

DOI: https://doi.org/10.21203/rs.2.17723/v1

Abstract

Background To describe a method to determine the effect size of an observed factor for a disease by evaluating the difference in the frequency of disease occurrence between exposed and unexposed groups.

Methods A model of multiple pathogenic factors was established by analyzing the number and distribution of observed factors in a study population. The difference in the incidence between two groups (exposed and unexposed) was calculated according to the model.

Results The distribution of observed factors in the population did not influence the difference in disease incidence between the two groups. However, the difference in incidence between the two groups was able to correctly reflect the number of factors combined in the models, and therefore, indicates that group differences in incidence is a reasonable indicator of effect size. Difference scores <0.25 indicate that one of four or more factors plays a role in a disease; scores >0.50 indicate one of two factors plays a role in disease and implies a high intensity level of the factor.

Conclusions A difference in incidence between two groups over 0.25 is suggested as an indicator of a substantial effect size.

Background

Complex events, those in which many factors exert synergetic effects, are frequently observed not only in medical practice but also in our everyday lives. Currently, the pathogenesis of most diseases is related to interactions among extrinsic and intrinsic risk factors [1–4]. Cohort studies, which observe the association between a specific factor and a disease, are considered to be the most reliable form of scientific evidence in the hierarchy of epidemiological evidence [5–7]. In such a study, a putative risk factor is used as an exposure variable, the exposed and unexposed study participants are observed until they develop the outcome of interest, and the difference in incidence between the two groups is evaluated by statistical methods. However, a statistical difference does not indicate the strength of the effect of an observed factor on a disease.

Quantitative variations in a particular event are normally distributed in terms of changes in ratios [8–10] and absolute values [11–13], such as relative risk (RR) [8–10] and absolute risk (AR) [11–13]. When the values of cardinal numbers are relatively small, an increase in a ratio may be very high although the absolute increase may not be. In contrast, when the values of cardinal numbers are relatively large, an increase in a ratio may not be high but the absolute increase may be highly significant. Thus, RR and AR are not comparable. Therefore, it is important to evaluate which risk indicator (RR or AR) is a better measure of the association of the risk factor with the disease. This can be done by comparing the difference in the frequency of disease occurrence between the exposed and unexposed groups, which indicates the strength of the effect of the observed factor (its effect size) on the disease. Here, we propose a multiple risk-factor model for cohort studies to evaluate more reliable measures of the strength

of the association, or functional intensity, between risk factors and outcomes. We believe such a model has the potential to solve the aforementioned problem.

Methods

Multiple risk-factor model

The basic assumptions of the analytical model are: (1) the prevalence of the different observed factors is independent of each other and play a role in a superimposed manner, regardless of interaction or weight function; and (2) a chronic disease is a continuous process of the superimposed manner of risk factors.

A four-factor model simulating pathogenic data was established. Four sets of random numbers with binomial distributions (P = 0.5, N = 100,000) were generated using SPSS statistical software. The four sets of data, which were independent of each other, were named A, B, C, and D. By adding the four sets of data to create group results for the ABCD group, group A can be regarded as a factor of the ABCD group, as shown in Figure 1. The highest value of ABCD was used as the denominator to convert value of ABCD from 0 to 1. A higher number of risk factors indicates a higher probability of disease. Hence, A can be regarded as a cause of ABCD; this model was named the four-factor model, which has a probability of 0.5.

Figure 1

In the same way, four-factor models simulating pathogenic data in which the probability of the risk factor was 0.01 and 0.001 in the study population (four-factor models with 0.01 and 0.001) were established to observe the influence of risk-factor distribution on differences in incidence between the two groups.

Evaluation of effect size

In a similar way, a three-factor model with 0.5 (A vs ACB) and a two-factor model with 0.5 (A vs AB) were established to evaluate the differences in the magnitude of the associations between the observed factors and outcomes. Using the A group as the cause group and ABCD, ABC, and AB as results, a cohort study was established to generate simulated results. The difference in the observed occurrence of disease between the two groups was then calculated to evaluate the effect sizes. That differences from the four-, three-, and two-factor models were approximately 0.25(1/4), 0.33(1/3), and 0.50(1/2), which can be considered to be reasonable measures of effect sizes.

Results

The differences in incidence between the two groups in the four-factors model with probabilities of 0.5, 0.01, and, 0.001 are listed in Table 1. A value of approximately 0.25 was obtained from all three of the models, indicating that the distribution of observed factors in the population did not influence the differences in disease incidence between the two groups.

Table 1

The risk indices (RR and AR) from the four-, three-, and two-factors models are shown in Table 2. The differences in incidence between the exposed and unexposed groups from the standardized analytical model correspond to the number of factors combined in the models, indicating that the difference between the two groups (AR) should be considered a reasonable measure of effect size.

Table 2

Discussion

The present study employed models of multiple pathogenic factors that examined the effects of the number factors and the distribution of factors in the population. The results of the models indicate this methodology can be used as a pragmatic, common-sense approach to intuitively understanding the roles of observed factors in complex biological events. We found the distribution of the observed factors in the population had no influence on the differences in the incidence of disease between two groups. We also found the difference in incidence between the two groups correctly reflected the number of factors combined in the models, and therefore, the difference in incidence between the two groups can be considered a reasonable indicator of effect size, which can be used to evaluate the intensity of an observed factor. We propose that an effect size (difference score): ranging from 0 to 0.25 indicates a weak intensity factor (which can be understood as one of more than four factors playing roles in a disease under the standard model); ranging from 0.25 to 0.50 indicates a moderate intensity factor (which implies that one or two of three factors plays a role in a disease); and ranging from 0.50 to 1.00 indicates a high intensity factor (which implies that one factor mainly plays a role in a disease); values in excess of 0.75 indicate only one observed factor plays a role in a disease. Thus, the intensity of a particular observed factor can reasonably be quantified by the obtained effect size.

The main contribution of this study is that we established a model of multiple pathogenic factors. The result showed that RR does not truly reflect the intensity of a risk factor on a disease outcome, based on the multiple pathogenic-factor model. When the values of the cardinal numbers are relatively small, the ratio may be very high even though the absolute difference may not be high. As disease occurrence is a small probability event, we think the RR may overestimate the effect of an observed factor on a disease. We propose that AR provides a better method of evaluating the risk represented by different conditions that arise in populations, even though AR is not widely used.

Conclusion

Because it is difficult to study the roles of more than four risk factors in a disease, we do not think a factor with an effect size (i.e., group difference) less than 0.25 should be considered a clinically significant factor, even if the observed difference is statistically significant. We suggest a group difference over 0.25 is a substantial effect size because such factors will have a strong effect on disease. Evaluating effect sizes using a model of multiple pathogenic factors could increase our understanding of quantitative variations in measures of association using new concepts.

Abbreviations

AR: absolute risk; RR: relative risk

Declarations

Ethics approval and consent to participate

Not applicable

Consent to publish

Not applicable

Availability of data and materials

The data used to support the findings of this study are available from the corresponding author upon request.

Competing interests

None declared

Funding

None

Authors' Contributions

L. H. conceived the analysis and wrote the final version of the manuscript. I have read and approved the manuscript.

Acknowledgments

We would like to thank the native English speaking scientists of Elixigen Company (Huntington Beach, California) for editing my manuscript.

References

1. Hui L. Quantifying the effects of aging and urbanization on major gastrointestinal diseases to guide preventative strategies. BMC Gastroenterol. 2018;18(1):145.
2. Xin G, Yang G, Hui L. Study to assess whether waist circumference and changes in serum glucose and lipid profile are independent variables for the CETP gene. Diabetes Res Clin Pract. 2014;106(1):95-100.
3. Hui L. Assessment of the role of ageing and non-ageing factors in death from non-communicable diseases based on a cumulative frequency model. Sci Rep. 2017;7(1):8159.
4. Wenbo L, Congxia B, Hui L. Genetic and environmental-genetic interaction rules for the myopia based on a family exposed to risk from a myopic environment. Gene. 2017; 626:305-8.
5. Durr-E-Sadaf. How to apply evidence-based principles in clinical dentistry. J Multidiscip Healthc. 2019;12:131-136.
6. Wallace DK. Evidence-based medicine and levels of evidence. Am Orthopt J. 2010;60:2-5.
7. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. Plast Reconstr Surg. 2011;128(1):305-10.
8. Lemans JVC, Wijdicks SPJ, Boot W, Govaert GAM, Houwert RM, Öner FC, Kruyt MC. Intrawound Treatment for Prevention of Surgical Site Infections in Instrumented Spinal Surgery: A Systematic Comparative Effectiveness Review and Meta-Analysis. Global Spine J. 2019;9(2):219-230.
9. Zhu X, Wu S. Risk of hypertension in Cancer patients treated with Abiraterone: a meta-analysis. Clin Hypertens. 2019;25:5.
10. Hui L, Jun T, Jing Y, Yu W. Screening of cerebral infarction-related genetic markers using a Cox regression analysis between onset age and heterozygosity at randomly selected short tandem repeat loci. J Thromb Thrombolysis. 2012;33(4):318-21.
11. Scribani M, Norberg M, Lindvall K, Weinehall L, Sorensen J, Jenkins P. Sex-specific associations between body mass index and death before life expectancy: a comparative study from the USA and Sweden. Glob Health Action. 2019;12(1):1580973.
12. Hydes TJ, Burton R, Inskip H, Bellis MA, Sheron N. A comparison of gender-linked population cancer risks between alcohol and tobacco: how many cigarettes are there in a bottle of wine? BMC Public Health. 2019;19(1):316.
13. Nelson M. Management of "Hypertension" Based on Blood Pressure Level Versus an Absolute Cardiovascular Risk Approach. Curr Hypertens Rep. 2019;21(1):6.

Tables

Table 1 Influence of the distribution of the observed factor in the population on the difference in the incidence between the exposed and unexposed groups

Observed factor			Incidence		Difference	
Distribution	Number combined	Groups	Mean	Median	Mean	Median
0.500	4	Exposed	0.624	0.500	0.250	0.250
		Unexposed	0.374	0.250		
0.010	4	Exposed	0.258	0.250	0.251	0.250
		Unexposed	0.007	0.000		
0.001	4	Exposed	0.250	0.250	0.249	0.250
		Unexposed	0.001	0.000		

Table 2 Evaluation of risk indicators, including ratios and differences in incidence, between the exposed and unexposed groups

		Results		RR (Exposed/Unexposed)		AR (Exposed-Unexposed)	
Model	Cause (A)	Mean	Median	Mean	Median	Mean	Median
Foure factor	Exposed	0.624	0.500	1.668	2.000	0.250	0.250
A vs ABCD	Unexposed	0.374	0.250				
Three factor	Exposed	0.665	0.333	2.003	2.003	0.333	0.334
A vs ABC	Unexposed	0.332	0.667				
Two factor	Exposed	0.749	0	3.008	-	0.500	0.500
A vs AB	Unexposed	0.249	0.500				

Figures

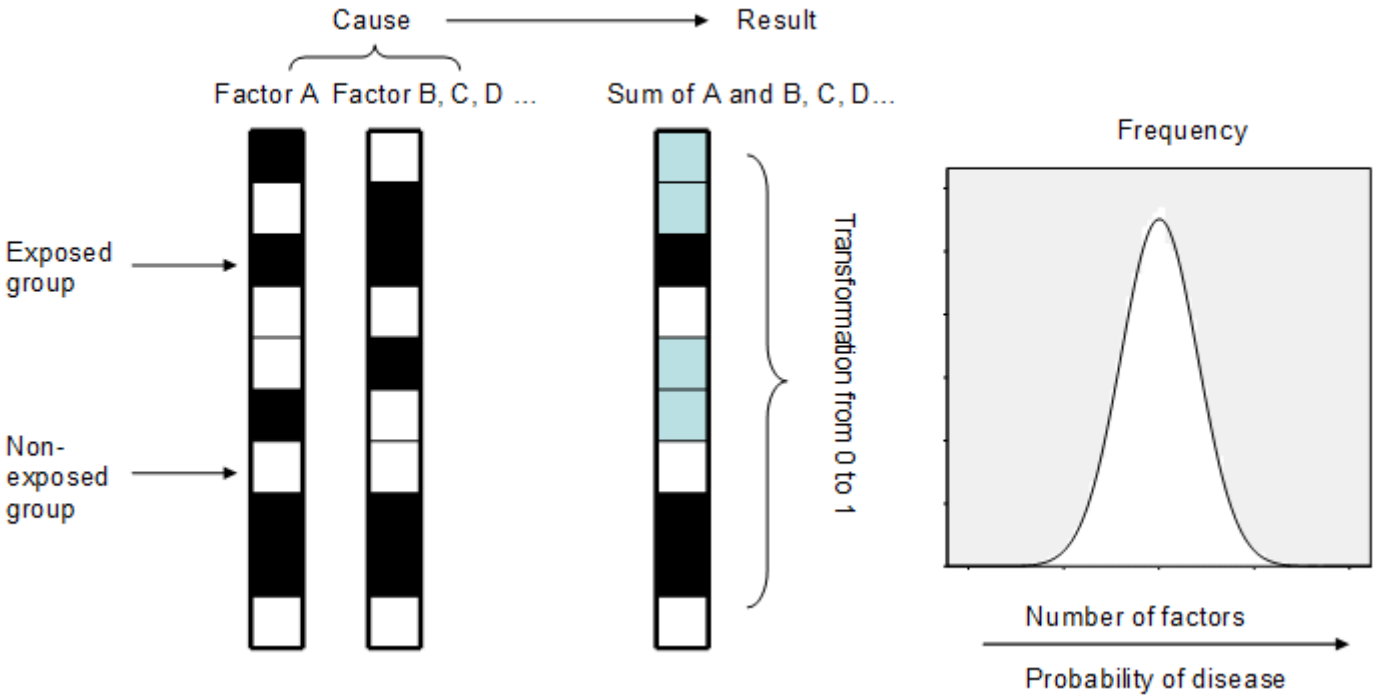

Figure 1

Model of multiple pathogenic factors to assess the association between a causal factor (observed factor) and an ABCD composite factor (disease).