

RESEARCH ARTICLE

A Bayesian approach for analysis of ordered categorical responses subject to misclassification

Ashley Ling^{1*}, El Hamidi Hay², Samuel E. Aggrey^{3,4}, Romdhane Rekaya^{1,3,5}

1 Department of Animal and Dairy Science, University of Georgia, Athens, Georgia, United States of America, **2** USDA Agricultural Research Service, Fort Keogh Livestock and Range Research Laboratory, Miles City, Montana, United States of America, **3** Institute of Bioinformatics, University of Georgia, Athens, Georgia, United States of America, **4** Department of Poultry Science, University of Georgia, Athens, Georgia, United States of America, **5** Department of Statistics, University of Georgia, Athens, Georgia, United States of America

* Asling@uga.edu



OPEN ACCESS

Citation: Ling A, Hay EH, Aggrey SE, Rekaya R (2018) A Bayesian approach for analysis of ordered categorical responses subject to misclassification. PLoS ONE 13(12): e0208433. <https://doi.org/10.1371/journal.pone.0208433>

Editor: Yong Deng, Southwest University, CHINA

Received: April 20, 2018

Accepted: November 10, 2018

Published: December 13, 2018

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: The full simulation data and programs used to generate the data can be found at http://nce.ads.uga.edu/~romdhane/data/validate_results_12132017_c/.

Funding: AL was funded by the United States Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) through the National Needs Grant, grant number 11754154 to RR, <https://nifa.usda.gov/>. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Ordinal categorical responses are frequently collected in survey studies, human medicine, and animal and plant improvement programs, just to mention a few. Errors in this type of data are neither rare nor easy to detect. These errors tend to bias the inference, reduce the statistical power and ultimately the efficiency of the decision-making process. Contrarily to the binary situation where misclassification occurs between two response classes, noise in ordinal categorical data is more complex due to the increased number of categories, diversity and asymmetry of errors. Although several approaches have been presented for dealing with misclassification in binary data, only limited practical methods have been proposed to analyze noisy categorical responses. A latent variable model implemented within a Bayesian framework was proposed to analyze ordinal categorical data subject to misclassification using simulated and real datasets. The simulated scenario consisted of a discrete response with three categories and a symmetric error rate of 5% between any two classes. The real data consisted of calving ease records of beef cows. Using real and simulated data, ignoring misclassification resulted in substantial bias in the estimation of genetic parameters and reduction of the accuracy of predicted breeding values. Using our proposed approach, a significant reduction in bias and increase in accuracy ranging from 11% to 17% was observed. Furthermore, most of the misclassified observations (in the simulated data) were identified with a substantially higher probability. Similar results were observed for a scenario with asymmetric misclassification. While the extension to traits with more categories between adjacent classes is straightforward, it could be computationally costly. For traits with high heritability, the performance of the methodology would be expected to improve.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Error in the measurement and recording of data, especially discrete data, is a prevalent issue that has the potential to affect the results of any analysis through the reduction of the statistical power and the increase in bias [1–6]. While measurement errors can occur in continuous and categorical variables, it has been shown that their impact on inference is more severe in discrete data [7]. When error occurs in categorical variables it is commonly called misclassification and is a topic that has drawn considerable research due to its negative effect on the inference. More specifically, misclassification occurs with the assignment of a discrete response to a class that does not accurately reflect its true state [8]. For complex and heterogeneous diseases, several subclasses with varying phenotypes often exist. Using typical diagnostic technologies, the different subclasses can be difficult to differentiate for various reasons; samples may look alike under microscopic analysis, individuals may share the same symptoms or markers used in the diagnostic, or one or more subclasses of the disease may simply be unknown [5]. This similarity can consequently lead to misdiagnosis [9]. It is well documented in the literature that misclassification in a response variable will have two primary inferential implications: 1) biased parameter estimates and 2) a reduction of statistical power [1,6,10–11]. Practically, this will manifest as inaccurate estimates of the model parameters and a reduced ability to identify truly influential explanatory factors. In recent years, there has been increased discussion of a lack of reproducible research across a wide range of fields. While many factors are likely to contribute to irreproducibility, inaccurate measurements have been implicated as a potential culprit [12–14]. Error can be introduced at various stages of data collection and analysis, and it has been estimated that error rates in many datasets can be expected to fall between 1–5% if measures to prevent them are not taken [15]. For human disease data, much higher error rates have been reported. In a 2010 report by the Pennsylvania Patient Safety Authority, diagnostic errors in acute care ranged between 3.1 and 49.8% [16].

The challenges presented by misclassification are not overcome by a large sample size as might be expected [17–18]. Therefore, as “Big Data” becomes more accessible and utilized, it will be essential to consider the potential for misclassification in these datasets [19–20]. It has even been suggested that some large datasets may be more susceptible to misclassification depending on the method and/or purpose of collection, such as electronic medical records [20].

Given the negative effects of misclassification on inference, ignoring its presence during analysis would not be a reasonable approach. A variety of methods for dealing with misclassification in a response have been reported [21–23]. The majority of these methods rely on the estimation of misclassification rates (e.g. false positives and false negatives). A double sampling approach has been suggested [10,24] that estimates rates of misclassification through the use of two tests: 1) an inexpensive fallible classifier for the full sample and 2) a more expensive infallible classifier (or “gold standard”) for a partial sample. In many situations, though, a perfect or even high accuracy classifier will either be inaccessible or nonexistent. Thus, misclassification has become largely a statistical problem and several methods for addressing it have been presented over the years. Several likelihood-based methods have been suggested for dealing with misclassification [3–4, 25–29]. Furthermore, [30] presented an elegant parametric and semi-parametric approach for analyzing speaking fluency of immigrants that contemplates errors due to subjective evaluations and scaling across individuals. Although these likelihood-based approaches are able to handle the misclassification problem, they seem not to perform well when the misclassification probabilities are small.

Advances in Markov Chain Monte Carlo (MCMC) simulation techniques have greatly facilitated the use of Bayesian methods in misclassification problems. Hierarchical Bayesian modeling with conjugate informative priors has become a method of choice for

analysis of misclassified binary data. Several studies by our group [5–6,11] and others [31–32] have been successful in dealing with misclassification errors in binary data under various sampling models. However, limited research has been carried out to scale those methods to the multinomial response situation. In this study, the methods presented by our group to deal with misclassified binary traits will be extended to multinomial response situations using a hierarchical Bayesian framework. Although our proposed approach was developed and implemented before [33] published their paper on dealing with unidirectional misclassification of ordinal covariates, both methods overlap on their approach to the problem. However, our method is unique as it addresses misclassification on response variables and it allows for asymmetric misclassification probabilities within a full Bayesian implementation.

Materials and methods

Let $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ be a vector of ordered categorical responses collected on n individuals with y_i taking the value of one of C mutually exclusive discrete responses. This observed data is considered to be a noisy representation of true unobserved data, $\mathbf{r} = (r_1, r_2, \dots, r_n)'$. The distribution of y_i , conditional on the probability of observing each response, is given by

$$p(y_i|\theta_i^*) = \prod_{k=0}^{C-1} (\theta_{ik}^*)^{I(y_i=k)}, \tag{1}$$

where $\theta_i^* = (\theta_{i0}^*, \theta_{i1}^*, \dots, \theta_{i(C-1)}^*)'$ is the vector of probabilities that observation y_i for individual i is observed as class k , $p(y_i = k)$, taking a value for each of the C discrete classes of the noisy data; and $I(y_i = k)$ is an indicator function that takes the value of one if $(y_i = k)$, otherwise equaling zero.

Obviously, the observed probabilities of the multinomial distribution in [1] are linear combinations of the true probabilities, θ_{ij} , and the misclassification probabilities, π_{jk} , which is the probability that an observation with true class j is classified as observed class k . Misclassification will occur when j does not equal k . Thus, for class k , θ_{ik}^* could be presented as

$$\theta_{ik}^* = \pi_{0k}\theta_{i0} + \pi_{1k}\theta_{i1} + \dots + \pi_{kk}\theta_{ik} + \dots + \pi_{(c-1)k}\theta_{i(c-1)}, \tag{2}$$

where θ_{ik}^* is the probability of observing class k for individual i , θ_{ij} is the true probability of observing the discrete response class j for individual i , and π_{jk} is the conditional probability of observing class k when the true class is j . When j is different from k , π_{jk} is the misclassification probability from true class j to observed class k . Thus,

$$\theta_{ik}^* = [\pi_{0k} \pi_{1k} \dots \pi_{kk} \dots \pi_{(c-1)k}] \theta_i = \boldsymbol{\pi}_k \theta_i. \tag{3}$$

In matrix notation and generalizing to all elements of the vector, θ_i^* , Eqs [2 and 3] can be rewritten as

$$\theta_i^* = \begin{bmatrix} \pi_{00} & \dots & \pi_{(c-1)0} \\ \vdots & \ddots & \vdots \\ \pi_{0(c-1)} & \dots & \pi_{(c-1)(c-1)} \end{bmatrix} \theta_i = \begin{bmatrix} \boldsymbol{\pi}_0 \\ \vdots \\ \boldsymbol{\pi}_{(c-1)} \end{bmatrix} \theta_i = \mathbf{P} \theta_i, \tag{4}$$

where the matrix \mathbf{P} assembles the conditional probabilities, π_{jk} , between all classes. The vectors θ_i^* and θ_i are specific to each individual and they could be modeled as a function of systematic and random effects. However, the matrix \mathbf{P} is common to all individuals. The matrix \mathbf{P} does not have to be symmetric and the only requirement is that columns sum to 1.

In the absence of misclassification, the matrix \mathbf{P} will be equal to the identity matrix, as expected, and $\theta^*_i = \theta_i$.

Assuming conditional independence, the joint distribution of the true data, \mathbf{r} , conditional on the vector of true probability, $\theta_i = (\theta_{i0}, \theta_{i1}, \dots, \theta_{i(C-1)})'$ is given by

$$p(\mathbf{r}|\theta) = \prod_{i=1}^n \prod_{j=0}^{C-1} (\theta_{ij})^{I(r_i=j)}, \tag{5}$$

where $I(r_i = j)$ is an indicator function that takes the value of one if $(r_i = j)$ and otherwise is equal to zero.

Let $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)'$, where α_i is an unknown random variable indicating if the observed discrete response for individual i arose from a switching (misclassification) event. Assuming that the observed (and potentially misclassified) discrete response of individual i is k ($k \in \{0, 1, \dots, C-1\}$), then α_i can be modeled following a multinomial process with probability vector $\pi_k = (\pi_{0k}, \pi_{1k}, \dots, \pi_{jk}, \dots, \pi_{(C-1)k})$,

$$p(\alpha|\mathbf{P}) = \prod_{i=1}^n \prod_{j=0}^{C-1} (\pi_{jk})^{I(\alpha_i=h)}, \tag{6}$$

where h is an index taking C discrete values and $I(\alpha_i = h)$ is an indicator function that takes the value of one if $(\alpha_i = h)$ and otherwise is equal to zero.

The joint distribution of α and \mathbf{r} given θ and π can be written as

$$p(\mathbf{r}, \alpha|\theta, \mathbf{P}) = \prod_{i=1}^n \prod_{j=0}^{C-1} (\pi_{jk})^{I(\alpha_i=h)} (\theta_{ij})^{I(r_i=j)}. \tag{7}$$

The joint distribution in Eq [7] depends on the true discrete data, \mathbf{r} , which is not available. However, such data could be generated as a function of the observed contaminated discrete responses, \mathbf{y} , and the vector α . Let $f()$ be a link function that relates the true discrete data to the noisy responses and the vector of indicator variables α

$$r_i = f(y_i, \alpha_i). \tag{8}$$

For example, if the observed and potentially noisy discrete response for individual i is $y_i = 0$ then the true unobserved response, r_i , can be generated using the relationship

$$r_i = (1 - \alpha_i)y_i + \alpha_i(1 - y_i),$$

with α_i taking the value of 0 (no misclassification), 1 (switching from $r_i = 1$ to $y_i = 0$) and 2 (switching from $r_i = 2$ to $y_i = 0$). A similar relationship could be built when the observed discrete response y_i is different from zero.

Using the relationship in Eq [8], the joint distribution in Eq [7] could be re-written as a function of the observed data,

$$p(\mathbf{y}, \alpha|\theta, \mathbf{P}) = \prod_{i=1}^n \prod_{j=0}^{C-1} (\pi_{jk})^{I(\alpha_i=h)} (\theta_{ij})^{I(f(y_i, \alpha_i)=j)}, \tag{9}$$

where $I(f(y_i, \alpha_i) = j)$ is an indicator function that takes the value of one if $(f(y_i, \alpha_i) = j)$ and otherwise is equal to zero.

It is often the case that the probability of observing a specific outcome of the multinomial process, θ_{ij} , is a function of a set of systematic (β) and random (u) effects. Thus, from hereafter θ_{ij} will be denoted by $\theta_{ij}(\beta, u)$ to indicate this relationship.

To finalize the Bayesian formulation, prior distributions are needed for the unknown parameters in the model. For the systematic and random effects the following priors were

specified:

$$p(\boldsymbol{\beta} | \boldsymbol{\beta}_{min}, \boldsymbol{\beta}_{max}) \sim U[\boldsymbol{\beta}_{min}, \boldsymbol{\beta}_{max}], \tag{10}$$

$$p(\mathbf{u} | \mathbf{A}, \sigma_u^2) \sim N(0, \mathbf{A}\sigma_u^2). \tag{11}$$

where $\boldsymbol{\beta}_{min}$ and $\boldsymbol{\beta}_{max}$ are known hyperparameters, \mathbf{A} is a known symmetric covariance matrix between the elements of the vector \mathbf{u} , and σ_u^2 is an unknown dispersion component for which a scaled inverse chi-squared with ν degrees of belief and scaling factor s^2 was used as the prior.

$$p(\sigma_u^2 | \nu, s^2) \sim \chi^{-2}(\nu, s^2) \tag{12}$$

$\boldsymbol{\beta}_{min}$ and $\boldsymbol{\beta}_{max}$ are hyperparameters to be specified by the researcher based on published estimates in the literature, based on the researcher’s own experience of the range of values the parameter might take, or by choosing an interval large enough to contain all reasonable values of the parameter.

For the misclassification probabilities vector, $\boldsymbol{\pi}_k$, a Dirichlet distribution was assumed,

$$p(\boldsymbol{\pi}_k | \boldsymbol{\tau}) = \frac{1}{B(\boldsymbol{\tau}_k)} \prod_{j=0}^{C-1} \pi_{jk}^{\tau_{jk}-1} = Di(\boldsymbol{\pi}_k; \boldsymbol{\tau}_k), \tag{13}$$

where $\boldsymbol{\tau} = (\tau_{0k}, \tau_{1k}, \dots, \tau_{(c-1)k}) > 0$ is the vector of concentration parameters of the Dirichlet distribution, $B(\boldsymbol{\tau}_j)$ is the multivariate Beta function that plays the role of a normalizing constant, and π_{jk} is the probability of misclassification (switching) from discrete response class j to class k as defined previously in Eqs [2] and [3].

The joint posterior distribution of all unknowns in the models is proportional to the product of the expressions in Eqs [9–13],

$$p(\boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\alpha}, \mathbf{P} | \mathbf{y}) \propto \prod_{i=1}^n \prod_{j=0}^{C-1} (\pi_{jk})^{I(\alpha_i=h)} [\theta_{ik}(\boldsymbol{\beta}, \mathbf{u})]^{I(f(y_i, \alpha_i)=j)} \cdot \left(\sigma_u^{-\frac{q+\nu}{2}} \right) \exp\left(-\frac{.5(\mathbf{u}'\mathbf{A}^{-1}\mathbf{u} + s^2)}{\sigma_u^2} \right) \cdot \prod_{k=0}^{C-1} \frac{1}{B(\boldsymbol{\tau}_k)} \prod_{j=0}^{C-1} \pi_{jk}^{\tau_{jk}-1} \text{ for } \boldsymbol{\beta}_{min} \leq \boldsymbol{\beta} \leq \boldsymbol{\beta}_{max}, \tag{14}$$

where q is the number of elements in the vector \mathbf{u} . In genetics studies, q will be equal to the number of individuals in the pedigree file, a file listing the ID of each animal in the population along with its sire and dam ID (where a zero is used for unknown sires and dams). Inclusion of pedigree information in the model is essential as it relates information concerning genetic merit of an individual through its relatives and information on the relatives.

A data augmentation algorithm as described by [34] and [5] will facilitate the implementation of the model in Eq [14]. It consists of assuming the existence of an unknown continuous random variable, l_i , that relates to the discrete response via the following relationship:

$$p(y_i = k | l_i, t_{k-1}, t_k) = \begin{cases} 1 & \text{if } t_{k-1} < l_i \leq t_k \\ 0 & \text{otherwise} \end{cases},$$

where t_j is an arbitrary threshold value. If y_i can take a value in one of C mutually exclusive ordered categorical responses, then there will be $C+1$ thresholds ($t_{-1}, t_0, t_1, \dots, t_{C-1}$) where t_{-1} and t_{C-1} are often set equal to negative and positive infinity.

The underlying liability for individual i could be modeled as a function of the systematic and random effects,

$$l_i = \mathbf{x}_i\boldsymbol{\beta} + \mathbf{z}_i\mathbf{u} + e_i,$$

Where \mathbf{x}_i and \mathbf{z}_i are known incidence matrices relating the liability of individual i to the set of systematic and random effects, respectively, and e_i is the error term. Augmenting the model in Eq [7] with the vector of liabilities results in a lack of identifiability. To make the augmented model identifiable [34], two restrictions are needed. It was assumed that $t_0 = 0$ and $var(e_i) = 1$.

At the liability scale, the full conditional distributions of $\boldsymbol{\beta}$, \mathbf{u} and σ_u^2 are identical to those obtained in a standard analysis of discrete data in the absence of misclassification and can be found in [34], [5] and [35]. These conditional distributions, needed for a Bayesian implementation of the model via the Gibbs sampler, were in closed form, and were truncated normal for the liabilities, normal for the position parameters, and scaled inverse chi-squared for σ_u^2 .

The conditional of α_i is a multinomial distribution,

$$p(\alpha_i | \boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\alpha}_{-i}, \mathbf{P}, \sigma_u^2, \mathbf{y}) \propto \prod_{k=0}^{C-1} (\pi_{jk})^{I(\alpha_i=h)} [\theta_{ij}(\boldsymbol{\beta}, \mathbf{u})]^{I(f(y_i, \alpha_i)=k)},$$

where $\boldsymbol{\alpha}_{-i}$ is the vector $\boldsymbol{\alpha}$ without the i^{th} position.

For the misclassification probabilities, $\boldsymbol{\pi}_j$, its conditional distribution is proportional to

$$p(\boldsymbol{\pi}_k | \boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\alpha}, \mathbf{P}_{-k}, \sigma_u^2, \mathbf{y}) \propto \prod_{k=0}^{C-1} \frac{1}{B(\boldsymbol{\tau}_k)} \prod_{j=0}^{C-1} \pi_{jk}^{\tau_{jk}-1} \prod_{i=1}^{n_k} \prod_{j=0}^{C-1} (\pi_{jk})^{I(\alpha_i=h)},$$

where \mathbf{P}_{-k} is the matrix \mathbf{P} without the k^{th} row and n_k is the number of individuals with an observed discrete response equal to the k^{th} category (class).

$$p(\boldsymbol{\pi}_k | \boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\alpha}, \mathbf{P}_{-k}, \sigma_u^2, \mathbf{y}) \propto \prod_{k=0}^{C-1} \frac{1}{B(\boldsymbol{\tau}_k)} \prod_{j=0}^{C-1} \pi_{jk}^{\tau_{jk}-1} \prod_{j=0}^{C-1} (\pi_{jk})^{\gamma_{jk}}$$

$$\propto \frac{1}{B(\boldsymbol{\tau}_k)} \prod_{j=0}^{C-1} \pi_{jk}^{(\tau_{jk} + \gamma_{jk} - 1)}.$$

where γ_{jk} is the number of discrete responses switched (misclassified) from true class j to the observed class k . Thus, $\boldsymbol{\pi}_k$ is distributed as $Di(\boldsymbol{\pi}_k; \boldsymbol{\tau}_k + \boldsymbol{\gamma}_k)$ with $\boldsymbol{\gamma}_k = (\gamma_{0k}, \gamma_{1k}, \dots, \gamma_{(C-1)k})$. The vector $\boldsymbol{\gamma}_k$ is easily obtained as a by-product of the sampling process.

Data

Calving ease, a measure of how easily a cow delivers her calf, is an important trait in beef cattle production. Selection on animals with a high growth rate is of importance to the industry for the sake of yielding a higher carcass weight per animal. However, growth rate and birth weight are positively correlated traits, and so selection on a higher growth rate indirectly selects for higher birth weights. It has been found that high birth weight is the largest contributing factor in rates of dystocia, or calving difficulty, an issue that leads to several health and economic concerns, such as increased mortality for calf and cow and additional veterinary costs. It is therefore of interest for producers to consider calving ease directly as a trait in their selection models. However, without strict protocols on the farm, measuring calving ease may involve a significant degree of subjectivity.

This study was largely motivated by a real beef cattle dataset. The data consisted of calving ease and related measurements in a composite breed of beef cattle obtained from the USDA Fort Keogh Range and Research Laboratory in Miles City, Montana [36–37]. After quality control, the USDA data included 955 calving ease phenotypes (or observations). Calving ease was scored on a four-category scale in the raw data, with a score of 1 for the easiest calving and 4 for the most problematic (Caesarian section required); however, category 4 contained few observations and so categories 3 and 4 were combined into a single bin to avoid an extreme-case problem (ECP), when the discrete response is the same for all individuals within a given

systematic effect class. The three resulting categories were scored as 0 for no difficulty, 1 for moderate difficulty and 2 for severe difficulty. The distribution of calving ease scores showed an unexpected pattern of few calvings for category 1 (moderate difficulty). Given the subjectivity of the trait (scoring carried out by farm employees) and the fact that some cows calve overnight under standard ranch conditions with limited presence or absence of ranch employees, it was reasonable to assume that the collected data could be subject to potential misclassification. In order to test the adequacy of the proposed method to deal with noisy ordered categorical data and to evaluate its robustness in the presence of a small sample size, two datasets of large and small (**D1** and **D2**, respectively) sample size were simulated.

Simulated data. Both datasets (**D1** and **D2**) were simulated following the structure of a real calving ease dataset. For both datasets (**D1** and **D2**), a discrete trait with three categories ($C = 3$), similar to calving ease, was simulated. The number of animals with phenotype (n) ranged between 8,939 and 9,042 for **D1** and 1,393 and 1,425 for **D2**. The pedigree files included 10,000 and 1,563 animals for **D1** and **D2**, respectively. A mixed linear model that includes three systematic effects and one random effect in addition to the error terms was used to model the liabilities,

$$l = X\beta + Zu + e,$$

where l is a vector of liabilities of size n , β and u are vectors of fixed and random (animal) effects, respectively, e is a vector of residuals, and X and Z are known incidence matrices with the appropriate dimensions.

In this model, the random effect accounts for the animal’s additive genetic contribution to the trait and is commonly called the animal additive effect. The genetic covariance between animals is accounted for through the average relationship matrix, commonly denoted as A , which is computed based on the known pedigree information. The element a_{ij} of the matrix A is simply the expected additive relationship between animals i and j . A detailed algorithm for constructing the relationship matrix can be found in [38]. Solving the mixed model equations requires inversion of the relationship matrix. While the relationship matrix is relatively sparse, it’s much more computationally efficient to directly build the inverse of the relationship matrix, A^{-1} using simple rules derived in [39].

Each systematic effect was drawn from a normal distribution. The number of levels and distributional parameters of each systematic effect are summarized in Table 1. The random effects and residual terms were generated from the following distributions:

$$u \sim N(\mathbf{0}, A\sigma_u^2),$$

$$e \sim N(\mathbf{0}, I\sigma_e^2),$$

where A is a known matrix of expected additive relationships between animals computed

Table 1. Number of classes and distributions used to simulate the systematic effects for the large (D1) and small (D2) datasets.

Fixed Effect	Number of Levels		Distribution
	D1	D2	
1	20	5	$N(-0.7,0.05)$
2	10	5	$N(-0.1,0.5)$
3	5	5	$N(0.1,0.2)$

<https://doi.org/10.1371/journal.pone.0208433.t001>

based on the information in the pedigree files [39] and σ_u^2 and σ_e^2 were set equal to 0.1 and 1, respectively, resulting in a heritability of 0.091.

The simulated liability was discretized to generate an ordered categorical trait with three categories ($C = 3$) using the following relationship:

$$y_i = \begin{cases} 0 & t_{min} < l_i \leq t_0 \\ 1 & t_0 < l_i \leq t_1 \\ 2 & t_1 < l_i \leq t_{max} \end{cases}$$

where y_i is the discrete response for animal i , taking the value of 0, 1, or 2; and $\mathbf{t} = (t_{min} = -\infty, t_0, t_1, t_{max} = \infty)$ is a vector of threshold values with t_0 and t_1 arbitrarily set to 0 and 1, respectively.

Incidence rates of the three categories (0, 1, and 2) were on average (0.39, 0.38, and 0.23) and (0.36, 0.46, and 0.18) for **D1** and **D2**, respectively. All scenarios were replicated 10 times and ECPs were avoided.

Misclassification was introduced to the simulated data by switching a subset of observations from true class $y_i = j$ to observed class $y_i = k$ for $(j, k) \in (0, 1, 2)$ and $j \neq k$. Two scenarios of misclassification were investigated, symmetrical misclassification, where the rate of misclassification (π_{jk}) between two classes is equal in both directions ($\pi_{jk} = \pi_{kj}$), and non-symmetrical misclassification, where the rate of misclassification from class j to class k is allowed to differ from class k to class j ($\pi_{jk} \neq \pi_{kj}$).

For the symmetrical misclassification, a misclassification rate of 2.5% for each direction was introduced to the data ($\pi_{jk} = 0.025$ for all $j \neq k$). This approach yielded a non-differential net rate of misclassification in the simulated data of 5%. For the nonsymmetrical scenario, a misclassification rate was set equal to 1, 3, 1.5, 1, 0.1, and 0.1% for $\pi_{12}, \pi_{21}, \pi_{23}, \pi_{32}, \pi_{13}$, and π_{31} , respectively.

Real data. The real data consisted of calving ease records from 955 first parity cows from a Composite Gene Combination breed (CGC; 50% Red Angus, 25% Charolais, 25% Tarentaise; [36–37]) born between 2002 and 2011 at USDA-ARS, Fort Keogh Livestock and Range Research Laboratory, Miles City, MT. The pedigree file was comprised of 1,357 animals, including 82 sires and 651 dams. Calving ease was scored on a discrete scale with three categories (0 = no difficulties; 1 = moderate difficulties; and 2 = severe difficulties), where the most severe category was composed of two separate classes that were collapsed together. The systematic effects consisted of sex (2 classes), feed treatment (2 classes) and year of birth (10 classes).

Data analysis

Each of the simulated data sets for the symmetric scenario was analyzed using the following models: true simulated discrete data analyzed with a classical threshold model (M1), noisy simulated data analyzed with a classical threshold model that does not contemplate misclassification (M2), noisy simulated data analyzed with a threshold model that contemplates misclassification following our proposed method and misclassification probabilities assumed known (M3), noisy simulated data analyzed with a threshold model that contemplates misclassification following our proposed method and misclassification probabilities assumed unknown (M4). An additional analysis was carried out to test the validity of our approach under a null model (when no misclassification is present in the data). For that purpose, the true simulated data was analyzed with a threshold model that contemplates misclassification following our proposed method and misclassification probabilities assumed unknown (M5).

The simulated data generated under the non-symmetric scenario was analyzed only with M4. Similarly, the real data was analyzed only with M2.

Implementation was carried out using the Gibbs sampler. A unique chain of 200,000 samples was implemented with the first 100,000 iterations discarded as burn-in period. The probability of each observation being misclassified was calculated as the ratio between the number of times the observation was found to be misclassified and the total number of iterations.

Results

Under the symmetric scenario, the estimate of the genetic variance, for the large simulated data and in the absence of misclassification in the discrete response (M1), was virtually equal to the true value used in the simulation (0.10). Furthermore, the 95% highest posterior density (HPD) interval is well-centered around the true value, indicating a reduction of bias (Table 2). However, when 5% of the discrete responses were artificially misclassified but no measures to address misclassification taken during the analysis (M2), the genetic variance was grossly underestimated with a posterior mean of roughly half the true value used in the simulation (0.052). Additionally, the 95% HPD interval did not include the true genetic variance (0.10), indicating significant bias in the estimate of the parameter (Table 2). This result demonstrates that when appropriate measures are not taken to correct or account for misclassification, there will be substantial bias in parameter estimation even with low to moderate noise in the input data.

When the same noisy data used in M2 was analyzed using the proposed methodology that accounts for potential misclassification, bias in estimates of genetic variance was significantly reduced. This was true both when the probability of misclassification, π_{jk} , between each pair of classes j ($j = 1,2,3$) and k ($k = 1,2,3$), was assumed known and equal to the true value used in the simulation (0.025) (M3) and when it was assumed unknown and estimated during the analysis (M4). In fact, the posterior mean of genetic variance was 0.112, and 0.0998, for M3, M4, respectively. In each of these cases, the true value of the parameter (0.10) was well within the 95% HPD interval, indicating a reduction of bias.

For the non-symmetric simulation scenario, the posterior mean (SD) of the genetic variance was equal to 0.118 (0.0196) and 0.103 (0.0243) for D1 and D2, respectively. These estimates are very similar to the true value used in the simulation (0.10) and based on the 95% HPD interval (information not presented), these estimates have minimal bias.

Estimates of the probabilities of misclassification under model M4 for the symmetric and non-symmetric simulation scenarios are presented in Table 3. For the symmetric scenario, the

Table 2. Posterior means, posterior standard deviations and the 95% highest posterior density interval (HPD) of the genetic variance (true value = 0.1) under different models¹ and datasets².

	Mean		Standard Deviation		95% HPD Interval	
	D1	D2	D1	D2	D1	D2
M1	0.106	0.135	0.0255	0.0670	0.0586–0.170	0.0469–0.319
M2	0.0521	0.0919	0.0143	0.0422	0.0288–0.0846	0.0364–0.198
M3	0.112	0.151	0.0355	0.0881	0.0540–0.191	0.0460–0.380
M4	0.0998	0.106	0.0188	0.0239	0.0688–0.142	0.0681–0.161
M5	0.113	0.108	0.0226	0.0251	0.0755–0.163	0.0692–0.167

¹ M1: True data analysis with a classical threshold model, M2: Noisy data analyzed with a classical threshold model, M3: noisy data analyzed with the proposed method assuming the misclassification probability is known, M4: same as M3 except the misclassification probability is assumed unknown, M5: Noise free data analyzed using our proposed method (null model);

² D1: large dataset, D2: small dataset

<https://doi.org/10.1371/journal.pone.0208433.t002>

Table 3. Posterior mean of the misclassification probability between the different categories of the discrete responses and datasets¹.

Parameter	Symmetrical Misclassification				
	True Value	M4		M5	
		D1	D2	D1	D2
π_{12}	0.025	0.0146	0.0128	0.0112	0.0124
π_{23}	0.025	0.0135	0.0062	0.0065	0.0057
π_{13}	0.025	0.0239	0.0067	0.0019	0.0029
	Nonsymmetrical Misclassification				
	True Value	M4			D2
		D1			
π_{12}	0.01	0.010		0.0095	
π_{21}	0.03	0.013		0.0100	
π_{23}	0.015	0.011		0.0098	
π_{32}	0.01	0.011		0.0099	
π_{13}	0.001	0.0013		0.0013	
π_{31}	0.001	0.0014		0.0011	

¹ D1: large dataset, D2: small dataset; π_{ij} is the misclassification probability between categories i and j

<https://doi.org/10.1371/journal.pone.0208433.t003>

misclassification probabilities between categories 1 and 2 (π_{12}) and 2 and 3 (π_{23}) of the discrete responses were underestimated with posterior means of 0.0146 and 0.0135, respectively, compared to the true value of 0.025. However, the misclassification probability between categories 2 and 3 (π_{23}) was estimated with minimal bias and the posterior mean (0.0239) is very similar to the true value (0.025). In spite of the underestimation of some of the misclassification probabilities, the results clearly indicate the ability of the proposed methodology to adjust for the misclassification in the data whether or not the true rate of misclassification is known. Similar performance was observed using data from the non-symmetric simulation scenario. In fact, five out of the six misclassification probabilities were accurately estimated and only the misclassification probability between categories 2 and 1 (π_{21}) was under estimated (Table 3).

To further evaluate the adequacy of the proposed methodology, we tested its performance under a null model (no misclassification in the data) under the symmetric simulation scenario. Thus, when the noise free data used in M1 was reanalyzed using our proposed model that contemplates potential misclassification (M5), the estimate of the genetic variance (0.113) was similar to the true value and minimal bias was observed (Table 2). Most importantly, the estimated misclassification probabilities were negligible, indicating that very few observations were detected as potentially misclassified (Table 3). This was expected given that the data was free of noise.

In order to evaluate the ability of the proposed method to correctly identify misclassified observations, we calculated the posterior probability of misclassification of each observation in the data set. The average misclassification probability of miscoded observations was over twenty times higher than that of the non-miscoded observations. For the 95% truly non-misclassified and 5% misclassified observations, their average misclassification probability was 0.026 and 0.554, respectively. Fig 1 presents the distribution of misclassification probability for the miscoded observations (Fig 1A) and the correctly coded observations (Fig 1B) for one replicate. The 85th percentile of correctly classified observations was 0.02613, while the 15th percentile of misclassified observations was 0.0297. This is important as it shows that the algorithm was able to distinguish between the two groups and the miscoded records were detected with a high probability.

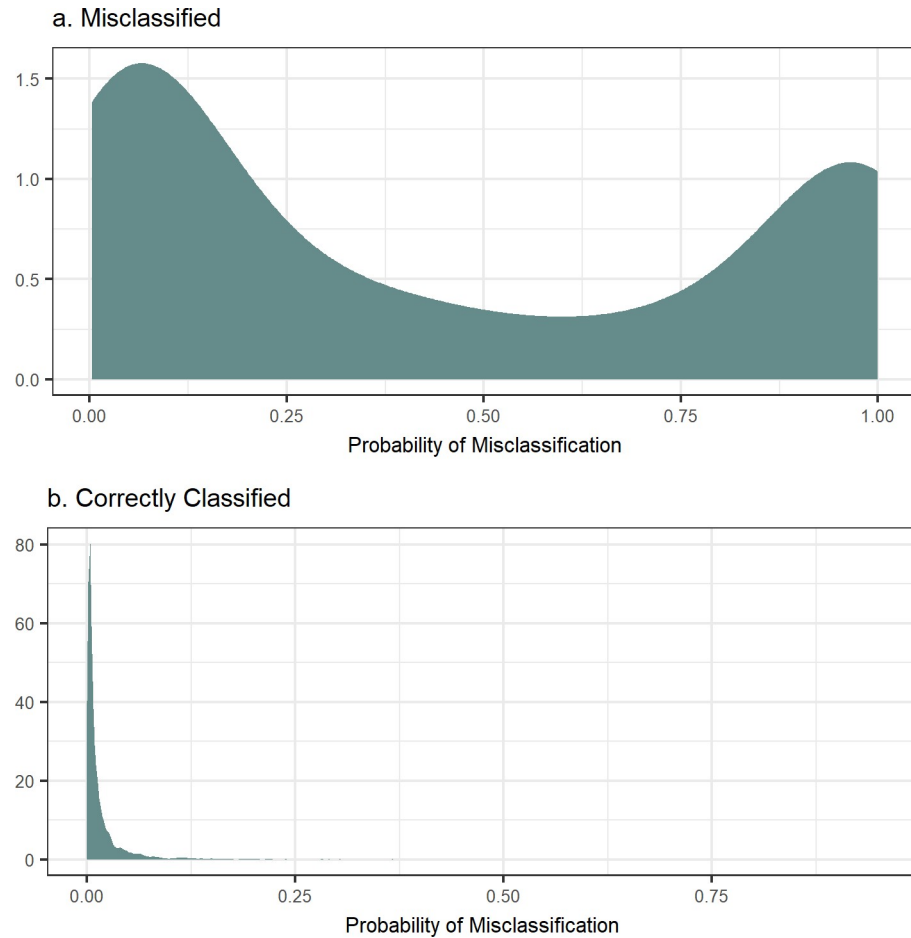


Fig 1. Misclassification probability densities for the a. miscoded observations and b. correctly coded observations.

<https://doi.org/10.1371/journal.pone.0208433.g001>

The correlation between the true and estimated breeding values under the different models for the symmetric scenario is presented in Table 4. As expected, the highest correlation (0.378) was obtained when the data was free of misclassification (M1), and the minimum (0.300) when the data was noisy and misclassification was ignored during the analysis (M2). Using the

Table 4. Pearson correlation between true and estimated breeding values under different models¹ and datasets².

	D1	D2
M1	0.378	0.370
M2	0.300	0.314
M3	0.350	0.346
M4	0.348	0.347
M5	0.377	0.370

¹ M1: True data analysis with a classical threshold model, M2: Noisy data analyzed with a classical threshold model, M3: noisy data analyzed with the proposed method assuming the misclassification probability is known, M4: same as M3 except the misclassification probability is assumed unknown, M5: Noise free data analyzed using our proposed method (null model);

² D1: large dataset, D2: small dataset

<https://doi.org/10.1371/journal.pone.0208433.t004>

noisy data and the proposed methods, the correlation was 0.350 and 0.348 for M3 and M4, respectively. Under the null model (M5), the correlation was the same as in M1. Under the non-symmetric simulation scenario, the correlation between true and estimated breeding values was equal to 0.36 and 0.356 for D1 and D2, respectively.

For the small simulated data, due to the fact that the true parameters of the model were estimated with large uncertainty, the performance of the different methods was compared to M1 as reference. This choice is motivated by the fact that results obtained using M1 (noise-free data and classical threshold model) are the best we can expect from the analysis of the small simulated data. Under the symmetric simulation scenario, methods M2, M4, and M5 underestimated the genetic variance by 32, 21, and 20% compared to M1, respectively, and M3 resulted in an overestimation of 12%. Although the proposed methods (M3, M4) for dealing with misclassification were unable to reproduce the results obtained using the noise-free data (M1), they reduced the bias by more than half compared to M2. For M4, misclassification probabilities were also underestimated with posterior means equal to 0.013, 0.007, and 0.007 for π_{12} , π_{23} , and π_{13} , respectively. Correlations between true and predicted breeding values (Table 4) followed the same trend as those for the large data set, although they were a bit smaller in magnitude.

The real data was analyzed assuming nonsymmetric misclassification probabilities between categories. The estimated genetic variance was 0.326 and 0.118 using a classical threshold model (M1) and our proposed method, respectively. Although there is no true value to compare with, estimates obtained using our method are more realistic for a trait characterized by low heritability [40–41]. Misclassification probabilities were around zero (0.001) for π_{31} and π_{13} , and approximately ten-fold higher (0.0098–0.0103) for π_{12} , π_{21} , π_{23} , and π_{32} . These results are expected given the unlikely misclassification between the third category (major calving difficulty) and the other two categories and any potential misclassification is likely the result of a posterior assessment of the calving event.

Discussions

It has been well-established in the literature that misclassification in the response variable, especially discrete responses, will lead to biased estimation of parameters in the model [1]. The results of this study showed clearly that a small noise level (5%) in an ordered categorical response (3 categories) resulted in a substantial bias in the estimation of genetic variance and breeding values. In animal and plant improvement programs such inaccurate estimates will negatively impact the efficiency of selection.

The methodology discussed here is an extension to the work carried out by our group on the analysis of binary data subject to misclassification [5–6,11]. It proposes to address misclassification by identifying individual observations that are miscoded followed by switching (reassignment) to their true class. In doing so, a ‘true’ dataset is produced from which estimates of parameters are inferred. While it is unlikely that all misclassified observations will be successfully detected in any practical application, the results presented here show that bias can be significantly reduced by the proposed method. When the rate of misclassification was assumed known, as may be the case in some diagnostic applications, bias in estimates of genetic variance was nearly eliminated and the resulting posterior mean (0.112) was very similar to the estimate obtained for the true large sample data (Table 2). In most applications there may be reason to suspect misclassification in the data, although its extent may not be known. Even in such a scenario, the proposed method was able to substantially reduce the bias in the estimation of the genetic variance (M4; Table 2). Both when the misclassification probabilities were known or unknown, estimates of the breeding values were virtually identical to those obtained

using the true noise-free data set. This is crucial for phenotype prediction and efficiency of selection. When the data was small, the proposed method was not able to completely eliminate the bias in the estimation of the genetic variance, but it helped reduce it by almost a half. This could be due to the complexity of estimating the genetic variance with small data even in the absence of misclassification. Additionally, with a small data set and a low noise level (5%) the number of misclassified observations is limited (in our case an average of 7 misclassified observations between each pair of categories), which complicates the learning of the true data generating process needed to identify potentially misclassified responses. However, the breeding values were estimated with an accuracy similar to the large data set (Table 4). For the large data set, the probability of misclassification between classes 1 and 3 (π_{13}) was accurately estimated. However, the probability of misclassification between classes 1 and 2 (π_{12}) and 2 and 3 (π_{23}) were underestimated. This is not surprising due to the marked differences in liabilities between observations in more distant categories. Some observations in adjacent categories tend to have very similar liabilities (close to the separation threshold), which enormously complicates the detection of misclassified observations and facilitates the switching of truly classified responses between categories. This behavior is illustrated in Fig 2, where the probability of misclassification for correctly classified observations is highest when the true liability is located near a threshold value separating two classes. In the case of misclassified observations, the probability of misclassification is highest (almost 1) when non-adjacent categories are involved (i.e., classes 1 and 3) as indicated in Fig 3. However, when the misclassification occurs between adjacent classes, the misclassification probability is lower, especially when the liability of the switched observation is closer to the threshold separating its true and observed classes (Fig 3).

Results of the analysis of the true dataset indicate the possibility of a low level of misclassification. This is supported by the significant change in the estimation of the genetic variance when misclassification was contemplated in the model. Furthermore, estimates of the misclassification probabilities are well supported by husbandry practices. In practice, it is unlikely that a misclassification could occur between category 3 (difficult calving) and the other two

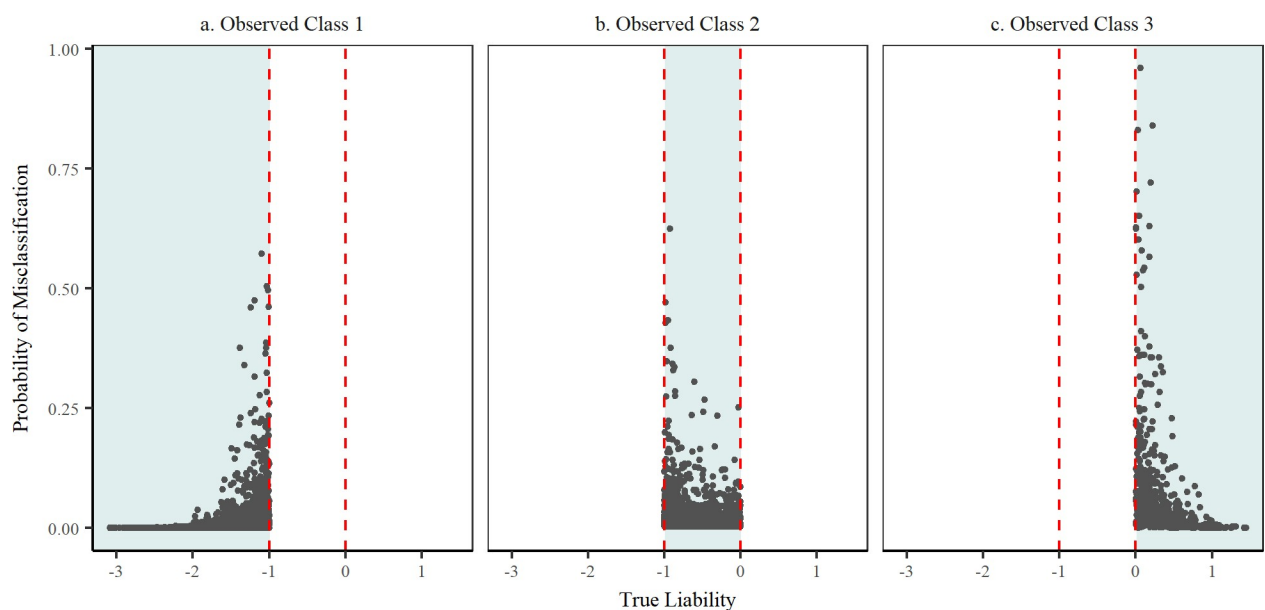


Fig 2. Probability of misclassification of correctly classified observations. Dashed red lines indicate the threshold values separating the three categories.

<https://doi.org/10.1371/journal.pone.0208433.g002>

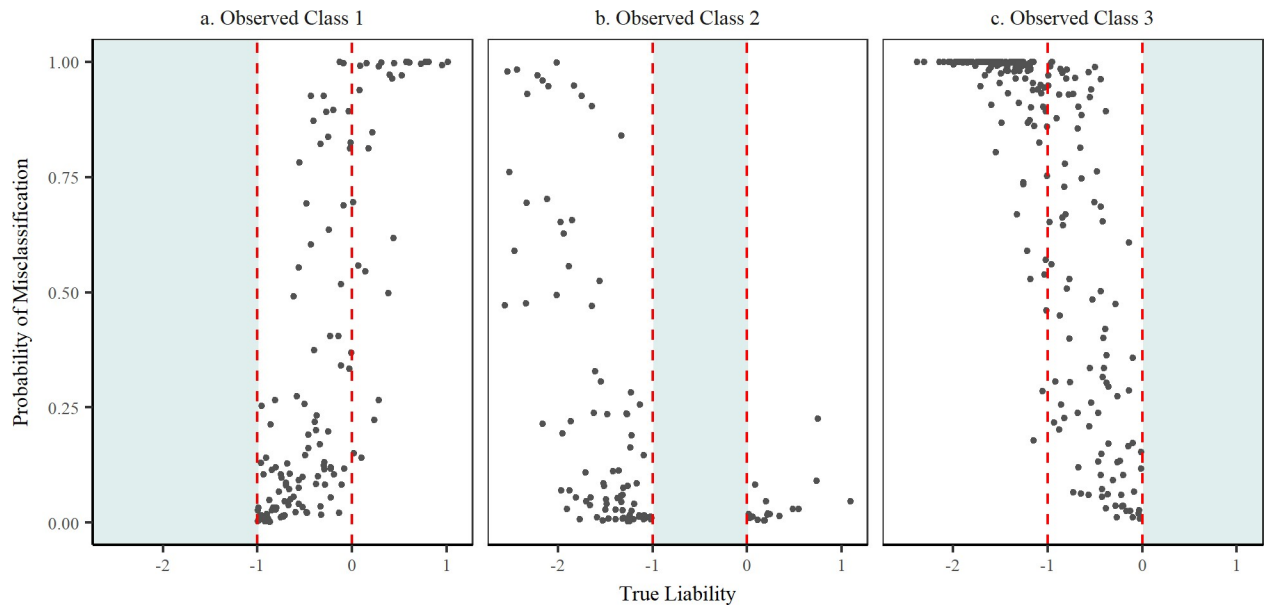


Fig 3. Probability of misclassification of miscoded observations. Dashed red lines indicate the threshold values separating the three categories. Individual plots are separated by the observed (miscoded) class of observations. Observations are plotted along the x-axis by their true liability, indicating their true class.

<https://doi.org/10.1371/journal.pone.0208433.g003>

categories (1 and 2), but owevemisclassification is likely between classes 1 and 2, which are more subjective.

In real applications, there is no absolute certainty about the presence of misclassification in a given data set. Thus, we tested our proposed methodology under a null model (no misclassification in the data). Under small and large data set scenarios, the estimated genetic variance and breeding values were virtually the same as when the true data was analyzed using a classical threshold model (Tables 2 and 4). Additionally, the probabilities of misclassification were small to negligible, indicating the absence of misclassification as expected.

In this study, the proposed methodology was tested using simulated datasets for an ordered categorical response with three categories and a heritability of 0.10. We assumed that the misclassification probabilities were symmetric between adjacent categories in order to facilitate the implementation. While the extension to traits with more categories and asymmetric misclassification between adjacent classes is straightforward, it could be computationally costly. For traits with higher heritability, and therefore higher predictability, performance of the methodology would be expected to improve.

Conclusions

Large data sets are being collected in several fields of research from human health to precision agriculture. Although this data will undoubtedly contribute towards answering complex scientific questions, the unavoidable noise, including the misclassification of discrete response variables, may result in biased inference, loss of statistical power and ultimately a less efficient decision-making process. The methodology proposed in this study provides an effective tool to at least reduce the negative impact of misclassification through the identification of potentially miscoded observations and the prediction of their true response class. Furthermore, it has been shown that the method is able to yield a similar reduction in bias whether the rate of misclassification in the data is known or to be estimated. As misclassification is seldom known

before the analysis, any proposed method to deal with misclassification has to perform well in the presence of noise-free data. As indicated by the results of the null model, our method was adequate. While the application demonstrated in the current study is on simulated beef cattle production data, the methodology is flexible and would be adaptable to any ordinal response.

Author Contributions

Conceptualization: El Hamidi Hay, Samuel E. Aggrey, Romdhane Rekaya.

Data curation: Ashley Ling.

Formal analysis: Ashley Ling.

Funding acquisition: Romdhane Rekaya.

Investigation: Ashley Ling.

Methodology: Samuel E. Aggrey, Romdhane Rekaya.

Project administration: Romdhane Rekaya.

Resources: El Hamidi Hay.

Software: Romdhane Rekaya.

Supervision: Romdhane Rekaya.

Validation: Ashley Ling.

Visualization: El Hamidi Hay, Samuel E. Aggrey, Romdhane Rekaya.

Writing – original draft: Ashley Ling, Romdhane Rekaya.

Writing – review & editing: Ashley Ling, El Hamidi Hay, Samuel E. Aggrey, Romdhane Rekaya.

References

1. Bross I. Misclassification in 2x2 tables. *Biometrics*. 1954 Dec; 10(4): 478–486. <https://doi.org/10.1016/j.jspi.2013.06.010>
2. Barron BA. The effects of misclassification on the estimation of relative risk. *Biometrics*. 1977 Jun; 33(2): 414–418. PMID: [884199](https://pubmed.ncbi.nlm.nih.gov/884199/)
3. Gaba A, Winkler RL. Implications of errors in survey data: a Bayesian model. *Manage Sci*. 1992 Jul; 38(7): 913–925.
4. Neuhaus JM. Bias and efficiency loss due to misclassified responses in binary regression. *Biometrika*. 1999; 86(4): 843–855.
5. Rekaya R, Weigel KA, Gianola D. Threshold model for misclassified binary responses with applications to animal breeding. *Biometrics*. 2001 Dec; 57(4): 1123–1129. PMID: [11764252](https://pubmed.ncbi.nlm.nih.gov/11764252/)
6. Smith S, Hay EH, Farhat N, Rekaya R. Genome wide association studies in presence of misclassified binary responses. *BMC Genet [Internet]*. 2013 Dec; 14(124): [10 p.]. <https://doi.org/10.1186/1471-2156-14-124> PMID: [24369108](https://pubmed.ncbi.nlm.nih.gov/24369108/)
7. Zhu X, Wu X. Class noise vs. attribute noise: a quantitative study of their impacts. *Artif Intell Rev*. 2004; 22: 177–210.
8. Poon WY, Wang HB. Analysis of ordinal categorical data with misclassification. *Br J of Math Stat Psychol*. 2010; 63: 17–42. <https://doi.org/10.1348/000711008X401314> PMID: [19364445](https://pubmed.ncbi.nlm.nih.gov/19364445/)
9. Newman-Toker DE. A unified conceptual model for diagnostic errors: underdiagnosis, overdiagnosis, and misdiagnosis. *Diagnosis (Berl)*. 2014 Jan; 1(1): 43–48. <https://doi.org/10.1515/dx-2013-0027> PMID: [28367397](https://pubmed.ncbi.nlm.nih.gov/28367397/)
10. Tenenbein A. A double sampling scheme for estimating from misclassified multinomial data with applications to sampling inspection. *Technometrics* 1972 Feb; 14(1): 187–202.

11. Rekaya R, Smith S, Hay EH, Farhat N, Aggrey SE. Analysis of binary responses with outcome-specific misclassification probability in genome-wide association studies. *The Appl Clin Genet* [Internet]. 2016 Nov 30; 9: 169–177. <https://doi.org/10.2147/TACG.S122250> PMID: 27942229
12. Chyou PH. Patterns of bias due to differential misclassification by case-control status in a case-control study. *Eur J Epidemiol*. 2007; 22: 7–17. <https://doi.org/10.1007/s10654-006-9078-x> PMID: 17225958
13. Huang Y, Gottardo R. Comparability and reproducibility of biomedical data. *Brief Bioinform*. 2012; 14(4): 391–401. <https://doi.org/10.1093/bib/bbs078> PMID: 23193203
14. Manchia M, Cullis J, Turecki G, Rouleau GA, Uher R, Alda M. The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. *PLoS One* [Internet]. 2013; 8(10): [8 p.]. <https://doi.org/10.1371/journal.pone.0076295> PMID: 24146854
15. Redman TC. The impact of poor data quality on the typical enterprise. *Commun ACM*. 1998 Feb; 41(2): 79–82.
16. Diagnostic error in acute care (Editorial). *Pa Patient Saf Advis* 2010 Sep; 7(3): 76–86. Available from: http://patientsafety.pa.gov/ADVISORIES/Pages/201009_76.aspx
17. Goldberg J. The effects of misclassification on the bias in the difference between two proportions and the relative odds in the fourfold table. *J Am Stat Assoc*. 1975 Sep; 70(351): 561–567.
18. Tavaré CJ, Sobel EL, Gilles FH. Misclassification of a prognostic dichotomous variable: sample size and parameter estimate adjustment. *Stat Med*. 1995 Jun 30; 14(12): 1307–1314. PMID: 7569489
19. Fan J, Han F, Liu H. Challenges of big data analysis. *Natl Sci Rev*. 2014 Jun; 1(2): 293–314. <https://doi.org/10.1093/nsr/nwt032> PMID: 25419469
20. Ladha KS, Eikermann M. Codifying healthcare—big data and the issue of misclassification. *BMC Anesthesiol*. 2015; 15(179): [2 p.]. <https://doi.org/10.1186/s12871-015-0165-y> PMID: 26667619
21. Hausman JA, Abrevaya J, Scott-Morton FM. Misclassification of the dependent variable in a discrete-response setting. *J Econom*. 1998; 87: 239–269.
22. Sapp RL, Spangler ML, Rekaya R, Bertrand JK. A simulation study for the analysis of uncertain binary responses: application to first insemination success in beef cattle. *Genet Sel Evol*. 2005; 37: 615–634. <https://doi.org/10.1186/1297-9686-37-7-615> PMID: 16277971
23. Joseph S, Robbins K, Zhang W, Rekaya R. Effects of misdiagnosis in input data on the identification of differential expression genes in incipient Alzheimer patients. *In Silico Biol*. 2008; 8: 545–554. PMID: 19374137
24. Tenenbein A. A double sampling scheme for estimating from binomial data with misclassifications. *J Am Stat Assoc*. 1970 Sep; 65(331): 1350–1361.
25. Lyles RH, Tang L, Superak HM, King CC, Celentano DD, Lo Y, et al. Validation data-based adjustments for outcome misclassification in logistic regression: an illustration. *Epidemiology*. 2011 Jul; 22(4): 589–597. <https://doi.org/10.1097/EDE.0b013e3182117c85> PMID: 21487295
26. Tang L, Lyles RH, King CC, Celentano DD, Lo Y. Binary regression with differentially misclassified response and exposure variables. *Stat Med*. 2015; 34: 1605–1620. <https://doi.org/10.1002/sim.6440> PMID: 25652841
27. Magder LS, Hughes JP. Logistic regression when the outcome is measured with uncertainty. *Am J of Epidemiol*. 1997; 146(2): 195–203. PMID: 9230782
28. Ogburn EL, VanderWeele TJ. Bias attenuation results for nondifferentially mismeasured ordinal and coarsened confounders. *Biometrika*. 2013; 100(1): 241–248. <https://doi.org/10.1093/biomet/ass054> PMID: 24014285
29. Wang D, Gustafson P. On the impact of misclassification in an ordinal exposure variable. *Epidemiol Methods*. 2014; 3(1): 97–106. Available from: <https://doi.org/10.1515/em-2013-0017>
30. Dustmann C, van Soest A. An analysis of speaking fluency of immigrants using ordered response models with classification errors. *J Bus Econ Stat*. 2004; 22(3): 312–321. <https://doi.org/10.1080/03610926.2014.887105>
31. Edwards JK, Cole SR, Troester MA, Richardson DB. Accounting for misclassified outcomes in binary regression models using multiple imputation with internal validation data. *Am J Epidemiol*. 2013; 177(9): 904–912. <https://doi.org/10.1093/aje/kws340> PMID: 24627573
32. Tennekoon V, Rosenman R. Systematically misclassified binary dependent variables. *Commun Stat Theory Methods*. 2016; 45(9): 2538–2555. <https://doi.org/10.1080/03610926.2014.887105> PMID: 27293307
33. Sun L, Xia M, Tang Y, Jones PG. Bayesian adjustment for unidirectional misclassification in ordinal covariates. *J Stat Comput Simul*. 2017; 87(18): 3440–3468. Available from: <https://doi.org/10.1080/00949655.2017.1370649>

34. Sorensen DA, Andersen S, Gianola D, Korsgaard I. Bayesian inference in threshold models using Gibbs sampling. *Genet Sel Evol.* 1995; 27: 229–249.
35. Korsgaard IR, Lund MS, Sorensen D, Gianola D, Madsen P, Jensen J. Multivariate Bayesian analysis of Gaussian, right censored Gaussian, ordered categorical and binary traits using Gibbs sampling. *Genet Sel Evol.* 2003; 35: 159–183. <https://doi.org/10.1186/1297-9686-35-2-159> PMID: 12633531
36. Newman S, MacNeil MD, Reynolds WL, Knapp BW, Urlick JJ. Fixed effects in the formation of a composite line of beef cattle: I. Experimental design and reproductive performance. *J Anim Sci.* 1993a; 71: 2026–2032. PMID: 8376225
37. Newman S, MacNeil MD, Reynolds WL, Knapp BW, Urlick JJ. Fixed Effects in the formation of a composite line of beef cattle: II. Pre- and postweaning growth and carcass composition. *J Anim Sci.* 1993b; 71: 2033–2039. PMID: 8376226
38. Wright S. Coefficients of inbreeding and relationship. *Am Nat.* 1922; 56(645): 330–338.
39. Henderson CR. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics.* 1976 Mar; 32(1): 69–83.
40. Mujibi FDN, Crews DH Jr. Genetic parameters for calving ease, gestation length and birth weight in Charolais cattle. *J Anim Sci.* 2009; 87: 2759–2766. <https://doi.org/10.2527/jas.2008-1141> PMID: 19465493
41. Vanderick S, Troch T, Gillon A, Glorieux G, Gengler N. Genetic parameters for direct and maternal calving ease in Walloon dairy cattle based on linear and threshold models. *J Anim Breed Genet.* 2014; 131: 513–521. <https://doi.org/10.1111/jbg.12105> PMID: 24965920