Multivariable Analysis

Marlies Wakkee[1], Loes M. Hollestein[1] and Tamar Nijsten[1]

Journal of Investigative Dermatology (2014) **134**, e20. doi:10.1038/jid.2014.132

In all observational research, one will sooner or later be confronted with the question of whether a certain exposure is related to an outcome. For example, is the risk of cutaneous melanoma affected by the use of nonsteroidal anti-inflammatory drugs (NSAIDs) or is psoriasis an independent predictor for the occurrence of cardiovascular diseases? Questions like these can be answered using multivariable regression analysis. This technique can be used in observational research to adjust for confounders, to assess the effect size of risk factors, or to develop prediction models.

Researchers with limited epidemiological background may feel uncertain how to appraise studies using multivariable regression analysis, let alone use multivariable regression analysis themselves. The objective of this article is therefore to provide a practical overview of the basic principles of multivariable analysis, illustrated with various examples.

WHAT MULTIVARIABLE REGRESSION ANALYSIS DOES

- Aims to explore how multiple risk factors are independently related to an outcome.
- Applies to various models depending on the distribution and temporal relationship of the outcome.
- Allows adjustment for known and available confounders.
- Enables accounting for statistical interaction between independent variables.

LIMITATIONS

- Multivariable regression analysis provides information on potential associations, but a significant association does not automatically imply causality.
- It only adjusts for measured confounding.

WHAT IS A MULTIVARIABLE ANALYSIS?

Multivariable analysis is a statistical technique that can be used to simultaneously explore whether multiple risk factors (referred to as independent variables) are related to a certain outcome (referred to as dependent variable). The type of regression model that is selected depends mainly on the outcome variable and the role of time in the available data (Table 1). In this article we will describe the three most frequently used types of regression analysis: linear regression, logistic regression, and Cox proportional hazards regression analysis, which are generally sufficient to answer most research questions.

Multivariable linear and logistic regression

In cross-sectional and case–control studies, you can use either linear or logistic regression to analyze the data. If the outcome is continuous (e.g., weight), linear regression can be applied and relationships will be represented by β-coefficients. For dichotomous outcomes, such as the presence or absence of a disease, logistic regression is used to calculate odds ratios (ORs). Continuous variables may also be transformed into a dichotomous variable, such as weight into the absence or presence of obesity. Transforming data by grouping results can lead to a loss of information and precision, but the resulting risk estimates may be easier to interpret. If cases and controls are matched for certain risk factors (e.g., age), a conditional logistic regression model should be used. Based on their matched criteria, the cases and controls are linked to form a set to which the multivariable analysis used to adjust for other confounders can be applied. Discarding the matched design by adjusting for matched factors in an unconditional analysis leads to a bias toward the null. Interactions with the matching variables can still be taken into account.

In all models it is assumed that the independent variables have a linear relationship with the continuous outcome (linear regression) or logarithm of odds of the dichotomous outcome (logistic regression). In the case of nonlinear relationships, the independent variables can be forced into a normal distribution by taking the natural logarithm or by creating multiple dichotomous variables.

Multivariable Cox proportional hazards analysis

In cohort studies, where the exposure precedes the outcome, data can be analyzed using multivariable Cox proportional

[1]*Department of Dermatology, Erasmus University Medical Center, Rotterdam, The Netherlands*

Correspondence: Marlies Wakkee, Postbus 2040, 3000 CA; Rotterdam, The Netherlands. E-mail: m.wakkee@erasmusmc.nl

© 2014 The Society for Investigative Dermatology

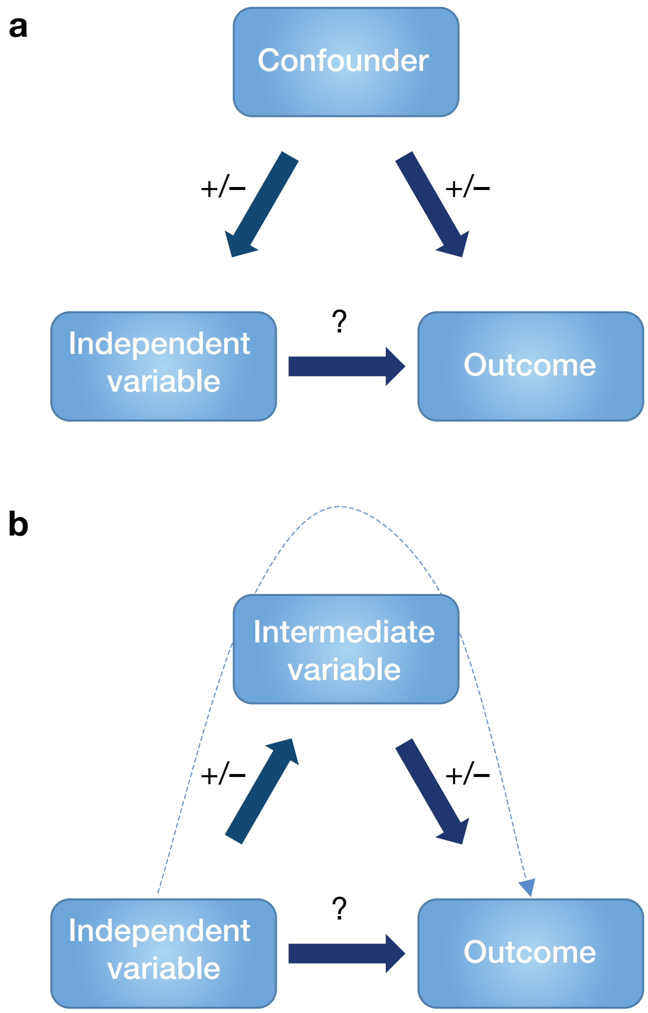

Figure 1. The role of a confounder versus an intermediate variable in the relationship between an independent variable and an outcome. (**a**) The association between an independent variable and an outcome may be confounded. That is, the confounder predicts the independent variable, predicts the outcome, and is not part of the causal pathway, leading to a triangular relationship. Thereby, the association between the independent variable and the outcome is (partly) explained by the confounder. (**b**) The role of an intermediate variable in the relation between an independent variable and an outcome. The intermediate variable is part of the causal pathway in the relationship between the independent variable and the outcome.

hazards regression analysis, also known as survival analysis. In Cox proportional hazards models the effect of independent variables on survival time is assessed and represented by hazard ratios (HRs). The advantage of Cox proportional hazards analysis is that it includes all observation-years available for each participant until the studied outcome (e.g., a cardiovascular event) or death, or otherwise to the end of the follow-up time, whichever comes first. It also takes into account the exposure time to a risk factor (e.g., days of sun exposure), which may be shorter than the total included follow-up time. The prospective means of data collection in cohort studies result in more precise data with a temporal component. This is in contrast to linear or logistic regression, where subjects are compared at one point in time or over a comparable timeframe (e.g., a two-year period), which implies retrospective data collection or excluding subjects without sufficient follow-up time and therefore losing valuable information. Finally, with Cox proportional hazards analysis it is assumed that all independent variables change linearly with the logarithm of the hazard.

Risk estimates

The β-coefficient obtained from linear regression is directly interpretable as the slope, which denotes the change in dependent variable per unit change in independent variable (Table 1). If the 95% confidence interval (CI) includes the value 0, this represents a nonsignificant association, because a slope of 0 means there is no association (or a non-linear relationship). In the case of logistic or Cox proportional hazards analysis, the ORs or HRs are an exponentiation of this β-coefficient, which results in an outcome that cannot extend below 0 but ranges from 0 to infinity. For ORs or HRs, the 95% CI must exclude the value 1 to demonstrate a significant association (Table 1) because a ratio of 1 means that the odds or hazards are the same for the two groups you are comparing. Reporting 95% CIs is preferred over reporting *P* values because reporting 95% CIs has the advantage of directly including both an effect size (point estimate) as well as the range of values in which the true value lies (width of 95% CI), rather than stating only whether a statistically significant difference is observed.

Examples of multivariable analysis

Multivariable conditional logistic regression was applied in an age- and gender-matched case–control study investigating the association between cutaneous melanoma (CM) as a dichotomous dependent variable and exposure to NSAIDs as an independent variable (Curiel-Lewandrowski *et al.*, 2011). Matching for age and gender was done by including at least one community-based control subject from the same 5-year age group and gender for every subject with CM in this study. The odds of CM were significantly lower among those using aspirin, with a crude OR of 0.75 and a 95% CI of 0.57–0.97, which actually represents an adjusted OR because it already includes adjustment for age and gender by matching the case and control subjects. This effect of aspirin was even stronger after adjusting for the number of sunburns during childhood in the multivariable model, resulting in an adjusted OR of 0.72 (95% CI 0.55–0.94) (Curiel-Lewandrowski *et al.*, 2011).

In a population-based cohort study assessing the association between psoriasis and cardiovascular events, multivariable Cox proportional hazards analysis was applied to adjust for other cardiovascular risk factors, including the significantly younger age of the psoriasis cohort (Dowlatshahi *et al.*, 2013). The crude HR showed a borderline significantly decreased cardiovascular risk with a HR of 0.69 and a 95% CI 0.48–1.00 for psoriasis patients compared to reference subjects. In this case the HR is called "borderline significant" because the upper limit of the CI includes 1; this is also reflected in the *P* value of 0.05. After adjusting for the total cardiovascular risk profile, the HR was 0.73 with a 95% CI between 0.50 and 1.06, which is not significant because the value 1 lies

© 2014 The Society for Investigative Dermatology

Download English Version:

https://daneshyari.com/en/article/3215309

Download Persian Version:

https://daneshyari.com/article/3215309

Daneshyari.com