Journal of Biometrics & Biostatistics
ISSN: 2155-6180

Jakobsen et al., J Biom Biostat 2015, 6:2
DOI: 10.4172/2155-6180.1000227

Review Article Open Access

Count Data Analysis in Randomised Clinical Trials

Jakobsen JC[1,2*], Tamborrino M[3], Winkel P[1], Haase N[4], Perner A[4], Wetterslev J[1] and Gluud C[1]

[1]*Copenhagen Trial Unit, Centre for Clinical Intervention Research, Department 7812, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark*
[2]*Emergency Department, Holbaek Hospital, Holbaek, Denmark*
[3]*Institute for Stochastics, Johannes Kepler University Linz, Austria*
[4]*Department of Intensive Care, Rigshospitalet, University of Copenhagen, Denmark*

Abstract

Choosing the best model for analysis of count data in randomised clinical trials is complicated. In this paper, we review count data analysis with different parametric and non-parametric methods used in randomised clinical trials, and we define procedures for choosing between the two methods and their subtypes. We focus on analysis of simple count data and do not consider methods for analyzing longitudinal count data or Bayesian statistical analysis. We recommend that: (1) a detailed statistical analysis plan is published prior to access to trial data; (2) if there is lack of evidence for a parametric model, both non-parametric tests (either the van Elteren test or the Tadap2 test, based on an aligned rank test with equal stratum weights) and bootstrapping should be used as default methods; and (3) if more than two intervention groups are compared, then the Kruskal–Wallis test may be considered. If our recommendations are followed, the risk of biased results ensuing from analysis of count data in randomised clinical trials is expected to decrease.

Keywords: Count data; Randomised clinical trials; Statistical methodology

Background

It is often optimal to use a dichotomous patient relevant outcome (for example, all-cause mortality) as primary outcome in a randomised clinical trial. Count data outcomes may be used as primary outcomes and are often used as secondary outcomes. Two types of count data are generally used in randomised clinical trials to investigate intervention effects: (a) observations expressed as discrete positive values arising from counting rather than ranking (for example, numbers of serious adverse events or days of intensive care) [1]; (b) number of events occurring in time intervals or space, with models focusing on the rate at which the events occur (for example, number of adverse events per day) [2].

It is possible in some circumstances (often with large number of counts) to analyse count data as continuous data (for example, using ANCOVA). However, it is our experience that count data in randomised clinical trials may rarely be validly analyzed as continuous data because the underlying statistical assumptions are rarely met. Several specific methods are available for analyzing count data, but we may fail in choosing the best method. For example, the underlying assumptions for choosing the model might not be fulfilled and the model may not fit the data properly. Furthermore, comparing groups with multiple tests showing different results will increase the risk of having at least one false positive significant result due to 'play of chance' (type I error) and makes it possible for trialists to choose specific tests based on whether the test show significance or not [3,4]. Therefore, a detailed procedure on how to choose the most reliable method for analysis of count data should be developed and published prior to access to the trial data [5].

In the following we present a review of different parametric and non-parametric methods to analyse count data in randomised clinical trials. We highlight their strengths and limitations, and define a procedure for choosing between the two methods with subtypes, depending on the data for analysis. The aim of our review is to provide guidance trialists in analysis of count data outcomes. We do not present or discuss methods for analysis of longitudinal count data or methods for Bayesian statistical analysis which are separate topics [1,6,7].

Methods

Based on a comprehensive literature search in PubMed and methodological and statistical considerations, selected experienced trialists, methodologists, and statisticians considered a variety of different models for analysing count data in randomised clinical trials.

Results

Design of the count data outcome measure

Our study of the literature, made us infer that the first thing to do is to consider how to design the count data outcome. For example, 'Number of exacerbations' may be analysed as count data without changing the design of the outcome. However, 'days of hospital admission' can, for example, be a problematic outcome in case of early mortality among the patients. In such a case, the patients who die early would contribute with low values for the number of days of admission, which otherwise would be perceived as a beneficial result. Furthermore, it can be difficult to gather count data observations within long-term periods of time, so for practical reasons data may only be collected within a limited time period. It is, therefore, often advantageous to design count data outcomes as, for example, 'days alive and out of hospital within 90 days', or to use rates instead of counts. If the observation period differs among the participants, for example when the first randomised participants are observed for a longer period of time than the participants randomised later during the trial, then the data from each participant can be adjusted according to the length of the observation period for each participant. 'Days alive and out of hospital' may be presented as the percentage of days alive spent outside

***Corresponding author:** Jakobsen JC, Department 7812, Rigshospitalet, Copenhagen University Hospital, Blegdamsvej 9, DK-2100 Copenhagen, Denmark, Tel: +45 2618 6242; E-mail: jcj@ctu.dk

Received April 24, 2015; **Accepted** May 29, 2015; **Published** June 05, 2015

Citation: Jakobsen JC, Tamborrino M, Winkel P, Haase N, Perner A, et al. (2015) Count Data Analysis in Randomised Clinical Trials. J Biom Biostat 6: 227. doi:10.4172/2155-6180.1000227

the hospital within a specified observation period considering not only the index admission, i.e., the outcome is analysed as a rate.

Simple Parametric Tests

Unadjusted count data may be analysed by two sample inferential procedures (for example, a Student's t-test).

The generalised parametric linear model

In many trials we want to adjust the analysis of an outcome using covariates in a regression analysis. It may be essential to adjust for stratification variables and potentially other design variables [8-10]. Therefore, it is preferable, if possible, to analyse count data using a parametric regression model.

The generalized linear model is a generalization of the linear regression model allowing response variables not to be normally distributed [2]. Moreover, the generalized linear model allows the linear model to be related to the response variable via a link function and the magnitude of the variance of each measurement to be a function of its predicted value [2]. Hence, a generalized linear model is made up of a linear predictor:

$$\eta_i = \beta_0 + \beta_1 \chi_{1i} + \ldots + \beta_p \chi_{pi}$$

And two functions:

A link function ($g(\mu_i)$) that describes how the mean, $E(Y_i)=\mu_i$, depends on the linear predictor (η_i)

$$g(\mu_i) = \eta_i$$

- A variance function ($var(Y_i)$) that describes how the variance, $var(Y_i)$, depends on the mean

$$var(Y_i) = \phi V(\mu)$$

A common way to analyse count data is to use Poisson regression models which are generalised linear models based on the Poisson distribution function with the logarithm as the link function [2]. A Poisson distribution is characterised by equal mean and variance and is right-skewed, i.e., the right tail of the distribution is longer [2]. An underlying assumption is that occurrences (count data events) are assumed to be independent [1]. As the mean increases, the skewness decreases and the distribution becomes more symmetric [2]. Depending on the properties of the observed data, for example, the number of observed zeros and the relation between the empirical mean and variance, other underlying distributions may be assumed: (a) a negative binominal may be used when the variance exceeds the mean and the observations are not independent [2]; (b) zero inflated (Poisson and negative binominal) models may be used when the count outcome exhibits more variation than that which is accommodated by the postulated model due to a preponderance of zero counts [1]; and (c) zero truncated (Poisson and negative binominal) models may be used when the value zero cannot occur [2].

The strengths of using these generalised linear models are that trial results may be adjusted for multiple stratification variables, and multiple intervention groups may be compared. Furthermore, odds ratios, incidence rate ratios, ratios of rate ratios, all with confidence intervals may present evidence of the intervention effects with a proper indication of uncertainty [8,9].

Standard approaches for both model checking and model selection need to be applied when choosing the model with the best data fit [11]. Analysis of residuals, for example scatterplots of residuals versus predicted values, is commonly used for model checking in regression analyses. If two or more models provide a good fit of the data via model checking, then the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) can be applied for model selection, provided that the likelihood function (L) of the model is available [11]. In particular, we have AIC=-2 log L+2p and BIC=-2 log L+plogT, where log L denotes the fitted log-likelihood of the model, p is the number of parameters, and T is the number of observations. Then, among the competing models, we shall choose the one with the smallest BIC and AIC (with most importance to BIC).

In practice, to deal with these models can be difficult. If the count data outcomes are designed as we recommend in this review (see '*Design of the count data outcome*') [12], then model checking will often show that the underlying assumptions of these generalised linear models are not fulfilled (Table 1) [13]. For example, a common problem arising with Poisson regression models is over dispersion, i.e., the count data outcomes exhibit more variation than the variation expected from the underlying model [12].

Generalised linear mixed models (GLMMs), for example, the Poisson generalised linear mixed model, extend the class of generalised linear models and they may be useful in dealing with over dispersion [2,14]. Generalised linear mixed models combine the properties of two statistical frameworks, i.e., linear mixed models (which incorporate random effects) and generalised linear models (which handle non-normal data by using link functions and exponential family (a set of probability distributions having a certain form) (for example, Poisson or negative binomial distributions) [15]. The use of the generalised linear mixed models allow for some degree of non-independence among observations, for example, when analysing longitudinal or clustered data [2,14,15]. The price to pay for these advantages is a low mathematical tractability since the integration over the random effect is not known in an analytic form, and mathematically-advanced algorithms such as the Markov chain Monte Carlo methods have to be used. Moreover, checking the underlying assumptions is often not possible [2,14,16].

Sandwich covariance matrix estimators are a popular tool in applied regression modelling [17]. Employing estimators for the covariance matrix based on this sandwich form can make inference for the parameters more robust against certain model misspecifications. However, sandwich covariance matrix estimators should not be used

Model type	Deviance/d.f.	Pearson Chi-square /d.f.	Standardised deviance residuals			
			Mean difference and p-value for H_0: mean difference=0	Median	p-value for H_0: normal distribution (via Kolmogorow-Smirnow test)	p-value of intervention
Poisson	1.88	1.28	-0.141, p=0.004	0.283	<0.0005	0.92
(link function: log)	-1.47	-1.06	(-0.083; p<0.0005)	(-0.000)	(<0.0005)	-0.56
Negative binomial (link function: log)	0.11	0.045	-0.035; p=0.004	0.068	<0.0005	0.79
	-0.097	-0.047	(-0.093; p<0.0005)	(-0.027)	(<0.0005)	-0.7

Table 1: Data from the Scandinavian starch for severe sepsis/septic shock (6S) trial on the rate of dialysis-free days within 90 days following randomisation.

for every model in every analysis [17]. First, if the model is correctly specified, the use of sandwich estimators leads to loss of power [17]. Second, if the model is not correctly specified, the sandwich estimators are only useful if the estimates of the parameters are still consistent, i.e., if the misspecification does not result in bias [17].

When analysing longitudinal data or data of a cluster randomised trial, another extension to the generalised linear model, namely generalised estimating equations (GEE) might be used [2]. GEE avoids modelling the within-patient (or cluster) covariance structure by treating the covariance structure as a nuisance [2]. Hence, the covariance structure does not need to be specified correctly in order to get valid results [2].

Non-parametric methods

When data are not normally distributed, non-parametric analytic methods ought to be considered. A number of non-parametric methods may be used to analyse count data. The Wilcoxon rank sum test, also known as the Wilcoxon-Mann-Whitney test or the Mann-Whitney U test, is a commonly used non-parametric test to detect differences in the distributions of the response of interests between two treatments [18]. In general, when data are normally distributed, non-parametric tests have lower power than parametric tests, but the power loss with, for example, the Wilcoxon rank sum test compared to that of the t-test is often limited [19]. When normality is violated, the Wilcoxon rank sum test can be three or four times more powerful than the independent samples t-test [19].

It is generally recommended to adjust for stratification variables (for example, for the effect of clinical site in a multi-center trial) in the primary analysis of a randomised clinical trial [8,9,18]. As it is not possible to adjust for stratum effects via the Wilcoxon rank sum test [18,20], we propose to use alternative tests. One option is the van Elteren test, which is a stratified Wilcoxon rank sum test [18,20].

The van Elteren test essentially tests [18]:

$H_0 : \pi = 0.5$

Where the competing probability $\pi = \sum_{k=1}^{d} c_k \, PR(Y_k > X_k)$ and $0 \le c_k \le 1$ ($\sum c_k = 1$) are weights and X_k and Y_k are random variables with distribution F_k and G_k.

Simulations have shown that the van Elteren test is more efficient by having higher power and lower risk of type I error than the Wilcoxon rank sum test when the number of strata is relatively low (below 50) [18]. The Wilcoxon rank sum test seems to be more efficient than the van Elteren test when the stratum effects are small and at the same time a large number of strata are used [18]. However, unless the sample size is very large, the number of strata in a randomised clinical trial should be much less than 50 [10]. The van Elteren test is available in different commercial statistical programs, for example STATA [21] and SAS [22].

A recent simulation study has shown that other non-parametric tests may outperform the van Elteren test in terms of statistical power, especially if the effects across strata differ significantly [23]. Among these tests, the best one in terms of power is the so-called T_{adap2} which uses the aligned rank test with equal stratum weights in a sequential manner via an adaptive multiple testing strategy [23]. However, presently the T_{adap2} test is not included in any standard statistical software, except for a code implemented in the R-software [24], kindly provided by Mehrotra et al. [23]. Further comparative studies examining the strengths and limitations of these different non-parametric tests are needed, but both the van Elteren and the T_{adap2} tests seem equally reliable methods for analysing count data.

Using the van Elteren and the T_{adap2} tests, it is possible to non-parametrically assess the hypothesis that the stratified effects are statistically the same. It may also be relevant and informative to estimate the difference between the medians (and/or the means) of the two interventions. To obtain a non-parametric estimate of the confidence interval of the intervention effect, different techniques may be applied [25,26]. Among others, we propose bootstrapping which estimates the confidence interval by sampling randomly from the observed data [20,25]. A limitation of using bootstrapping to obtain a non-parametric estimate of the confidence interval of the intervention effect is that the uncertainty of the estimated non-parametric confidence interval might be considerable if the sample size is limited or small. Other valid methods which can provide confidence intervals are the Kendall's τa, Somers' D, and the Hodges–Lehmann median difference methods [26].

There are two limitations in using the van Elteren and the T_{adap2} tests. First, it is only possible to adjust the data for one stratification variable, for example clinical site in a multicenter trial. Second, only two intervention groups can be compared using these tests. If more than two intervention groups need to be compared, then the Kruskal-Wallis test, which is the non-parametric version of the one-way analysis of variance test (ANOVA), may be used. A stratified version of the Kruskal-Wallis test is available in R [27].

Discussion

In contrast to most observational studies [28], obtaining a perfect model fit when analysing count data outcomes of randomised clinical trials is often not necessary. A given trial population might not be similar to the population of another trial studying the same disease, intervention, and outcomes. If data driven transformations are used to optimise a model fit in a given trial, the same transformations may not lead to an optimal model fit in future trials assessing the effects of the same intervention on comparable populations. It will, therefore, be difficult to replicate a given trial result if outcomes or covariates are transformed (e.g., square root, square, or inverse) in different statistical ways to optimise a model fit. Furthermore, the primary aim in a randomised clinical trial is to assess whether an intervention works or not, and precise estimations of covariates, coefficients, etc., might not be of primary interest. On the other hand, to ensure the validity of the trial results and under some circumstances to optimise their power, choosing correctly the optimal statistical method is essential. The optimal choice of tests of assumptions and analysis methods should bring about optimal balance between obtaining a model fit and using a methodology which would allow the trial results to be replicated and generalised. We believe that our present review may help trialists to find this balance.

The aim of our review was to provide a simplified, valid, hands-on guide for trialists faced with simple count data. We have deliberately not included in-depth descriptions of theoretical issues and statistical details, and we have not included considerations of longitudinal count data analysis or Bayesian statistical analysis [1,6,7]. Therefore, we have focused on the most common aspects of count data analysis in randomised clinical trials. An overview of the strengths and limitations of using simple count data in analysis methods is presented in Table 2.

As with all statistical analysis, analysis of count data ought to rest on transparently published, detailed statistical plans for the conduct of the analyses [29]. Such detailed analysis plans ought to be formulated before

Analysis method	Strengths	Limitations	Notes
The van Elteren test and the T_{adap2} tests	No underlying assumptions about the distribution of the observed data are necessary. It is possible to adjust for stratum effects.	Only possible to adjust the data for one stratification variable. Only two intervention groups can be compared. The T_{adap2} test is available only in a code written in R.	Non-parametric test. Often the optimal choice of analysis method.
Bootstrapping	No underlying assumptions about the distribution of the observed data are necessary.	The uncertainty of the non-parametric confidence interval provided via bootstrapping could be large if it is based on a limited number of observations.	Non-parametric method to estimate the confidence interval of, e.g., means difference or median difference of two treatments.
The Kruskal–Wallis tests	Multiple intervention groups can be compared. No underlying assumptions about the distribution of the observed data are necessary.	The stratified version of the test is available in R only.	Non-parametric method. A valid method but only if more than two groups have to be compared.
The Wilcoxon rank sum test	No underlying assumptions about the distribution of the observed data are necessar	Not possible to adjust for stratum effects.	Non-parametric test. The van Elteren test is often a better choice.
Generalised linear model	Trial results can be adjusted for multiple stratification variables. Odds ratios and confidence intervals can demonstrate the intervention effects. Multiple intervention groups can be compared.	It can be impossible to deal with some data due to over dispersion [12]. Model checking will often show that the underlying assumptions of the models are not fulfilled.	Parametric method. Often not an optimal choice of analysis method.
Generalised linear mixed model	Trial results can be adjusted for multiple stratification variables. Odds ratios and confidence intervals can demonstrate the intervention effects. Multiple groups can be compared. Can often handle overdispersion.	Complicated analysis. Checking the underlying assumptions is not easy.	Parametric method. Often not an optimal choice of analysis method.

Table 2: An overview of the strengths and limitations when using the different count data analysis methods.

the data are collected or at least before investigators or statisticians get access to the trial data [29]. If such detailed statistical analysis plans demonstrate defects during the conduct of the analysis, then, of course, the plans must be transparently amended and reported [29].

Conclusion

The design of the count data outcome and evaluation should be carefully considered, and a detailed statistical analysis plan ought to be published before the analysis of the trial results begins, preferably before data are collected or at least before access to data are allowed. A detailed selection procedure of the opted model should be described in the analysis plan. Unless there is evidence for using a parametric model, we recommend, as default methods, the use of either the van Elteren or the Tadap2 tests and bootstrapping for estimation of the confidence intervals for the medians or the mean differences. A stratified version of the Kruskal-Wallis one-way analysis of variance test may be used if more than two trial groups have to be compared. The risk of biased results when analysing count data in randomised clinical trials should decrease if our recommendations are followed.

Contributors

JCJ critically reviewed the available literature and wrote the first draft. MT critically reviewed the available literature, contributed with statistical expertise, participated in the design of the study, and contributed to drafting the manuscript. PW, NH, AP critically reviewed the available literature, participated in the design of the study, and contributed to draft the manuscript. JCJ, JW, and CG conceived the study, critically reviewed the available literature, and contributed to drafting the manuscript. All authors have approved the manuscript.

Acknowledgement

We thank Dimitrinka Nikolova for her patient copyediting and linguistic suggestions. JCJ, JW, PW, and CG were partly funded by The Copenhagen Trial Unit. Otherwise, we have received no funding. We thank the Scandinavian Starch for Severe Sepsis/Septic Shock (6S) trial investigators for providing the data.

References

1. Karazsia BT, Dulmen MH (2008) Regression models for count data: illustrations using longitudinal predictors of childhood injury. J Pediatr Psychol 33: 1076-1084.
2. Agresti A (2007) An Introduction to Categorical Data Analysis (2ndedn.) Wiley-Interscience.
3. Dmitrienko A, Ajit C, Tamhane AC, Bretz F (2009) Multiple testing problems in pharmaceutical statistics (Chapman and Hall/CRC Biostatistics Series): Chapman and Hall/CRC.
4. The European Agency for the Evaluation of Medical Products (2002) Points to consider on multiplicity issues in clinical trials.
5. Schulz KF, Altman DG, Moher D (2010) Consort 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Int Med 152: 726-732.
6. Sutradhar BC (2011) Dynamic Mixed Models for Familial Longitudinal Data. Springer Series in Statistics.
7. Winkelmann R (2008) Bayesian Analysis of Count Data. In: Econometric Analysis of Count Data. Springer, Berlin, Heidelberg, pp. 241-250.
8. Kahan BC, Morris TP (2011) Improper analysis of trials randomised using stratified blocks or minimisation. Stat Med 31: 328-340.
9. Kahan BC, Morris TP (2012) Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis. BMJ 345: e5840.
10. Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI (1999) Stratified randomization for clinical trials. J Clin Epidemiol 52: 19-26.
11. Zucchini W (2000) An introduction to model selection. J Math Psychol 44:41-61.
12. Berk R, MacDonald JM (2008) Over dispersion and Poisson regression. J Quant Criminol 24: 269-284.
13. Perner A, Haase N, Guttormsen AB, Tenhunen J, Klemenzson G, et al. (2012) Hydroxyethyl starch 130/0.42 versus Ringer's acetate in severe sepsis. N Eng J Med 367: 124-134.
14. McCulloch CE (2006) Generalized Linear Mixed Models. Encyclopedia of Environmetrics.

15. Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, et al. (2009) Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol 24: 127-135.

16. Gilks WR, Richardson S, Spiegelhalter D (1995) Markov Chain Monte Carlo in Practice (Chapman & Hall/CRC Interdisciplinary Statistics). Chapman and Hall/CRC: 504p.

17. Zeileis A (2006) Object-oriented Computation of Sandwich Estimators. J Stat Software.

18. Qu Y, Zhao YD, Rahardja D (2008) Wilcoxon-Mann-Whitney test: stratify or not? J Biopharm Stat 18: 1103-1111.

19. Sawilowsky SS (2005) Misconceptions leading to choosing the t test over the Wilcoxon Mann-Whitney U test for shift in location parameter. J Mod Appl Stat Methods 4: 598-600.

20. Edmans J, Bradshaw L, Franklin M, Gladman J, Conroy S (2013) Specialist geriatric medical assessment for patients discharged from hospital acute assessment units: randomised controlled trial. BMJ 2013.

21. Statistical Consulting Group UCLA Academic Technology Services (2007) R Relative to Statistical Packages: Comment 1 on Technical Report Number 1 (Version 1.0) strategically using General Purpose Statistics Packages: A Look at Stata, SAS and SPSS. Technical Report Series.

22. Dmitrienko A, Molenberghs G, Offen W, Chuang-Stein C (2005) Analysis of Clinical Trials Using SAS: A Practical Guide: SAS Institute.

23. Mehrotra DV, Lu X, Li X (2010) Rank-Based Analyses of Stratified Experiments: Alternatives to the van Elteren Test. Am Stat 64: 1-27.

24. R Foundation for Statistical Computing (Vienna A: R: a language and environment for statistical computing.

25. Baverel PG, Savic RM, Karlsson MO (2011) Two bootstrapping routines for obtaining imprecision estimates for nonparametric parameter distributions in nonlinear mixed effects models. J Pharmacokinet Pharmacodyn 38: 63-82.

26. Newson R (2002) Parameters behind “nonparametric” statistics: Kendall’s tau, Somers’ D and median differences. Stata J 2: 45-64.

27. Kruskal WH, Wallis WA (1952) Use of ranks in one-criterion variance analysis. J Am Stat Assoc 47: 583-621.

28. Jakobsen JC, Gluud C (2013) The necessity of randomized clinical trials. Br J Med Res 3:1453-1468.

29. The Nordic Trial Alliance Working Group 6 (2015) Report on transperancy and registration in clinical research in the Nordic countries. Project WP 6: Transparency and registration 2015.