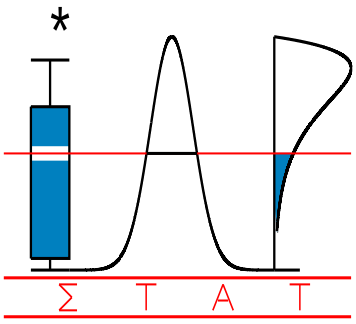
T E C H N I C A L

R E P O R T

0622

ON THE ESTIMATION OF THE MISCLASSIFICATION COUNT TABLE FOR FINITE DATA WITH AN APPLICATION IN CARIES RESEARCH

Emmanuel Lesaffre, Helmut Küchenhoff, Samuel M. Mwalili & Dominique Declerck



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

SUMMARY. We look at the correction for misclassification of possibly corrupted finite count data in epidemiological studies. The general strategy is to estimate the misclassification probabilities from a validation study and use the estimated misclassification matrix to correct for the distortion. However, most often the validation study is quite small implying that the misclassification probabilities are estimated with high variability if based on the multinomial distribution. To increase efficiency we propose an approach based on the fact that to determine a count the examiner needs to evaluate all items that make up that count. In the simplest case, two independent binomial distributions are modelled and therefore we call it the double binomial approach. We also suggest various extensions of the double binomial approach which might mimic better the scoring behavior of the examiner relative to a gold standard. We evaluate the relative efficiency of our approach(es) to estimate the misclassification probabilities in comparison to the multinomial approach in an analytical way and in a simulation study. Finally, the practical use of our methods is exemplified on an oral health survey examining caries experience in seven-year old Flemish children.

KEY WORDS: Count data; Logistic Regression; Misclassification; Prevalence; Response Error

1. Introduction

Large epidemiological studies often involve different scorers or examiners. Unless mortality is the outcome, assessing (the severity of) the disease is nearly always prone to misclassification. Here we will assume that the severity of the disease is expressed as a finite count. Some well-known examples are: (a) in rheumatology: the number of sensitive joints; (b) in cardiology:

k-vessel disease where $k = 1, 2, 3, \ldots$ indicates the number of coronary arteries that are atherosclerotic; or (c) in dentistry: the dmft(DMFT)-index expresses the total number of deciduous (permanent) teeth that are either decayed, missing due to extraction because of caries, or filled.

The effect of misclassification has been studied in various models and in different circumstances. In general, the conclusion is that misclassification results in a biased estimate of the prevalence (or incidence) of the disease. In regression models it has been shown that in the most simple cases misclassification causes attenuated regression parameters (biased towards the null value), but in general its effect is unpredictable (see e.g. Gustafson, 2004). When the misclassification probabilities can be estimated unbiasedly, then a suitable correction mechanism can result in (nearly) unbiasedly estimated parameters. These probabilities can be estimated from a validation study, if available, or alternatively from expert knowledge.

In a validation study the different examiners score the (severity of the) disease and their score is compared to the score of a gold standard or benchmark scorer. However, from our experience in oral health research we must conclude that validation studies are often too small implying that the corrective terms are estimated with high variability. Especially with (finite) count data the misclassification matrix is most often sparse. Hence when estimated from a multinomial model, some of the misclassification probabilities cannot be determined or the estimated misclassification probabilities will be quite unreliable yielding corrected estimated parameters with high variability and thus unclear scientific conclusions.

Here, we suggest to base the estimation of the misclassification proba-

bilities on the *double binomial* (DB) approach or extensions thereof. The DB approach exploits the nature of a count. Namely, in order to obtain a count the sum needs to be made of binary scores, each of which is prone to misclassification. Thus, one could estimate the misclassification probabilities for count data from the misclassification table of the binary scores making up the count. This approach can only be applied when the misclassification process of the binary scores is done independently and does not depend on the label of the binary score. These assumptions have been implicitly taken for granted in a number of applications (see e.g. Paulino et al., 2003; Paulino et al., 2005). However, when the counts are determined within a subject, it is not immediately clear that these simplifying assumptions hold in practice. Further, when the misclassification process is differential, this also needs to be taken account. This is exemplified by Luan et al. (2005). Finally, Paulino et al. (2003) and Paulino et al. (2005) assumed a (random-effects) binomial regression model, which easily combines with the misclassification model. However, in the dental example examined here, a more complex model had to be assumed. Thus, we argue that there is a need for an approach to analyze finite count data in the presence of misclassification that: (a) expresses the misclassification process in an appropriate manner; (b) maximizes the efficiency in estimating the misclassification probabilities; and (c) combines the misclassification process with an appropriate and possibly complex main model. The elegant approach of Albert et al. (1997), developed for ordinal responses, has been applied by Mwalili et al. (2005a) but showed not be the best choice here.

In Section 2, the double binomial approach and its extensions are intro-

duced. In Section 3 we compare analytically the sampling variability of the multinomial and the basic double binomial approach. Sometimes it is sufficient to just estimate the prevalence of the disease even with count data. Therefore, we examine in Section 4 the relative efficiency of estimating the sensitivity and specificity using the double binomial method to the direct estimate obtained from the 2×2 contingency table of the dichotomized true count and the possibly corrupted count. In Section 5 the results of a simulation study are summarized. More details of the simulation study can be found in Lesaffre et al. (2006). In Section 6 we describe how the estimated misclassification probabilities can be used to correct for possible misclassification and how to incorporate the variability with which these correction terms have been estimated. In Section 7 the approaches are applied to the Signal-Tandmobiel® study. In the final section we discuss the practical implications of our results.

2. Estimating the misclassification probabilities of a finite count

Let $Y = \sum_{k=1}^{K} Z_k$ be the "true" count as determined by the gold standard and $Y^* = \sum_{k=1}^{K} Z_k^*$ be the possibly corrupted observed count established by an examiner. Z_k and Z_k^* are the true and possibly corrupted binary scores, respectively which make up the respective counts. Further, let $\pi_{rs}(\boldsymbol{x}) = P(Y^* = r|Y = s, \boldsymbol{x})$ $(r, s = 0, \ldots, K)$ with $\sum_{r=0}^{K} \pi_{rs}(\boldsymbol{x}) = 1$ represent the misclassification probabilities constituting the vector $\boldsymbol{\pi}_s(\boldsymbol{x}) = (\pi_{0s}(\boldsymbol{x}), \pi_{1s}(\boldsymbol{x}), \cdots, \pi_{Ks}(\boldsymbol{x}))^T$ and the misclassification matrix $(\boldsymbol{\pi}_0(\boldsymbol{x}), \ldots, \boldsymbol{\pi}_K(\boldsymbol{x}))$. When the misclassification probabilities depend on a vector $\boldsymbol{x}$, the misclassification process is called differential, otherwise it is called non-differential. Suppose a $(K+1) \times (K+1)$ misclassification table is

obtained from validation data with entries n_{rs} with $\sum_{r,s=0}^{K} n_{rs} = n$ whereby n_{rs} represents the number of subjects classified as $Y = s$ by the gold standard and $Y^* = r$ by the examiner and n is the total number of subjects involved in the validation study.

In this paper we assume that interest lies in Y and not in the individual Z_ks. In another paper, we will consider models for multivariate binary responses subject to misclassification.

2.1 *The multinomial approach*

Let us assume independence of the subjects. Then the sth column, $\boldsymbol{n}_s$, of the misclassification table with entries n_{rs} follows a multinomial distribution:

$$\boldsymbol{n}_s \sim \text{Multinomial}\big(n_s, \boldsymbol{\pi}_s(\boldsymbol{x})\big). \tag{1}$$

For a non-differential misclassification process, the multinomial estimate of π_{rs} is $\hat{\pi}_{rs} = n_{rs}/\sum_{r=0}^{K} n_{rs}$ and has variance $\pi_{rs}(1-\pi_{rs})/\sum_{r=0}^{K} n_{rs}$. However, the variance can be high and the estimate does not exist when the gold standard does not score 's' in the validation data. When the misclassification process is differential the dependence on the covariates needs to be modelled.

2.2 *The double binomial approach*

In order to obtain a count, one needs to score the binary indicators Z_k $(k = 1, \ldots, K)$. Hence, it is likely that the validation data provide a misclassification table for each Z_k. Suppose for a non-differential misclassification process that $\alpha_k = P(Z_k^* = 1|Z_k = 1)$, $\beta_k = P(Z_k^* = 0|Z_k = 0)$ $(k = 1, \ldots, K)$ represent the sensitivity and specificity for Z_k, respectively of the examiner relative to a gold standard. α_k and β_k can be estimated from

the corresponding 2×2 misclassification table established in the validation study with entries $n_{k,rs}$ with $\sum_{r,s=0}^{1} n_{k,rs} = n$ as follows: $\hat{\alpha}_k = \frac{n_{k,11}}{n_{k,01}+n_{k,11}}$ and $\hat{\beta}_k = \frac{n_{k,00}}{n_{k,00}+n_{k,10}}$ $(k = 1, \ldots, K)$.

The above assumptions imply a binomial model for the sensitivity and for the specificity on the binary score. Further, it is assumed in first instance that the misclassification process is non-differential. Therefore, the basic double binomial approach (see below) will be based on the following three simplifying assumptions:

- Assumption A1: scoring Z_k is done independently from scoring Z_l with $k \neq l$
- Assumption A2: the scoring behavior of the examiner does not depend on k
- Assumption A3: the scoring behavior of the examiner does not depend on the subject (non-differential misclassification process)

For the dental example, the assumptions $A1$ and $A2$ imply that the scoring of teeth is done equally well or bad for all teeth and that in an independent manner. Hence, when $A1$ to $A3$ are satisfied, $\alpha_k = \alpha_Z$ and $\beta_k = \beta_Z$ and are estimated by $\hat{\alpha}_Z = \frac{\sum_{k=1}^{K} n_{k,11}}{\sum_{k=1}^{K}[n_{k,01}+n_{k,11}]}$ and $\hat{\beta}_Z = \frac{\sum_{k=1}^{K} n_{k,00}}{\sum_{k=1}^{K}[n_{k,00}+n_{k,10}]}$, respectively.

Under the above simplifying assumptions, one can determine the misclassification table for Y based on the misclassification table for Z_k $(k = 1, \ldots, K)$, which is assumed to be equal for all k. Namely

$$\pi_{rs} = \sum_{m=M_0}^{M_1} \binom{s}{m} \binom{K-s}{r-m} \alpha_Z^m (1-\alpha_Z)^{(s-m)} (1-\beta_Z)^{(r-m)} \beta_Z^{(K-s-r+m)}, \quad (2)$$

where the bounds $M_0 = \max(r - (K - s), 0)$ and $M_1 = \min(r, s)$ arise from the fact that expression (2) is derived from the distribution of two independent binomial distributions, $\text{Bin}(s, \alpha_Z)$ and $\text{Bin}(K - s, 1 - \beta_Z)$. In the dental example, the first binomial distribution expresses the probability that the examiner scores m teeth as decayed from the s teeth that the gold standard has scored decayed. The second binomial distribution expresses the probability that the examiner scores $(r - m)$ teeth as decayed from the $(K - s)$ teeth that the gold standard has scored not decayed. Plugging the estimates $\hat{\alpha}_Z$, $\hat{\beta}_Z$ in expression (2) yields estimates $\tilde{\pi}_{rs}$ $(r, s = 0, \ldots, K)$ and hence the vectors $\tilde{\boldsymbol{\pi}}_s = (\tilde{\pi}_{0s}, \ldots, \tilde{\pi}_{Ks})^T$ $(s = 0, \ldots, K)$.

2.3 *Extensions of the DB approach*

Assumptions $A1$ to $A3$ might not hold in practice. But, since the DB approach is based on two binomial distributions in principle all types of extensions of binomial models could be used. We discuss below some natural extensions of the DB model. The DB model can be described as follows:

$$P(Z_1^*, \ldots, Z_K^* | Z_1, \ldots, Z_K) = \prod_{k=1}^{K} P(Z_k^* | Z_k), \tag{3}$$

with $P(Z_k^* = 1 | Z_k = 1) = \alpha_Z$ and $P(Z_k^* = 0 | Z_k = 0) = \beta_Z$. Thus, extensions of the DB approach can be formulated as extensions of (3).

- Extension $E1$: the sensitivity and specificity of the binary scores depend on covariates (Begg, 1987). For instance, when diagnosing oral cancer the sensitivity of detecting the disease might be higher for smokers than for non-smokers because the physician is more alerted for a smoker. Thus, we assume that (3) holds, but that $\alpha_Z \equiv \alpha_Z(\boldsymbol{x})$,

$\beta_Z \equiv \beta_Z(\boldsymbol{x})$. In other words, we assume that $A1$ and $A2$ are satisfied, but not $A3$. In this case, the misclassification process is called differential.

- Extension $E2$: model (3) is extended to

$$P(Z_1^*, \ldots, Z_K^* | Z_1, \ldots, Z_K) = \prod_{k=1}^{K} P(Z_k^* | Z_k, f(Z_1, \ldots, Z_K)).$$

Thus, the misclassification process depends on a global summary statistic of the binary scores, i.e. $\alpha_Z \equiv \alpha_Z(f(Z_1, \ldots, Z_K))$,
$\beta_Z \equiv \beta_Z(f(Z_1, \ldots, Z_K))$. The motivation behind this extension is best seen in the caries example. When $f(Z_1, \ldots, Z_K) = \sum_{k=1}^{K} Z_k$ is large there is much caries in the mouth. It is conceivable that in such a mouth there is might be some confusion of when a tooth is decayed or not.

- Extension $E3$: the scoring is dependent, i.e. $P(Z_1^*, \ldots, Z_K^* | Z_1, \ldots, Z_K)$ does not split up in a product. Thus, assumption $A1$ is relaxed, for a motivation in oral health studies see e. g. Hujoel et al. (1990). There exists a variety of models for correlated binary random variables, see e.g. Rudolfer (1990). A convenient way to introduce correlation is to assume that, given a subject, the scoring is independent but that the sensitivity and specificity depend on the subject (unknown) characteristics. More formally, assume that α and β have a distribution depending on the subject's unknown characteristics, given by the random vector $\boldsymbol{b}$ and that

$$P(Z_1^*, \ldots, Z_K^* | Z_1, \ldots, Z_K, \boldsymbol{b}) = \prod_{k=1}^{K} P(Z_k^* | Z_k, \boldsymbol{b}),$$

and $P(Z_1^*, \ldots, Z_K^*|Z_1, \ldots, Z_K) = \int \prod_{k=1}^{K} P(Z_k^*|Z_k, \boldsymbol{b}) f(\boldsymbol{b}) d\boldsymbol{b}$. There are two natural candidates for the distribution of α and β. Firstly, assume that α and β each have a Beta density and that they are independent of each other. In that case $P(Z_1^*, \ldots, Z_K^*|Z_1, \ldots, Z_K) = \int \prod_{Z_k=1} P(Z_k^*|Z_k, \alpha) B(\alpha) d\alpha \int \prod_{Z_k=0} P(Z_k^*|Z_k, \beta) B(\beta) d\beta$, where $B(\cdot)$ represents a Beta-density. Thus, $\alpha \equiv b_1$ and $\beta \equiv b_2$, where $\boldsymbol{b}^T = (b_1, b_2)$. This generalizes the binomial distribution $\text{Bin}(N, \pi)$ to a beta-binomial distribution $\text{BB}(N, \pi, \tau)$, with mean π and variance τ (see Section 7 for an expression). Further, this implies that in expression (2), $\text{Bin}(s, \alpha_Z)$ and $\text{Bin}(K - s, 1 - \beta_Z)$ are replaced by $\text{BB}(s, \alpha_Z, \tau_\alpha)$ and $\text{BB}(K - s, \beta_Z, \tau_\beta)$, respectively. Secondly, since α and β have a distribution depending on the subject, it is natural to assume that they are also correlated. Correlation can be introduced by first taking the logit transform of α and β, i.e. $b_1 = \text{logit}(\alpha)$, $b_2 = \text{logit}(\beta)$ and then assuming that $\boldsymbol{b}^T = (b_1, b_2) \sim N(\boldsymbol{\mu_b}, \Sigma_{\boldsymbol{b}})$.

- Extension $E4$: sensitivity and specificity depend on k, i.e. α_k, β_k ($k = 1, \ldots, K$). For instance, in caries research it is known that detecting caries experience in molars is more difficult than in other teeth.

In the sequel the basic DB model will also be denoted by $E0$. To test whether the E0 model needs to be extended to Ex ($x = 1, \ldots, 4$) a likelihood ratio test can be employed. To choose between the extensions, Akaike's Information Criterion might be used. Further, the above extensions could be combined making the misclassification process even more general. The availability of a battery of models for the misclassification process helps

in obtaining unbiased estimates of the misclassification probabilities while maintaining high efficiency. For all of the extensions, the estimated misclassification probabilities will still be denoted by $\tilde{\boldsymbol{\pi}}_s$.

In the literature it is often recommended that the validation data constitute a random sample of the main study, thereby establishing an internal validation data set. However, it is sufficient that the misclassification probabilities π_{rs} are estimated unbiasedly to obtain consistent estimates of the main model parameters. This allows other sampling strategies for the validation data. An example is given by the method of conditional sampling as suggested by Haitovsky and Rapp (1992). In caries research an alternative strategy than simple random sampling is also to be recommended especially when using the multinomial method.

3. Efficiency considerations

In this section we compare the efficiency of the multinomial and the simple DB approach. The improvement in efficiency is an upper bound of the efficiency that one could get by modelling the misclassification probabilities as above since the extensions of the DB approach involve more parameters to fit.

3.1 *Sampling variability of the multinomial approach*

The covariance matrix of $\hat{\boldsymbol{\pi}}_s = (\hat{\pi}_{0s}, \ldots, \hat{\pi}_{Ks})^T$ is equal to $\Gamma_s = \frac{1}{n_s}(Diag(\boldsymbol{\pi}_s) - \boldsymbol{\pi}_s\boldsymbol{\pi}_s^T)$ estimated by $\hat{\Gamma}_s$ by plugging $\hat{\boldsymbol{\pi}}_s$ into Γ_s. The elements of the matrix Γ_s are denoted as $\gamma_{gh,s}$ $(g, h = 0, \ldots, K)$.

3.2 *Sampling variability of the DB approach*

Recall that the estimates $\hat{\alpha}_Z$, $\hat{\beta}_Z$ are plugged in expression (2) to yield $\tilde{\boldsymbol{\pi}}_s$. Further, the variance of $\hat{\alpha}_Z$ ($\hat{\beta}_Z$) equals $\tau_1 = \frac{\alpha_Z(1-\alpha_Z)}{\sum_{k=1}^{K}[n_{k,01}+n_{k,11}]}$ ($\tau_0 = \frac{\beta_Z(1-\beta_Z)}{\sum_{k=1}^{K}[n_{k,00}+n_{k,10}]}$). Together with the (bivariate) Delta Method the covariance matrix of $\tilde{\boldsymbol{\pi}}_s$ can be obtained. The kth component of $\tilde{\boldsymbol{\pi}}_s$ is a function $f_{s,k}$ of α_Z and β_Z. Thus, the covariance matrix of $\tilde{\boldsymbol{\pi}}_s$ is equal to

$$\Sigma_s = \left(\frac{\partial f_{s,k}}{\partial \alpha_Z}, \frac{\partial f_{s,k}}{\partial \beta_Z}\right) \Sigma_{\hat{\alpha}_Z,\hat{\beta}_Z} \begin{pmatrix} \frac{\partial f_{s,m}}{\partial \alpha_Z} \\ \frac{\partial f_{s,m}}{\partial \beta_Z} \end{pmatrix}, \tag{4}$$

where $\Sigma_{\hat{\alpha}_Z,\hat{\beta}_Z} = diag\,(\tau_1, \tau_0)$. The elements of the matrix Σ_s are denoted by $\sigma_{gh,s}$ $(g, h = 0, \ldots, K)$. By plugging in the estimates $\hat{\alpha}_Z$ and $\hat{\beta}_Z$ the matrix $\tilde{\Sigma}_s$ is obtained.

3.3 *Asymptotic relative efficiency of the DB approach with respect to the multinomial approach*

In this subsection we work under the assumptions $A1$ to $A3$. Let us denote the asymptotic relative efficiency of the double binomial method to the multinomial method for estimating π_{rs} by $\text{ARE}_{rs} = \frac{\sigma_{rr,s}}{\gamma_{rr,s}}$. The expression for ARE_{rs} can be derived from the multinomial variance and expression (4). We show in Figure 1 the values of $\log(\text{ARE})_{rs}$ for $K = 5$, $\alpha_Z = 0.90$, $\beta_Z = 0.90$ and prevalence equal to 0.10 where $r = s$ ranges from 0 to 5. From this figure we can see that the relative efficiency is always bigger than 1. It actually varies here from 1.9 for $r = s = 0$ to $12,483$ for $r = s = 5$. Equally so, the relative efficiency for the off-diagonal elements of the 6×6 table of misclassification probabilities shows for most values of r and s a tremendous gain (results not shown). Further, the efficiency increases with K, see also next section. Furthermore, it is important to note that the efficiency also

depends on the marginal distribution of Y.

[Figure 1 about here.]

4. Estimating the prevalence of a disease

Suppose that only the prevalence of the disease is of interest. In that case the interest lies in detecting $Y > 0$. Define the binary variable W as $W = 0$ when $Y = 0$ (not diseased) and $W = 1$ when $Y > 0$ (diseased). Further, let W^* be the corresponding score given by the examiner. The specificity and sensitivity of the examiner vis-a-vis the gold standard in scoring prevalence is $\beta_W = P(W^* = 0|W = 0) = P(Y^* = 0|Y = 0)$,

$$\alpha_W = P(W^* = 1|W = 1) = \sum_{r=1}^{K}\sum_{s=1}^{K} \pi_{rs} P(Y = s|Y > 0), \tag{5}$$

respectively. α_W and β_W can be estimated in two ways. Firstly, a direct estimate is obtained from the 2×2 misclassification table of W and W^*, i.e. $\hat{\beta}_W = \frac{n_{00}}{\sum_{r=0}^{K} n_{k0}}$ and $\hat{\alpha}_W = \frac{\sum_{r=1}^{K}\sum_{s=1}^{K} n_{rs}}{\sum_{r=0}^{K}\sum_{s=1}^{K} n_{rs}}$. The expression for $\hat{\alpha}_W$ also follows from taking the multinomial estimates of π_{rs} and of $P(Y = s|Y > 0)$, i.e. $\frac{n_{rs}}{\sum_{r=0}^{K} n_{rs}}$, $\frac{\sum_{r=0}^{K} n_{rs}}{\sum_{r=0}^{K}\sum_{s=1}^{K} n_{rs}}$, respectively.

Secondly using the DB approach the specificity is equal to $\beta_W = \beta_Z^K$ yielding the estimate $\tilde{\beta}_W = \hat{\beta}_Z^K = \left(\frac{\sum_{k=1}^{K} n_{k,00}}{\sum_{k=1}^{K}[n_{k,00}+n_{k,10}]}\right)^K$. For the sensitivity one obtains $\tilde{\alpha}_W = \sum_{r=1}^{K}\sum_{s=1}^{K} \tilde{\pi}_{rs} \frac{\sum_{r=0}^{K} n_{rs}}{\sum_{r=0}^{K}\sum_{s=1}^{K} n_{rs}}$.

Using the Delta Method the variance of $\tilde{\beta_W}$ is equal to $[K\hat{\beta_Z}^{K-1}]^2 \frac{\beta_Z(1-\beta_Z)}{\sum_{k=1}^{K}[n_{k,00}+n_{k,10}]}$. In a similar manner the variance of $\tilde{\alpha}_W$ is obtained, but a more complicated expression arises and is omitted here.

The estimate $\hat{\beta}_W$ only requires an unbiased estimate of π_{00}, while $\tilde{\beta}_W$ also requires that assumptions $A1$ to $A3$ hold. For $\hat{\alpha}_W$ and $\tilde{\alpha}_W$ the probability

$P(Y = s|Y > 0)$ needs to be estimated unbiasedly from the validation data. Thus, the validation data need to constitute an unbiased subsample from the main study. For $\tilde{\alpha}_W$ also the simplifying assumptions $A1$ to $A3$ need to hold.

[Table 1 about here.]

Clearly, the asymptotic relative efficiency for estimating the specificity equals ARE_{00}. In Table 1, we calculated the relative efficiency in estimating the specificity β_W and sensitivity α_W for a variety of situations. The relative efficiency varies greatly but is always bigger than 1. Further, we observe that the relative efficiency in estimating β_W increases with K and decreases with increasing specificity of Z_k. Since the estimated specificity equals $\tilde{\beta}_W = \hat{\beta}_Z^K$ its relative efficiency does not depend on the sensitivity of Z_k. The relative efficiency in estimating α_W increases with K and decreases with increasing sensitivity and specificity of Z_k.

5. Simulation study

A simulation study was set up to evaluate the performance of the DB approach and its extensions. More specifically, we have set up two types of simulations: (1) Evaluating the efficiency of the $E0$ model (basic DB approach) when the true model is in fact Ex ($x = 0, \ldots, 4$) in comparison to the true extension and in comparison to the multinomial model. This evaluates the basic DB approach in estimating the correction terms; (2) Evaluating the efficiency of the DB approach as above, but when estimating the main model parameters. This efficiency comparison was set up to evaluate the practical gain in estimating the measurement model parameters when using the most efficient procedure for estimating the correction terms.

5.1 *Setup of the simulation study*

5.1.1 First simulation study In the first simulation study, the sample size of the validation study was fixed to $N = 100$. Each scenario was sampled 1000 times. We formulated below the four extensions of the basic DB approach as logistic models. Namely:

$$L[P(Z^*_{k,i} = 1)] = \gamma_0 + \gamma_1 Z_{k,i} \quad (6)$$

$$L[P(Z^*_{k,i} = 1)] = \gamma_0 + \gamma_1 Z_{k,i} + \xi_1 X_{1,i} + \ldots + \xi_q X_{q,i} \quad (7)$$

$$L[P(Z^*_{k,i} = 1)] = \gamma_0 + \gamma_1 Z_{k,i} + \delta f(Z_{1i}, \ldots, Z_{Ki}) \quad (8)$$

$$L[P(Z^*_{k,i} = 1)] = \gamma_0 + \gamma_1 Z_{k,i} + \gamma_2 T_{2,ki} + \ldots + \gamma_K T_{K,ki} \quad (9)$$

where $L(t) = \ln(t/(1-t))$ is the logit, $i = 1, \ldots, N$ and $k = 1, \ldots, K$ and the dependence on the covariates is omitted from the expression for clarity. The true and possibly misclassified counts are obtained by making the sums $\sum_{k=1}^{K} Z_{k,i}$ and $\sum_{k=1}^{K} Z^*_{k,i}$, where $K = 8$ has been taken.

Two values for the prevalence p_M at mouth level were considered, i.e. $p_M = 0.10$ and $p_M = 0.30$. This implies a prevalence at tooth level equal to $p_T = 1 - (1 - p_M)^{(1/K)}$ yielding $p_T = 0.013$ and $p_T = 0.043$, respectively.

Model $E0$, i.e. the basic DB approach, corresponds to model (6) with $\gamma_0 = -2.94$ and $\gamma_1 = 5.14$, so we obtain $\alpha_Z = 0.90$ and $\beta_Z = 0.95$.

In extension $E1$, see equation (7), we have taken $q = 2$. The regression vector ξ is taken equal to $(0.7, -0.3, 0.4)^T$ for a minimal variation of the sensitivities and specificities and equal to $(1.5, -0.6, 0.9)^T$ for a moderate variation. Further, for the two covariates we assume that $X_1 \sim N(3, 0.2)$ and $X_2 \sim Bernoulli(0.6)$. These two distributions were inspired by the covariates considered in the Signal-Tandmobiel® study (namely, "age at start brushing"

and "gender"). α_Z and β_Z vary over the subjects and hence assumption $A3$ is violated here.

Extension $E2$, see equation (8), expresses that α_Z and β_Z depend on the values of $Z_1, \ldots, Z_K$ and thus assumption $A1$ is violated now. More specifically, we have taken $f(Z_1, \ldots, Z_K) = \sum_{k=1}^{K} Z_k$. For a minimal variation of the sensitivities and specificities we have taken $\delta = 0.1$ and for a moderate variation $\delta = 0.2$.

In extension $E3$, assumption $A1$ is relaxed by allowing the α_Z and β_Z to have a distribution varying by subject. For a minimal variation of the sensitivities and specificities we take $\boldsymbol{\mu}_{\boldsymbol{b}}' = (\text{logit}(\alpha_Z), \text{logit}(\beta_Z))$, e.g. $\boldsymbol{\mu}_{\boldsymbol{b}}' = (2.197, 2.944)$ for $\alpha_Z = 0.90$ and $\beta_Z = 0.95$, and

$$\Sigma_{\boldsymbol{b}} = \begin{pmatrix} 0.050 & 0.035 \\ 0.035 & 0.030 \end{pmatrix}.$$

For a moderate variation the values of $\Sigma_{\boldsymbol{b}}$ are doubled.

Finally, in extension $E4$, see equation (9), the variables $T_{j,k} = 1$ if $j = k$ and 0 otherwise for $j, k = 1, \ldots, K$. They express the fact that the α_k and β_k differ over k and hence that assumption $A2$ is violated. For a small variation, we have taken for $K = 6$, $\gamma = (-0.3, -0.3, 0.3, 0.3, 0.3)^T$ and for $K = 8$ we have taken $\gamma = (-0.3, -0.3, -0.3, 0.3, 0.3, 0.3, 0.3)^T$. For a moderate variation, these values are doubled.

We have assumed that the true binary scores Z_k are independent ($\rho = 0$) and as well as that they are related ($\rho = 0.7$). However, since the simulation results for the two values of ρ are similar, we report only the results for $\rho = 0.7$. To generate the correlated binary scores we used the method of Dunn and Davies (1998).

For the models Ex ($x = 0, \ldots, 4$) and the multinomial model, we calculated for the estimated α_W and β_W: the average, the median, the SD, the Mean Squared Error (MSE) and the 95% range. We calculated also the estimates for all elements of the misclassification matrix, i.e. π_{rs}, but reporting all of these results would be overwhelming. Alternatively, we calculated the discrepancy measure $D = \sum_{s=0}^{K} \sum_{r=0}^{K} (\pi_{rs} - p_{rs})^2/(K+1)$, where π_{rs} is the true misclassification probability and $p_{rs} = \hat{\pi}_{rs}, \overline{\pi}_{rs}, \tilde{\pi}_{rs}$, respectively. A chi-square type of statistic is possible too but it would give too much weight on the small (and unimportant) true misclassification probabilities and is therefore not reported here.

To determine the true misclassification probabilities and consequently the value of α_W and β_W an approximative method was used. Namely, we approximate the true misclassification probabilities using the multinomial method determined on a validation study of size $200,000$.

We report below only the case of moderate variability together with $K = 8$ since these results are sufficient to deliver the message.

5.1.2 Second simulation study In the second simulation study, we examined the effect of the basic DB approach and its extensions on the estimation of the main model parameters and compared their performance to the multinomial approach. For all cases, the size of the main study is 1000 and 1000 simulations were performed for each scenario. We considered as measurement model a binomial regression model given by $Y \sim \text{Binom}(K, p_Y)$ with $\text{logit}(p_Y) = \beta_0 + \beta_X X + \beta_Z Z$, where X is a binary covariate with success probability p_X and Z is an independent continuous normal variate with

mean 0 and standard deviation SD_Z. We have varied the values of p_X and SD_Z. This was done to examine the effect of the precision with which the regression coefficient is estimated in the model without scoring errors on the relative gain of the DB approach. For instance, when SD_Z is large it is known that the regression coefficient of Z is estimated with more precision than when SD_Z is small. Therefore, it is expected that the gain of the DB approach will be better seen for a relatively large value of SD_Z.

The size of the validation study was fixed at 100. Further, we sampled the validation data such that :

- there is *equal* probability for scoring $Y = s$, i.e. $P(Y = s) = 1/(K+1)$;
- there is *unequal* probability for $Y = s$, i.e. $P(Y = s) = [2(K + 1 - s)]/[(K + 1)(K + 2)]$ which is decreasing in s;
- the validation study is a random sample of the main study.

Sampling and estimation was done under the different DB approaches. More specifically, we sampled from extension Ex and estimated the parameters with $E0$ and Ex. Thus, when sampling was done under extension Ex, estimation was done under the same model.

5.2 *Simulation results*

5.2.1 First simulation study In Table 2 the simulation results are shown for the sensitivity (α_W) and in Table 3 the simulation results for the specificity (β_W) are given. More specifically, we show the estimated sensitivity and specificity for the two values of the prevalence when estimated with the basic DB approach and the multinomial approach when sampling is done under the models Ex ($x = 0, \ldots, 4$). We observe that in all cases α_W and β_W are

estimated unbiasedly for both approaches. For a low prevalence the specificity is estimated with less variability than the sensitivity, while the reverse is true for the higher prevalence. In all cases the variability in estimation is lower for the DB approach than with the multinomial approach.

[Table 2 about here.]

[Table 3 about here.]

The discrepancy measure D of the misclassification probabilities is given in Table 4. More specifically, the measure D is calculated for each simulated scenario and descriptive statistics over the 1000 simulations are reported. We observe that again the basic DB approach is the winner under all scenarios, with the most important gain for the higher value of the prevalence.

[Table 4 about here.]

5.2.2 Second simulation study Tables 5 and 6 show the simulation results for binomial regression with equal and unequal probability, respectively, of scoring $Y = 0, 1, \cdots, K$ in the validation data.

The simulation results indicate that when SD_Z is relatively low, i.e. when the precision of estimating the true regression coefficient in the data set without scoring errors is relatively low, then the DB approaches are roughly equivalent (although practically always better than) to the multinomial approach. In contrast, when the precision of estimating the true regression is high, there is much gain in DB approaches as compared to multinomial method. Further, the multinomial method shows a serious bias in estimating the regression coefficients when SD_Z is high.

We also observe that the variability with which the parameters are estimated with the DB approach does not depend on the marginal probability distribution of Y. The same seems to be true for the multinomial approach. However, the latter approach clearly suffers from computational difficulties when the marginal probability of Y is not uniform. The results are sometimes dramatic when the validation study is a random sample of the main study, see Table 7.

Finally, we observe that when the correct extension is used the performance of the DB approach is best. That is, when sampling is done under Ex and estimation is done under Ex, then the MSE is the lowest. However, the performance of the basic DB approach is relatively close to the extension, certainly in view of its difference with the multinomial approach.

[Table 5 about here.]

[Table 6 about here.]

[Table 7 about here.]

6. Correcting for misclassification in the main study

6.1 *Correcting for misclassification in a regression model*

Interest lies in relating Y to covariates, but if Y^* is observed instead, then the relationship will be distorted. The relationship between the model for Y^* and the model for Y, assuming non-differential misclassification, is:

$$P(Y^* = r|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\theta}; A) = \sum_{s=0}^{K} P(Y = s|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\theta})\pi_{rs}, \qquad (10)$$

where $\boldsymbol{\theta}$ contains the vector of regression coefficients and model parameters relating the true counts to the regressors, and $A \equiv (\pi_{rs})_{rs}$. In some simple cases, a closure property under misclassification holds. Indeed, when $P(Y = s|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\theta})$ is a (random effects) binomial response model, one can show that combined with a DB misclassification model $P(Y^* = r|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\theta}; A)$ is again a (random effects) binomial response model, see e.g. Paulino et al. (2003) and Paulino et al. (2005). However, this property does not hold in general and one needs expression (10) to fit more complex models as encountered in our caries example.

Suppose that there are m observations in the main data, i.e. $\{Y_1^*, \ldots, Y_m^*\}$, and that an extra n pairs of observations $\{(Y_i^*, Y_i), i = (m+1), \ldots, (m+n)\}$ constitute the validation data set, either being a random subsample from the main data or sampled to increase the efficiency in unbiasedly estimating the misclassification probabilities π_{rs} $(r, s = 0, \ldots, K)$. The estimated probabilities $\hat{\pi}_{rs}$ (multinomial) or $\tilde{\pi}_{rs}$ $(r, s = 0, \ldots, K)$ (DB approach) are imputed in equation (10), to estimate the parameter vector $\boldsymbol{\theta}$ using, e.g. a maximum likelihood procedure, yielding $\hat{\boldsymbol{\theta}}$ or $\tilde{\boldsymbol{\theta}}$, respectively.

6.2 *Variability of the corrected estimates*

For a likelihood-based method, the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ can be derived from the second order derivatives of the log-likelihood at the final iteration, where the likelihood is derived from expression (10) replacing the unknown misclassification probabilities by their estimates obtained from the validation study. However, this approach does not take the sampling variability of $\hat{\pi}_{rs}$ $(r, s = 0, \ldots, K)$ into account. The same remark applies to $\tilde{\boldsymbol{\theta}}$.

The total likelihood, combining the main data and the validation data, is given by (the dependence on covariates is omitted for convenience):

$$\prod_{i=1}^{m} P(Y_i^*|\boldsymbol{\theta}, A) \prod_{i=m+1}^{m+n} P(Y_i^*, Y_i|\boldsymbol{\phi}, A), \tag{11}$$

where the first term is obtained from (10). Further, the second term splits up in the products $\prod_{i=m+1}^{m+n} P(Y_i^*|Y_i; A)$ and $\prod_{i=m+1}^{m+n} P(Y_i|\boldsymbol{\phi})$. The first product represents the misclassification probabilities and the second product pertains to the true counts. When the validation study is a random sample of the main study, $\boldsymbol{\phi} \equiv \boldsymbol{\theta}$.

The second derivative matrix at the ML estimate of $\boldsymbol{\theta}$, obtained from likelihood (11), yields the asymptotic covariance matrix of the estimate of $\boldsymbol{\theta}$ taking the variability into account with which π_{rs} is estimated. This approach has been implemented by Mwalili et al. (2005b). For a non-likelihood approach, such as the MC-SIMEX method of Küchenhoff et al. (2006), a bootstrap procedure could be used. Finally, a Bayesian approach (see e.g. Mwalili et al., 2005a and Mwalili et al., 2005b) together with flexible software such as WinBUGS (Spiegelhalter et al., 1996), can take the variability of the correction terms into account in an elegant manner. More specifically, the parameters determining the sensitivity (α) and specificity (β) in the validation data set are sampled in parallel with the parameters of the main model ($\boldsymbol{\theta}$). The posterior distribution of the parameters $\boldsymbol{\theta}$ will then automatically take into account the uncertainty with which the main model parameters have been estimated. This procedure is described in more detail in Mwalili et al. (2005a). However, now the WinBUGS Development Interface (WBDev) was used, which enables the implementation of user defined functions

into the WinBUGS system via compiled Pascal code. The WBDev code that was used to compute the zero-inflated beta-binomial (ZIBB) distribution can be obtained from

`http://med.kuleuven.be/biostat/software/software.htm`

7. Application to the Signal-Tandmobiel® study

7.1 *The Signal-Tandmobiel® study*

The Signal-Tandmobiel® survey is a longitudinal oral health study involving 4468 children conducted in Flanders (Belgium). Detailed dental data were collected on oral hygiene level using established criteria. Questionnaire data were also obtained from the parents on the dietary and brushing behavior of their child. The children were examined annually for a period of six years (1996-2001) and at entry their average age was 7.1 years ($sd = 0.4$). Here we will look at caries experience in the first year of the study. For more details on this study, we refer to Vanobbergen et al. (2000).

A popular measure for caries experience is the dmft-score, which is the sum of the number of decayed (d), missing due to caries (m) and filled (f) deciduous teeth with 0 as minimal value (no caries experience) and 20 as maximal value (all teeth affected). Here, we look at the dmft-score restricted to the 8 deciduous molars (teeth x4 and x5, with x=5,6,7,8), denoted by $\text{dmft}_{4,5}$-score. We are interested in the effect of dietary and brushing behavior on the $\text{dmft}_{4,5}$-score.

The regression model expressing $\text{dmft}_{4,5}$-score as a function of the dietary and brushing behavior covariates is based on 3303 children with no-missing information on these covariates. 96.2% of these children had 8 deciduous mo-

lars, for the remaining children one or more deciduous molars were replaced by their corresponding permanent molars. We regarded such a deciduous molar as sound.

7.2 *Analysis of the validation study*

In the Signal-Tandmobiel® study sixteen dental examiners were involved. Three calibration exercises for scoring caries experience were organized and at the end of each exercise the scoring behavior of the dental examiners was compared to that of a gold standard (last author) yielding each time a misclassification table for each examiner. Here we look at the validation data of the first calibration exercise.

[Table 8 about here.]

[Table 9 about here.]

The validation study was based on ninety-two children, but the children were not sampled at random from the main study. Rather, a school was selected (and all seven-year old children examined) where a relatively high prevalence for caries experience could be expected. Although the validation study is not internal, since the children belong to the same population as those of the main study and the dental examiners are also the same, the misclassification probabilities can be unbiasedly estimated using the validation data. But, for these children no questionnaire data were available so their true scores could not be included in likelihood (11). In this analysis we will pool over the examiners. In Table 8 the observed misclassification table for $\text{dmft}_{4,5}$ with respect to the gold standard is given. Clearly, this is a very sparse table. In Table 9 scoring caries experience by the dental examiners on

tooth level is compared with the scores of the gold standard. From this table we can obtain $\hat{\alpha}_Z = 133/154 = 0.86$ and $\hat{\beta}_Z = 418/433 = 0.97$. These values are plugged in (2) together with $K = 8$ and yields the estimated misclassification probabilities for the basic DB method. Combined with the marginal totals from Table 8 the estimated frequencies of misclassification are obtained and compared to those of Table 8. The fitted table (not shown) indicates that the expected frequency of cell $(0, 0)$ under the basic DB method is 26 which is about 25% lower than the observed frequency.

[Figure 2 about here.]

In Figure 2 the observed tooth-specific specificities and sensitivities are plotted as a function of the $\text{dmft}_{4,5}$-index. From this figure there is some evidence that the specificity and the sensitivity depend on the actual value of the $\text{dmft}_{4,5}$-index, namely they are higher for $\text{dmft}_{4,5} = 0$ and 1. A possible explanation for the dependence of sensitivity and specificity on $\text{dmft}_{4,5}$, is that when there is (almost) no caries experience in the mouth caries might be easier to distinguish from no-caries, while in a mouth with considerable caries experience the dental examiner might be distracted somewhat easier. We fitted a logistic regression model predicting the scoring behavior of the examiners as a function of the true score and $\text{dmft}_{4,5}$:

$$\begin{aligned} \text{logit}(\alpha_Z) &= 2.6 - 0.14 \times \text{dmft}_{4,5}, \\ \text{logit}(\beta_Z) &= 4.1 - 0.45 \times \text{dmft}_{4,5}, \end{aligned} \tag{12}$$

where we have omitted the subscript k in $\text{dmft}_{4,5}$ for convenience. Model (12) corresponds to extension $E2$. The regression coefficient (SE) indicates

that the specificity decreases with $\text{dmft}_{4,5}$, but the negative dependence of the sensitivity on $\text{dmft}_{4,5}$ is not so pronounced. Hence, we used model (12) to provide the correction terms for the main model.

[Table 10 about here.]

7.3 *Analysis of the main study*

Since the response $\text{dmft}_{4,5} \equiv Y = Z_1 + \ldots + Z_K$ where $K = 8$ is a finite count, an obvious candidate for the distribution of Y is the binomial distribution. However, the Z_k $(k = 1, \ldots, K)$ are correlated. In that case, the beta-binomial distribution is a possible choice, given by Prentice (1986):

$$P(Y = y) \equiv BB(y|N, \pi, \tau) = \frac{\left(\begin{array}{c} K \\ y \end{array} \right) \prod_{h=0}^{y-1} (\pi + \tau h) \prod_{h=0}^{K-y-1} (1 - \pi + \tau h)}{\prod_{h=0}^{K-1} (1 + \tau h)}, \quad (13)$$

with mean $K\pi$ and variance $K\pi(1-\pi)[1 + (K-1)\delta]$, where $\delta = \tau/(1+\tau)$. In the Signal-Tandmobiel® study, there is an excess of zeros. Therefore, we have chosen for the zero-inflated beta-binomial (ZIBB) model assuming for the distribution of the count a mixture of a beta-binomial distribution and a point mass at zero. The ZIBB distribution is given by

$$P(Y = y) = \begin{cases} p + (1-p)BB(0|N, \pi, \tau) & \text{if } y = 0; \\ (1-p)BB(y|N, \pi, \tau) & \text{if } y > 0. \end{cases} \quad (14)$$

The ZIBB regression model relates the parameters π and p of the ZIBB distribution to covariates as follows:

$$\text{logit}(\pi_i) = \boldsymbol{x}_i'\boldsymbol{\beta} \text{ and } \text{logit}(p_i) = \boldsymbol{z}_i'\boldsymbol{\gamma}, \ (i = 1, \cdots, m) \quad (15)$$

where $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ are d- and q-dimensional vectors of covariates pertaining to the ith subject, and with $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ the corresponding vector of regression coefficients, respectively.

7.3.1 Estimating the true prevalence The observed prevalence is 55%. The corrected estimate of prevalence (and 95% credibility interval) assuming a ZIBB distribution for $\text{dmft}_{4,5}$ are: (a) using the multinomial method: 60.1% (54.2 – 66.7) and (b) using extension $E2$: 59.4% (55.5 – 64.4). The estimated prevalence from both methods are quite close, but the estimate of $E2$ presents less variability than the multinomial estimate.

7.3.2 Correcting for misclassification when regressing caries experience on covariates We now consider the impact of age, gender, dietary and reported brushing behavior on $\text{dmft}_{4,5}$. As dietary covariates we consider the binary variables 'intake of sugar-containing drinks' (yes = 1) and 'snacks in-between meals' ($> 2 = 1$). As brushing covariates 'frequency of brushing' ($< 2/day = 1$) and 'age at start brushing' are taken. Additionally, the use of systemic fluorides (yes=1) was taken in the model. In Table 11 we show the fitted ZIBB regression model without correction. The positive regression coefficient for age in the beta-binomial part implies that the older the children the higher the degree of caries experience. In addition, the later the children start brushing the higher the degree of caries experience. Further, the use of systematic fluoride supplements lowers the degree of caries experience whereas the consumption of sugar containing drinks and intake of in-between-meals increases the risk of caries experience. For the degenerate

part (explaining 0) only age at start of brushing was important. However, we omitted this variable from the model to enhance the comparability between this model and the two fitted models corrected for misclassification (see below).

[Table 11 about here.]

Table 12 shows the corrected ZIBB regression models based on the two approaches. First, observe that the fitted value for p is much lower for extension $E2$ than for the multinomial approach. This implied that no covariates turned out to be important to predict the extra-zeros in the DB approach and hence we dropped covariate age at start of brushing in the degenerate part from all models. For both correction methods the regression coefficients increased in absolute value, with the largest increase for the multinomial correction method. On the other hand, the 95% credibility interval increased in size with respect to the uncorrected model. But, the median increase in size is about 1.5 for extension $E2$, while for the multinomial correction this increase is about 2.5.

[Table 12 about here.]

8. Discussion

Correction for misclassification can only work efficiently if the correction terms are estimated with high precision. This necessitates that the validation study is large enough. To increase the efficiency with which the misclassification probabilities are estimated some modelling of these probabilities seems necessary. In this paper, we have suggested to describe the misclassification process in a simple statistical way by the double binomial method.

The gain in efficiency but also the decrease in bias, compared to multinomial modelling, can be large if certain assumptions are roughly satisfied. We admit that it is not clear how often these assumptions will hold in practice. But from the simulations we can conclude that moderate violation of the assumptions $A1$ seems of less importance. Further, we have shown that our approach can easily be extended when the assumptions do not hold.

In a series of papers (Espeland and Odoroff, 1985; Espeland, 1986; Espeland and Hui, 1987), Espeland and coworkers suggested to fit categorical data in the presence of misclassification errors using a general class of models incorporating log-linear models. Their approach also allows easily calculation of variance estimates of the main model parameters, taking into account the sampling variability of the correction terms. Thus, when the response and the covariates are categorical, i.e. frequency tables are modelled, their approach is an elegant way to deal with misclassification errors. In contrast, the DB approach is a particular misclassification model for finite count data that yields a misclassification matrix, which can be used to correct for scoring errors in the count used as a response of a regression model or as a regressor. Thus the DB approach involves only a misclassification model combined in a second step with the measurement model in a general manner. Therefore, the DB model can be combined with the measurement model in a likelihood way or in combination with a structural approach, e.g. the MC-SIMEX (Küchenhoff et al., 2006). The approach of Espeland integrates the measurement and misclassification model into one model. However, the latter approach does not exploit the nature of the count (and was not developed for it) and therefore might suffer from similar problems as the multinomial

model, e.g. if the true count has not been observed in the validation study computational problems will occur.

Acknowledgements

The third author is supported by Research Grants OT/00/35 and OT/05/60, Catholic University Leuven. The first and fourth author acknowledge partial support from Research Grants OT/00/35 and OT/00/60, Catholic University Leuven. Data collection was supported by Unilever, Belgium. The Signal-Tandmobiel® project comprises following partners: D. Declerck (Dental School, Catholic University Leuven), L. Martens (Dental School, University Ghent), J. Vanobbergen (Oral Health Promotion and Prevention, Flemish Dental Association), P. Bottenberg (Dental School, University Brussels), E. Lesaffre (Biostatistical Centre, Catholic University Leuven) and K. Hoppenbrouwers (Youth Health Department, Catholic University Leuven; Flemish Association for Youth Health Care). We also acknowledge support from the Interuniversity Attraction Poles Program P5/24 − Belgian State − Federal Office for Scientific, Technical and Cultural Affairs.

References

Albert, P. S., Hunsberger, S, A. and Biro, F. M. (1997). Modeling repeated measures with monotonic ordinal responses and misclassification, with applications to studying maturation. *Journal of the American Statistical Association* **92**, 1304–1311.

Begg, C. B. (1987). Biases in the assessment of diagnostic tests. *Statistics in Medicine* **6**, 411–423.

Dunn, A. D. and Davies, S. J. (1998). A note on generating correlated binary variables. *Biometrika* **85**, 487 – 490.

Espeland, M. A. (1986). A general class of models for discrete multivariate data. *Commun. Statist.-Simula* **15**, 405–424.

Espeland, M. A. and Hui, S. L. (1987). A general approach to analyzing epidemiological data that contain misclassification errors. *Biometrics* **43**, 1001–1012.

Espeland, M. A. and Odoroff, C. L. (1985). Log-linear models for doubly sampled categorical data fitted by the EM algorithm. *Journal of the American Statistical Association* **80**, 663–670.

Gustafson, P. (2004). *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments*. Chapman & Hall, New York.

Haitovsky, Y. and Rapp, J. (1992). Conditional resampling for misclassified multinomial data with applications to sampling inspection. *Technometrics* **34**, 473–483.

Hujoel, P. P., Moulton, L. H. and Loesche, W. J. (1990). Estimation of sensi-

tivity and specificity of site-specific diagnostic tests. *Journal of Peridontal Research* **25**, 193–196.

Küchenhoff, H., Mwalili, S. and Lesaffre, E. (2006). A general method for dealing with misclassification in regression: the Misclassification SIMEX. *(Biometrics)* **62**, 85–96.

Lesaffre, E., Küchenhoff, H., Mwalili, S. and Declerck, D. (2006). On the estimation of the misclassification table for finite count data with an application in caries research. *Report TR0622* .

Luan, X., Pan, W., Gerberich, S. and Carlin, B. (2005). Does it always help to adjust for misclassification of a binary outcome in logistic regression? *Statistics in Medicine* **24**, 2221–2234.

Mwalili, S., Lesaffre, E. and Declerck, D. (2005a). A Bayesian ordinal logistic regression model to correct for inter-observer measurement error in a geographical oral health study**1**, 77–93.

Mwalili, S., Lesaffre, E. and Declerck, D. (2005b). The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research. *(accepted in Statistical Methods in Medical Research)* .

Paulino, C., Silva, G. and Achcar, J. (2005). Bayesian analysis of correlated misclassified binary data. *Computational Statistics and Data Analysis* **49**, 1120–1131.

Paulino, C., Soares, P. and Neuhaus, J. (2003). Binomial regression with misclassification. *Biometrics* **59**, 670–675.

Prentice, R. L. (1986). Binary regression using an extended beta-binomial distribution, with discussion of correlation included by covariate measure-

ment errors. *Journal of the American Statistical Association* **81**, 321 – 327.

Rudolfer, S. M. (1990). A Markov Chain model of extrabinomial variation. *Biometrika* **77**, 255 – 264.

Spiegelhalter, D., Thomas, A., Best, N. and Gilks, W. (1996). *Bayesian Inference Using Gibbs Sampling Manual (version ii).* Cambridge, UK.

Vanobbergen, J., Martens, L., Lesaffre, E. and Declerck, D. (2000). The Signal-Tandmobiel® project – a longitudinal intervention health promotion study in Flanders (Belgium): baseline and first year results. *European Journal of Paediatric Dentistry* **2**, 87–96.

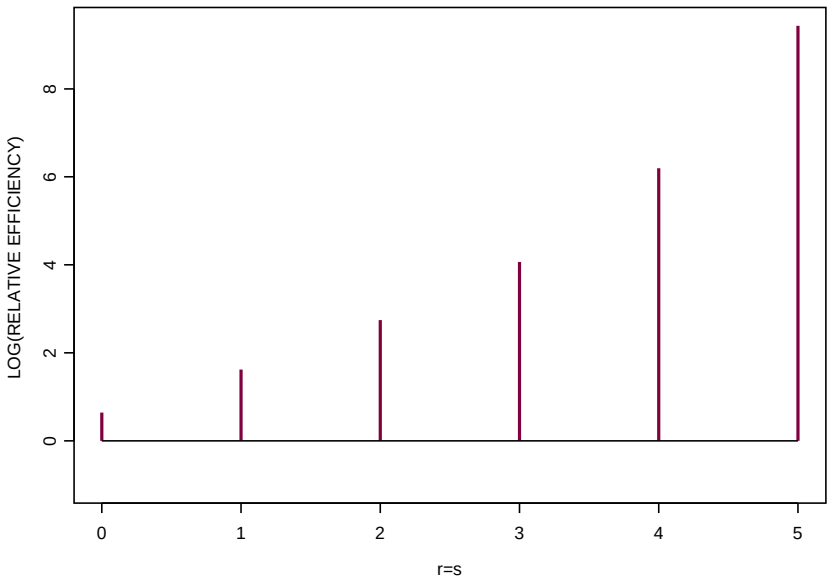


Figure 1. Log(Relative efficiency) $\log(\text{ARE})_{rs}$ with $r = s$ for $K = 5$, $\alpha_Z = 0.90$, $\beta_Z = 0.90$ and prevalence equal to 0.10

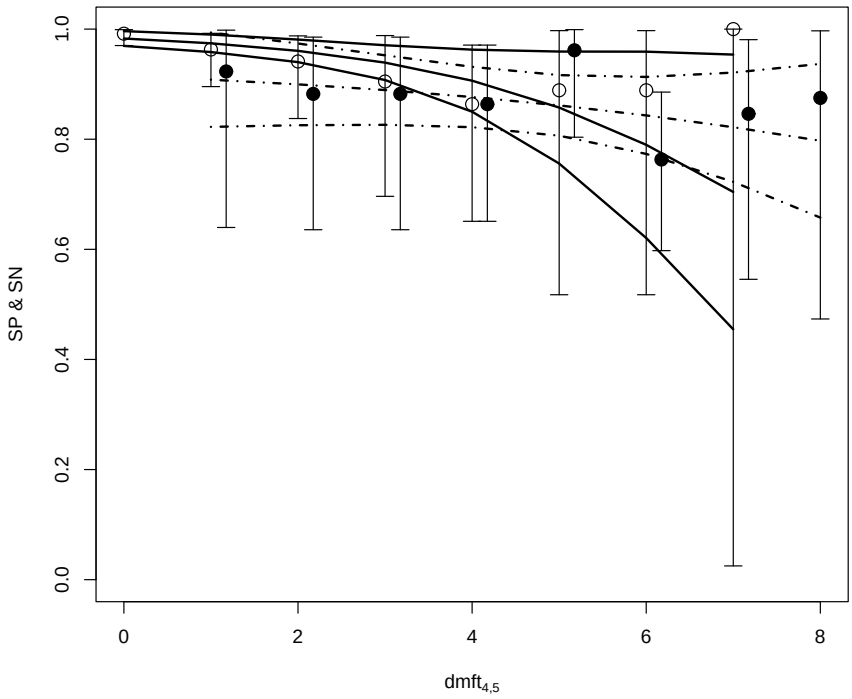


Figure 2. Signal-Tandmobiel® Study: Tooth-specific specificity and sensitivity (+ 95% CI) as a function of $\text{dmft}_{4,5}$. Open (filled) circles represent the observed specificities (sensitivities). The lines correspond to fitted values from model (12), the solid lines represent specificity and the dashed lines sensitivity. The inner lines represent the average curve, the outer lines show the 95% pointwise boundary values

Table 1

Relative efficiency as a function of K, the prevalence and the sensitivity and specificity of Z_k

				Relative efficiency	
K	Prevalence	Sensitivity	Specificity	Sensitivity	Specificity
5	0.10	0.90	0.90	3.34	1.90
10	0.10	0.90	0.90	5.56	4.34
20	0.10	0.90	0.90	26.23	24.07
5	0.20	0.90	0.90	3.35	3.05
10	0.20	0.90	0.90	12.79	12.53
20	0.20	0.90	0.99	226.59	225.61
5	0.30	0.90	0.90	5.36	5.20
10	0.30	0.90	0.90	41.91	41.66
20	0.30	0.90	0.90	2845.58	2852.35
5	0.10	0.95	0.90	2.16	1.90
10	0.10	0.95	0.90	4.49	4.34
20	0.10	0.95	0.90	24.43	24.07
5	0.20	0.95	0.90	3.14	3.05
10	0.20	0.95	0.90	12.67	12.53
20	0.20	0.95	0.99	226.59	225.61
5	0.30	0.95	0.90	5.27	5.20
10	0.30	0.95	0.90	41.87	41.66
20	0.30	0.95	0.90	2856.67	2852.35
5	0.10	0.90	0.95	3.20	1.69
10	0.10	0.90	0.95	4.20	3.29
20	0.10	0.90	0.95	13.27	12.58
5	0.20	0.90	0.95	3.03	2.71
10	0.20	0.90	0.95	9.68	9.49
20	0.20	0.90	0.95	118.33	117.96
5	0.30	0.90	0.95	4.80	4.63
10	0.30	0.90	0.95	31.73	31.55
20	0.30	0.90	0.95	1487.43	1491.40
5	0.10	0.95	0.95	1.97	1.69
10	0.10	0.95	0.95	3.40	3.29
20	0.10	0.95	0.95	12.74	12.58
5	0.20	0.95	0.95	2.80	2.71
10	0.20	0.95	0.95	9.59	9.49
20	0.20	0.95	0.95	118.41	117.96
5	0.30	0.95	0.95	4.69	4.63
10	0.30	0.95	0.95	31.70	31.55
20	0.30	0.95	0.95	1493.05	1491.40

Table 2

Simulation results for sensitivity: K= 8, N = 100 with moderately varying sensitivity α_k and specificity β_k around 90%, 95% respectively for each of the four extensions (E0 corresponds to the basic DB approach). Prev represents the prevalence at mouth level. All values are expressed in percentages.

			Double binomial (E0)		Multinomial		
Ext	Prev	SN†	Mean(Median)/ 95% Range	SD/ MSE	Mean(Median)/ 95% Range	SD/ MSE	RE‡
E0	10	93.3	93.3(94.1)	6.59	93.5(100.0)	8.36	
			[78.0, 100.0]	43.4	[72.7, 100.0]	69.8	161
E1	10	94.1	94.5(95.1)	5.95	94.2(100.0)	7.75	
			[80.5, 100.0]	35.5	[75.0, 100.0]	60.1	169
E2	10	92.5	91.8(92.8)	7.54	91.9(92.3)	9.11	
			[74.3, 100.0]	57.4	[70.0, 100.0]	83.2	145
E3	10	93.3	93.3(94.0)	6.45	93.4(95.0)	8.03	
			[77.9, 100.0]	41.6	[75.0, 100.0]	64.5	155
E4	10	92.8	92.5(93.3)	7.45	92.8(93.8)	8.83	
			[75.0, 100.0]	55.6	[70.0, 100.0]	77.9	140
E0	30	94.0	94.0(94.4)	3.30	94.0(94.3)	4.42	
			[86.7, 100.0]	10.9	[83.3, 100.0]	19.6	180
E1	30	94.8	95.0(95.3)	2.86	94.8(95.8)	4.12	
			[88.2, 100.0]	8.2	[85.2, 100.0]	16.9	206
E2	30	93.1	93.0(93.4)	3.70	93.1(93.3)	4.73	
			[85.1, 98.5]	13.7	[82.9, 100.0]	22.3	163
E3	30	94.1	93.8(94.0)	3.12	94.0(94.2)	4.32	
			[86.9, 98.9]	9.8	[84.0, 100.0]	18.7	191
E4	30	93.4	92.6(92.8)	3.58	93.5(93.9)	4.53	
			[85.2, 98.6]	13.4	[83.3, 100.0]	20.5	153

†SN = true sensitivity at mouth level.

‡RE = $\text{MSE}_{mult}/\text{MSE}_{DB} \times 100$

Table 3

Simulation results for specificity: K= 8, N = 100 with moderately varying sensitivity α_k and specificity β_k around 90%, 95% respectively for each of the four extensions (E0 corresponds to the basic DB approach). Prev represents the prevalence at mouth level. All values are expressed in percentages.

			Double binomial (E0)		Multinomial		
Ext	Prev	SP†	Mean(Median)/ 95% Range	SD/ MSE	Mean(Median)/ 95% Range	SD/ MSE	RE%‡
E0	10	66.1	66.6(66.5)	4.33	66.4(66.3)	4.95	
			[58.6, 75.7]	18.9	[57.0, 76.3]	24.6	130
E1	10	61.2	61.1(61.2)	4.33	61.3(61.4)	5.12	
			[52.7, 69.6]	18.8	[51.6, 71.4]	26.2	139
E2	10	71.6	71.9(71.8)	4.40	72.0(72.2)	4.88	
			[63.2, 80.6]	19.4	[62.5, 81.0]	24.0	123
E3	10	65.7	65.3(65.3)	4.47	65.5(65.6)	4.92	
			[57.0, 73.4]	20.1	[56.2, 74.7]	24.2	120
E4	10	71.6	71.8(71.8)	4.27	71.7(71.8)	4.81	
			[63.7, 80.6]	18.3	[61.9, 81.1]	23.1	126
E0	30	66.5	66.2(66.4)	4.34	66(66.2)	5.82	
			[57.6, 75.0]	18.9	[54.3, 77.0]	34.1	180
E1	30	61.4	60.9(60.9)	4.49	61.4(61.4)	5.82	
			[52.2, 70.1]	20.3	[50.0, 72.5]	33.9	167
E2	30	71.9	71.1(71.1)	4.48	71.7(72.0)	5.54	
			[62.3, 80.0]	20.6	[60.0, 82.6]	30.7	149
E3	30	65.6	65.3(65.0)	4.43	65.5(65.3)	5.58	
			[57.5, 74.5]	19.7	[55.2, 77.1]	31.2	158
E4	30	71.5	71.9(71.8)	4.42	71.6(71.7)	5.57	
			[63.9, 80.9]	19.7	[60.9, 82.4]	31.0	157

†SP = true specificity at mouth level.

‡RE = $\text{MSE}_{mult}/\text{MSE}_{DB} \times 100$

Table 4

Simulation results for misclassification probabilities: estimate (over 1000 simulation samples) of the discrepancy measure $D = \sum\sum(\pi_{rs} - p_{rs})^2/(K+1)$, *where* π_{rs} *is the true and* p_{rs} *the estimated misclassification probability over of the four extensions (*$K = 8$, $N = 100$; *E0 corresponds to the basic DB approach).*

		Double binomial (E0)				Multinomial			
			Quantile				Quantile		
Ext	Prev %	Mean	25%	50%	75%	Mean	25%	50%	75%
E0	10	9.3	7.8	8.4	9.8	17.2	15.0	16.3	17.9
E1	10	12.9	12.2	12.5	13.2	23.9	21.2	22.6	25.3
E2	10	17.5	14.9	16.4	18.9	25.0	22.5	23.9	26.0
E3	10	9.4	8.3	8.7	9.8	18.6	16.0	17.1	19.8
E4	10	9.3	7.6	8.5	10.1	15.1	12.2	14.0	16.4
E0	30	1.1	0.6	0.8	1.2	17.1	14.4	16.2	18.8
E1	30	1.4	1.1	1.2	1.5	18.3	15.3	17.2	20.5
E2	30	1.7	0.9	1.4	2.2	20.3	16.6	19.0	23.2
E3	30	1.3	0.8	1.1	1.5	18.9	16.0	17.8	20.8
E4	30	0.7	0.3	0.5	0.8	16.4	13.4	15.6	19.1

Table 5: Simulation results for binomial regression: K= 8, N = 100 with moderately varying sensitivity α_k and specificity β_k around 90%, 95% respectively for each of the four extensions (E0 corresponds to the basic DB approach). Case of $P(Y = s) = 1/(K + 1)$ in the validation data. p_X is the success probability of the binary regressor, SD_Z is the standard deviation of the normal continuous regressor.

		β_0		β_X		β_Z	
Ext (p_X,SD_Z)		Mean(SD)	MSE	Mean(SD)	MSE	Mean(SD)	MSE
	True	0.000(0.031)	—	−0.998(0.046)	—	0.996(0.238)	—
E0 (0.5, 0.1)	Naive	−0.100(0.031)	0.011	−0.849(0.045)	0.024	0.835(0.242)	0.084
	Mult	0.001(0.072)	0.005	−0.992(0.082)	0.007	0.982(0.289)	0.084
	E0	0.002(0.059)	0.003	−0.999(0.061)	0.004	0.992(0.289)	0.084
	True	0.000(0.039)	—	−0.999(0.051)	—	1.001(0.030)	—
E0 (0.6, 1)	Naive	−0.114(0.039)	0.015	−0.806(0.050)	0.040	0.803(0.028)	0.040
	Mult	−0.005(0.078)	0.006	−0.981(0.082)	0.007	0.982(0.066)	0.005
	E0	0.001(0.072)	0.005	−1.002(0.070)	0.005	1.004(0.050)	0.002
	True	−0.001(0.075)	—	−1.001(0.095)	—	1.000(0.024)	—
E0 (0.7, 5)	Naive	−0.181(0.061)	0.036	−0.475(0.076)	0.282	0.470(0.016)	0.281
	Mult	−0.032(0.117)	0.015	−0.816(0.145)	0.055	0.812(0.107)	0.047
	E0	−0.000(0.121)	0.015	−1.001(0.132)	0.018	0.996(0.072)	0.005
	True	−0.001(0.130)	—	−1.001(0.152)	—	1.002(0.033)	—
E0 (0.8, 10)	Naive	−0.200(0.089)	0.047	−0.273(0.104)	0.541	0.271(0.010)	0.534
	Mult	−0.044(0.153)	0.025	−0.624(0.202)	0.183	0.622(0.135)	0.162
	E0	−0.001(0.192)	0.035	−0.994(0.208)	0.043	0.997(0.086)	0.007
	True	0.001(0.031)	—	−1.002(0.046)	—	1.009(0.236)	—
E1 (0.5, 0.1)	Naive	−0.058(0.031)	0.004	−0.842(0.047)	0.028	0.842(0.238)	0.085
	Mult	−0.002(0.073)	0.005	−0.991(0.085)	0.007	0.994(0.291)	0.085
	E0	0.000(0.059)	0.003	−1.003(0.063)	0.004	1.013(0.289)	0.083
	E1	0.003(0.044)	0.002	−1.004(0.060)	0.004	0.997(0.281)	0.080
	True	−0.000(0.040)	—	−1.000(0.054)	—	1.002(0.031)	—
E1 (0.6, 1)	Naive	−0.066(0.037)	0.006	−0.798(0.051)	0.043	0.794(0.028)	0.044
	Mult	−0.004(0.076)	0.006	−0.978(0.087)	0.008	0.978(0.069)	0.005
	E0	−0.001(0.071)	0.005	−1.003(0.073)	0.005	1.005(0.052)	0.003

Continued on next page

Table 5 – continued from previous page

Ext (p_X,SD_Z)		β_0		β_X		β_Z	
		Mean(SD)	MSE	Mean(SD)	MSE	Mean(SD)	MSE
	E1	0.003(0.049)	0.002	−1.001(0.070)	0.005	0.998(0.040)	0.002
	True	0.001(0.072)	—	−1.001(0.091)	—	1.002(0.025)	—
E1 (0.7, 5)	Naive	−0.105(0.057)	0.015	−0.465(0.070)	0.292	0.464(0.016)	0.289
	Mult	−0.015(0.119)	0.014	−0.811(0.150)	0.058	0.810(0.106)	0.048
	E0	0.004(0.125)	0.016	−1.004(0.140)	0.020	1.002(0.071)	0.005
	E1	0.014(0.124)	0.026	−1.016(0.131)	0.017	0.996(0.066)	0.004
	True	0.000(0.127)	—	−1.004(0.151)	—	1.002(0.033)	—
E1 (0.8, 10)	Naive	−0.118(0.087)	0.021	−0.269(0.099)	0.55	0.268(0.010)	0.538
	Mult	−0.037(0.155)	0.026	−0.594(0.195)	0.205	0.594(0.134)	0.184
	E0	−0.009(0.188)	0.036	−0.992(0.213)	0.045	0.994(0.086)	0.007
	E1	0.006(0.172)	0.029	−1.005(0.198)	0.039	1.006(0.051)	0.003
	True	−0.001(0.032)	—	−0.999(0.047)	—	0.998(0.235)	—
E2 (0.5, 0.1)	Naive	−0.050(0.033)	0.003	−0.897(0.049)	0.013	0.887(0.242)	0.071
	Mult	−0.003(0.065)	0.004	−0.993(0.077)	0.006	0.988(0.280)	0.078
	E0	−0.005(0.055)	0.003	−1.034(0.062)	0.005	1.032(0.283)	0.082
	E2	−0.003(0.037)	0.001	−0.998(0.055)	0.003	0.996(0.277)	0.078
	True	−0.001(0.041)	—	−1.002(0.055)	—	1.002(0.032)	—
E2 (0.6, 1)	Naive	−0.055(0.041)	0.005	−0.865(0.055)	0.022	0.861(0.031)	0.021
	Mult	−0.004(0.070)	0.005	−0.985(0.079)	0.006	0.985(0.061)	0.004
	E0	−0.005(0.069)	0.005	−1.043(0.074)	0.007	1.045(0.051)	0.005
	E2	−0.006(0.046)	0.002	−0.993(0.064)	0.004	0.997(0.038)	0.001
	True	0.002(0.074)	—	−1.004(0.088)	—	1.001(0.025)	—
E2 (0.7, 5)	Naive	−0.082(0.063)	0.011	−0.566(0.076)	0.198	0.562(0.019)	0.194
	Mult	−0.008(0.108)	0.012	−0.866(0.140)	0.039	0.859(0.100)	0.03
	E0	0.006(0.118)	0.014	−1.078(0.129)	0.022	1.072(0.066)	0.009
	E2	0.003(0.090)	0.008	−0.998(0.106)	0.011	1.000(0.039)	0.001
	True	0.001(0.128)	—	−1.004(0.144)	—	1.003(0.032)	—
E2 (0.8, 10)	Naive	−0.094(0.094)	0.018	−0.338(0.105)	0.454	0.339(0.014)	0.442
	Mult	−0.026(0.148)	0.023	−0.680(0.194)	0.142	0.683(0.138)	0.122
	E0	−0.003(0.189)	0.036	−1.075(0.213)	0.050	1.077(0.078)	0.011
	E2	0.014(0.172)	0.030	−1.013(0.192)	0.037	1.00(0.047)	0.002
	True	−0.001(0.032)	—	−0.999(0.048)	—	1.006(0.227)	—

Continued on next page

Table 5 – continued from previous page

Ext (p_X,SD_Z)		β_0 Mean(SD)	β_0 MSE	β_X Mean(SD)	β_X MSE	β_Z Mean(SD)	β_Z MSE
E3 (0.5, 0.1)	Naive	−0.104(0.033)	0.012	−0.843(0.049)	0.026	0.841(0.227)	0.079
	Mult	−0.005(0.071)	0.005	−0.990(0.084)	0.007	0.995(0.282)	0.080
	E0	−0.001(0.059)	0.003	−1.001(0.064)	0.004	1.008(0.274)	0.075
	E3	−0.003(0.037)	0.001	−0.995(0.057)	0.003	1.016(0.271)	0.074
	True	0.001(0.039)	—	−1.002(0.052)	—	1.002(0.031)	—
E3 (0.6, 1)	Naive	−0.117(0.038)	0.015	−0.801(0.050)	0.043	0.796(0.029)	0.043
	Mult	−0.004(0.078)	0.006	−0.980(0.084)	0.008	0.979(0.069)	0.005
	E0	0.001(0.074)	0.005	−1.005(0.074)	0.005	1.006(0.053)	0.003
	E3	0.003(0.049)	0.002	−1.001(0.066)	0.004	1.002(0.041)	0.002
	True	0.005(0.074)	—	−1.005(0.092)	—	1.000(0.024)	—
E3 (0.7, 5)	Naive	−0.183(0.059)	0.039	−0.466(0.073)	0.296	0.462(0.016)	0.290
	Mult	−0.031(0.117)	0.015	−0.806(0.145)	0.061	0.800(0.100)	0.050
	E0	0.002(0.121)	0.015	−1.000(0.135)	0.018	0.994(0.070)	0.005
	E3	0.006(0.095)	0.009	−1.004(0.122)	0.015	1.000(0.041)	0.002
	True	−0.008(0.128)	—	−0.993(0.145)	—	1.005(0.033)	—
E3 (0.8, 10)	Naive	−0.210(0.092)	0.049	−0.257(0.104)	0.553	0.265(0.010)	0.548
	Mult	−0.058(0.156)	0.027	−0.594(0.188)	0.195	0.605(0.129)	0.176
	E0	−0.004(0.195)	0.035	−0.990(0.222)	0.042	1.001(0.087)	0.008
	E3	−0.014(0.249)	0.029	−1.009(0.196)	0.039	1.008(0.056)	0.003
	True	0.001(0.031)	—	−1.001(0.047)	—	0.986(0.239)	—
E4 (0.5, 0.1)	Naive	−0.214(0.032)	0.047	−0.837(0.048)	0.029	0.810(0.240)	0.089
	Mult	−0.006(0.082)	0.007	−0.990(0.090)	0.008	0.968(0.297)	0.088
	E0	−0.017(0.065)	0.005	−1.008(0.064)	0.004	0.988(0.295)	0.087
	E4	0.002(0.041)	0.002	−1.000(0.059)	0.003	0.977(0.283)	0.080
	True	0.003(0.038)	—	−1.003(0.050)	—	1.000(0.030)	—
E4 (0.6, 1)	Naive	−0.231(0.038)	0.056	−0.794(0.051)	0.046	0.786(0.029)	0.046
	Mult	−0.002(0.085)	0.007	−0.986(0.092)	0.009	0.981(0.072)	0.006
	E0	−0.006(0.078)	0.006	−1.021(0.077)	0.006	1.017(0.056)	0.003
	E4	−0.003(0.053)	0.003	−0.994(0.071)	0.005	1.001(0.042)	0.002
	True	0.003(0.075)	—	−1.005(0.091)	—	1.001(0.024)	—
E4 (0.7, 5)	Naive	−0.332(0.060)	0.116	−0.449(0.073)	0.314	0.447(0.015)	0.308
	Mult	−0.046(0.124)	0.018	−0.815(0.143)	0.057	0.810(0.102)	0.047

Continued on next page

Table 5 – continued from previous page

Ext (p_X,SD_Z)		β_0 Mean(SD)	β_0 MSE	β_X Mean(SD)	β_X MSE	β_Z Mean(SD)	β_Z MSE
	E0	0.007(0.131)	0.017	−1.042(0.138)	0.020	1.038(0.073)	0.007
	E4	0.007(0.095)	0.009	−1.008(0.115)	0.013	1.004(0.046)	0.002
	True	0.005(0.125)	—	−1.005(0.142)	—	1.002(0.032)	—
E4 (0.8, 10)	Naive	−0.360(0.089)	0.141	−0.250(0.102)	0.580	0.254(0.010)	0.560
	Mult	−0.087(0.160)	0.034	−0.604(0.197)	0.199	0.610(0.134)	0.172
	E0	0.009(0.221)	0.036	−1.021(0.264)	0.043	1.032(0.083)	0.008
	E4	−0.008(0.179)	0.032	−1.004(0.239)	0.041	1.006(0.059)	0.003

Table 6: Simulation results for binomial regression: K= 8, N = 100 with moderately varying sensitivity α_k and specificity β_k around 90%, 95% respectively for each of the four extensions ((E0 corresponds to the basic DB approach). Case of $P(Y = s)$ descending in s in the validation data. p_X is the success probability of the binary regressor, SD_Z is the standard deviation of the normal continuous regressor.

		β_0		β_X		β_Z	
Ext (p_X,SD_Z)		Mean(SD)	MSE	Mean(SD)	MSE	Mean(SD)	MSE
	True	0.002(0.031)	—	−1.002(0.048)	—	0.989(0.242)	—
E0 (0.5, 0.1)	Naive	−0.098(0.031)	0.011	−0.854(0.047)	0.024	0.836(0.232)	0.077
	Mult (889)†	−0.001(0.075)	0.006	−0.994(0.081)	0.007	0.977(0.283)	0.080
	E0	0.003(0.062)	0.004	−1.005(0.062)	0.004	0.994(0.278)	0.077
	True	−0.001(0.038)	—	−1.000(0.053)	—	1.002(0.031)	—
E0 (0.6, 1)	Naive	−0.116(0.037)	0.015	−0.806(0.051)	0.040	0.803(0.029)	0.040
	Mult (882)†	−0.019(0.080)	0.007	−0.964(0.085)	0.009	0.965(0.069)	0.006
	E0	0.002(0.076)	0.006	−1.000(0.072)	0.005	1.003(0.051)	0.003
	True	0.004(0.070)	—	−1.008(0.091)	—	1.002(0.024)	—
E0 (0.7, 5)	Naive	−0.177(0.060)	0.037	−0.477(0.074)	0.287	0.470(0.015)	0.282
	Mult (869)†	−0.134(0.129)	0.036	−0.697(0.153)	0.120	0.691(0.123)	0.112
	E0	0.001(0.129)	0.017	−1.008(0.139)	0.019	0.999(0.073)	0.005
	True	0.001(0.129)	—	−1.004(0.147)	—	1.002(0.033)	—
E0 (0.8, 10)	Naive	−0.206(0.091)	0.051	−0.268(0.100)	0.552	0.272(0.011)	0.534
	Mult (887)†	−0.185(0.155)	0.058	−0.453(0.181)	0.336	0.455(0.126)	0.316
	E0	0.000(0.195)	0.038	−1.004(0.214)	0.046	0.999(0.092)	0.009
	True	−0.000(0.031)	—	−1.000(0.049)	—	1.003(0.233)	—
E1 (0.5, 0.1)	Naive	−0.059(0.031)	0.004	−0.842(0.049)	0.027	0.833(0.228)	0.081
	Mult (891)†	−0.006(0.071)	0.005	−0.990(0.083)	0.007	0.994(0.276)	0.076
	E0	0.001(0.062)	0.004	−1.004(0.064)	0.004	1.003(0.275)	0.076
	E1	0.001(0.036)	0.001	−1.005(0.058)	0.003	0.974(0.270)	0.073
	True	−0.000(0.038)	—	−1.002(0.051)	—	1.002(0.031)	—
E1 (0.6, 1)	Naive	−0.068(0.037)	0.006	−0.798(0.051)	0.044	0.792(0.029)	0.045
	Mult (888)†	−0.022(0.080)	0.007	−0.964(0.081)	0.008	0.963(0.065)	0.006
	E0	−0.002(0.074)	0.006	−1.002(0.072)	0.005	1.002(0.051)	0.003

Continued on next page

Table 6 – continued from previous page

Ext (p_X,SD_Z)		β_0 Mean(SD)	β_0 MSE	β_X Mean(SD)	β_X MSE	β_Z Mean(SD)	β_Z MSE
	E1	−0.003(0.047)	0.002	−0.998(0.063)	0.004	0.999(0.042)	0.002
	True	0.001(0.077)	—	−1.003(0.093)	—	1.001(0.025)	—
E1 (0.7, 5)	Naive	−0.105(0.062)	0.015	−0.467(0.076)	0.293	0.465(0.015)	0.289
	Mult (885)†	−0.122(0.138)	0.034	−0.678(0.157)	0.130	0.679(0.119)	0.118
	E0	0.006(0.133)	0.018	−1.003(0.140)	0.020	0.999(0.073)	0.005
	E1	−0.009(0.130)	0.017	−0.988(0.133)	0.018	0.987(0.070)	0.005
	True	−0.000(0.127)	—	−1.002(0.146)	—	1.002(0.033)	—
E1 (0.8, 10)	Naive	−0.118(0.091)	0.022	−0.268(0.102)	0.55	0.268(0.010)	0.538
	Mult (884)†	−0.159(0.164)	0.052	−0.452(0.184)	0.337	0.451(0.126)	0.319
	E0	0.010(0.225)	0.047	−1.003(0.242)	0.060	1.003(0.090)	0.008
	E1	−0.003(0.145)	0.021	−0.998(0.166)	0.028	1.004(0.051)	0.003
	True	0.001(0.032)	—	−1.002(0.047)	—	0.997(0.231)	—
E2 (0.5, 0.1)	Naive	−0.047(0.033)	0.003	−0.900(0.048)	0.013	0.889(0.238)	0.068
	Mult (878)†	0.001(0.068)	0.005	−0.994(0.077)	0.006	0.976(0.275)	0.076
	E0	0.023(0.058)	0.004	−1.037(0.061)	0.005	1.033(0.277)	0.078
	E2	−0.001(0.038)	0.001	−0.998(0.054)	0.003	0.989(0.271)	0.073
	True	−0.001(0.037)	—	−0.999(0.052)	—	1.000(0.031)	—
E2 (0.6, 1)	Naive	−0.056(0.038)	0.004	−0.862(0.053)	0.022	0.858(0.031)	0.021
	Mult (876)†	−0.015(0.071)	0.005	−0.973(0.079)	0.007	0.972(0.063)	0.005
	E0	0.023(0.069)	0.005	−1.041(0.073)	0.007	1.043(0.051)	0.004
	E2	−0.007(0.051)	0.003	−0.991(0.065)	0.004	0.994(0.043)	0.003
	True	−0.001(0.075)	—	−1.000(0.093)	—	1.000(0.025)	—
E2 (0.7, 5)	Naive	−0.086(0.062)	0.011	−0.561(0.078)	0.199	0.561(0.018)	0.194
	Mult (882)†	−0.106(0.124)	0.026	−0.736(0.149)	0.092	0.741(0.122)	0.082
	E0	0.035(0.125)	0.017	−1.079(0.136)	0.025	1.078(0.071)	0.011
	E2	0.003(0.093)	0.009	−1.007(0.116)	0.014	1.002(0.041)	0.002
	True	0.003(0.128)	—	−1.002(0.150)	—	1.002(0.031)	—
E2 (0.8, 10)	Naive	−0.096(0.096)	0.019	−0.335(0.106)	0.455	0.339(0.014)	0.440
	Mult (895)†	−0.133(0.159)	0.044	−0.520(0.194)	0.269	0.526(0.152)	0.250
	E0	0.042(0.192)	0.039	−1.080(0.213)	0.052	1.088(0.082)	0.014
	E2	0.001(0.166)	0.027	−1.006(0.185)	0.034	1.003(0.055)	0.003
	True	0.000(0.032)	—	−1.002(0.047)	—	0.998(0.239)	—

Continued on next page

Table 6 – continued from previous page

Ext (p_X,SD_Z)		β_0 Mean(SD)	β_0 MSE	β_X Mean(SD)	β_X MSE	β_Z Mean(SD)	β_Z MSE
E3 (0.5, 0.1)	Naive	−0.102(0.032)	0.012	−0.847(0.048)	0.026	0.838(0.241)	0.084
	Mult (901)†	−0.006(0.075)	0.006	−0.988(0.081)	0.007	0.983(0.286)	0.082
	E0	0.002(0.063)	0.004	−1.005(0.062)	0.004	1.004(0.290)	0.084
	E3	0.001(0.034)	0.001	−1.000(0.056)	0.003	0.992(0.269)	0.072
	True	0.000(0.039)	—	−1.000(0.051)	—	1.001(0.031)	—
E3 (0.6, 1)	Naive	−0.118(0.038)	0.015	−0.799(0.051)	0.043	0.796(0.029)	0.043
	Mult(888)†	−0.023(0.081)	0.007	−0.962(0.086)	0.009	0.964(0.069)	0.006
	E0	0.002(0.077)	0.006	−1.003(0.071)	0.005	1.006(0.054)	0.003
	E3	0.003(0.052)	0.003	−0.996(0.063)	0.004	0.997(0.042)	0.002
	True	−0.001(0.075)	—	−1.003(0.093)	—	1.003(0.025)	—
E3 (0.7, 5)	Naive	−0.189(0.060)	0.039	−0.458(0.072)	0.301	0.463(0.016)	0.292
	Mult(888)†	−0.148(0.131)	0.039	−0.660(0.149)	0.14	0.662(0.120)	0.131
	E0	−0.000(0.131)	0.017	−0.999(0.143)	0.02	1.002(0.077)	0.006
	E3	0.001(0.097)	0.009	−0.998(0.123)	0.015	1.000(0.041)	0.002
	True	−0.002(0.134)	—	−1.000(0.152)	—	1.004(0.032)	—
E3 (0.8, 10)	Naive	−0.201(0.094)	0.049	−0.267(0.105)	0.548	0.265(0.010)	0.546
	Mult (877)†	−0.183(0.157)	0.057	−0.436(0.181)	0.351	0.437(0.116)	0.334
	E0	−0.001(0.201)	0.036	−1.002(0.208)	0.043	0.999(0.090)	0.008
	E3	0.005(0.162)	0.026	−1.011(0.188)	0.036	1.006(0.055)	0.003
	True	−0.001(0.033)	—	−0.999(0.047)	—	0.991(0.237)	—
E4 (0.5, 0.1)	Naive	−0.216(0.034)	0.047	−0.835(0.050)	0.029	0.826(0.245)	0.087
	Mult (880)†	−0.011(0.086)	0.007	−0.986(0.088)	0.008	0.987(0.300)	0.090
	E0	−0.007(0.072)	0.005	−1.005(0.066)	0.004	1.007(0.299)	0.090
	E4	0.004(0.048)	0.002	−1.005(0.063)	0.004	0.982(0.287)	0.083
	True	0.001(0.038)	—	−1.003(0.053)	—	0.999(0.031)	—
E4 (0.6, 1)	Naive	−0.233(0.038)	0.057	−0.793(0.053)	0.047	0.786(0.029)	0.046
	Mult (891)†	−0.024(0.084)	0.008	−0.969(0.087)	0.009	0.964(0.069)	0.006
	E0	0.005(0.082)	0.007	−1.021(0.080)	0.007	1.018(0.054)	0.003
	E4	0.002(0.054)	0.003	−1.004(0.071)	0.005	1.004(0.042)	0.002
	True	−0.005(0.072)	—	−0.995(0.090)	—	1.001(0.025)	—
E4 (0.7, 5)	Naive	−0.337(0.062)	0.114	−0.446(0.076)	0.308	0.447(0.015)	0.307
	Mult (900)†	−0.174(0.142)	0.048	−0.660(0.154)	0.136	0.662(0.121)	0.130

Continued on next page

Table 6 – continued from previous page

Ext (p_X,SD_Z)		β_0		β_X		β_Z	
		Mean(SD)	MSE	Mean(SD)	MSE	Mean(SD)	MSE
	E0	0.022(0.148)	0.023	−1.034(0.150)	0.024	1.038(0.079)	0.008
	E4	0.001(0.103)	0.011	−1.009(0.123)	0.015	1.001(0.041)	0.002
	True	−0.007(0.130)	—	−0.999(0.151)	—	1.003(0.032)	—
E4 (0.8, 10)	Naive	−0.358(0.088)	0.131	−0.252(0.098)	0.567	0.254(0.010)	0.560
	Mult (875)†	−0.235(0.153)	0.075	−0.420(0.176)	0.366	0.417(0.114)	0.356
	E0	0.019(0.226)	0.042	−1.033(0.226)	0.050	1.031(0.081)	0.007
	E4	−0.002(0.212)	0.028	−0.995(0.225)	0.050	1.007(0.058)	0.003

†Number of simulation samples (over 1000) of which the multinomial approach was estimable

Table 7

Simulation results for binomial regression: K= 8, N = 100 with moderately varying sensitivity α_k and specificity β_k around 90%, 95% respectively for E0 (basic DB approach), when the validation data is a random sub-sample of the main data.

		β_0		β_X		β_Z	
Ext (p_X,SD_Z)		Mean(SD)	MSE	Mean(SD)	MSE	Mean(SD)	MSE
	True	0.000(0.030)	—	−1.001(0.046)	—	0.992(0.241)	—
(0.5, 0.1)	Naive	−0.101(0.031)	0.011	−0.851(0.046)	0.025	0.832(0.240)	0.083
	Mult (227)†	−0.016(0.064)	0.004	−0.977(0.074)	0.006	0.934(0.274)	0.079
	E0	−0.001(0.059)	0.003	−1.002(0.061)	0.004	0.989(0.285)	0.081
	True	−0.002(0.038)	—	−0.999(0.051)	—	1.000(0.031)	—
(0.6, 1)	Naive	−0.117(0.037)	0.015	−0.806(0.050)	0.04	0.802(0.030)	0.040
	Mult (887)†	−0.018(0.072)	0.005	−0.964(0.081)	0.008	0.964(0.067)	0.006
	E0	−0.002(0.070)	0.005	−0.999(0.069)	0.005	1.001(0.052)	0.003
	True	0.001(0.076)	—	−1.005(0.091)	—	1.002(0.025)	—
(0.7, 5)	Naive	−0.181(0.062)	0.037	−0.473(0.074)	0.289	0.471(0.015)	0.282
	Mult (931)†	−0.014(0.150)	0.023	−0.908(0.150)	0.032	0.906(0.102)	0.020
	E0	0.003(0.124)	0.015	−1.006(0.132)	0.017	1.003(0.071)	0.005
	True	0.002(0.127)	—	−1.007(0.147)	—	1.002(0.034)	—
(0.8, 10)	Naive	−0.198(0.090)	0.048	−0.279(0.103)	0.541	0.271(0.011)	0.534
	Mult (563)†	−0.005(0.237)	0.056	−0.834(0.247)	0.091	0.822(0.173)	0.063
	E0	0.001(0.182)	0.033	−1.011(0.205)	0.042	1.004(0.082)	0.007

†Number of simulation samples (over 1000) of which the multinomial approach was estimable

Table 8

Signal-Tandmobiel® study: Overall misclassification table for $dmft_{4,5}$, column= gold standard, row= (pool of) dental examiner(s)

	Y									
Y^*	0	1	2	3	4	5	6	7	8	
0	32	1	3	0	0	0	0	0	0	36
1	2	13	2	1	0	0	0	0	0	18
2	0	1	5	2	3	0	1	0	0	12
3	0	0	2	4	1	1	1	0	0	9
4	0	0	0	0	2	1	2	0	0	5
5	0	0	0	0	1	3	1	0	0	5
6	0	0	0	0	0	1	2	2	0	5
7	0	0	0	0	0	0	0	1	1	2
8	0	0	0	0	0	0	0	0	0	0
	34	15	12	7	7	6	7	3	1	92

Table 9

Signal-Tandmobiel® Study: Misclassification table at tooth level, column= gold standard, row= (pool of) dental examiner(s)

	Z		
Z^*	0	1	
0	418	21	439
1	15	133	148
	433	154	587

Table 10

Signal-Tandmobiel® Study: Expected misclassifications (probability of misclassification $\times 100$) of $dmft_{4,5}$ with specificities & sensitivities estimated from model (12). Column = gold standard, row = dental examiner(s)

	Y									
Y^*	0	1	2	3	4	5	6	7	8	
0	30(88)	1(7)	0(1)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	31
1	4(11)	12(78)	2(14)	0(2)	0(1)	0(0)	0(0)	0(0)	0(0)	18
2	0(1)	2(14)	8(68)	1(20)	0(5)	0(1)	0(0)	0(0)	0(0)	11
3	0(0)	0(1)	2(16)	4(59)	2(24)	1(9)	0(3)	0(1)	0(1)	9
4	0(0)	0(0)	0(1)	1(17)	4(50)	2(28)	1(13)	0(7)	0(4)	8
5	0(0)	0(0)	0(0)	0(2)	1(17)	3(43)	2(32)	1(20)	0(14)	7
6	0(0)	0(0)	0(0)	0(0)	0(3)	1(16)	3(37)	1(35)	0(29)	5
7	0(0)	0(0)	0(0)	0(0)	0(0)	0(3)	1(13)	1(30)	1(34)	3
8	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(2)	0(7)	0(18)	0
	34	15	12	6	7	7	7	3	1	92

Table 11

Signal-Tandmobiel® study: Posterior estimates of the ZIBB regression model (without correction) fitted to the $dmft_{4,5}$ (results from WinBUGS (version 1.4)).

Parameter	Mean(SD)	2.5 %	97.5%
Intercept	−1.414(0.116)	−1.647	−1.184
Gender (girl)	0.059(0.063)	−0.062	0.183
Age	0.306(0.075)	0.162	0.455
Freq. Brushing	0.087(0.087)	−0.080	0.253
Age start brush	0.150(0.028)	0.095	0.205
Fluoride suppl.	−0.429(0.064)	−0.551	−0.303
Sugar drinks	0.382(0.067)	0.256	0.517
Between meals	0.160(0.065)	0.032	0.284
p	0.177(0.035)	0.099	0.242
τ	0.582(0.040)	0.506	0.667

Table 12

Signal-Tandmobiel® study: Posterior estimates of the corrected ZIBB regression model fitted to the $dmft_{4,5}$ (results from WinBUGS (version 1.4)).

	Double binomial			Multinomial		
Parameter	Mean(SD)	2.5%	97.5%	Mean(SD)	2.5%	97.5%
Intercept	−1.592(0.143)	−1.871	−1.322	−1.779(0.333)	−2.362	−1.080
Gender (girl)	0.053(0.097)	−0.131	0.255	0.056(0.223)	−0.382	0.515
Age (years)	0.337(0.081)	0.183	0.495	0.477(0.164)	0.187	0.837
Brushing frequency (< 2)	0.111(0.118)	−0.131	0.342	0.174(0.256)	−0.338	0.695
Age start brushing (years)	0.157(0.043)	0.073	0.244	0.237(0.099)	0.049	0.443
Systemic fluoride (yes)	−0.465(0.086)	−0.635	−0.298	−0.683(0.230)	−1.183	−0.289
Sugary drinks (yes)	0.388(0.088)	0.221	0.556	0.519(0.229)	0.117	1.017
Between meals (> 2)	0.171(0.080)	0.016	0.329	0.212(0.169)	−0.135	0.545
p	0.004(0.017)	0.000	0.060	0.128(0.161)	0.000	0.463
τ	1.302(1.124)	0.777	2.335	1.143(3.810)	0.160	2.736
$^{\dagger}\hat{\beta}_0$	−4.128(0.417)	−5.030	−3.383			
$^{\dagger}\hat{\beta}_Z$	6.655(0.791)	5.204	8.356			
$^{\dagger}\hat{\beta}_{dmft_{4,5}}$	0.453(0.130)	0.203	0.715			
$^{\dagger}\hat{\beta}_{Z*dmft_{4,5}}$	−0.589(0.180)	−0.953	−0.249			

†Coefficients of the logistic regression model (12) from double binomial correction.