Title

Overdispersion models for correlated multinomial data: Applications to blinding assessment

Journal

Statistics in Medicine, 38(25)

ISSN

0277-6715

Authors

Landsman, V
Landsman, D
Li, CS
et al.

Publication Date

2019-11-10

DOI

10.1002/sim.8344

Peer reviewed

HHS Public Access

Author manuscript

Stat Med. Author manuscript; available in PMC 2020 November 10.

Published in final edited form as:

Stat Med. 2019 November 10; 38(25): 4963–4976. doi:10.1002/sim.8344.

Overdispersion models for correlated multinomial data: applications to blinding assessment

V. Landsman[1,2,*], **D. Landsman**[3], **C.S. Li**[4], **H. Bang**[5]

[1]Institute for Work and Health, Toronto, ON, Canada

[2]Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada,

[3]Department of Computer Science, University of Toronto, Toronto, ON, Canada,

[4]School of Nursing, The State University of New York, University at Buffalo, Buffalo, NY, USA,

[5]Division of Biostatistics, Department of Public Health Sciences, University of California, Davis, CA, USA

Abstract

Overdispersion models have been extensively studied for correlated normal and binomial data but much less so for correlated multinomial data. In this work, we describe a multinomial overdispersion model that leads to the specification of the first two moments of the outcome and allows the estimation of the global parameters using generalized estimating equations (GEE). We introduce a Global Blinding Index as a target parameter and illustrate the application of the GEE method to its estimation from 1) a clinical trial with clustering by practitioner and 2) a meta-analysis on psychiatric disorders. We examine the impact of a small number of clusters, high variability in cluster sizes and the magnitude of the intra-class correlation on the performance of the GEE estimators of the Global Blinding Index using the data simulated from different models. We compare these estimators with the inverse-variance weighted estimators and a maximum-likelihood estimator, derived under the Dirichlet-multinomial model. Our results indicate that the performance of the GEE estimators was satisfactory even in situations with a small number of clusters, whereas the inverse-variance weighted estimators performed poorly, especially for larger values of the intra-class correlation coefficient. Our findings and illustrations may be instrumental for practitioners who analyze clustered multinomial data from clinical trials and/or meta-analysis.

Keywords

Blinding index; Dirichlet-multinomial; GEE; Meta-analysis

[*]Correspondence to: V. Landsman, Institute for Work and Health, 481 University Avenue, Suite 800, Toronto, Ontario M5G 2E9, Canada. vlandsman@iwh.on.ca.

1. Introduction

Various types of clustered data are often collected in epidemiological and clinical studies aimed at evaluating the efficacy and safety of certain medical interventions. For example, outcomes of individuals may be clustered within a practitioner, group therapy or center in randomized clinical trials (RCTs). In a meta-analysis, clusters are represented by different published studies that are analyzed jointly. In settings, where the patients are randomly assigned to the treatment arms within clusters, cluster-specific estimators of the effects, such as mean, odds ratio, relative risk, or risk difference, may be combined into one global estimator to improve the efficiency of the estimated effect size and increase the power of statistical tests. Potential correlation between the individual level outcome measures within the same cluster can lead to overdispersion (or extra variation) [1]. Proper accounting for extra variation is important for obtaining correct variances of the estimators of a global parameter and correct inference on test statistics [2]. Multilevel and marginal models are two major modeling approaches that can be used to model overdispersion in correlated data.

In multilevel models, the correlation is introduced by including a cluster-specific random effect in the assumed conditional mean model of an outcome. In the estimation of a global parameter, these models usually require some type of ‘marginalization’ ([3],[4]), which, in turn, often entails the use of computationally intensive algorithms to integrate over the random effects. For certain link functions and mixing distributions of random effects, as is the case with the multinomial overdispersion model (MOM) introduced in Section 2, the unconditional (on random effects) mean and variance of an outcome can be expressed analytically as functions of a global parameter and additional heterogeneity/overdispersion parameter(s). In such cases, the estimation of the unknown parameters of interest can be handled via generalized estimating equations (GEE) ([5], [6]).

In the marginal approach, the (unconditional) expectation of an outcome is modeled directly as a function of global parameters. Under mild regularity conditions and correct specification of the marginal mean model of an outcome, the GEE approach will lead to consistent estimators of a global parameter. The (‘working’) correlation matrix should be prespecified and its choice may affect the efficiency of the resulting estimators, especially if the correlations between the individual outcomes within a cluster are strong. An ‘exchangeable’ correlation structure is frequently used with clustered data [7]. In the absence of reliable information about the correlation structure, the identity matrix is assumed. In this case, the estimation of a global parameter is handled via independence estimating equations, a special case of GEE. Potential correlations among the observations within clusters should be taken into account when estimating the variances of the estimated global effects. The robust (‘sandwich’) variance estimator is used to obtain consistent estimators of the variances of the estimated global parameters. This approach is widely used in survey sampling to estimate standard errors from clustered data [2].

In this paper, we introduce MOM and study the performance of the GEE estimators for estimating global probabilities of outcomes from clustered multinomial data. Our study was motivated by two studies: 1) an RCT with multiple practitioners, and 2) a meta-analysis of RCTs [8], both aiming at evaluating the extent of blinding in the trials ([9], [10]). As will be

described in the following sections, the data collected for blinding assessment can be viewed as a random sample from a multinomial distribution of order two (binomial), three (trinomial) or higher, and the target parameter can be presented as a linear combination of the global probabilities.

The GEE method has been applied to estimating the regression coefficients from longitudinal data with nominal and ordinal outcomes (e.g., [7], [11]). Krewski and Zhu [12] and Zhu *et al.* [13] estimated the regression parameters in dose-response models for clustered trinomial data arising from developmental toxicity studies. Robust variance estimators of the GEE estimators for global parameters have been shown to perform well in settings with a large number of clusters and a relatively small and equal number of observations within each cluster, as is typical in longitudinal studies. However, they tend to underestimate the true variance of the estimators with a small number of clusters and highly variable cluster size, as is common in meta-analyses and multi-center RCTs. Various corrections aimed at improving small-sample properties of the robust variance estimators in different settings have been proposed ([14], [15]). The performance of corrected variance estimators has been studied extensively for continuous and binary outcomes [16], but not for a multinomial outcome. Finally, the existing software to analyze clustered multinomial data using the GEE method are not very general or flexible.

This paper is organized as follows. In Section 2 we introduce MOM and show its connection with the Dirichlet-multinomial (DM) distribution and the exchangeable correlation model. We provide a description of the GEE method to estimate the global probabilities in MOM in Section 3. In Section 4 we introduce a Global Blinding Index (GBI) as a linear combination of the multinomial probabilities and propose five possible estimators of the GBI: two GEE estimators, two inverse-variance weighted (IVW) estimators, widely used in meta-analysis, and a maximum-likelihood estimator (MLE) derived under the DM model. In Section 5, we use simulations to compare the performance of the proposed estimators under different models and with special attention to small-sample properties of the robust variance estimators. The two real studies that motivated this paper are described and analyzed in Section 6. We conclude our paper with a discussion.

2. Multinomial Overdispersion Model

In this section we describe the overdispersion model that can accommodate the extra variation in clustered multinomial data and discuss the connection of this model with the exchangeable correlation model and the Dirichlet-multinomial (DM) model.

In what follows, we assume that the data consist of K clusters, with n_i participants' responses in the i-th cluster ($i = 1, 2,\ldots, K$) being classified into $m + 1$ mutually exclusive categories. Let $\mathbf{Z}_i = (Z_{i1}, Z_{i2},\ldots, Z_{im})'$ denote an m-dimensional vector of counts that corresponds to the first m categories such that each entry Z_{il} represents the number of units (out of n_i) that were classified into the l-th category, $l = 1, 2,\ldots,m$, and $Z_{i,m+1} = n_i - \sum_{l=1}^{m} Z_{il}$. Next, we define a vector of cluster-specific probabilities $\mathbf{P}_i = (P_{i1}, P_{i2}, \ldots, P_{im})'$ such that $0 < P_{il} < 1$; $\sum_{l=1}^{m} P_{il} < 1$ and $P_{i,m+1} = 1 - \sum_{l=1}^{m} P_{il}$. To introduce

the overdispersion, it can be assumed that $\boldsymbol{Z}_i|\boldsymbol{P}_i$ has a multinomial distribution of size n_i and parameters $\boldsymbol{P}_i$ such that

$$E(\boldsymbol{P}_i) = \boldsymbol{\pi};\ Var(\boldsymbol{P}_i) = \rho^2[\mathrm{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}'], \tag{1}$$

where $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_m)'$ such that $0 < \pi_l < 1$; $\sum_{l=1}^{m} \pi_l < 1$ and $\pi_{m+1} = 1 - \sum_{l=1}^{m} \pi_l$. Here $0 \le \rho^2$ is an overdispersion parameter and it is usually treated as a nuisance parameter; $\rho^2 = 0$ corresponds to the case of no overdispersion. The global probabilities π_l's, or any continuously differentiable function of them, can be defined as a target parameter.

Using the laws of conditional expectation and variance and the assumptions above, it can be shown that the first two moments (unconditional) of $\boldsymbol{Z}_i$ are:

$$E(\boldsymbol{Z}_i) = n_i\boldsymbol{\pi}, \tag{2}$$

and

$$Var(\boldsymbol{Z}_i) = n_i\left[1 + (n_i - 1)\rho^2\right][\mathrm{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}']. \tag{3}$$

This model is referred to as a multinomial overdispersion model (MOM) throughout this paper.

The full marginal (unconditional) distribution of $\boldsymbol{Z}_i$ can be derived under additional assumptions. Neerchal and Morel [17] discussed two overdispersed multinomial distributions of $\boldsymbol{Z}_i$ that have the first two moments as defined by (2) and (3): a DM distribution and a random-clumped multinomial (RCM) distribution. In this work, we restrict our attention to the DM distribution. If the random variables $\boldsymbol{P}_i$ are assumed to be distributed by a Dirichlet distribution with parameters $\boldsymbol{\pi}$ and ρ^2, then their probability density function is expressed as

$$f(\boldsymbol{p}_i;\boldsymbol{\pi},\rho)=\frac{\Gamma(c)}{\prod_{l=1}^{m+1}\Gamma(c\pi_l)}\prod_{l=1}^{m+1}p_{il}^{c\pi_l-1}, \quad (4)$$

where $\boldsymbol{p}_i$ is a realization of $\boldsymbol{P}_i$; $c=\rho^{-2}(1-\rho^2)$; $0<\rho<1$; and $\Gamma(\cdot)$ is the gamma function. In this case, the marginal distribution of $\boldsymbol{Z}_i$ is DM [18]. The DM distribution is a multivariate extension of the beta-binomial distribution, which has been used in the literature for modeling the overdispersed binomial data ([19], [20]).

It is worth noting the connection between MOM and the exchangeable correlation model [7]. Toward this end, for each participant j in cluster i ($j=1,2,\ldots,n_i$), let us define the vector $\boldsymbol{Y}_{ij}=(Y_{ij1},Y_{ij2},\ldots,Y_{ijm})'$ of binary random variables, such that Y_{ijl} takes value 1 with probability π_l, if the category l was selected among the first m mutually exclusive categories, and value 0 otherwise; and $Y_{ij,m+1}=1-\sum_{l=1}^{m}Y_{ijl}$ takes value 1 with probability π_{m+1} and 0 otherwise. Since $\boldsymbol{Y}_{ij}$ is a multinomial random variable of size 1, the within-unit correlation is given by,

$$\gamma_{ll'}:=Corr\left(Y_{ijl},Y_{ijl'}\right)=-\frac{\pi_l\pi_{l'}}{\sqrt{\pi_l(1-\pi_l)\pi_{l'}(1-\pi_{l'})}},\quad \text{whenever } l\neq l',$$

and $\gamma_{ll}=1$. In addition, denote $\rho_{jj'}:=Corr(Y_{ijl},Y_{ij'l})$ for any two individuals $j\neq j'$ and assume that $Corr(Y_{ijl},Y_{ij'l'})=\rho_{jj'}\cdot\gamma_{ll'}$ for any two categories $l\neq l'$. Using the above notation, $Z_{il}=\sum_{j=1}^{n_i}Y_{ijl}$, $j=1,2,\ldots,n_i$ for a given i, counts the number of times the category l was selected by individuals in cluster i. Under exchangeability assumption, $\rho_{jj'}=\rho^*$ for $j\neq j'$, implying $Var(\boldsymbol{Z}_i)=n_i[1+(n_i-1)\rho^*][\mathrm{diag}(\boldsymbol{\pi})-\boldsymbol{\pi}\boldsymbol{\pi}']$ [21]. The latter expression is equal to (3) with ρ^2 replaced by ρ^*. Notice that ρ^* can be negative as long as $\rho^*>-1/[\max\{n_i\}-1]$. The factor $\phi_i:=1+(n_i-1)\rho^*$ is often referred to as the design effect due to clustering [22] and the parameter ρ^* is an intra-class correlation coefficient (ICC).

3. Estimation of MOM parameters and their variance

3.1. GEE

Let $\boldsymbol{\theta}=(\boldsymbol{\pi},\rho^2)$ define the vector of unknown parameters of interest. Then, the GEE method ([5], [6]) can be used to estimate $\boldsymbol{\theta}$. For fixed ρ^2, the estimating equations for $\boldsymbol{\pi}$ have the form

$$\sum_{i=1}^{K} \boldsymbol{D}_i' \boldsymbol{V}_i^{-1} \boldsymbol{R}_i = \boldsymbol{0},$$

(5)

where $\boldsymbol{D}_i$ is a matrix of the first partial derivatives of $E(\boldsymbol{Z}_i)$ with respect to $\boldsymbol{\pi}$; $\boldsymbol{R}_i = \boldsymbol{Z}_i - n_i\boldsymbol{\pi}$ are the residuals; and $\boldsymbol{V}_i = n_i[1 + (n_i - 1)\rho^2][\mathrm{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}']$. Also, under (2), $\boldsymbol{D}_i = n_i\boldsymbol{I}_{m\times m}$, where $\boldsymbol{I}_{m\times m}$ is the identity matrix.

To estimate ρ^2, an additional estimating equation should be added. One possible option is by equating the Pearson chi-square statistic $\sum_{i=1}^{K} \boldsymbol{R}_i \boldsymbol{V}_i^{-1} \boldsymbol{R}_i$ to its expected value [23]. Based on the fact that $E\left(\boldsymbol{R}_i \boldsymbol{V}_i^{-1} \boldsymbol{R}_i\right) = m$, the equation has the form,

$$\sum_{i=1}^{K} \boldsymbol{R}_i \boldsymbol{V}_i^{-1} \boldsymbol{R}_i = mK,$$

(6)

The desired estimate $\hat{\theta} = \left(\hat{\boldsymbol{\pi}}, \hat{\rho}^2\right)$ is obtained by alternating between the equations (5) and (6) until convergence is achieved.

The same estimation approach can be used to estimate $\boldsymbol{\pi}$ from a marginal model under the assumption (2). This model also requires an assumption about the correlation structure of $\boldsymbol{Z}_i$, which may not be known in many practical applications. In these cases, $\boldsymbol{V}_i$ in (5) serves as a 'working' covariance matrix, such that $\boldsymbol{V}_i \approx Var(\boldsymbol{Z}_i)$. The closer the approximation between the two matrices is, the more efficient $\hat{\boldsymbol{\pi}}$ is likely to be [6]. In the absence of a sensible approximation, the independence working model, namely $\boldsymbol{V}_i := n_i[\mathrm{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}']$, is assumed for a multinomial outcome, leading to the 'independence estimating equations' in (5), a special case of GEE [5].

3.2. Robust variance estimator

Under mild regularity conditions, $\sqrt{K}$-consistency of $\hat{\rho}^2$ given $\boldsymbol{\pi}$, and correct specification of the marginal mean model (2), the estimator $\hat{\boldsymbol{\pi}}$ is consistent and $\sqrt{K}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi})$ converges in distribution to a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $Var(\hat{\boldsymbol{\pi}})$, which can be consistently estimated using a robust variance estimator given by

$$\widehat{Var}(\hat{\boldsymbol{\pi}}) = \hat{\boldsymbol{H}}^{-1}\hat{\boldsymbol{G}}\hat{\boldsymbol{H}}^{-1},$$

(7)

$$\hat{\boldsymbol{H}} = \sum_{i=1}^{K} \boldsymbol{D}_i(\hat{\boldsymbol{\pi}})\boldsymbol{V}_i^{-1}\left(\hat{\boldsymbol{\pi}}, \hat{\rho}^2\right)\boldsymbol{D}_i(\hat{\boldsymbol{\pi}}),$$

$$\hat{\boldsymbol{G}} = \frac{N-1}{N-m}\frac{K}{K-1}\sum_{i=1}^{K}\left(\boldsymbol{d}_i - \bar{\boldsymbol{d}}\right)\left(\boldsymbol{d}_i - \bar{\boldsymbol{d}}\right)',$$

$$\boldsymbol{d}_i = \boldsymbol{D}_i(\hat{\boldsymbol{\pi}})\boldsymbol{V}_i^{-1}\left(\hat{\boldsymbol{\pi}}, \hat{\rho}^2\right)\boldsymbol{R}_i(\hat{\boldsymbol{\pi}}) \text{ and } \bar{\boldsymbol{d}} = \frac{\sum_{i=1}^{K}\boldsymbol{d}_i}{K}.$$

The variance estimator (7) originates from expanding the left-hand side of the GEEs (5) using a Taylor series. In the usual expansion, $\bar{\boldsymbol{d}} = \boldsymbol{0}$, which holds if $\hat{\boldsymbol{\pi}}$ is an exact solution of (5), and there is no factor $\frac{N-1}{N-m}$. Both adjustments have been proposed to reduce small sample bias associated with the estimated deviations in $\hat{\boldsymbol{G}}$ [15].

Once $\hat{\boldsymbol{\pi}}$ and $\widehat{Var}(\hat{\boldsymbol{\pi}})$ have been obtained, further inference about various parameters of interest defined as continuously differentiable functions of $\boldsymbol{\pi}$ is straightforward. One example, in which the target parameter is a linear combination of the entries of the vector $\boldsymbol{\pi}$ is described in the next section.

The GEE approach described above can be generalized for non-identity link functions which might be of interest in other applications; see [7] and [11] for examples in longitudinal context.

4. Application of MOM to blinding assessment in clinical trials

4.1. Global Blinding Index

Blinding is a crucial component in RCTs and its assessment is increasingly conducted by means of collecting empirical data and performing appropriate statistical analysis. Quantifying the amount of potential unblinding in a trial can be important from a quality control perspective (for example, in a pilot study) and the interpretation of study findings (e.g., effectiveness of treatment). For example, Wood *et al.* [24] have shown smaller effects in blinded than in unblinded trials. Blinding Indexes (Bis) have been proposed to quantify the degree of potential unblinding objectively and systematically ([25], [9]). The BI by Bang

et al. [9], used in this study, attempts to estimate the percentage of excess (beyond chance) of correct guesses for a given arm (T or C) of the trial.

Typical blinding data are collected, usually at the exit survey and separately for each arm, by asking a participant to provide the best guess regarding their treatment allocation among three possible choices: 1 (I was in T arm; T), 2 (I was in C arm; C), 3 (I don't know; DKN). These data can be summarized in a 2 × 3 table (see Table 1) with two allocation arms and three choices: $n_{l|a}$ are the observed counts in cell (a, l) and n_a are the row totals, where a =T or C is the allocation arm and l = 1, 2, 3 is the guess. The two other popular formats are 2 × 2 and 2 × 5. In the former, an I don't know option is not allowed, in which case the data can be viewed as a sample from a binomial distribution for each arm. In the latter, the degree of belief is assessed in a multinomial format: 1 (Strongly believe I was in T arm), 2 (Somewhat believe I was in T arm), 3 (Somewhat believe I was in C arm), 4 (Strongly believe I was in C arm), 5 (I don't know).

In the general case, the observed data in arm a can be viewed as a realization of a multinomial random variable with parameters n_a and $\boldsymbol{P}_a = (P_{1|a}, P_{2|a}, \ldots, P_{m|a})$, where $P_{l|a} := Pr(\text{guessed } l/(\text{assigned } a)$ for $l = 1,\ldots,m$, and m is the number of all possible guesses minus one. Under this model, BI is defined as $BI_a = \boldsymbol{\nu}_a' \boldsymbol{P}_a$, where $\boldsymbol{\nu}_a$ is a vector of the corresponding constants/weights $\nu_{l|a}$ used to define a contrast in arm a. For example, for data collected in a 2 × 3 format (i.e., a trinomial case): $m = 2$, $BI_T = (1, -1)(P_{1|T}, P_{2|T})' = P_{1|T} - P_{2|T}$ and $BI_C = (-1, 1)(P_{1|C}, P_{2|C})' = P_{2|C} - P_{1|C}$. For 2×5 data: $m = 4$, $BI_T = (1, \nu_{2|T}, -\nu_{3|T}, -1)$ $(P_{1|T}, P_{2|T}, P_{3|T}, P_{4|T})'$ and $BI_C = (-1, -\nu_{2|C}, \nu_{3|C}, 1)$ $(P_{1|C}, P_{2|C}, P_{3|C}, P_{4|C})'$. It has been suggested to use $\nu_{2|T} = \nu_{3|T} = \nu_{2|C} = \nu_{3|C} = 0.5$ but other choices along with ancillary/validation data can be explored for sensitivity analysis ([9],[26]).

A Global Blinding Index (GBI) may be of interest when blinding assessment data have been collected from multiple centers, practitioners or studies as in meta-analysis. To address overdispersion due to clustering, assume that cluster-specific conditional probabilities $\boldsymbol{P}_{i,a}$ $(i = 1,\ldots, K)$ are a vector of random variables with the first two moments as defined by (1). Then, the GBI for arm a is defined by

$$GBI_a = \boldsymbol{\nu}_a' \boldsymbol{\pi}_a.$$

4.2. Five estimators for GBI

Estimation of BI_a and its variance is discussed elsewhere ([9], [26]). In this section we focus on estimation of the GBI and propose five estimators: three estimators that utilize the estimated global probabilities $\hat{\boldsymbol{\pi}}_a$, and two estimators that utilize the estimated cluster-specific indexes $\widehat{BI}_{i,a}$.

The GBI can be estimated directly from the definition as $\widehat{GBI}_a = \boldsymbol{\nu}_a' \hat{\boldsymbol{\pi}}_a$, which, in turn, requires to estimate the global probabilities $\hat{\boldsymbol{\pi}}_a$. The global probabilities can be estimated in three different ways using the ideas outlined in the earlier sections. The estimated global

probabilities under MOM are denoted as $\hat{\boldsymbol{\pi}}_a^{\text{exch}}$ and the corresponding GBI estimator is denoted as $\widehat{GBI}_a^{\text{exch}}$. Assuming $\boldsymbol{V}_i := n_i[\text{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}']$ in (5) leads to the independence estimating equations. The resulting estimated probabilities are denoted as $\hat{\boldsymbol{\pi}}_a^{\text{indep}}$ and the corresponding GBI estimator is denoted as $\widehat{GBI}_a^{\text{indep}}$. Finally, the global probabilities can also be estimated using a maximum-likelihood method under the assumption that the outcomes follow the DM distribution, i.e., $\boldsymbol{Z}_{i,a} \sim \text{DM}(n_{i,a}; \boldsymbol{\pi}_a; \rho^2)$. In this case, we denote the estimated probabilities as $\hat{\boldsymbol{\pi}}_a^{\text{DM}}$ and the corresponding GBI estimator as $\widehat{GBI}_a^{\text{DM}}$. In all three cases, the variance of $\widehat{GBI}_a$ is estimated by $Var\left(\widehat{GBI}_a\right) = \boldsymbol{\nu}_a' Var\left(\hat{\boldsymbol{\pi}}_a\right)\boldsymbol{\nu}_a$. The robust variance estimator (7) is used to estimate the covariance matrix of the estimated probabilities $\hat{\boldsymbol{\pi}}_a^{\text{exch}}$ and $\hat{\boldsymbol{\pi}}_a^{\text{indep}}$. The asymptotic covariance matrix of $\hat{\boldsymbol{\pi}}_a^{\text{DM}}$ is obtained from the inverse of the Fisher information matrix.

Another approach to estimating the GBI would be to utilize the cluster-specific estimates of BI, i.e., $\widehat{BI}_{i,a}$ for $i = 1, \ldots, K$, as is commonly done in meta-analysis. To do this, a multilevel model underlying the distribution of $\widehat{BI}_{i,a}$ can be formulated with $BI_{i,a}$ defined as random effects with expected value GBI_a. In this case, GBI_a is frequently estimated as

$$\widehat{GBI}_a^{\text{IVW}} = \frac{\sum_{i=1}^{K} w_{i,a}\widehat{BI}_{i,a}}{\sum_{i=1}^{K} w_{i,a}}, \tag{8}$$

where $w_{i,a} = 1/Var\left(\widehat{BI}_{i,a}\right)$ are the inverse-variance weights. The estimator (8) is called the inverse-variance weighted (IVW) estimator. Under assumptions of independence between the model errors and the random effects, and normality of the random effects, the IVW estimator can be derived as an MLE [27]. The variances of the IVW estimators are computed under the assumption that the weights $w_{i,a}$ are fixed quantities ([28], [20]). Under this assumption, $Var\left(\widehat{GBI}_a^{\text{IVW}}\right) = 1/\sum_{i=1}^{K} w_{i,a}$ and the IVW estimator is the estimator with minimal variance among all weighted estimators [29]. The IVW estimators deserve our attention as they are intuitive, straightforward in their calculation and have been widely used as a generic method in meta-analysis [30].

Assumptions about the distribution/moments of $\widehat{BI}_{i,a}$ will determine the shape of the weights $w_{i,a}$. In this work, we consider two IVW estimators for illustration and comparison purposes: (1) $\widehat{GBI}_a^{\text{IVW},\rho^*}$ is the IVW estimator (8) with weights $w_{i,a} = \left\{\left[1 + (n_i - 1)\rho^*\right]Var\left(\widehat{BI}_{i,a}\right)\right\}^{-1}$, where ρ^* is the ICC [31]; and (2) $\widehat{GBI}_a^{\text{IVW},0}$ is a special case of, $\widehat{GBI}_a^{\text{IVW},\rho^*}$ when $\rho^* = 0$ (no overdispersion). The latter estimator is also referred to in meta-analysis as a ‘naive’

estimator. The former estimator, $\widehat{GBI}_a^{\mathrm{IVW},\rho^*}$, can be referred to as an 'upgraded' IVW estimator and its weights can be justified by assuming that observed data in arm a in study i is a realization of the DM distribution, i.e., $\boldsymbol{Z}_{i,a} \sim \mathrm{DM}(n_{i,a}; \boldsymbol{\pi}_a; \rho^*)$ [31].

The ICC parameter ρ^* required for $\widehat{GBI}_a^{\mathrm{IVW},\rho^*}$ can be estimated using a variety of methods. In this study, we used a moment estimator based on Cochran's Q statistic derived in [20] using the same ideas as were used by DerSimonian and Laird [32] to derive a moment estimator for τ^2, a heterogeneity parameter, closely related to ρ^*.

4.3. Implementation

All five estimators have been implemented in the simulations (Section 5) and in the applications to real studies (Section 6). The GEE estimators were obtained using an R package *OMGEE* [33]. The MLE estimators were obtained using an R package *OverdispersionModelsinR* [34]. The CIs of the GEE estimators were computed using a t-distribution with degrees of freedom equal to the number of independent clusters minus one since the GBI is estimated for each arm independently ([35], [36]). Satterthwaite-type degrees of freedom approximations have also been proposed in various settings ([14], [37], [38]) but were not explored in this paper. For other estimators the normal approximation was used.

5. Simulation study

In this section we evaluated the performance of the five estimators for GBI, described in Section 4.2, in a simulation study. In particular, we were interested in exploring the relative bias (RB) of the different estimators as well as the impact of the number of clusters, the coefficient of variation (CV) of the cluster sizes and the magnitude of the overdispersion parameter ρ^2 on the performance of the robust variance estimators for the GBIs estimated using the GEE method. We also studied the robustness of the five estimators under different models used to generate clustered multinomial data.

5.1. Data generation

Clustered trinomial data for a given arm was generated from the following four models: (1) DM model with constant ρ^2; (2) DM model with cluster-specific $\rho^2 : \rho_i^2 \sim U(a, b)$ for $0 < a < b < 1$; (3) RCM model with common ρ^2; (4) RCM model with cluster-specific $\rho^2 : \rho_i^2 \sim U(a, b)$ for a and b as above; $i = 1, \ldots, K$. To generate the data from the DM distribution with constant ρ^2, a vector of cluster-specific probabilities $\boldsymbol{P}_i = (P_{i1}, P_{i2}, P_{i3})'$ was generated from a Dirichlet distribution (4) with parameters $\boldsymbol{\pi}$ and ρ^2 at the first step. At the second step, the cluster-level outcome variable $\boldsymbol{Z}_i$ was generated from the trinomial distribution of size n_i and vector of parameters $\boldsymbol{P}_i$. In all models we set $\boldsymbol{\pi} = (0.5, 0.4, 0.1)'$, which implies $GBI = 0.1$ in arm T. The data generation process for the RCM distribution follows from the model definition as described in [17].

We generated data for all possible combinations of the four different models, three different values of the overdispersion parameter ρ^2 ($\rho^2 = 0.1, 0.3, 0.5$), three different values of K (K

= 8, 15, 30), and two different values of the CV of the cluster sizes ($CV \approx 90\%, \approx 40\%$), for a total of 72 scenarios. The cluster sizes n_i were generated from a negative binomial distribution with probability of success p and size parameter r. In this case, the average cluster size and the CV of the cluster sizes are equal to $\frac{r(1-p)}{p}$ and $\frac{1}{\sqrt{r(1-p)}}$ respectively. The average cluster size ($\bar{n}$) was set to 50 and the clusters have been constrained to contain at least 5 participants.

For each scenario, we generated 5,000 datasets and calculated the five GBI estimates, defined in Section 4.2, for each dataset. We obtained the RB, the sampling standard error of the estimator (SSE = standard deviation of the 5,000 estimated GBIs), the estimated standard error of the estimator (ESE = mean of the 5,000 estimated standard errors of the estimated GBIs) and the 95% coverage probability (CP) for the true GBI parameter. The SSE/ESE ratios were used to assess the performance of the robust variance estimators of the GEE estimators, under different scenarios, paying special attention to cases with a small number of clusters and a high CV of the cluster sizes. Ratios close to 1 indicate good performance, whereas ratios greater than 1 indicate underestimated variances.

5.2. Results

The results from the three scenarios that correspond to $\rho^2 = 0.1; 0.3; 0.5$ for the data generated from the DM model with constant ρ^2 are presented in Tables 2a, 2b and 2c, respectively. To save space, the results for the data generated from all four models are presented only for the most challenging scenario with $\rho^2 = 0.5$, $K = 8$ and $CV \approx 90\%$; see Table 2d.

The results in Tables 2a–2c showed good performance of $\widehat{GBI}^{\text{exch}}$: low RBs, SSE/ESE ratios close to 1 and good coverage probabilities were observed for all values of ρ^2, K and CV that we studied. The performance of $\widehat{GBI}^{\text{exch}}$ was comparable with the performance of the estimator $\widehat{GBI}^{\text{DM}}$, an MLE estimator under the DM model with constant ρ^2, which was the model used to generate the data for Tables 2a–2c. With respect to $\widehat{GBI}^{\text{indep}}$, it was nearly unbiased in all the scenarios and had SSE/ESE ratios slightly higher than 1 for scenarios with $K = 8$, $CV \approx 90\%$ and higher values of ρ^2. However, the SSE/ESE ratios for $\widehat{GBI}^{\text{indep}}$ improved with an increase in the number of clusters and/or a decrease in the CV of the cluster sizes. The CPs for $\widehat{GBI}^{\text{indep}}$ were slightly lower than the nominal value for scenarios with a small number of clusters and a high CV of the cluster sizes. At the same time, the CP values around 90%, observed for this estimator in scenarios with only 8 clusters and $CV \approx 90\%$, are still reasonably good for practical needs.

It is worth mentioning that the estimating equation (6) may not have a solution when ρ^2 is very close to 0 and the number of clusters is very small, thus causing numerical errors for the GEE estimator under MOM. Rarely, numerical problems have also occurred with the MLE, again, for the cases where ρ^2 was very close to 0 and the number of clusters was very small. In these situations, switching to the GEE estimator under the independence working assumption, which tended to be very stable numerically and also performed very well in scenarios with small ρ^2 (Table 2a), should be considered. In addition to being numerically

stable, $\widehat{GBI}^{\text{indep}}$ was nearly as efficient as $\widehat{GBI}^{\text{exch}}$, with the exception of some scenarios with a high CV of the cluster sizes.

With regards to the IVW estimators, it should be noted that the 'naive' weighted estimator did not perform well even for very small values of the ICC, and its performance got markedly worse with an increase in ρ^2, regardless of the model used to generate the data. The performance of the 'upgraded' estimator, $\widehat{GBI}^{\text{IVW},\rho^*}$, was comparable with the performance of $\widehat{GBI}^{\text{indep}}$ for $\rho^2 = 0.1$. However, the performance of $\widehat{GBI}^{\text{IVW},\rho^*}$ deteriorated for larger values of ρ^2: RB values up to 60% and SSE/ESE ratios greater than 1.5 were observed in scenarios with $\rho^2 = 0.5$. Overall, we can say that for higher values of ρ^2, the 'upgraded' IVW estimator was primarily biased, whereas the 'naive' IVW estimator was biased and also had underestimated variances. As a result, the CPs for the 'naive' IVW estimator were very low for large values of ρ^2. Given the popularity of the IVW estimators in meta-analysis, these findings should serve as an important lesson for practical users. Based on our findings, we recommend using these estimators with great caution.

Furthermore, we evaluated the robustness of the five estimators under misspecified models. Keeping in mind that the DM and the RCM distributions have the same two first moments, we expected to see almost identical performance of the estimator $\widehat{GBI}^{\text{exch}}$ for the data generated from these distributions with constant ρ^2. This was confirmed by the results in Table 2d. A slight decrease in the performance measures was observed for the estimators $\widehat{GBI}^{\text{indep}}$ and $\widehat{GBI}^{\text{DM}}$ when applied to the data generated from the RCM distribution. Models with varying versus constant ρ^2 seemed to have little, if any, impact on the performance of the three estimators $\widehat{GBI}^{\text{indep}}$, $\widehat{GBI}^{\text{exch}}$, $\widehat{GBI}^{\text{DM}}$. Interestingly, the 'upgraded' IVW estimator had an improved performance under the RCM models, with both constant and varying ρ^2. The performance of the 'naive' IVW estimator was suboptimal across all four models.

6. Applications to real studies

6.1. Example 1: RCT with clustering by practitioner

Vernon and his colleagues [39] used the BI as a measure of blinding assessment of a sham manipulative procedure for the cervical spine in a pilot clinical trial. Recently, they designed a new RCT with multiple practitioners and defined the GBI as their 'primary' outcome measure (https://clinicaltrials.gov/ct2/show/NCT01772966).

In the follow-up study, the eligible participants were assigned randomly to one of two groups, Typical (T) or Alternative (C) intervention for mechanical neck pain. In this trial only the clinician providing the treatment and the data manager were aware of the group assignment. Neither the independent assessors performing baseline and outcome measurements nor the statistical analyst were aware of the group assignment. The treatment protocol included three identical treatments over a period of up to two weeks delivered by the same clinician for a given participant. The number of patients treated by different clinicians is highly variable, ranging from a few patients to 100 patients per clinician. At the end of the study, the patients were asked to provide the best guess about their treatment

allocation. The DKN option was not offered as a possible answer and, therefore, the blinding data were collected in a 2 × 2 format.

The real data from this study are not approved for public use at this time. Therefore, artificial data was generated for the purpose of illustration. We generated two datasets. First, we generated the data of the hypothetical responses in a 2 × 2 format collected from 411 patients treated by 10 clinicians: 206 patients from Alternative and 205 from Typical intervention arms. These data was generated using the DM model with constant ρ^2 as described in Section 5.1. We set the CV of the cluster sizes to 75%, the ICCs to 0.1 in arm T and 0.2 in arm C, and the GBIs to 0.1 in arm T and 0.2 in arm C. These values of the parameters were selected to closely match the corresponding parameters that were estimated from the real data. Next, we created an additional dataset of the hypothetical responses of the same participants that would have been obtained if they were asked to provide the best guess about their treatment allocation among three possible choices (i.e., data in a 2 × 3 format). To do this, for each clinician, we randomly selected approximately 20% to 25% of the T and C responses proportionally to the size of each group in the generated 2 × 2 data and changed them to the DKN option. This proportion of the DKN guesses was obtained from an earlier, single clinician study in a 2 × 3 format [40]. The distribution of patients' choices for each clinician along with clinician-level estimates for Bis for both settings are presented in Table 3a.

The five GBI estimates defined in Section 4.2 and their corresponding 95% CIs are summarized in Table 3b. Overall, we observed that the CIs of all the estimates, with the exception of $\widehat{GBI}_{\mathrm{C}}^{\mathrm{IVW},0}$, contain 0. Also, the GBI estimates from all the methods were in the range (−0.2; 0.2), which suggests that blinding was satisfactory in both arms [41], [26]. The CIs of the estimates did contain values larger than 0.2 for many cases, which can indicate that blinding may not be ideal (i.e., random guessing). At the same time, it is important to keep in mind that the proposed threshold values of ±0.2 are only ad-hoc reference points and should not be used as a strict rule. In general, the focus of blinding assessment is on estimation rather than hypothesis testing, partly because blinding is a tool, not a goal. In addition, the values of $\widehat{GBI}^{\mathrm{exch}}$ and $\widehat{GBI}^{\mathrm{DM}}$ were fairly close as it would be expected, based on our findings from the simulation studies. Finally, it is interesting to notice that the GBI estimates in the T arm were closer to zero than the GBI estimates in the C arm. In the case of real data with a similar pattern, this finding may reflect the fact that patients were able to guess the sham procedure more easily, possibly because of lower treatment effect. Also, the GBIs estimated from a hypothetical 2 × 3 data were closer to zero in both arms and for all the methods, which indicates a tendency of attenuation of the GBI estimates in a 2 × 3 format.

6.2. Example 2: Meta-analysis

In this section, we illustrate the estimation of the GBI in a meta-analysis setting. The systematic review of blinding assessment in RCTs was performed in Baethge *et al.* [42]. The review included 61 RCTs of affective disorders and schizophrenia that were published in either English or German during 2000 – 2010. Out of these 61 studies, 40 studies have provided sufficient data to assess blinding quantitatively: 10 studies provided the blinding

data in a 2 × 3 format, whereas 30 studies provided data in a 2 × 2 format. A meta-analysis of the BIs was performed by Freed *et al.* [8], where the authors used a 'naive' IVW estimator to estimate the GBI pooled over all 40 studies. The data are available upon request from the authors.

We reanalyzed the data from [8] and obtained the GBI estimates in two ways: 1) by analyzing all 40 studies together ('joint' analysis) and 2) by analyzing 10 studies in a 2 × 3 format and 30 studies in a 2 × 2 format separately ('separate' analysis). For a joint analysis of all 40 studies, the studies in a 2 × 2 format were converted to a 2 × 3 format by adding a column of zeros for the DKN choice. Therefore, the joint approach relies on the assumption that the participants of studies in a 2 × 2 format would never select the DKN option if offered. In our opinion, this is a strong and, perhaps, unrealistic assumption and, therefore, we believe that the separate analysis is more sound. A downside of the separate analysis is that it will result in two GBI estimates for each arm, one for each format. We present the results from both approaches for illustration and comparison; see Tables 4a and 4b, respectively.

As follows from the results in the tables, the GBI estimates from the joint analysis were roughly equal to the sum of the corresponding estimates from the 2 × 2 and 2 × 3 formats. However, the two separate analyses yielded quite different estimates for two different formats. The GBI estimates obtained from a separate analysis of the studies in the 2 × 2 format were positive for T arm and negative for C arm. Also, the magnitude of the GBI estimates for T arm was approximately two times higher than those for C arm. These results might reflect some degree of a psychological scenario of "wishful thinking" [26] and/or a placebo effect. The GBI estimates obtained from a separate analysis of the studies in a 2 × 3 format were close to zero in T arm and to 0.1 in C arm. Similarly to the results obtained from the RCT with multiple practitioners, this finding may indicate a low treatment effect.

Overall, most estimates were within the range (−0.2; 0.2), so we might conclude that the studies 'passed' the blinding test. The observed difference in the results from studies that used a 2 × 2 data format vs. studies that used a 2 × 3 data format might imply systematic difference or that the DKN guess option does not reflect the real 'I don't know' state. This may support the importance of collecting ancillary validation data, e.g., by asking those who originally chose the option DKN to choose the treatment assignment (T or C) a second time (see [9] for more details).

7. Discussion

In this paper, we defined the multinomial overdispersion model and studied the performance of the GEE estimators for estimating global probabilities included in the specification of this model. The GEE estimation approach is very attractive because it requires specification of only the first two moments rather than the full parametric distribution of an underlying outcome, and it leads to consistent estimators under a correct marginal mean model of an outcome of interest. The robust variance estimator should be used for variance estimation. We illustrated the application of the methods to the estimation of the Global Blinding Index

(defined as a linear combination of the global probabilities) in simulations and two real studies that involve potential clustering.

We compared the performance of the five estimators for the Global Blinding Index: two GEE estimators, two inverse-variance weighted estimators, commonly used in meta-analysis, and a maximum likelihood estimator derived under the Dirichlet-multinomial distribution. The performance of the GEE estimators was contrasted with the performance of the inverse-variance weighted estimators in simulations, while the maximum likelihood estimator served as a 'gold standard' when the data was generated from the Dirichlet-multinomial distribution. We also studied the robustness of the five estimators to possible distortions from the DM model.

The GEE estimator under the independence working assumption showed satisfactory performance in all scenarios and under different models used to generate the raw data. This estimator may be appealing in practice when a good approximation to the correlation structure in the outcome variable is not available to researchers. Assuming a working correlation matrix to be the identity matrix means that the potential correlations among the observations within a cluster are not used to update the estimates during the iterative process of solving the GEEs. However, the correlations are taken into account in the robust variance estimator. If a sensible approximation to the true correlation matrix can be suggested for the working correlation matrix, it may be possible to obtain more efficient point estimators. At the same time, our results demonstrated that the GEE estimator with the independence working correlation was nearly as efficient as the GEE estimator with the exchangeable working correlation even for large values of the overdispersion parameter.

Concerns have been raised in the literature that a small number of clusters, high variability in cluster sizes and a high intra-class correlation may influence the performance of the robust variance estimator, which, in turn, may result in underestimated variances of the GEE estimators and suboptimal coverage probabilities. The two types of adjustments that have been proposed to address this problem are: adjustments that attempt to reduce the bias of the sandwich estimator and adjustments that use a *t*-distribution rather than a normal distribution to calculate significance levels or construct CIs [14]. In this study, we implemented the small-sample corrections in the robust variance estimator as suggested by Morel *et al.* [15]. We also used a *t*-distribution with degrees of freedom approximated by the number of clusters minus one when constructing the CIs for the GBI estimated using the GEE methods. These simple adjustments were straightforward in implementation and showed satisfying results in the scenarios that we studied.

Our illustrations indicate that for the designs with moderate variability in cluster sizes ($CV \approx$ 40%), the performance of the GEE estimator under the independence working assumption was satisfactory and comparable with the performance of the GEE estimator under the exchangeable working model and an MLE, which are expected to have superior performance by design. We observed that high variability in cluster sizes in combination with a small number of clusters may decrease the performance of the GEE estimator under the independence working assumption. It is possible that additional, more sophisticated adjustments can be beneficial for these scenarios, and their exploration may be a good

direction for future research. Finally, we found that the inverse-variance weighted estimators performed poorly, especially in scenarios with higher values of ρ^2, and, therefore, these estimators should be interpreted with caution in real applications.

In conclusion, our results re-emphasized the importance of accounting for extra-variation to obtain correct variances of the estimators and maintain appropriate coverage probabilities. We suggest using the GEE estimator under the independence working assumption when a reliable approximation to the true correlation structure is not available.

Acknowledgment:

Heejung Bang was partly supported by the U.S. National Institutes of Health through grant ULI TR001860. Authors thank Drs. Barry Graubard and Majnu John for providing useful advice and insight.

References

1. Hinde J, Demetrio CGB. Overdispersion: models and estimation. Computational Statistics & Data Analysis 1998; 27(2): 151–170.
2. Graubard BI, Korn EL. Regression analysis with clustered data. Statistics in Medicine 1994; 13(5–7): 509–522. [PubMed: 8023032]
3. Heagerty PJ, Zeger SL. Marginalized multilevel models and likelihood inference (with comments and a rejoinder by the authors). Statistical Science 2000; 15(1): 1–26.
4. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. Biometrics 1988; 44(4): 1049–1060. [PubMed: 3233245]
5. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986; 73(1): 13–22.
6. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986; 42(1): 121–130. [PubMed: 3719049]
7. Lipsitz SR, Kim K, Zhao L. Analysis of repeated categorical data using generalized estimating equations. Statistics in Medicine 1994; 13(11): 1149–1163. [PubMed: 8091041]
8. Freed B, Assail OP, Panagiotakis G, et al. Assessing blinding in trials of psychiatric disorders: a meta-analysis based on blinding index. Psychiatry Research 2014; 219(2): 241–247. [PubMed: 24930582]
9. Bang H, Ni L, Davis CE. Assessment of blinding in clinical trials. Controlled Clinical Trials 2004; 25(2): 143–156. [PubMed: 15020033]
10. Landsman V, Fillery M, Vernon H, Bang H. Sample size calculations for blinding assessment. Journal of Biopharmaceutical Statistics 2017: 1–13.
11. Touloumis A, Agresti A, Kateri M. GEE for multinomial responses using a local odds ratios parameterization. Biometrics 2013; 69(3): 633–640. [PubMed: 23724948]
12. Krewski D, Zhu Y. Applications of multinomial dose-response models in developmental toxicity risk assessment. Risk Analysis 1994; 14(4): 613–627. [PubMed: 7972962]
13. Zhu Y, Krewski D, Ross WH. Dose-response models for correlated multinomial data from developmental toxicity studies. Applied Statistics 1994: 583–598.
14. Fay MP, Graubard BI. Small-sample adjustments for Wald-type tests using sandwich estimators. Biometrics 2001; 57(4): 1198–1206. [PubMed: 11764261]
15. Morel JG, Bokossa MC, Neerchal NK. Small sample correction for the variance of GEE estimators. Biometrical Journal: journal of mathematical methods in biosciences 2003; 45(4): 395–409.
16. Wang M, Kong L, Li Z, Zhang L. Covariance estimators for generalized estimating equations (GEE) in longitudinal analysis with small samples. Statistics in Medicine 2016; 35(10): 1706–1721. [PubMed: 26585756]

17. Neerchal NK, Morel JG. Large cluster results for two parametric multinomial extra variation models. Journal of the American Statistical Association 1998; 93(443): 1078–1087.
18. Brier SS. Analysis of contingency tables under cluster sampling. Biometrika 1980; 67(3): 591–596.
19. Young-Xu Y, Chan KA. Pooling overdispersed binomial data to estimate event rate. BMC Medical Research Methodology 2008; 8(1): 58. [PubMed: 18713448]
20. Kulinskaya E, Olkin I. An overdispersion model in meta-analysis. Statistical Modelling 2014; 14(1): 49–76.
21. Zhu Y. Correlated multinomial data In: Encyclopedia of Envirometrics, John Wiley & Sons, Ltd 2nd ed. 2012.
22. Lohr SL. Sampling: Design and Analysis. Boston, Mass: Brooks/Cole 2nd ed 2010.
23. Moore DF. Modelling the extraneous variance in the presence of extra-binomial variation. Applied Statistics 1987: 8–14.
24. Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008; 336(7644): 601–605. [PubMed: 18316340]
25. James KE, Bloch DA, Lee KK, Kraemer HC, Fuller RK. An index for assessing blindness in a multi-centre clinical trial: disulfiram for alcohol cessation - a VA cooperative study. Statistics in Medicine 1996; 15(13): 1421–1434. [PubMed: 8841652]
26. Bang H, Flaherty SP, Kolahi J, Park J. Blinding assessment in clinical trials: a review of statistical methods and a proposal of blinding assessment protocol. Clinical Research and Regulatory Affairs 2010; 27(2): 42–51.
27. Viechtbauer W. Hypothesis tests for population heterogeneity in meta-analysis. British Journal of Mathematical and Statistical Psychology 2007; 60(1): 29–60. [PubMed: 17535578]
28. Kulinskaya E, Morgenthaler S, Staudte RG. Combining statistical evidence. International Statistical Review 2014; 82(2): 214–242.
29. Shahar DJ. Minimizing the variance of a weighted average. Open Journal of Statistics 2017; 7(02): 216.
30. Schwarzer G, Carpenter JR, Rucker G. Meta-Analysis with R (Use R!). Cham Heidelberg New York Dordrecht London: Springer 2015.
31. Bakbergenuly I, Kulinskaya E. Beta-binomial model for meta-analysis of odds ratios. Statistics in Medicine 2017; 36(11): 1715–1734. [PubMed: 28124446]
32. DerSimonian R, Laird N. Meta-analysis in clinical trials. Controlled Clinical Trials 1986; 7(3): 177–188. [PubMed: 3802833]
33. Landsman D. Overdispersion Models using Generalized Estimating Equations (OMGEE) Package, https://github.com/Express50/omgee; 2019.
34. Raim AM. Overdispersion Models in R Package, https://githubcom/andrewraim/OverdispersionModelsInR; 2017.
35. Korn EL, Graubard BI. Analysis of Health Surveys. New York: Wiley 1999.
36. Li P, Redden DT. Small sample performance of bias-corrected sandwich estimators for cluster-randomized trials with binary outcomes. Statistics in Medicine 2015; 34(2): 281–296. [PubMed: 25345738]
37. Fay MP, Graubard BI, Freedman LS, Midthune DN. Conditional logistic regression with sandwich estimators: application to a meta-analysis. Biometrics 1998: 195–208. [PubMed: 9544516]
38. Pan W, Wall MM. Small-sample adjustments in using the sandwich variance estimator in generalized estimating equations. Statistics in Medicine 2002; 21(10): 1429–1441. [PubMed: 12185894]
39. Vernon H, Triano JJ, Ross K, Tran S, Soave D, Dinulos M. Validation of a novel sham cervical manipulation procedure. The Spine Journal 2012; 12(11): 1021–1028. [PubMed: 23158966]
40. Vernon H, Triano JT, Soave D, Dinulos M, Ross K, Tran S. Retention of blinding at follow-up in a randomized clinical study using a sham-control cervical manipulation procedure for neck pain: secondary analyses from a randomized clinical study. Journal of Manipulative and Physiological Therapeutics 2013; 36(8): 522–526. [PubMed: 24011656]

41. Park J, Bang H, Caette I. Blinding in clinical trials, time to do it better. Complementary Therapies in Medicine 2008; 3(16): 121–123.

42. Baethge C, Assail OP, Baldessarini RJ. Systematic review of blinding assessment in randomized controlled trials in schizophrenia and affective disorders 2000–2010. Psychotherapy and Psychosomatics 2013; 82(3): 152–160. [PubMed: 23548796]

Table 1.

Typical blinding assessment data

Allocation (*a*)	**Guess (*l*)**			**Total**
	1 (=T)	**2 (=C)**	**3 (=DKN)**	
T	$n_{1\|T}$	$n_{2\|T}$	$n_{3\|T}$	n_T
C	$n_{1\|C}$	$n_{2\|C}$	$n_{3\|C}$	n_C

Table 2a.

Simulation results for $\rho^2 = 0.1$ over 5,000 repetitions under the DM model with constant ρ^2. RB = relative bias; SEE = sampling standard error of the estimator [= standard deviation of the 5,000 estimated BIs]; ESE = estimated standard error of the estimator [= mean of the 5,000 estimated standard errors]; 95% CP = 95% coverage probability

Estimator		***CV* ≈ 90%, $\bar{n} \approx 50$**			***CV* ≈ 40%, $\bar{n} \approx 50$**		
		***K* = 8**	***K* = 15**	***K* = 30**	***K* = 8**	***K* = 15**	***K* = 30**
$\widehat{GBI}^{\text{indep}}$	RB, %	−1.66	−0.44	−0.09	−0.08	0.99	0.40
	SSE/ESE	1.160	1.092	1.049	1.056	1.049	1.009
	ESE	0.056	0.044	0.032	0.057	0.042	0.030
	95% CP	90.82	93.14	93.24	94.06	94.54	94.82
$\widehat{GBI}^{\text{exch}}$	RB, %	−3.69	−0.84	−0.68	−0.74	1.09	0.41
	SSE/ESE	0.968	0.992	1.007	0.949	1.009	0.996
	ESE	0.065	0.046	0.032	0.062	0.044	0.030
	95% CP	95.72	95.47	94.81	96.73	95.70	95.25
$\widehat{GBI}^{\text{DM}}$	RB, %	−1.84	−0.79	−0.56	−0.63	1.11	0.46
	SSE/ESE	1.089	1.057	1.034	1.048	1.067	1.018
	ESE	0.059	0.044	0.031	0.056	0.042	0.030
	95% CP	95.47	94.84	94.72	96.49	94.96	94.92
$\widehat{GBI}^{\text{IVW},\rho^*}$	RB, %	4.17	6.10	6.72	7.42	6.51	7.25
	SSE/ESE	1.095	1.110	1.084	1.060	1.058	1.050
	ESE	0.065	0.047	0.033	0.060	0.044	0.031
	95 % CP	91.06	91.50	91.80	92.44	92.72	92.96
$\widehat{GBI}^{\text{IVW},0}$	RB,%	3.65	5.30	5.72	5.65	7.22	6.76
	SSE/ESE	1.429	1.460	1.479	1.342	1.382	1.348
	ESE	0.048	0.034	0.024	0.047	0.034	0.024
	95% CP	89.74	85.82	82.36	91.90	87.22	86.88

Table 2b.

Simulation results for $\rho^2 = 0.3$ over 5,000 repetitions under the DM model with constant ρ^2. RB = relative bias; SEE = sampling standard error of the estimator [= standard deviation of the 5,000 estimated GBIs]; ESE = estimated standard error of the estimator [= mean of the 5,000 estimated standard errors]; 95% CP = 95% coverage probability

Estimator		***CV* ≈ 90%, $\bar{n} \approx 50$**			***CV* ≈ 40%, $\bar{n} \approx 50$**		
		***K* = 8**	***K* = 15**	***K* = 30**	***K* = 8**	***K* = 15**	***K* = 30**
$\widehat{GBI}^{\text{indep}}$	RB, %	1.79	−0.64	0.73	−1.85	−1.47	−0.99
	SSE/ESE	1.217	1.152	1.065	1.068	1.051	1.017
	ESE	0.113	0.091	0.069	0.109	0.082	0.059
	95% CP	89.18	90.32	93.06	93.66	93.58	94.72
$\widehat{GBI}^{\text{exch}}$	RB, %	−1.75	−2.35	−1.39	−1.78	−0.86	−0.87
	SSE/ESE	1.046	1.041	0.995	1.028	1.021	1.006
	ESE	0.112	0.083	0.059	0.108	0.080	0.057
	95% CP	94.69	93.54	95.30	94.96	94.54	95.12
$\widehat{GBI}^{\text{DM}}$	RB, %	−1.54	−2.50	−1.56	−1.64	−0.81	−0.91
	SSE/ESE	1.111	1.070	1.007	1.084	1.052	1.020
	ESE	0.105	0.080	0.058	0.101	0.077	0.056
	95% CP	94.91	94.10	95.44	95.40	94.72	95.18
$\widehat{GBI}^{\text{IVW},\rho^*}$	RB, %	18.11	20.61	24.14	22.57	28.18	30.05
	SSE/ESE	1.261	1.241	1.245	1.238	1.244	1.251
	ESE	0.116	0.088	0.064	0.114	0.086	0.062
	95% CP	85.14	86.14	85.86	86.04	86.16	85.46
$\widehat{GBI}^{\text{IVW},0}$	RB,%	22.23	23.24	26.51	22.14	24.64	26.62
	SSE/ESE	3.599	3.966	4.077	3.266	3.247	3.502
	ESE	0.045	0.032	0.022	0.044	0.032	0.023
	95% CP	49.12	41.16	37.08	53.06	46.00	42.68

Table 2c.

Simulation results for $\rho^2 = 0.5$ over 5,000 repetitions under the DM model with constant ρ^2. RB = relative bias; SEE = sampling standard error of the estimator [= standard deviation of the 5,000 estimated GBIs]; ESE = estimated standard error of the estimator [= mean of the 5,000 estimated standard errors]; 95% CP = 95% coverage probability

Estimator		***CV* ≈ 90%, $\bar{n} \approx 50$**			***CV* ≈ 40%, $\bar{n} \approx 50$**		
		***K* = 8**	***K* = 15**	***K* = 30**	***K* = 8**	***K* = 15**	***K* = 30**
$\widehat{GBI}^{\text{indep}}$	RB, %	−0.59	−5.53	−1.87	3.30	−1.44	0.20
	SSE/ESE	1.216	1.121	1.070	1.065	1.044	1.016
	ESE	0.180	0.145	0.110	0.170	0.129	0.093
	95% CP	88.46	90.68	92.46	92.82	93.56	94.28
$\widehat{GBI}^{\text{exch}}$	RB, %	−5.51	−9.38	−6.27	3.55	−1.59	−0.81
	SSE/ESE	1.031	1.018	1.013	1.027	1.021	1.003
	ESE	0.166	0.123	0.087	0.166	0.123	0.088
	95% CP	94.17	94.50	94.90	94.20	94.50	94.82
$\widehat{GBI}^{\text{DM}}$	RB, %	−7.73	−12.09	−9.29	1.01	−4.48	−4.08
	SSE/ESE	1.085	1.031	1.008	1.092	1.042	1.007
	ESE	0.152	0.116	0.084	0.150	0.115	0.083
	95% CP	94.30	94.64	95.22	94.08	94.78	94.84
$\widehat{GBI}^{\text{IVW},\rho^*}$	RB, 95%	41.89	49.78	49.99	51.62	60.16	62.93
	SSE/ESE	1.470	1.508	1.541	1.488	1.551	1.567
	ESE	0.180	0.140	0.105	0.182	0.142	0.105
	95% CP	76.82	76.84	75.98	75.72	74.26	73.16
$\widehat{GBI}^{\text{IVW},0}$	RB,%	45.46	51.66	67.89	61.10	60.57	69.98
	SSE/ESE	8.363	9.966	12.021	7.847	8.982	9.715
	ESE	0.038	0.027	0.018	0.037	0.026	0.018
	95% CP	22.42	17.32	13.90	22.68	18.22	15.48

Table 2d.

Simulation results for data generated under different models over 5,000 repetitions. DM = Dirichlet-multinomial distribution, RCM = Random Clumped Multinomial distribution. RB = relative bias; SEE = sampling standard error of the estimator [= standard deviation of the 5,000 estimated GBIs]; ESE = estimated standard error of the estimator [= mean of the 5,000 estimated standard errors]; 95% CP = 95% coverage probability. In all scenarios: $\boldsymbol{\pi} = (0.5, 0.4, 0.1)$, CV ≈ 90%, $\bar{n} \approx 50$, $K = 8$.

Estimator		Data generation mechanism			
		DM, $\rho^2 = 0.5$	**DM, $\rho_i^2 \sim U(0.2, 0.8)$**	**RCM, $\rho^2 = 0.5$**	**RCM, $\rho_i^2 \sim U(0.2, 0.8)$**
$\widehat{GBI}^{\text{indep}}$	RB, %	−0.59	−2.17	−0.24	−6.27
	SSE/ESE	1.216	1.263	1.244	1.196
	ESE	0.180	0.186	0.249	0.247
	95% CP	88.46	88.44	83.30	85.22
$\widehat{GBI}^{\text{exch}}$	RB, %	−5.51	−6.13	−8.20	−9.83
	SSE/ESE	1.031	1.006	0.998	0.966
	ESE	0.166	0.180	0.234	0.230
	95% CP	94.17	94.31	93.58	94.07
$\widehat{GBI}^{\text{DM}}$	RB, %	−7.73	−10.35	−13.82	−16.89
	SSE/ESE	1.085	1.144	1.216	1.158
	ESE	0.152	0.154	0.166	0.166
	95% CP	94.30	93.74	91.64	92.96
$\widehat{GBI}^{\text{IVW}, \rho^*}$	RB, %	41.89	33.02	32.64	29.93
	SSE/ESE	1.470	1.776	1.145	1.236
	ESE	0.180	0.187	0.227	0.230
	95% CP	76.82	68.86	85.18	83.06
$\widehat{GBI}^{\text{IVW}, 0}$	RB, %	45.46	37.56	41.30	34.96
	SSE/ESE	8.363	11.249	9.489	10.899
	ESE	0.038	0.035	0.033	0.032
	95% CP	22.42	15.14	14.54	13.58

Table 3a.

Example 1: Two sets of simulated data that closely mimic the data in a multi-clinician RCT.

Arm	Chiropractor	2×2 study			2×3 study				Total
		$n_{2\|C}$	$\hat{BI}_C$		$n_{1\|C}$	$n_{2\|C}$	$\hat{BI}_C$		n_C
Alternative	A	6	−0.14		6	4	−0.14		14
	B	2	−0.33		3	1	−0.33		6
	C	3	0.20		2	2	0.00		5
	D	20	0.67		3	15	0.50		24
	E	11	0.29		5	8	0.18		17
	F	2	−0.20		2	2	0.00		5
	G	13	−0.16		13	10	−0.10		31
	H	4	0.33		1	3	0.33		6
	J	29	0.35		10	22	0.28		43
	I	27	−0.02		31	10	−0.38		55
Total									206

		$n_{1\|T}$	$\hat{BI}_T$		$n_{1\|T}$	$n_{2\|T}$	$\hat{BI}_T$		n_T
Typical	A	4	−0.43		3	8	−0.36		14
	B	3	0.00		2	3	−0.17		6
	C	3	0.00		3	2	0.17		6
	D	7	−0.39		5	13	−0.35		23
	E	6	−0.25		4	8	−0.25		16
	F	4	0.33		3	2	0.17		6
	G	18	0.16		14	10	0.13		31
	H	3	0.20		2	2	0.00		5
	J	24	0.12		18	15	0.07		43
	I	41	0.49		32	11	0.38		55
Total									205

$n_{2|C}$ and n_C represent the count of the correct guesses and the total number of patients in arm C, respectively; $n_{1|T}$ and n_T represent the count of the correct guesses and the total number of patients in arm T, respectively; $\hat{BI}_C$ and $\hat{BI}_T$ are the estimates of clinician-level BIs for arms T and C, respectively.

Table 3b.

Estimated GBI values and their 95% confidence intervals (CI) from the RCT with multiple practitioners in 2 × 2 format and 2 × 3 format

Method		2 × 2 format		2 × 3 format	
		Alternative arm	Typical arm	Alternative arm	Typical arm
GEE (indep)	$\widehat{GBI}^{\text{indep}}$	0.136	0.108	0.005	0.059
	95% CI	(−0.106, 0.378)	(−0.192, 0.408)	(−0.316, 0.325)	(−0.194, 0.311)
GEE (exch)	$\widehat{GBI}^{\text{exch}}$	0.135	0.050	0.020	0.059
	95% CI	(−0.107, 0.377)	(−0.270, 0.371)	(−0.303, 0.343)	(−0.210, 0.186)
	$\hat{\rho}^2$	0.029	0.034	0.010	n/a
DM	$\widehat{GBI}^{\text{DM}}$	0.134	0.040	0.030	0.030
	95% CI	(−0.096, 0.364)	(−0.207, 0.288)	(−0.153, 0.213)	(−0.142, 0.203)
	$\hat{\rho}^2$	0.206	0.232	0.172	0.111
IVW, ρ^*	$\widehat{GBI}^{\text{IVM}, \rho^*}$	0.175	0.034	0.049	−0.012
	95% CI	(−0.045, 0.394)	(−0.203, 0.272)	(−0.186, 0.284)	(−0.210, 0.186)
	$\hat{\rho}^*$	0.079	0.100	0.155	0.077
IVW ($\rho^* = 0$)	$\widehat{GBI}^{\text{IVM}}$	0.183	0.121	0.013	0.062
	95% CI	(0.056, 0.311)	(−0.008, 0.249)	(−0.096, 0.122)	(−0.053, 0.176)

Table 4a.

Estimated GBI values and their 95% confidence intervals (CI) from meta-analysis of psychiatric disorders (‘joint’ analysis of 40 studies)

Method		Control arm	Treatment arm
GEE (indep)	$\widehat{GBI}^{\text{indep}}$	−0.040	0.120
	95 % CI	(−0.132; 0.052)	(0.007; 0.233)
GEE (exch)	$\widehat{GBI}^{\text{exch}}$	−0.020	0.121
	95 % CI	(−0.108, 0.067)	(0.019, 0.224)
	$\widehat{\rho}$	0.294	0.261
DM	$\widehat{GBI}^{\text{DM}}$	0.000	0.103
	95 % CI	(−0.158; 0.159)	(−0.051; 0.258)
	$\widehat{\rho}$	0.492	0.486
IVW, ρ^*	$\widehat{GBI}^{\text{IVM}, \rho^*}$	−0.007	0.163
	95% CI	(−0.083; 0.070)	(0.056; 0.270)
	$\widehat{\rho}$	0.061	0.152
IVW ($\rho^* = 0$)	$\widehat{GBI}^{\text{IVM}}$	−0.016	0.138
	95% CI	(−0.052; 0.021)	(0.105; 0.171)

Table 4b.

Estimated GBI values and their 95% confidence intervals (CI) from meta-analysis of psychiatric disorders ('separate' analysis of 30 2 × 2 studies and 10 2 × 3 studies)

Method		2 × 2 format		2 × 3 format	
		Control arm	Treatment arm	Control arm	Treatment arm
GEE (indep)	$\widehat{GBI}^{\text{indep}}$	−0.092	0.151	0.107	0.031
	95% CI	(−0.200, 0.016)	(0.004, 0.297)	(−0.006, 0.219)	(−0.073, 0.136)
GEE (exch)	$\widehat{GBI}^{\text{exch}}$	−0.076	0.140	0.148	0.052
	95% CI	(−0.180, 0.029)	(0.012, 0.268)	(0.029, 0.267)	(−0.076, 0.181)
	$\hat{\rho}^2$	0.219	0.300	0.113	0.065
DM	$\widehat{GBI}^{\text{DM}}$	−0.076	0.143	0.121	0.040
	95% CI	(−0.178, 0.026)	(0.015, 0.270)	(−0.064, 0.306)	(−0.097, 0.178)
	$\hat{\rho}^2$	0.204	0.304	0.339	0.224
IVW, ρ^*	$\widehat{GBI}^{\text{IVM}, \rho^*}$	−0.103	0.232	0.098	0.038
	95% CI	(−0.206, −0.001)	(0.091, 0.372)	(0.007, 0.188)	(−0.082, 0.159)
	$\hat{\rho}^*$	0.057	0.169	0.033	0.059
IVW ($\rho^* = 0$)	$\widehat{GBI}^{\text{IVM}}$	−0.105	0.221	0.068	0.023
	95% CI	(−0.157, −0.053)	(0.177, 0.264)	(0.018, 0.118)	(−0.028, 0.074)