Published in final edited form as:

Stat Med. 2013 December 30; 32(30): 5241–5259. doi:10.1002/sim.5944.

A multilevel model for spatially correlated binary data in the presence of misclassification: an application in oral health research

Timothy Mutsvari^a, Dipankar Bandyopadhyay^b, Dominique Declerck^c, and Emmanuel Lesaffre^{a,d,*},†

^aL-Biostat, KU Leuven, Belgium ^bDivision of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, U.S.A ^cDepartment of Oral Health Sciences, KU Leuven, Belgium ^dDepartment of Biostatistics, Erasmus University Rotterdam, Rotterdam, The Netherlands

Abstract

Dental caries is a highly prevalent disease affecting the tooth's hard tissues by acid-forming bacteria. The past and present caries status of a tooth is characterized by a response called caries experience (CE). Several epidemiological studies have explored risk factors for CE. However, the detection of CE is prone to misclassification because some cases are neither clearly carious nor noncarious, and this needs to be incorporated into the epidemiological models for CE data. From a dentist's point of view, it is most appealing to analyze CE on the tooth's surface, implying that the multilevel structure of the data (surface–tooth–mouth) needs to be taken into account. In addition, CE data are spatially referenced, that is, an active lesion on one surface may impact the decay process of the neighboring surfaces, and that might also influence the process of scoring CE. In this paper, we investigate two hypotheses: that is, (i) CE outcomes recorded at surface level are spatially associated; and (ii) the dental examiners exhibit some spatial behavior while scoring CE at surface level, by using a spatially referenced multilevel autologistic model, corrected for misclassification. These hypotheses were tested on the well-known Signal Tandmobiel® study on dental caries, and simulation studies were conducted to assess the effect of misclassification and strength of spatial dependence on the autologistic model parameters. Our results indicate a substantial spatial dependency in the examiners' scoring behavior and also in the prevalence of CE at surface level.

Keywords

binary; autologistic; multilevel; spatial

*Correspondence to: Emmanuel Lesaffre, Department of Biostatistics, Erasmus University Rotterdam, Rotterdam, The Netherlands.
†emmanuel.lesaffre@med.kuleuven.be

1. Introduction

Oral health has improved tremendously over the last century, but the prevalence of dental caries in children remains a significant clinical problem that needs attention. Many epidemiological and clinical studies have explored the risk factors that influence dental caries. The most commonly used outcome that quantifies the present and past caries status of a tooth or tooth's surface is caries experience (CE). However, the process of detecting CE is challenging, especially during the early stages of the disease. This leads to misclassified outcomes resulting in possibly poor quality data. A lot of effort has been devoted to improve CE detection. For example, guidelines for conducting a dental examination were developed by the (WHO) [1] and the British Association for the Study of Community Dentistry, (BASCD) [2] and more recently the (ICDAS) [3] system was proposed. All of these underline the need for training the examiners and measuring the reliability of the obtained scores. This is achieved by comparing the scores of dental examiners with the gold standard (or a benchmark scorer in the absence of a gold standard). However, for various reasons, see, for example, [4] and [5], the process of CE detection remains prone to misclassification error even after adhering to these standardized criteria. Therefore apart from the assessment of CE prevalence, the scoring process of CE needs further investigation and should be incorporated into epidemiological models.

In a dental study, data is recorded at surface level but can be analyzed at mouth, tooth, or surface level. However, from a dentist's point of view, the tooth's surface is most appealing, implying that the multilevel structure of the data (surface–tooth–mouth) needs to be taken into account. Indeed, the CE status of surfaces from the same tooth is more correlated than the CE status of surfaces from different teeth. Moreover, caries on one tooth surface may cause caries on the neighboring teeth, that is, there is spatial dependency in the CE status. Spatial modeling in dental caries has been studied by, for example, Zhang *et al.* [6] and Bandyopadhyay *et al.* [7]. In addition, the anatomical structure of a tooth or tooth surface, for example, pit, fissured, proximal, or smooth surfaces influences the development of dental caries. In this paper, we wish to investigate the factors that influence the prevalence of CE while taking care of the spatial dependency of CE. We also investigate whether the scoring behavior of examiners is spatially induced, that is, whether the scoring is more similar for surfaces on the same tooth or adjacent surfaces than surfaces further apart. This necessitates the investigation of spatial dependency of the scoring behavior of dental examiners, which to our knowledge has never been carried out before. Further, an exploration of scoring CE multilevel data coupled with spatial dependency has never been explored. It is clear that analyzing this type of data with its complex correlation structure provides a serious statistical challenge for researchers in the dental field.

There is a vast amount of literature that addresses the analysis of spatial associations [8–11]. In a typical disease mapping framework, the unit of spatial association is the geographical entity, say county, and subjects are clustered within a particular county. On the contrary, the spatial topology within a mouth is unique, because here the spatial association is manifested at the tooth, or site level clustered within a mouth/subject, and subjects are independent. Although spatial models are widely used with geographical data, its use in dental applications for binary data is scarce; see, for example, Bandyopadhyay *et al.* [7]. Under a

Markov random field framework, (where the conditional distribution of the spatial unit depends on only the outcomes of the specified set of neighbors), one of the most common approaches to deal with spatial data is the conditional autoregressive (CAR) model [11]. The CAR model has desirable properties, namely one can implement the estimation by using full likelihood [10]. An alternative and plausible approach to deal with spatially correlated binary data (as in our case) is via the autologistic regression model; see, for example, Besag [8]. The autologistic model extends the usual logistic regression accounting for spatial correlation. Here, one computes the sum of the outcomes for the neighbors for each site and includes the sum as a covariate in the model to capture the spatial effect. However, the major drawback of this approach is that the standard maximum likelihood estimation methods are untenable due to the intractable normalizing constant, except for small data sets [10]. We address this problem by using a pseudo-likelihood (PL) estimation technique [12], where only the autocovariate term is included in the regression model without the normalizing constant. This is a more flexible but approximate approach, which allows the inclusion of multilevel random effects. Also, standard software can be used to carry out the estimation. The PL approach leads to unbiased and consistent estimates that are asymptotically normal [11, 13]; however, the efficiency of the PL estimate decreases with increasing spatial dependence.

In this paper, we examine two spatial hypotheses on CE outcomes recorded at surface level. First, we explore the spatial dependence of CE on proximal surfaces (adjacent surfaces), and compare it with the spatial dependence of the nonproximal surfaces (on the same tooth). Next, we explore whether there is evidence of spatial referencing for scoring CE by dental examiners. More specifically, we compare the spatial dependency on the misclassification process for proximal and nonproximal surfaces. For both hypotheses, we adopt a Bayesian framework for our autologistic model where the PL is used as an approximation to the full likelihood. Our analysis considers the surface level of tooth, thereby exploiting the full multilevel structure of CE data into account. For the first hypothesis, we fit the *main* model to the main cross-sectional data. For the second hypothesis, we adjust the autologistic model for misclassification and fit the *misclassification* model to the validation data. The main and validation data sets are obtained from the well-known Signal Tandmobiel® (ST) study, whereby children in the main study were examined by 16 dental examiners and also by a benchmark scorer in the validation study.

This paper is organized as follows. Section 2 provides a general description of the motivating data set and the study population. Statistical models for the main data (error-free) and validation data are outlined in Section 3. In Section 4, we present a model for the main data adjusted for misclassification. Details on the PL estimation technique appear in Section 5. Section 6 applies the methodology to the ST data set, while Section 7 presents a limited simulation study to investigate the impact of misclassification and spatial association on model parameters. Section 8 concludes, with some further discussion.

2. Motivating data set

2.1. Main data set

The ST study is a longitudinal (1996–2001) oral health intervention project that took place in Flanders (north of Belgium). At the first examination, the average age of the children was 7.1 years ($SD = 0.4$). Sixteen trained dentists (examiners) conducted annual oral health examinations on 4468 children (2315 boys and 2153 girls) from 179 primary schools, after parental consent was obtained. Data on oral hygiene and dietary habits were obtained through structured questionnaires completed by the parents. The children received a clinical examination based on the diagnostic criteria for caries prevalence surveys published by the BASCD [2]. The clinical examinations took place in a mobile dental clinic, with a standard dental chair and artificial dental light. Detection was performed using a visual–tactile method by using a mouth mirror and a World Health Organization/Community Periodontal Index of Treatment Needs type E probe (Prima Instruments, Gloucester, UK). No radiographs were taken. For the analysis of the main data, we consider only the data collected at the last visit, that is, year 2001. The outcome of interest is binary CE ($CE = 1$ if the surface shows CE, and 0 otherwise) subject to misclassification.

2.2. Validation data set

Training sessions for scoring CE were organized and the scoring behavior measured by sensitivity (SE) and specificity (SP), of each of the 16 dental examiners was compared with that of the benchmark scorer, that is, the third author (Dominique Declerck) who was trained in BASCD scoring (with regular updates) and has extensive experience in scoring CE in children. Because scoring was carried out at d3 level (irreversible caries), her scores were assumed to be close to perfect. Nevertheless, we realize that a benchmark scorer is never without error, but we had to ignore this in our analysis in order not to further increase the complexity of our model. SE and SP measure the capability of the examiners in detecting CE or its absence, respectively. During the study period (1996–2001), three calibration exercises for scoring CE (1996, 1998, and 2000) involving 92, 32, and 24 children respectively were organized, vis-à-vis the benchmark scorer. In the present work, data from the three calibration exercises were combined into one validation data set. More details of the ST study are given in Vanobbergen *et al.* [14].

To analyze the spatial scoring behavior of the examiners, we defined a misclassification binary score $W = 1$ if the benchmark and examiner agree in the CE score and 0 otherwise. A statistical model is fitted later to determine the factors that influence this outcome while accounting for spatial dependency.

3. Statistical models for error-free responses

3.1. Spatial dependency of caries experience data

In this section, we introduce the metric for the spatial dependency structure that will be investigated in this paper. Our main interest for the main and validation data is to investigate the spatial dependency at surface level. Note first that a tooth consists of four to five surfaces. Four surfaces are found on incisors and canines: (i) buccal surface, which faces the

front of the mouth; (ii) lingual surface, which faces the tongue side; and (iii) the in-between teeth surfaces, mesial and distal. The mesial side faces the middle of the mouth while the distal side faces the back. The molars and premolars have an extra occlusal surface, which is the surface that makes contact with the vertically opposing tooth. (Table I)

Motivated by the work of Leroy *et al.* [15], we distinguish spatial dependency for proximal surfaces (distal and mesial) and the other for nonproximal surfaces (buccal, lingual, and occlusal). Figure 1 shows these two types of spatial dependency. In both models, we let γ_1 represent the spatial dependency for the nonproximal surfaces belonging to the same tooth and γ_2 represent the spatial dependency for the proximal surfaces, that is, adjacent mesial and distal surfaces.

3.2. The multilevel autologistic model for main data

The binary outcome variable $Y_{M,mts}$ equals one if the s th surface ($s = 1, \dots, S_{tm}$) in tooth $t = 1, \dots, T_m$ of subject $m = 1, \dots, M$ has CE and $Y_{M,mts} = 0$ otherwise. The subscript M stands for ‘main data’. Note that, following [8], our proposition takes the most natural route to model spatially correlated binary data. We consider a generalized linear mixed model framework to model the probability of CE and incorporate the fixed effects, heterogeneous random effects and spatial effects via the *autologistic* specification, that is,

$$\text{logit}(\pi_{mts}) = \mathbf{x}'_{mts} \boldsymbol{\beta}_M + \mathbf{z}'_{mts} \mathbf{u}_m + \gamma_{M,1} \sum_{s': s \sim s'} y_{mts'} + \gamma_{M,2} \sum_{\tilde{s}: s \simeq \tilde{s}} y_{mt\tilde{s}}, \quad (1)$$

where $y_{mts'}$ is the observed outcome of surface s' considered a neighbor to surface s on the same tooth, $\sim$ refers to neighboring nonproximal surfaces, and $\simeq$ refers to neighboring proximal surfaces, $y_{mt\tilde{s}}$ is the observed outcome of surface $\tilde{s}$, a neighbor of surface s but from an adjacent tooth, $\mathbf{x}_{mts}$ is a d -dimensional vector of covariates for surface s within tooth t in mouth m with $\boldsymbol{\beta}_M$ the corresponding vector of regression coefficients, and

$\mathbf{u}'_m = (u_m, u_{tm})$ is the random effects vector, where u_m and u_{tm} are random intercepts at mouth and tooth (nested in mouth) level, distributed independently, with mean zero and variances σ_m^2 and σ_{tm}^2 , respectively, that is, $\mathbf{u}_m \sim N(\mathbf{0}, \mathbf{D}_M)$ where $\mathbf{D}_M = \text{diag}(\sigma_m^2, \sigma_{tm}^2)$, and $\mathbf{z}_{mts}$ the corresponding q -dimensional design matrix.

The probability of a surface s belonging to tooth t and mouth m having CE is given by $\pi_{mts} = \Pr(Y_{M,mts} = 1 | y_{mts'}, y_{mt\tilde{s}}, \mathbf{u}_m, \mathbf{x}_{mts}, \boldsymbol{\beta}_M, \gamma_{M,1}, \gamma_{M,2})$. In this model, we specified two spatial parameters $\gamma_{M,1}$ and $\gamma_{M,2}$ to determine the degree or strength of spatial correlation, that is, it explains how the CE status of a surface is influenced by the CE of neighboring surfaces. Specifically, $\gamma_{M,1}$ controls for the spatial association of surfaces within the same tooth while $\gamma_{M,2}$ controls for spatial dependency of surfaces on adjacent teeth on the same jaw.

The interpretation of the autologistic spatial parameters $\gamma_{M,1}$ and $\gamma_{M,2}$ is not straight forward. In the presence of positive spatial dependence, it can be illustrated that the traditional autologistic parameterization leads to the conditional expectation of $Y_{M,mts}$ given

its neighbors that always increases as long as neighbors have nonzero outcomes and never decreases. This is unreasonable if most of the neighbors are zeros and could bias the realizations toward 1. Hence, the interpretation of parameters is difficult across varying levels of statistical dependence. Because of this problem, Caragea and Kaiser [16] recently proposed a reparametrization of the autologistic regression in order to have interpretable parameters. This necessitates an adjustment, whereby $\Sigma_{s':s \sim s'} y_{mts'}$ and $\Sigma_{\tilde{s}:s \sim \tilde{s}} y_{mt\tilde{s}}$ are substituted by $\Sigma_{s':s \sim s'} (y_{mts'} - \lambda_{mts})$ and $\Sigma_{\tilde{s}:s \sim \tilde{s}} (y_{mt\tilde{s}} - \lambda_{mts})$, respectively, and

$\lambda_{mts} = \exp(x'_{mts} \beta_M + z'_{mts} u_m) / (1 + \exp(x'_{mts} \beta_M + z'_{mts} u_m))$. Given this parameterization, the odds of $Y_{M,mts} = 1$ increases if the number of positive neighbor surfaces is greater than the number of neighbors expected under the independence model and decreases if the neighbor surfaces is less than the number expected under the independence model.

3.3. The multilevel autologistic model for validation data

Let $Y_{V,mts}^*$ and $Y_{V,mts}$ be the binary CE outcome variables in the validation data for surface $s = 1, \dots, S_{tm}$ in tooth $t = 1, \dots, T_m$ of subject $m = 1, \dots, M$ according to examiner and benchmark scorer, respectively. The subscript V stands for 'validation data'. Let W_{mts} be the binary outcome variable for surface $s = 1, \dots, S_{tm}$ in tooth $t = 1, \dots, T_m$ of subject $m = 1, \dots, M$, with $W_{mts} = 1$ if the examiner and the benchmark score the surface differently for CE and $W_{mts} = 0$ otherwise, that is, $W_{mts} = 1$ if $Y_{V,mts}^* = 0$ & $Y_{V,mts} = 1$ or $Y_{V,mts}^* = 1$ & $Y_{V,mts} = 0$, and 0 otherwise. Then, the autologistic model that relates the discrepancy in scoring of the examiner and the benchmark scorer to determinants is given by

$$\text{logit}(\mu_{mts}) = x'_{mts} \beta_V + z'_{mts} u_m + \gamma_{V,1} \sum_{s':s \sim s'} w_{mts'} + \gamma_{V,2} \sum_{\tilde{s}:s \sim \tilde{s}} w_{mt\tilde{s}},$$

where μ_{mts} is the probability of agreement (between examiner and benchmark scorer) in scoring CE for the m tooth, t th tooth, and the s th surface, $w_{mts'}$ is the agreement outcome on the neighboring surface s' (on the same tooth) while $w_{mt\tilde{s}}$ is the outcome of surface $\tilde{s}$ (on the adjacent tooth). The vector β_V contains the fixed effects regression coefficients, $\gamma_{V,1}$ measures the degree of spatial dependency in scoring the CE status of surfaces of the same tooth and $\gamma_{V,2}$ measures the degree of spatial dependency in misclassifying the CE status of a proximal surface, that is, mesial (distal) surface and adjacent distal (mesial) surface. Similarly to the main model, we denote the covariance matrix of the random effects u_m as D_V .

4. Correction for misclassification

In Section 3.2, we presented a statistical model relating the score of the gold standard (or benchmark scorer) to covariates taking into account the spatial structure of the data. However, in the main data set the dental examiners scored the children and their scores are likely to be prone to error. Ignoring in the statistical analysis that the binary CE in the main data is prone to misclassification might lead to wrong conclusions. In this section, we present an approach for correcting the autologistic regression model for misclassification

error. We denote $Y_{M,mts}^*$ the binary CE score obtained from the examiner(s) for surface $s = 1, \dots, S_{tm}$ in tooth $t = 1, \dots, T_m$ of subject $m = 1, \dots, M$, with $Y_{M,mts}^* = 1$ if the dental examiner noticed CE, and is 0 otherwise. In the main data set, the true score $Y_{M,mts}$ (from the benchmark scorer) is not available. However, the SE and SP values obtained from the validation data can be used to correct for misclassification. Note that we assumed in this paper that all dental examiners scored similarly, this to reduce the complexity of the model. However, there is no theoretical difficulty (only computational burden in view of the large data set) to extend the model taking into account 16 different SE and SP estimates. Given $Y_{V,mts}$ the true (benchmark score) CE from the validation data, we estimate the correction terms, that is, $\tau_{11} = Pr(Y_{V,mts}^* = 1 | Y_{V,mts} = 1)$ and $\tau_{00} = Pr(Y_{V,mts}^* = 0 | Y_{V,mts} = 0)$, which are the SE and SP, respectively. The corrected autologistic regression model is then given by

$$\begin{aligned} Pr(Y_{M,mts}^* = 1 | \dots) &= \sum_y Pr(Y_{M,mts}^* = 1 | Y_{M,mts} = y) Pr(Y_{M,mts} = y | \dots) \\ &= (1 - \tau_{00}) + (\tau_{11} + \tau_{00} - 1)g^{-1}(\eta) \end{aligned} \quad (2)$$

where g is a link function and

$\eta = (x_{mts}\beta_M + z'_{mts}u_m + \gamma_{M,1} \sum_{s': s \sim s'} y_{mts'} + \gamma_{M,2} \sum_{\tilde{s}: s \sim \tilde{s}} y_{mts\tilde{s}})$, '...' represents $\beta_M, x_{mts}, D_M, \tau_{00}, \tau_{11}, \gamma_{M,1}$ and $\gamma_{M,2}$. Note that we did not use a misclassification spatial model for correction because it is not compatible with the correction terms needed in the main model. In the next section, we give the details of PL estimation approach.

5. The pseudo-likelihood approach

The full likelihood for model (1), that is, the joint probability distribution for all binary outcomes y_{mts} is known only up to a normalization constant [17]. The contribution to the likelihood of the m th mouth and t th tooth is given by

$$L(\beta_M, \gamma_M, u_m | y_{mt}) = \frac{\exp \left(\sum_s y_{mts} \left[x'_{mts}\beta_M + z'_{mts}u_m + \frac{\gamma_{M,1}}{2} \sum_{s': s \sim s'} y_{mts'} + \frac{\gamma_{M,2}}{2} \sum_{\tilde{s}: s \sim \tilde{s}} y_{mts\tilde{s}} \right] \right)}{k(\beta_M, \gamma_M, u_m)},$$

where y_{mt} is the vector of responses for the t th tooth in the m th mouth, and $\gamma' = (\gamma_{M,1}, \gamma_{M,2})$ is the vector of the spatial parameters. The constant

$$k(\beta_M, \gamma_M, u_m) = \sum_{\{y_{mts}\}} \exp \left(\sum_s y_{mts} \left[x'_{mts}\beta_M + z'_{mts}u_m + \frac{\gamma_{M,1}}{2} \sum_{s'} y_{mts'} + \frac{\gamma_{M,2}}{2} \sum_{\tilde{s}} y_{mts\tilde{s}} \right] \right)$$

involves $\Sigma_{\{y_{mts}\}}$, which is the sum of all possible realizations in the observed data. This is the source of difficulty in fitting the autologistic model, because it is an analytically intractable function of the regression parameters for large data sets. For a data set with 1000

observations, the full likelihood evaluation would require $2^{1000} = 1,071,509 \times 10^{301}$ possible cases.

Certainly, a CAR specification [18] of the spatial terms under Markov random field assumptions would lead to a tractable likelihood. The CAR model is ideal and popular in a disease mapping [19] context but based on Gaussian assumptions that can be unrealistic in the context of its latent variable formulations to handle categorical data. As mentioned earlier in Section 3.2, our autologistic proposition is in the spirit of the celebrated 1972 paper by Besag [8] for discrete (binary) spatial data and directly addresses spatial association by estimating how much the response variable at any one site reflects response values at surrounding sites [20]. In addition, initial attempts to fit the CAR model turned out to be negative in the sense that convergence of some of the model parameters were not achieved, for example, for the variance of the spatial random effect. Henceforth, we proceeded with our analysis by using the autologistic specification.

Several techniques have been suggested to address the issue of intractable normalizing constants, see, for example, the Huffer and Wu approach [17], path sampling [21], Møller *et al.* algorithm [22], and so on. However, adapting these techniques to a multilevel data structure with various types of spatial associations manifested via $\gamma_{M,1}$ and $\gamma_{M,2}$, and a large number of observed data points, would involve formidable computations as in our case. In addition, implementing such techniques in standard Bayesian statistical software such as WinBUGS is infeasible. In contrast, the Bayesian approach based on PL is straightforward, and does not rely on the normal approximations or nonlinear maximization steps involved in the previous suggested methods. Besag [9] proposed the PL approximation in which the autocovariate term ($\Sigma_{s'} y_{mts'}$ and $\Sigma_{\tilde{s}} y_{mts\tilde{s}}$) is included in the regression, but the likelihood function is constructed as if the responses were independent. Hereby, standard software can be used. The PL for the CE responses is then given by

$$\begin{aligned} \text{PL}(\beta_M, \gamma_M, \mathbf{u}_M | \mathbf{y}_M) &= Pr(y_{mts}=1 | y_{mts'}, y_{mts\tilde{s}}, \mathbf{u}_m, \mathbf{x}_{mts}, \beta_M, \gamma_M) \\ &= \prod_{m=1}^M \prod_{t=1}^{T_m} \prod_{s=1}^{S_{tm}} \frac{\exp(\mathbf{x}'_{mts} \beta_M + \mathbf{z}'_{mts} \mathbf{u}_m + \gamma_{M,1} \sum_{s'} y_{mts'} + \gamma_{M,2} \sum_{\tilde{s}} y_{mts\tilde{s}})}{1 + \exp(\mathbf{x}'_{mts} \beta_M + \mathbf{z}'_{mts} \mathbf{u}_m + \gamma_{M,1} \sum_{s'} y_{mts'} + \gamma_{M,2} \sum_{\tilde{s}} y_{mts\tilde{s}})}, \end{aligned} \quad (3)$$

with $\mathbf{y}_M$ representing all CE outcomes. The evaluation of the marginal PL for the m th mouth involves integrating out the random effects $\mathbf{u}_m$ and this is accomplished via MCMC techniques [23].

5.1. Bayesian inference

We adopt a Bayesian paradigm for parameter estimation in the models for validation data and the main data. By using this approach, we can incorporate prior knowledge about the parameters combined with the observed data to yield a posterior distribution of the parameters by using the MCMC approach [23]. For the main model corrected for misclassification, we opted to process the main data and the validation data simultaneously, that is, at each iteration of the Markov chain of the validation data, we obtained estimates of τ_{11} and τ_{00} , and these estimates were imputed into the Markov chain pertaining to the main

data as priors. This presents a simple way to take into account the variability with which τ_{11} and τ_{00} are estimated from the validation data. In the next subsections, we provide further details on the choice of our prior distributions and the related posterior distribution.

5.1.1. Prior distributions—For both the main and misclassification model, we chose non-informative priors for the model parameters. For the fixed effects β_M and β_V , vague independent normal priors were assumed, that is, $\beta_i \sim N(0, 10^6)$, $i = 0, \dots, p$. Independent Uniform(0, 100) priors were assumed for the standard deviations σ_m and σ_{tm} of the random effects [24]. Note that the prior of D_M and of D_V is determined by the priors of σ_m and σ_{tm} . For the spatial random effects $\gamma_{M,1}$, $\gamma_{M,2}$ prior belief of a positive (spatial) association suggested to take U(0, 3) priors. Independent Beta(1, 1) priors were taken for the misclassification probabilities τ_{11} and τ_{00} .

5.1.2. Posterior distribution—Denoting the prior distributions on β_M , γ_M , D_M (the variance–covariance matrix) by $p(\beta_M)$, $p(\gamma_M)$ and $p(D_M)$, respectively, the joint posterior distributions for the main model (assuming error-free) is $p(\beta_M, \gamma_M, D_M | Y_M) \propto \text{PL}(\beta_M, \gamma_M, D_M) p(\beta_M) p(\gamma_M) p(D_M | Y_M)$. For the misclassification model (M replaced by V), the joint posterior distributions become

$p(\beta_V, \gamma_V, D_V | Y_V^*, Y_V) \propto \text{PL}(\beta_V, \gamma_V, D_V | Y_V^*, Y_V) p(\beta_V) p(\gamma_V) p(D_V)$, where Y_V^* and Y_V are vectors of misclassification responses in the validation data obtained from the examiner and benchmark, respectively.

We now look at the posterior distribution for the main corrected model (2). Note that we assumed that all dental examiners score similarly. This is an assumption that can be relaxed, but given the complexity of the suggested model, we preferred to make this simplifying assumption. For given values of SE and SP, the analysis of the cross-sectional main data yields $p(\beta_M, D_M, \gamma_M | Y_M^*, Y_V^*, Y_V, \tau_{00}, \tau_{11})$, that is, the posterior estimates obtained are conditioned on the imputed τ_{00} and τ_{11} , where Y_M^* is a vector of error-prone CE responses (from examiners) in the main data. Specifically, the prior distributions of τ_{00} and τ_{11} , that is $p(\tau_{00})$ and $p(\tau_{11})$ yield the following posterior distributions:

$$\begin{aligned}\tau_{10} &\sim \text{Beta}(n_{10}+1, n_{00}+1), \\ \tau_{01} &\sim \text{Beta}(n_{01}+1, n_{11}+1)\end{aligned}$$

where $\tau_{00} = 1 - \tau_{10}$ and $\tau_{11} = 1 - \tau_{01}$, n_{ij} is the frequency of scoring ‘ i ’ for CE in the validation data by the examiner when the benchmark scorer scores ‘ j ’. This posterior distribution is imputed into the model for the main data to correct for misclassification. Given the posterior estimates of the misclassification probabilities $\tau = (\tau_{10}, \tau_{01})$ from the validation data, the pseudo-likelihood Equation (3) and the prior information for the parameters in the main corrected model, the posterior distribution of the main model is given by

$$p(\beta_M, \gamma_M, D_M | Y_M^*, Y_V^*, Y_V, \tau_{00}, \tau_{11}) \propto \text{PL}(\beta_M, \gamma_M, D_M | Y_M^*, Y_V^*, Y_V) p(\beta_M) p(\gamma_M) p(D_M) p(\tau)$$

Note that the posterior estimates of misclassification probabilities are used as priors in the main corrected model and the process is carried out simultaneously.

5.2. Model selection and assessment

The Bayesian toolbox offers various model selection criteria. One of the most popular one is the deviance information criterion (DIC) [25], which is based on the deviance $D(\Theta) = -2\text{Log}(\text{data}|\Theta)$, where Θ is the parameter space. DIC is defined as $DIC = \bar{D} + p_D$, that is, posterior mean deviance plus the ‘effective number of parameters’ $p_D = \bar{D} - D(\bar{\Theta})$, whereby $D(\bar{\Theta})$ is the deviance evaluated at the posterior mean of the parameters. Models with smaller DIC are better supported by the data. Note that, in our proposition, the PL approximates the true likelihood, hence the reported DIC will be an approximate/pseudo-DIC. Because of our PL route, we wanted to avoid using Bayes factors [26], which is fundamentally based on full-likelihood computations.

To strengthen our model comparisons, we also explored some cross-validators approaches on the basis solely of predictive performances of competing models. If y_{pr} denotes the predictive data vector and y denotes the observed data vector, then the posterior predictive distribution is given as $p(y_{pr}|y) = \int p(y_{pr}|\Theta)p(\Theta|y)d\Theta$. One can obtain predictive data easily from a converged posterior sample and samples from the posterior predictive distribution are replicates of the observed model generated data. Cross-validation algorithms [27] involve fitting the model to a randomly selected subset of the data, predicting the remaining data, and finally comparing the error in this prediction (misclassification rates for our binary response) across alternative models. The model with the lowest estimated predictive error is chosen. In our example, our cross-validated measure randomly removes 10% of the data and monitors the total predictive error of incorrectly classified deleted binary responses. This is given as $CV_1 = \sum\{y(1 - y_{pr})\}$ and $CV_0 = \sum\{(1 - y)y_{pr}\}$, where y is the observed response, y_{pr} is the predicted response, and the summations are taken over the respected deleted observations. Under a Bayesian paradigm, this measure is obtained as a posterior mean of the total predictive errors over all parameters and the data, computed as ECV_0 and ECV_1 where the expectation is over all other parameters and data.

We note, however, that the previous measures for model comparison measures do not make sense for the evaluation of the corrected (for misclassification) models, because these corrected models provide parameter estimates on the basis of the (unobserved) true responses. Evidently the fit of these models to the observed (but possibly corrupted) responses must be worse than the models that aim to predict the true responses.

Postmodel comparisons, model diagnostics to assess goodness-of-fit of the best model uses a criterion on the basis of the posterior predictive distribution. The discrepancy measure between model and data is computed as a summary statistic [28] by using model parameters and data defined as $T(y, \Theta) = \sum \frac{(y - E\{y|\Theta\})^2}{\text{Var}\{y|\Theta\}}$, where the summation is taken over all observations. The Bayesian p -value/posterior predictive p -value [28] is then defined as the number of times $T(y_{pr}, \Theta)$ exceeds $T(y, \Theta)$ out of L simulated draws. A very large p -value (> 0.95) or a very small (< 0.05) both signals model misspecification, that is, the observed pattern would be unlikely to be seen in replications of the data under the true model [28]. In

a similar spirit as explained previously, we computed the Bayesian p -value for the best (uncorrected) model.

6. Application to the Signal Tandmobiel® study

In this section, we summarize the results obtained from the ST main and validation study. The analysis was carried out in WinBUGS 1.4.3 [29]. For both models, we ran three MCMC chains each for 10,000 iterations with a burn-in period of 2000. The MCMC chains show adequate mixing and convergence, as revealed from the trace plots, ACF plots and the Gelman–Rubin potential scale-reduction factor and Rubin's diagnostic ($\hat{R}$) [28], which were less than 1.1 for all the parameters and with a Monte Carlo error less than 5% of the posterior standard deviation (see e.g., [23]).

Table II presents the comparison among the three competing models fitted for the main data by using the DIC, CVO, and CV1. All the model selection tools (DIC, CV0, and CV1) favored the multilevel autologistic as the best model. This motivated the extension of multilevel autologistic model to correct for misclassification, and henceforth we discuss parameter estimates on the basis of this model. The Bayesian p -value for the multilevel autologistic model was 0.574, and so we do not suspect any inadequacy of model fit.

6.1. Analysis of the main data

Four models were fitted to the main data: that is, (i) logistic regression model that assumes that the data are independent; (ii) multilevel model that assumes that the data are dependent without considering the spatial random effects; (iii) multilevel autologistic model that incorporates unstructured random effects as well as spatial random effects; and (iv) the multilevel autologistic model in (iii) but corrected for misclassification.

For all the previously mentioned models, the following covariates were considered; the effect of gender (girls versus boys), dentition type (permanent versus deciduous), tooth type (canine versus incisor, molar or premolar versus incisor), jaw (mandible versus maxilla) and quadrant (left versus right), age of the child, and the geographical location (represented by the standardized (x,y) coordinate of the municipality of the school to which the child belongs). The latter covariate was included because it turned out to be a significant predictor in previous analyses; see for example, Mwalili *et al.* [30]. The reason for its importance is, though, still not clear. Table III shows the posterior means, standard deviations and 95% equal tail credible intervals (CI) for the fixed effects (in log-odds), random effects and spatial parameters obtained for models (i), (ii), and (iii). Generally, the three models gave similar statistical significant conclusions.

6.2. Analysis of the validation data

Three models were fitted to the validation data: that is, (i) logistic regression model that assumes that the data are independent; (ii) multilevel model that assumes that the data are dependent without considering the spatial random effects; and (iii) multilevel autologistic model that incorporates unstructured random effects as well as spatial random effects. The covariates considered now are gender (girls versus boys), dentition type (permanent versus deciduous), tooth type (canine versus incisor, molar or premolar versus incisor), jaw

(mandible versus maxilla), and quadrant (left versus right) were considered. Table IV shows the posterior means, standard deviations, and 95% equal tail CIs for the fixed effects (in log-odds), variance of random effects and spatial parameters obtained from the validation data.

There was a consensus for the three models regarding the influence of gender, that is, a lower probability of misclassification for girls compared to boys. The probability of misclassifying a surface belonging to permanent teeth was lower compared with that of deciduous teeth in all the models and higher for canine and pre(molar) teeth compared with incisors. The position of the surface in the mouth also has an impact on the probability of misclassifying CE, but this was only observed in the standard logistic regression model. Finally, for all models misclassifying CE on surfaces on the right quadrant happens more than with surfaces on the left quadrant.

Because the 95% equal tail CI of the difference $\gamma_1 - \gamma_2$ does not contain zero (-0.735 ; -0.050), we concluded that proximal spatial dependency is significantly greater than nonproximal spatial dependency. This motivated the extension of our model in order to investigate examiners' specific spatial dependency on proximal and nonproximal surfaces. The results of the two types of spatial dependencies for each examiner are shown in Figure 2. The left panel indicates the examiners' specific spatial dependency estimates on proximal surfaces and the right panel indicates estimates on nonproximal surfaces. A high estimate of γ_1 means that the examiner scores adjacent distal and mesial surface either both right or both wrong. This situation is observed for examiner 14. Similarly, a high estimate of γ_2 is an indication of an examiner who scores surfaces on one tooth (excluding distal and mesial) all wrong or all right, which was noticed for examiners 6, 12, and 14. An obvious conclusion from the analysis of validation data is that a flaw in scoring CE of a surface is more likely to be repeated on the neighboring surface.

6.3. Analysis of the main data (corrected for misclassification)

The results for the main corrected model (2), that is, after correcting for misclassification are shown in Table V. By using the SE and SP from the validation data, we proceeded to correct the main model for possible misclassification. Substantial changes were observed in the parameter estimates of the fixed effects and standard deviations. Also, the spatial effects parameters and variance of the random effects increased in the corrected model. However, the conclusions did not change qualitatively. The parameter estimates obtained with differential correction are generally higher in absolute value compared to the models without correction. Also, the 95 % equal tail CI for differential correction are often wider than when no correction and nondifferential correction was applied.

The probability of a surface having CE decreased as the age increases, which is a biologically awkward result. Also, the probability of CE of a surface of a permanent tooth was lower compared with that of a deciduous tooth. Further, there was a higher probability of CE on surfaces on the right quadrant than surfaces on the left quadrant (from the patient's perspective). There was a remarkably high proximal and nonproximal spatial dependency of CE. The comparison of γ_1 and γ_2 revealed a nonsignificant (95% equal tail CI = $[-0.729$; $0.538]$) difference between the two spatial dependencies. Although we reported the DIC

(1979,31) for the corrected autologistic model, we cannot compare this model with the other models since in this case we are modeling the latent (unobserved response).

6.4. Sensitivity analysis

We performed a limited sensitivity analysis in order to investigate the stability of our model because of prior distribution choices for β , γ_1 , γ_2 , σ_m^2 , and σ_{tm}^2 . More specifically, we changed the prior distributions of α_m and α_{tm} from Uniform(0, 100) to IG (10^{-4} , 10^{-4}) for σ_m^2 and σ_{tm}^2 . Further we changed the prior distribution of the regression coefficients from a normal to a t -distribution with 4 degrees of freedom. A `dfLat()` prior was also tested for the regression coefficients. We have also chosen a Uniform(0,100) prior and a $N(0, 10^6)$ prior for the spatial parameters. Although slight changes were observed in the parameter estimates (both fixed effects, spatial, and variance components), results were robust overall, and did not change any conclusions regarding the strength of the spatial parameters, and inference on the fixed effects.

7. Simulation study

In this section, we present a simulation study to investigate the performance of the corrected autologistic regression model. In order to assess the performance, we include also the logistic regression, corrected logistic model, and uncorrected autologistic in the simulation. Triggered by the criticism expressed in the literature about biased parameter estimates for the autologistic specification as compared with a non-spatial logistic regression [31], we wished to illustrate here the effect of misclassification and the impact of the strength of spatial dependency on parameter estimation.

7.1. Generating the simulation data sets

We set the true parameters for β_0 , β_1 , β_2 , and γ to generate spatial data on a lattice of size 20×20 , that is, 400 observations by using the autologistic model given by

$$\text{logit}(\pi_j) = \beta_0 + \beta_1 x_j + \beta_2 y_j + \gamma \sum_{j': j \sim j'} y_{j'},$$

where x and y are the coordinates on the lattice. Note that here we did not use γ_1 and γ_2 in order to avoid unnecessary complexity. The adjacency matrix consists of the four nearest neighbor adjacent cells and sites on the corner have three neighbors. Because in the autologistic model the outcome is also a covariate, we used a Gibbs Sampler to generate the equilibrium lattice or map. For more technical details on generating spatially correlated data we refer to Huffer and Wu [17] and Hughes *et al.* [32]. Initial binary values $y_{k,l}$ where $k, l = 1, \dots, 10$ for site $j, j = 1, \dots, 400$ on the lattice were arbitrarily assigned. After n completed sweeps of updating the map, it was assumed that an equilibrium state has been reached. The last update was then taken as the fixed observed binary data for map Y . Examples of generated maps for different values of the spatial parameter are shown in Figure 3. The Gibbs sampling algorithm to generate the data for iteration i is as follows:

1. Pick arbitrary values of $y_{k,l}$ to create map $Y_0^{(i)}$.
2. Given map $Y_0^{(i)}$ and parameters $(\beta_0, \beta_1, \beta_3, \gamma)$, generate the next map $Y_1^{(i)}$, that is, where each observation in map $Y_1^{(i)}$ is obtained by predicting using the autologistic model (7.1). In the general map $Y_{(t+1)}^{(i)}$ is created from map $Y_{(t)}^{(i)}$.
3. Generate sufficient maps ($t \rightarrow \infty$) until equilibrium state is attained, and keep the last updated map.

7.2. Simulation set-up

We generated 10,000 equilibrium maps. For each of the generated equilibrium maps, we applied the misclassification process in order to generate the corrupted outcome. Let Y be the true outcome generated from the previous algorithm and τ_{00} and τ_{11} be the SP and SE, respectively. We generate the corrupted outcome Y^* by using the following command:

If $Y=1$ then generate $Y^* \sim \text{Bin}(1, \tau_{11})$
 else if
 If $Y=0$ then generate $Y^* \sim \text{Bin}(1, 1-\tau_{00})$

Next, by using Y^* as the outcome and x and y as covariates, we fitted the following models: (i) logistic model; (ii) corrected logistic model; (iii) autologistic model without correction; and (iv) corrected autologistic model. We then calculated the bias and 95% coverage rates for the parameters $\beta_0, \beta_1, \beta_3$, and γ from each model.

7.3. Simulation results

The results obtained from the previous simulation study are shown in Table VI. The data were generated for three different scenarios of γ and two different scenarios of SE and SP . (i) $SE=0.7, SP=0.7, \gamma=0.1$, (ii) $SE=0.7, SP=0.7, \gamma=0.6$, and (iii) $SE=0.7, SP=0.7, \gamma=1.0$, (iv) $SE=0.8, SP=0.8, \gamma=0.1$, (v) $SE=0.8, SP=0.8, \gamma=0.6$, and (vi) $SE=0.8, SP=0.8, \gamma=1.0$. The corresponding generated values are plotted in Figure 4. Plots for the other scenarios are given in Figures 5 and 6. From Table VI, it is clear that the logistic regression and autologistic regression produce biased estimates in the presence of misclassification. One of the important findings is that the bias increases with increasing misclassification. This is also confirmed by the 95% coverage rates of the 95% equal tail CIs. Also, as the spatial dependency increases, the bias in the parameter estimation increases.

The simulation study confirmed that ignoring the spatial structure and ignoring misclassification implies biased estimates for the regression coefficients. This limited simulation study, however, does not confirm the large bias in case of a strong spatial dependence. An advantage of using the PL approach is that it is computationally efficient even for data with high spatial dependency such as that shown in Figure 3. We also repeated the simulation by using SE, SP , and γ values from the Signal Tandmobiel® study (Table VII). From this simulation, we also concluded that the corrected autologistic gave unbiased estimates compared with all the other models.

8. Discussion

In this paper, we proposed an autologistic regression framework for the multilevel validation and main caries experience data. For the validation data, our aim was to model the probability of misclassifying the CE status of a surface as a function of covariates. By using this model, we have addressed the issue of spatial dependency in examiners' scoring behavior. For the main data, we were interested in investigating the factors that influence the prevalence of CE. Also, we investigated if the prevalence of CE is spatially distributed in the mouth. Previous research has confirmed the existence of spatial dependency in CE data [6, 7]. An analysis closely related to our current work is by Leroy *et al.* [15], where the authors concluded that cavity formation in permanent first molars was clearly influenced by the status of the adjacent primary molars, confirming the findings of this study. Our analysis is unique in a sense that the correction for misclassification of the outcome within a multilevel autologistic framework has never been attempted before. Note that the spatial structure within a mouth is unique and must be carefully chosen.

When correcting for misclassification, a fundamental assumption in a validation study is that true scores are recorded by a gold standard, that is, an approach (instrument or examiner) that is 100% error free. However, in practice, the scores are often generated by a benchmark scorer, that is, an experienced examiner who is assumed to be error free or nearly so. A benchmark scorer is also referred to in the literature as an 'alloyed gold standard'. The question of how 'alloyed' a benchmark can be in order to maintain valid statistical inferences is subjective. More discussion on this matter is given by Wacholder *et al.* [33]. Indeed, it is possible that methods correcting for the effect of misclassification that make use of information gathered by comparing with a benchmark scorer might introduce more bias than they are correcting. The effect of the violation of the assumption on the corrected estimate depends on the magnitudes of the errors in the two measures and on their correlation. In particular, when the errors are negatively correlated, independent, or weakly positively correlated, the corrected estimate will tend to overcorrect beyond the true value. In this work, we are more confident because the benchmark scorer had vast experience in CE surveys and followed a special training.

Although the autologistic framework is widely used to model spatially correlated binary data, the statistical inference is wrought with complications due to the intractable normalizing constant. Hence, likelihood approaches cannot be directly implemented. The consequence becomes severe for large data sets [11, 17, 34]. We avoided these immense computational complexities encountered in the various propositions that tries to handle the normalizing constant by using the straightforward PL framework. Because the PL is an approximation to the full likelihood, the methodology may not be considered a full Bayes approach, but a 'working/pseudo' Bayes approach, providing us with a computationally feasible scheme in this multilevel framework. The relevant computations can also be carried out in JAGS or WinBUGS software and can be easily tuned to adapt to other data sets by any practitioner. Extensions to include temporal/longitudinal patterns of data are also feasible with some modifications.

Results from this analysis advances the subject area knowledge in various directions. We showed that there is spatial referencing among CE statuses of surfaces of the same tooth and for surfaces from adjacent teeth. Another interesting finding is that the scoring behavior of examiners also exhibits some degree of spatial dependency. This information could be useful in training dental examiners to promote independent examination of CE for clustered multilevel data.

Acknowledgments

This investigation was supported by research grant OT/05/60, KU Leuven; data collection was supported by Unilever, Belgium. The Signal Tandmobiel project has the following partners: D. Declerck (Department of Oral Health Sciences, KU Leuven), L. Martens (Dental School, University Ghent), J. Vanobbergen (Dental School, University Ghent), P. Bottenberg (Dental School, University Brussels), E. Lesaffre (Department of Biostatistics, Erasmus Medical Center, Rotterdam, the Netherlands and L-Biostat, KU Leuven), and K. Hoppenbrouwers (Youth Health Department, KU Leuven (University of Leuven); and Flemish Association for Youth Health Care). Bandyopadhyay acknowledges funding support from grants P20RR017696-06, R03-DE020114, and R03-DE021762 from the US National Institutes of Health.

References

1. World Health Organization. Oral health surveys. Basic methods. World Health Organization; Geneva: 1997.
2. Pine CM, Pitts NB, Nugent ZJ. British Association for the Study of Community Dentistry (BASCD) guidance on the statistical aspects of training and calibration of examiners for surveys of child dental health. A BASCD coordinated dental epidemiology programme quality standard. *Community Dent Health*. 1997; 14(Suppl 1):18–29.
3. International Caries Detection and Assessment System Coordinating Committee. Criteria manual - International Caries Detection and Assessment System (ICDAS II). 2005
4. Assaf AV, Meneghim MC, Zanin L, Mialhe FL, Pereira AC, Ambrosano GM. Assessment of different methods for diagnosing dental caries in epidemiological surveys. *Community Dent Oral Epidemiology*. 2004; 32:418–425.
5. Heifetz SB, Brunelle JA, Horowitz HS, Leske GS. Examiner consistency and group balance at baseline of a caries clinical trial. *Community Dent Oral Epidemiology*. 1985; 13:82–85.
6. Zhang Y, Todem D, Kim K, Lesaffre E. Bayesian latent variable models for spatially correlated tooth-level binary data in caries research. *Statistical modelling*. 2011; 11(1):25–47. [PubMed: 22701345]
7. Bandyopadhyay D, Reich BJ, Slate EH. Bayesian modeling of multivariate spatial binary data with applications to dental caries. *Statistics in Medicine*. 2009; 28:3492–3508. [PubMed: 19902498]
8. Besag JE. Nearest-neighbor systems and autologistic model for binary data. *Journal of Royal Statistical Society, Series B*. 1972; 34:75–83.
9. Besag J. Spatial interaction and the statistical analysis of lattice systems. *Journal of Royal Statistical Society, Series B*. 1975; 2:192–236.
10. Besag J, Kooperberg C. On conditional and intrinsic autoregressions. *Biometrika*. 1995; 82(4): 733–746.
11. Cressie, NA. *Statistics for Spatial Data*. John Wiley & Sons; New York: 1993.
12. Besag J, Tantrum J. *Likelihood analysis of binary data in space and time*. Oxford Statistical Science Series. 2003; 9A:289–294.
13. Guyon, X. *Random Fields on a Network*. Springer; Berlin: 1995.
14. Vanobbergen J, Martens L, Lesaffre E, Declerck D. The Signal Tandmobiel® project, a longitudinal intervention health promotion study in Flanders (Belgium): Baseline and first year results. *European Journal of Paediatric Dentistry*. 2000; 2:87–96.

15. Leroy R, Bogaerts K, Lesaffre E, Declerck D. Effect of caries experience in primary molars on cavity formation in the adjacent permanent first molar. *Caries Research*. 2005; 39:342–349. [PubMed: 16110204]
16. Caragea CP, Kaiser MS. Autologistic models with interpretable parameters. *Journal of Agriculture, Biological, and Environmental Statistics*. 2009; 14(3):281–300.
17. Huffer WF, Wu H. Markov chain Monte Carlo for autologistic regression models with application to the distribution of plant species. *Biometrics*. 1998; 54:509–524.
18. Banerjee, S., Gelfand, AE., Carlin, BP. *Hierarchical Modeling and Analysis for Spatial Data*. Vol. 101. Chapman and Hall/CRC; London/CRC: 2003.
19. Lawson, AB. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. Vol. 32. Chapman & Hall; Boca Raton, FL: 2009.
20. Dormann FC, McPherson MJ, Araújo BM, Bivand R, Bolliger J, Carl G, Davies GR, Hirzel A, Jetz W, Daniel Kissling W, et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*. 2007; 30(5):609–628.
21. Gelman A, Meng X. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statistical Science*. 1998; 13(2):163–185.
22. Møller J, Pettitt A, Reeves R, Berthelsen K. An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika*. 2006; 93(2):451–458.
23. Lesaffre, E., Lawson, AB. *Bayesian Biostatistics*. John Wiley & Sons; New York: 2012.
24. Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*. 2006; 1(3):515–534.
25. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2002; 64(4): 583–639.
26. Kass RE, Raftery AE. Bayes factors. *Journal of the American Statistical Association*. 1995; 90(430):773–795.
27. Vehtari A, Lampinen J. Bayesian model assessment and comparison using cross-validation predictive densities. *Neural Computation*. 2002; 14(10):2439–2468. [PubMed: 12396570]
28. Gelman, A., Carlin, JB., Stern, HS., Rubin, DB. *Bayesian Data Analysis*. 2. Chapman & Hall/CRC; New York: 2004.
29. Plummer, M. JAGS: A program for analysis Bayesian graphical models using Gibbs sampling. In: Hornik, K., Leisch, F., Zeileis, A., editors. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC, 2003)*; Vienna, Austria: Technische Universität Wien; 2003. <http://www.ci.tuwien.ac.at/Conferences/DSC.html>
30. Mwalili S, Lesaffre E, Declerck D. A Bayesian ordinal logistic regression model to correct for inter-observer measurement error in a geographical oral health study. *Journal of the Royal Statistical Society, Series C*. 2005; 54:77–93.
31. Dormann C. Assessing the validity of autologistic regression. *Ecological Modeling*. 2007; 207:234–242.
32. Hughes J, Haran M, Caragea CP. Autologistic models for binary data on a lattice. *Environmetrics*. 2011; 22(7):857–871.
33. Wacholder SBA, Hartge P. Validation studies using an alloyed gold standard. *American Journal of Epidemiology*. 1993; 137:1251–1258. [PubMed: 8322765]
34. Weir, I., Pettitt, A. Alternative to the autologistic model using a hidden conditional autoregressive Gaussian process. *Proceedings 28th Symposium on the Interface, SISC-96.(a)(b)*; Citeseer. 1996.

Figure 1.
Representation of the two types of spatial dependency for caries experience data.

Figure 2. Plots of examiner-specific spatial correlations and 95% equal tail credible intervals (CIs) for proximal surfaces (γ_1) and nonproximal surfaces (γ_2) from the validation data.

Figure 3. Adjacent matrices for different values of spatial dependency; $\square$ represents the disease-free site and $\blacksquare$ represents the presence of the disease.

Figure 4. Density plots of simulated results from logistic, autologistic, and corrected autologistic models ($SE = 0.8$, $SP = 0.8$, $\gamma = 0.1$).

Figure 5.

Density plots of simulated results from logistic, autologistic, and corrected autologistic models ($SE = 0.8$, $SP = 0.8$, $\gamma = 0.6$).

Figure 6. Density plots of simulated results from logistic, autologistic, and corrected autologistic models ($SE = 0.8$, $SP = 0.8$, $\gamma = 1.0$).

Table 1

Tooth surfaces neighbor structure.

Neighbor surfaces for five-surfaced teeth						
Surface	Buccal	Lingual	Distal	Mesial	Occlusal	Distal ²
Buccal	-		✓	✓	✓	
Lingual		-	✓		✓	
Distal	✓	✓	-		✓	
Mesial	✓	✓		-	✓	✓
Occlusal	✓	✓	✓		-	
Neighbor surfaces for four-surfaced teeth						
Surface	Buccal	Lingual	Distal	Mesial		Distal ²
Buccal	-	✓	✓	✓		
Lingual	✓	-	✓	✓		
Distal	✓	✓	-			
Mesial	✓	✓		-		✓

✓ Indicates a neighbor.

¹ Indicates that the mesial surface on the next tooth is a neighbor.

² Indicates that the distal surface on the next tooth is a neighbor.

- Indicates that a surface cannot be a neighbor to itself.

Table II

Model comparison by using deviance information criterion (DIC), cross-validation and L-measure for the main data.

Model	DIC	CV0	CV1
Logistic	2402.800	29.827	31.518
Multilevel logistic	2517.012	27.534	28.071
Multilevel autologistic	2017.613	24.783	18.214
Corrected multilevel autologistic	1979.312	24.551	28.915

Table III

Parameter estimates, standard deviation and 95% equal tail credible intervals (CIs) for standard logistic, multilevel, and multilevel autologistic models for the main data.

Parameter	Standard logistic			Multilevel			Multilevel autologistic		
	Estimate (SD)	95% CI		Estimate (SD)	95% CI		Estimate (SD)	95% CI	
FIXED EFFECTS:									
Intercept	-2.519(0.136)	[-2.797; -2.259]		-3.997(0.417)	[-4.827; -3.182]		-3.541(0.249)	[-4.047; -3.073]	
Gender									
Girls	0.099(0.124)	[-0.141; 0.341]		0.412(0.431)	[-0.408; 1.276]		0.208(0.253)	[-0.274; 0.723]	
Boys	—	—		—	—		—	—	
Age	-0.191(0.067)	[-0.317; -0.058]		-0.249(0.218)	[-0.691; 0.176]		-0.151(0.130)	[-0.413; 0.101]	
Dentition Type									
Permanent	-1.823(0.125)	[-2.073; -1.577]		-2.611(0.240)	[-3.101; -2.155]		-1.474(0.172)	[-1.817; -1.145]	
Deciduous	—	—		—	—		—	—	
Jaw									
Upper	0.089(0.126)	[-0.153; 0.338]		0.239(0.190)	[-0.121; 0.617]		0.170(0.136)	[-0.095; 0.435]	
Lower	—	—		—	—		—	—	
Quadrant									
Right	0.429(0.120)	[0.197; 0.663]		0.594(0.191)	[0.222; 0.974]		0.335(0.135)	[0.072; 0.600]	
Left	—	—		—	—		—	—	
Geographical location									
x-coordinate	-0.010(0.062)	[-0.131; 0.111]		0.114(0.229)	[-0.327; 0.573]		0.054(0.133)	[-0.202; 0.328]	
y-coordinate	-0.150(0.062)	[-0.271; -0.028]		-0.178(0.206)	[-0.578; 0.230]		-0.123(0.125)	[-0.369; 0.127]	
SPATIAL EFFECTS:									
γ_1							1.100(0.106)	[0.892; 1.302]	
γ_2							1.202(0.323)	[0.553; 1.821]	
RANDOM EFFECTS:									
σ_m^2				2.811(0.811)	[1.585; 4.679]		0.815(0.324)	[0.328; 1.583]	
$a2_{im}$				2.312(0.466)	[1.485; 3.308]		0.024(0.065)	[0.001; 0.138]	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Parameter	Standard logistic			Multilevel			Multilevel autologistic		
	Estimate (SD)	95% CI		Estimate (SD)	95% CI		Estimate (SD)	95% CI	
σ_e^2				0.099(0.352)	[0.000; 1.157]		0.066(0.130)	[0.000; 0.371]	

Table IV

Parameter estimates, standard deviation, and 95% equal tail credible intervals (CIs) for standard logistic, multilevel, and multilevel autologistic models for the validation data.

Parameter	Standard logistic			Multilevel			Multilevel autologistic		
	Estimate (SD)	95% CI		Estimate (SD)	95% CI		Estimate (SD)	95% CI	
FIXED EFFECTS:									
Intercept	-4.604(0.229)	[-5.089; -4.172]		-7.595(0.577)	[-8.783; -6.562]		-7.381(0.515)	[-8.410; -6.409]	
Gender									
Girls	-0.549(0.087)	[-0.720; -0.373]		-0.762(0.328)	[-1.412; -0.130]		-0.723(0.294)	[-1.312; -0.150]	
Boys	—	—		—	—		—	—	
Dentition Type									
Permanent	-0.769(0.093)	[-0.951; -0.592]		-1.036(0.235)	[-1.486; -0.582]		-0.933(0.228)	[-1.383; -0.491]	
Deciduous	—	—		—	—		—	—	
Tooth Type									
Canine	0.960(0.253)	[0.478; 1.476]		1.371(0.494)	[0.448; 2.403]		1.292(0.460)	[0.409; 2.207]	
Pre(molar)	2.017(0.215)	[1.625; 2.461]		3.219(0.429)	[2.449; 4.148]		2.879(0.390)	[2.124; 3.669]	
Incisor	—	—		—	—		—	—	
Jaw									
Upper	-0.182(0.086)	[-0.341; -0.010]		-0.308(0.202)	[-0.701; 0.074]		-0.241(0.192)	[-0.618; 0.136]	
Lower	—	—		—	—		—	—	
Quadrant									
Right	-0.246(0.100)	[-0.435; -0.051]		-0.267(0.237)	[-0.730; 0.185]		-0.239(0.230)	[-0.692; 0.213]	
Left	—	—		—	—		—	—	
RANDOM EFFECTS:									
σ_m^2				1.665(0.471)	[0.902; 2.722]		1.218(0.386)	[0.589; 2.092]	
a_{im}^2				4.538(0.605)	[3.466; 5.859]		4.103(0.551)	[3.114; 5.284]	
DIC	5148.213			4229.123			3892.617		

DIC, deviance information criterion.

Table V

Parameter estimates, standard deviation and 95% equal tail credible intervals (CIs) for multilevel autologistic model for the main data corrected for misclassification.

Parameter	Corrected multilevel autologistic	
	Estimate (SD)	95% CI
FIXED EFFECTS:		
Intercept	−3.616(0.345)	[−4.325; −2.961]
Gender		
Girls	−0.107(0.386)	[−0.893; 0.644]
Boys	—	—
Age	−0.138(0.200)	[−0.555; 0.252]
Dentition Type		
Permanent	−2.188(0.290)	[−2.783; −1.663]
Deciduous	—	—
Jaw		
Upper	0.203(0.203)	[−0.203; 0.587]
Lower	—	—
Quadrant		
Right	0.476(0.199)	[0.101; 0.867]
Left	—	—
Geographical location		
x-coordinate	0.027(0.205)	[−0.346; 0.443]
y-coordinate	−0.196(0.202)	[−0.619; 0.174]
SPATIAL EFFECTS:		
γ_1	1.406(0.161)	[1.106; 1.739]
γ_2	1.350(0.411)	[0.527; 2.144]
RANDOM EFFECTS:		
σ_m^2	1.636(0.795)	[0.576; 3.423]
$\alpha_{2,m}$	0.061(0.054)	[0.002; 0.190]
σ_e^2	0.152(0.239)	[0.000; 0.837]

Table VI

Simulation results: bias (95% coverage) of logistic, autologistic, and corrected autologistic models for different sensitivity (SE), specificity (SP), and spatial parameters.

Model	$SE = 0.8, SP = 0.8, \gamma = 0.1$				$SE = 0.7, SP = 0.7, \gamma = 0.1$			
	β_0	β_1	β_2	γ	β_0	β_1	β_2	γ
Logistic	0.207 (67.22)	0.447 (55.71)	0.447 (55.38)	—	0.193 (65.27)	0.632 (29.08)	0.627 (29.89)	—
Corrected logistic	0.147 (91.73)	0.066 (98.14)	0.071 (97.91)	—	0.128 (94.89)	0.106 (97.58)	0.117 (97.28)	—
Autologistic	0.127 (91.41)	0.475 (52.91)	0.475 (52.64)	0.035 (89.75)	0.141 (84.26)	0.650 (28.50)	0.645 (28.71)	0.048 (80.08)
Corrected autologistic	0.006 (98.41)	0.013 (98.40)	0.019 (98.28)	0.003 (98.37)	0.007 (97.86)	0.071 (97.49)	0.083 (97.27)	0.005 (97.61)
$SE = 0.8, SP = 0.8, \gamma = 0.6$								
Logistic	1.136 (0.24)	0.093 (95.17)	0.088 (95.19)	—	1.141 (0.42)	0.275 (77.52)	0.282 (76.05)	—
Corrected logistic	1.060 (16.44)	0.385 (91.73)	0.397 (91.30)	—	1.057 (48.77)	0.410 (95.73)	0.399 (95.48)	—
Autologistic	0.576 (32.09)	0.294 (79.69)	0.291 (80.30)	0.261 (7.98)	0.7799 (14.03)	0.405 (61.43)	0.412 (59.49)	0.364 (1.65)
Corrected autologistic	0.015 (98.24)	0.000 (98.41)	0.009 (98.25)	0.008 (98.06)	0.0570 (97.83)	0.012 (97.66)	0.005 (97.32)	0.032 (97.59)
$SE = 0.8, SP = 0.8, \gamma = 1.0$								
Logistic	1.762 (0.00)	0.193 (87.85)	0.189 (88.52)	—	1.841 (0.00)	0.025 (95.12)	0.021 (94.94)	—
Corrected logistic	1.584 (1.60)	0.567 (72.32)	0.559 (72.94)	—	1.587 (15.52)	0.570 (85.69)	0.567 (87.07)	—
Autologistic	0.968 (1.29)	0.151 (93.06)	0.153 (93.51)	0.461 (0.00)	1.329 (0.00)	0.197 (87.53)	0.205 (86.61)	0.638 (0.00)
Corrected autologistic	0.0419	0.0030	0.0017	0.0218	0.126	0.018	0.007	0.058

	γ	β_2	β_1	β_0	γ	β_2	β_1	β_0	Model
$T0 = \gamma, \gamma, SP = 0.7, SE = 0.7, SP = 0.8, \gamma, SP = 0.8, SE = 0.8, SP = 0.8, \gamma, SP = 0.8, SE = 0.8, SP = 0.8$									
	(97.07)	(96.55)	(97.59)	(97.53)	(0.9781)	(98.35)	(98.39)	(97.81)	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table VII

Simulation results: bias (95% coverage) of logistic, autologistic and corrected autologistic models using sensitivity (SE), specificity (SP), and spatial parameters from the Signal Tandmobiel® study.

Model	SE = 0.79, SP = 0.99, $\gamma = 1.35$				SE = 0.79, SP = 0.99, $\gamma = 1.40$			
	β_0	β_1	β_2	γ	β_0	β_1	β_2	γ
Logistic	1.057 (31.90)	0.226 (76.05)	0.200 (76.25)	—	1.012 (37.30)	0.138 (80.60)	0.117 (81.05)	—
Corrected logistic	1.267 (28.40)	0.266 (75.40)	0.237 (75.75)	—	1.190 (34.25)	0.172 (80.15)	0.148 (80.70)	—
Autologistic	0.040 (99.20)	0.028 (99.55)	0.020 (99.65)	0.272 (55.50)	0.041 (99.30)	0.020 (99.75)	0.03269 (99.75)	0.261 (67.00)
Corrected autologistic	0.036 (99.80)	0.010 (99.65)	0.028 (99.90)	0.003 (99.30)	0.0275 (99.65)	0.002 (99.90)	0.020 (99.75)	0.033 (99.65)