Logistic regression and odds ratios as means to adjust for baseline incomparability's in randomised controlled trials

Description of the disadvantages, alternatives and frequency of use

Ruben G. Duijnhoven

Under the supervision of dr. Mirjam J. Knol

Julius Center for Health Sciences and Primary Care
University Medical Center Utrecht

Contents

Abstract 3
Abbreviations 4
Introduction 5
Alternatives 6
Objective 8
Methods 9
Data extraction 9
Analysis 10
Results 11
Discussion 13
Bibliography 15

Abstract

Background

In randomised controlled trials (RCTs) adjustment for baseline characteristics is often done. When logistic regression is used, the measure of association will be the odds ratio (OR). If the incidence of the outcome is high (>10%) the odds ratio is incomparable to the risk ratio. Furthermore, the larger the effect of the intervention (OR < 0.5 or > 2.0) the larger the difference between the odds ratio and risk ratio. Additionally, we reviewed how often ORs are reported which are incomparable to risk ratios in RCTs and how frequently researchers use logistic regression in RCTs.

In this report we present alternative methods to adjust for baseline incomparability's with which it is possible to calculate RRs directly. By analysing RCT we determine to what extent alternatives for logistic regression are used.

Methods

We reviewed all 298 RCTs published between January 1st and December 31st 2008 in the British Medical Journal, Annals of Internal Medicine, The Lancet, The Journal of the American Medical Association and The New England Journal of Medicine. All articles were scored on a standardised form. We calculated the percentage of RCTs that presented an OR below 0.5 or above 2.0, an incidence above 10%, or both.

Results

In 78 RCTs (26.2%) logistic regression was used for any purpose and 55 RCTs (18.5%) reported one or more ORs. ORs smaller than 0.5 or/and larger than 2.0 were reported in 37 RCTs (12.4%), and in 20 RCTs (6.7%) at least one overall incidence reported was above or equal to 10%. In 9 RCTs (3.0%) an OR below 0.5 or above 2.0 with a corresponding incidence above 10% was reported, this is 13% of the RCTs that reported ORs.

Alternative methods for baseline adjustment in RCTs are used only 34 times out of the 298 publications that were analysed.

Conclusion

Our results show that in 1 in 33 RCTs published in five major general clinical journals the presented odds ratio cannot be interpreted as a risk ratio. The use of logistic regression occurs in 1 in 4 RCTs while there are more appropriate models available.

The available alternative methods are very scarcely used, although they can be easily applied with modern statistical computer programs.

A misinterpretation of risks or benefits of e.g. surgical procedures or pharmacotherapy may have life threatening consequences and could cost huge amounts of money in clinical settings.

Abbreviations

Ann Int Med	Annals of Internal Medicine	N Engl J Med	New England Journal of Medicine
BMJ	British Medical Journal	OR	Odds ratio
CI	Confidence interval	RCT	Randomised controlled trial
GLM	Generalised Linear Model	RR	Risk ratio
JAMA	Journal of the American Medical Association		

Introduction

In randomised controlled trials it is often necessary to adjust for baseline incomparability's such as differences in comorbidities between individuals, or use of a specific drug. A method particularly popular to be used for baseline adjustment is logistic regression. However, logistic regression has the important disadvantage of yielding an odds ratio as the effect measure instead of a risk ratio (also called a relative risk).

Odds ratios (ORs) are approximately the same as risk ratios when the incidence of the event of interest overall or in the control or intervention group, is rare (it occurs in less than approximately 10% of the cases) [1, 2]. If the incidence is larger there will be an overestimation of the effect when the OR is interpreted as if it is a risk ratio. A second circumstance under which the OR should not be interpreted as a RR, is when the OR is smaller than 0.5 or larger than 2.0. These cut-off values are chosen arbitrarily but based on knowledge about the relation between the OR and RR. This relation is shown visually in the graph that is shown as Figure 1. In the case of an OR below 0.5 an underestimation of the effect would be made and if the OR is larger than 2.0 an overestimation of the effect.

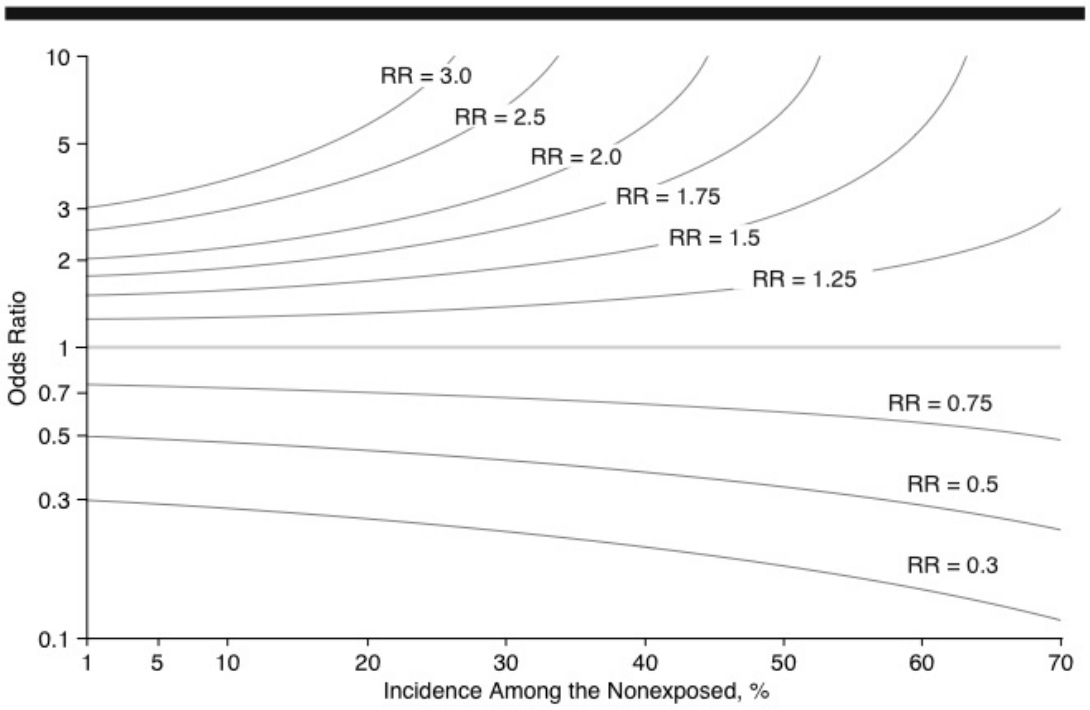

Figure 1 | The relation of the difference between the value of the OR and the RR by a varying incidence percentage [1].

Alternatives

Zhang and Yu have reported a method to transform the odds ratio and its corresponding confidence interval into a risk ratio [1]. Because the formula is relatively simple this method quickly became quite popular. However, this method is criticised as it is unreliable and the new risk ratios and confidence intervals are often different from the risk ratio's directly calculated using a correct method [2-5]. Better solutions are found in the use of Mantel-Haenszel stratified analysis, log-binomial regression, Poisson regression and Cox regression [2, 5, 6].

Mantel-Haenszel analysis

The simplest method to adjust for baseline incomparability's in RCTs is by using a Mantel-Haenszel analysis. Some researchers have used the method as if it were a gold standard for baseline adjusted risk ratios as it is reliable and easy [2, 6]. A disadvantage of the method is that it can't adjust for variables that are not, or in very small numbers present in every group that is created by stratification. In the Mantel-Haenszel analysis these missing variables would result in empty cells, hampering the analysis [7]. This limitation makes the method sometimes unsuitable for the adjustment of trial results as some trails are performed with small numbers of patients what will be a limiting factor for stratification [7, 8].

Log-binomial regression

The use of log-binomial regression to calculate risk ratios and simultaneously adjust for multiple covariates in epidemiology was already described in 1986 [9]. It was then described for the statistical analysis program GLIM, but it is possible to perform the calculation with SAS, as well as several other programs [4, 6].

Binomial regression is of the type of the 'generalised linear models' using a binomial distribution and a log link. In logistic regression the binomial distribution with a logit link is used. Modern computers are sufficiently powerful to fit a risk ratio to the model and statistical programs overcome the difficulties of the calculations that used to be a drawback for the average researcher. A possible disadvantage of the binomial regression is the outcome; the calculation does not yield a value between zero and one by definition as is necessary for risks, but can take a value between zero and infinity above zero. Log-binominal regression overcomes this problem. The logarithmic transformation makes result values below zero impossible.

Robbins showed that binomial regression is a better method to calculate risks ratios and the corresponding confidence intervals than is the method of Zhang and Yu [4]. In some cases (log-) binomial regression does not converge, which makes it impossible to calculate risk ratios [5]. Explanations why (log-) binomial regression sometimes does not converge are usually related to starting values that approach the boundaries of the possible values, i.e. values approach 0 or 1 or are even larger than 1. Lack of convergence may also be caused by inclusion of a continuous variable, or may be due an inappropriate default setting in statistical software [2, 6].

Poisson regression

Poisson regression is a generalised linear model with a Poisson distribution and a log link. Usually this regression model is used in analysing count data. Its use is normally limited to data with a rare incidence of the outcome measure. When Poisson regression is applied to binomial data with a frequent outcome (not a rare event) an overestimation of the error for the risk ratio will be made. To prevent a false error, the calculation can be expanded with a robust error variance estimation procedure; a generalised estimating equation (GEE). This GEE (also called 'sandwich') procedure applied to Poisson regression by Zou [5] results in a modified Poisson regression, also known as robust Poisson regression.

The outcome of a robust Poisson regression model is reliable; it is comparable to that yielded by a Mantel-Haenszel stratified analysis, but also applicable if correction for a larger number of variables is needed [3, 5, 6]. Unlike it is the case in log-binomial regression, when using Poisson regression convergence has never been reported to pose any problems related to convergence and the estimated risk ratio will always remain between 0 and 1 [3, 5].

Cox regression

Cox regression (also known as a proportional hazards model or a Cox model) [10] is usually applied to analyse time-to-event data e.g. survival over time after treatment or time to first hospitalisation. If time is kept equal for all persons in this model the results that are in Cox regression called hazard ratios, are in fact just risk ratios.

Just like log-binomial regression and Poisson regression a Cox model needs a robust sandwich estimate (GEE) to calculate correct CIs for the risk ratios, the results are exactly equal if these methods are compared [11, 12].

Cox regression is not a regression model based on the generalised linear model. The model does not follow a specific distribution and thus no a priori assumption about the course of the survival curve is assumed [7]. There are no circumstances reported under which the calculation would fail to produce a risk ratio and the outcome will be between 0.0 and 1.0 under all conditions. However, a theoretical but tenuous disadvantage to the method is that it

is not a very intrinsic approach to use Cox regression for this purpose. That is to say Cox regression is normally associated with survival analysis and its use to calculate a risk ratio alone might be preposterous.

Barros et al. showed that Cox regression has a difference in the risk ratio of less than 5.0% compared to Mantel-Haenszel [6].

Objective

In this review we aim to analyse the randomised controlled trials published in five major general medical journals investigating the frequency of the use of logistic regression and ORs. The number of RCT in which these ORs were calculated for outcomes that were not rare will be counted to gain knowledge about possible misinterpretation.

Furthermore we want to measure the frequency of the use of alternative methods as were described in the introductory part of this report; Mantel-Haenszel stratified analysis, log-binomial regression, Poisson regression and Cox regression for the calculation of RRs.

Methods

Data extraction

We included all randomised controlled trials (RCTs) published in 2008 in the British Medical Journal, Annals of Internal Medicine, The Lancet, The New England Journal of Medicine and the Journal of the American Medical Association. We searched PubMed with the search line (([Journal] AND [Volume]) AND "randomized controlled trial" [Publication type]), for example (("BMJ" [Journal]) AND 337[Volume]) AND "randomized controlled trial" [Publication type]).

Subsequently a standardised data extraction form to score all articles was made and piloted on six articles. On the form the first author and first page of the article, the journal, the sample size (total number of patients included), the type of research (pharmacotherapeutic, surgical, physiotherapy, behavioural therapy or other), the use of logistic regression, Cox regression, log-binomial regression, Poisson regression and Mantel-Haenszel methods or other statistical methods were scored. Finally if ORs were reported they were written down together with the incidence of the events relating to the reported OR for the group overall and the control group alone. The overall group was defined as the groups composed of the combination of the both therapeutic groups in the trial. The control group is the group that was used as the 'control' in the odds ratio calculation, meaning it served as the group used as the denominator if the OR would be calculated directly by hand.
A maximum of ten ORs were noted. If more than ten ORs were reported in the article this was noted on the data extraction form.

The type of research was based on the intervention; categories were 'pharmacotherapeutic', 'surgical', 'physiotherapy' or 'behavioural therapy' or 'other'. If several categories were combined the article was scored in both categories. Some trials had vitamins or supplements as determinants; they were scored as 'pharmacotherapeutic'. The sample size was read from the flowchart describing the patient selection process, or from the baseline table if a flowchart was not available.
In the same way all articles reporting the use of two or more selected methods (Mantel-Haenszel, (log-)binomial regression, Poisson regression and Cox regression) were scored on all applicable methods. If only statistical methods but the five aforementioned were used, e.g. χ^2 tests, linear mixed model, Student t-tests or Mann-Whitney tests they were categorised as 'other'.

Analysis

All analyses were performed using SPSS 15 (SPSS Inc., Chicago, IL) for Windows. We calculated the percentage of types of research as well as of the use of logistic regression, Cox regression, log-binomial regression and Poisson regression and of the reporting of ORs. Further more we counted the number of published RCTs in the selected journals. We also determined the median and the maximum and minimum for the study population size.

We have chosen boundaries in order to evaluate the comparability of the values of ORs as if the were RRs. For these boundaries we choose to consider below 0.5 and above 2.0 as relatively low or high, respectively. ORs that were below 0.5 and above 2.0 were considered to present a markedly different effect measure than would be the case if they were interpreted as if they were RRs.

We also categorised all incidences; recoding them 1 if the incidence was above 10% and 0 for all incidences that were below 10%. Finally, we determined how many articles were published with either one or more incidences above 10% in the 'control' group and ORs below 0.5 or above 2.0 corresponding to the same outcome measure. The control group was chosen because there were less missing incidences in the control group as compared to the overall group.

Results

The search yielded 328 articles, of which 30 appeared to be other publications than RCTs falsely coded as being a RCT in PubMed (most of them were letters or comments on RCTs) and were excluded. Therefore the total number of RCTs published in 2008 in the British Medical Journal, the Lancet, the New England Journal of Medicine, the Journal of the American Medical Association and the Annals of Internal Medicine was considered to be 298.

Table 1. Characteristics of the included articles		
Type intervention in RCT	Frequency	(%)
Pharmacotherapeutic	203	(68.1 %)
Surgical	17	(5.7 %)
Physiotherapy	10	(3.4 %)
Behavioural therapy	19	(6.4 %)
Other	60	(20.1 %)
Total	309	(103.7 %) [1]
Number of RCTs [2]		
BMJ	49	(16.4 %)
JAMA	52	(17.4 %)
N Engl J Med	95	(31.9 %)
Lancet	82	(27.5 %)
Ann Intern Med	20	(6.7 %)
Total	298	(100 %)
Sample size		
Median: 756,5		
Min.: 16	Max.: 106000	

The various articles type of research, the journal and sample size are described in Table 1.

Of all articles that were included in the analysis (results also in Table 2) there were 78 (26.3 %) that reported the use of logistic regression for any purpose (either to make a model or to calculate an adjusted OR). Of these 78 trials there were 47 that also reported ORs (15.4 %). In total 55 articles reported ORs as a measure of association. The difference of eight articles in which no logistic regression was used but ORs were reported indicates that the OR was used eight times as a first measure of association in RCTs, in which it does not originate from baseline adjustment through logistic regression.

The analysis of the regression models and Mantel-Haenszel methods used (results summarised in Table 2) show that logistic regression is indeed a widely used model.

[1] Due to the possibility to be in more than one category only, the total is >100 %

[2] Number of RCTs published in journal in 2008

Table 2. Statistical methods used in RCTs		
Method	Frequency	(%)
Logistic regression	78	26,2
Binomial regression	11	3,7
Poisson regression	21	7,0
Cox regression	93	31,2
Mantel-Haenszel method	29	9,7
No adjustment baseline	106	34,8
Total	338	112,6

Binomial regression includes all binomial regression methods such as negative binomial regression as well as log-binomial regression that we propose as alternative for logistic regression. Log-binomial regression to calculate RRs as proposed by McNutt et al. [2] based on the ideas of Wacholder [9], was described only two times.

Cox regression was not used to calculate RRs in any of the RCTs. Poisson regression was used as a method to calculate RRs as proposed by Zhou [5] in three articles. The only method that is commonly used is the Mantel-Haenszel method that was used in 29 articles.

Out of the 55 articles that reported ORs, 37 times one or more of the ORs was below 0.5 or above 2.0. An incidence of over 10% was seen in 26 or 32 in the overall or control incidence in the publications, respectively. This difference arises because there are less missing values of the incidence in the control groups than in the overall groups. An important other remark about the incidences is that in many articles one or more incidences corresponding to the reported ORs were not available, because authors did not include them in the report. In 14 RCTs it was impossible to relate an incidence to one or more of the ORs that were presented in the article.

When reviewing the number of RCTs that reported both an OR of over 2,0 or below 0,5 and a corresponding incidence of over 10 % we identified 13 articles (4,7 %).

Reports presenting ORs that were considered out of the range in which they can be interpreted as RRs are not very common relatively to all published RCTs, but are a substantial part of the publications that report ORs. In almost one out of every four publications that describe ORs (13 out of 55) the ORs should not be interpreted as if they were RRs.

Discussion

In this evaluation of RCTs published over the past year we show that logistic regression is frequently used. We have also shown that it is still quite common to present ORs even though it is possible without too many difficulties to calculate RRs. Furthermore we show that of these ORs almost one out of four should not be interpreted as a RR because the outcome is not rare and because the OR is relatively small or large.

The frequency of the use of alternatives for logistic regression as a means to adjust for baseline incomparability's is very low; only the Mantel-Haenszel method is used frequently.

It is interesting that in eight articles the ORs were used as a crude measure of association only, i.e. there was no adjustment for baseline characteristics used. This means that it would have been possible and just as easy to calculate the RR.

Possible explanations why logistic regression is still the foremost statistical method to adjust for baseline incomparability's might be that it is easy to use and well-known, also by today's (statistics) teachers in health sciences.

There has been little attention for the problems associated with the use of ORs. Articles have focused mainly on the 'rare event assumption' when using ORs and the disadvantages of ORs and those are usually letters as a reaction on articles publishing distorted results [12-14]. Some articles paid attention to alternatives [2, 4-6] but all except one (Zhang, [1]) were published in epidemiological journals. Zhang's article on the method to transform ORs in RRs was published in the JAMA in 1998, but the proposed method was criticised, as it is unreliable and yields erroneous results [2-5].

The absence of articles on easy to use alternatives for logistic regression in the large medical journals (such as the Lancet, the N Engl J of Med, the BMJ, the JAMA and the Ann of Int Med) make it very difficult for the average doctor to become familiar with the problem and alternative methods. By making this problem understandable to a broader audience and give accessible descriptions of alternatives it should be able to gradually replace the ORs for RRs in RCTs.

It is interesting that out of all 298 articles there were two that paid attention to either their use of the OR or explained their choice not to use it. Both justified their choice on the 'rare event assumption', in one article the 'outcome was too common to use the OR' in the other the 'use of the OR is possible because the outcome is very rare'.

Background information on ORs was inadequate in a great number of articles. Fourteen times ORs were presented of which neither the incidence was described nor were the values that were used for calculating the OR. Without this information it was impossible to judge if the OR could be read as a RR.
In case no prevalence could be found or calculated from the numbers one only has the value of the OR (< 0,5 or > 2,0) to appraise the OR and this is less informative.

In the discussion of the alternative methods to adjust for baseline incomparability's, we have tried to limit the mathematical background information as much as possible. We hope this makes the alternative methods to adjust for baseline incomparability's understandable an accessible for a broader audience.

We conclude that there is considerable distortion and possible misinterpretation of trial results when the effect size is presented as an OR. Misinterpretation of RCT outcomes due to unsuitable use of ORs can have life threatening consequences or can result in wrong policy wasting money. Alternative methods that yield a RR have become very easy to use, but are still scarcely used in RCTs published in major general medical journals. Clear information on the ease of alternatives for logistic regression to calculate RRs with the most widely used statistical packages is very welcome and should be made available for medical researchers through one or more of the important general medical journals.

Bibliography

1. Zhang, J. and K.F. Yu, *What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes.* JAMA, 1998. **280**(19): p. 1690-1.
2. McNutt, L.A., et al., *Estimating the relative risk in cohort studies and clinical trials of common outcomes.* Am J Epidemiol, 2003. **157**(10): p. 940-3.
3. Greenland, S., *Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies.* Am J Epidemiol, 2004. **160**(4): p. 301-5.
4. Robbins, A.S., S.Y. Chao, and V.P. Fonseca, *What's the relative risk? A method to directly estimate risk ratios in cohort studies of common outcomes.* Ann Epidemiol, 2002. **12**(7): p. 452-4.
5. Zou, G., *A modified poisson regression approach to prospective studies with binary data.* Am J Epidemiol, 2004. **159**(7): p. 702-6.
6. Barros, A.J. and V.N. Hirakata, *Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio.* BMC medical research methodology, 2003. **3**: p. 21.
7. Armitage, P., G. Berry, and J.N.S. Matthews, *Statistical Methods in Medical Research*. 4th ed. 2002, Oxford: Blackwell Science. 817.
8. Mantel, N. and W. Haenszel, *Statistical aspects of the analysis of data from retrospective studies of disease.* J Natl Cancer Inst, 1959. **22**(4): p. 719-48.
9. Wacholder, S., *Binomial regression in GLIM: estimating risk ratios and risk differences.* American Journal of Epidemiology, 1986. **123**(1): p. 174-84.
10. Cox, D., *Regression Models and Life Tables.* J R Stat Soc. B (Methodological), 1972. **34**(2).
11. Greenland, S., *Interpretation and choice of effect measures in epidemiologic analyses.* American Journal of Epidemiology, 1987. **125**(5): p. 761-8.
12. Nurminen, M., *To use or not to use the odds ratio in epidemiologic analyses?* Eur J Epidemiol, 1995. **11**(4): p. 365-71.
13. Altman, D.G., J.J. Deeks, and D.L. Sackett, *Odds ratios should be avoided when events are common.* BMJ, 1998. **317**(7168): p. 1318.
14. Deeks, J., *When can odds ratios mislead? Odds ratios should be used only in case-control studies and logistic regression analyses.* BMJ, 1998. **317**(7166): p. 1155-6; author reply 1156-7.