

Theses of doctoral (PhD) dissertation

Handling misclassification in statistical analysis

Dr. Hársfalvi Péter

Supervisor: Dr. Reiczigel Jenő



UNIVERSITY OF VETERINARY MEDICINE
BUDAPEST

Doctoral School of Veterinary Sciences

Budapest, 2025

Supervisors

.....

Dr. Reiczigel Jenő

Department of Biostatistics

University of Veterinary Medicine Budapest

Doctoral candidate

.....

Dr. Hársfalvi Péter

Produced in ... copies. This is copy number

Table of Contents

<i>1. Background and objectives of the doctoral thesis</i>	4
<i>2. Summary of the results</i>	6
2.1 Profile likelihood confidence interval for the prevalence assessed by an imperfect diagnostic test	6
2.2 Confidence limits for risk differences and risk ratios adjusted for estimated sensitivity and specificity	6
2.3 The effect of misclassification on sample size for one and two- sample tests with binary endpoints	7
2.4 Applying the promising zone method for sample size re- estimation in clinical trials when the binomial endpoint is based on a diagnostic test	7
2.5 Logistic regression with covariate-dependent probability of misclassification	8
<i>3. Main findings of the research</i>	9
<i>4. Publications related to the thesis research</i>	10
4.1 Peer-reviewed journals	10
4.2 Referred journals	10
4.3 Conference presentations	10

1. Background and objectives of the doctoral thesis

Measurement errors are common in medical research, behavioural sciences, and other fields. For continuous variables, they are called "measurement errors," whereas for categorical variables, such as binary diagnoses, they are referred to as "misclassification". While extensive methods exist for addressing measurement errors in continuous data, misclassification in categorical data has received less attention, with fewer studies and methods available. Despite recent calls to account for misclassification, appropriate analytical methods for complex models remain limited.

The aim of our research was twosome. Firstly, by combining theoretical-analytical and simulation technics we extended some already existing methods not performing well in the presence of misclassification so they become also available for situations when misclassification might be present. We used this approach to extend the already available method of logistic regression and the combination of some methods used for confidence interval (CI) construction so that both have new applications for various misclassification problems.

Secondly, we developed new methods for a few complex research problems including misclassification. We developed a new CI based on the profile likelihood method for the true prevalence when Sensitivity (Se) and Specificity (Sp) of the diagnostic test are estimated from independent validation samples. We also developed sample size calculation methods for several different tests on binary data and a sample size re-calculation method for the popular adaptive clinical trial method the "promising zone" design.

Our research has explored innovative methods to address the challenges posed by misclassification in statistical analysis, with a focus on applications in medical research and epidemiology. Through the development of novel methodologies we have extended the statistical tools available for researchers working with imperfect diagnostic tests and binary outcomes. Each method was rigorously

evaluated through simulations and real-world examples, demonstrating their practical utility and robustness.

The findings underscore the critical importance of accounting for misclassification, not only during data analysis but also at the design stage of studies. Neglecting this factor can lead to biases, reduced statistical power, and potentially inconclusive results, particularly in resource-limited settings where sample sizes are constrained. By providing computationally feasible and statistically sound solutions, this research bridges gaps in existing methodologies and offers tools that can be readily applied in a variety of fields, from clinical trials to public health.

2. Summary of the results

2.1 Profile likelihood confidence interval for the prevalence assessed by an imperfect diagnostic test

We created a Profile Likelihood Confidence Interval (PLCI) for the estimation of the disease prevalence for the specific case of using independent validation samples for the estimation of Se and Sp. PLCI uses the likelihood ratio test and adjusts for extreme observed values, ensuring better performance under challenging conditions. The method's performance was compared to existing intervals (Lang-Reiczigel and Flor et al.) through simulations evaluating coverage probability (CP) and expected length (EL).

The adjusted prevalence-linked confidence interval (PLCI) demonstrates comparable or superior performance to both available alternatives, the Lang-Reiczigel CI and the Flor interval in terms of CP and EL. Application examples illustrate the practical relevance of these comparisons. For large sample size studies, all three methods yield similar results. However, for studies with smaller prevalences, PLCI and Flor provide notably shorter intervals compared to Lang-Reiczigel. In summary, the PLCI combines strong CP with shorter or comparable EL, making it an effective alternative to the Lang-Reiczigel CI and a robust competitor to the Flor interval.

2.2 Confidence limits for risk differences and risk ratios adjusted for estimated sensitivity and specificity

We proposed a confidence interval, integrating the two approaches by Zou and Donner and Lang and Reiczigel for the two disease association metrics, the risk ratio (RR) and risk difference (RD). The method account for the variability of Se and Sp of the diagnostic tests used for the estimation of the measures and is crucial for studies involving rapid diagnostic tests, such as COVID-19 field evaluations, where differential classification errors and small sample sizes are common.

The method is easy to use and computationally feasible, the coverage probabilities are close to the nominal confidence level except when at least one of the prevalences (or one of the response rates in a clinical trial) is extreme (close to 0). Application examples highlighted the practical impact of the interval and the method is recommended when Se and Sp variability cannot be ignored, especially with smaller sample sizes.

2.3 The effect of misclassification on sample size for one and two-sample tests with binary endpoints

We introduced methods for calculating the sample size for the one- and two-sample tests for binary endpoints that account for potential misclassification, using either fixed parameters for Se and Sp or considering their uncertainty as estimated parameters.

The findings emphasize the critical importance of considering potential misclassification during study design, as ignoring these factors can lead to underpowered studies. Even with high Se and Sp values (98-99%), necessary sample sizes may be several times larger than those calculated under idealized assumptions.

2.4 Applying the promising zone method for sample size re-estimation in clinical trials when the binomial endpoint is based on a diagnostic test

We proposed adjusted sample size formulas for the interim analysis of the adaptive clinical trial design, the so called “promising zone” sample size re-calculation design for the scenario when a binomial outcome based on a diagnostic test with known Se and Sp causing potential misclassification.

By simulations we showed that our adjustments are easy to implement, yielding an unbiased estimate of the treatment effect and an accurate conditional power at the interim analysis of clinical trials using the “promising zone” method.

2.5 Logistic regression with covariate-dependent probability of misclassification

We proposed an extension of the logistic model that accounts for misclassification in the outcome, allowing either Se or Sp to depend on covariates, while assuming the other remains constant. Simulation results confirmed that the model performs well, with the sample size and statistical power depending on the true parameters and distributions of predictors.

The model is useful for analyzing sensitive survey data, where misclassification and response biases depend on covariates. However, limitations include the reliance on the logistic form, potential bias from model misspecification, and instability due to strong correlations between predictors. Despite these challenges, the model shows promise as a powerful tool for data analysis, especially in fields where misclassification is an issue.

3. Main findings of the research

1. I proposed a new confidence interval for the estimation of disease prevalence when Se and Sp is estimated from validation samples which outperforms the other available intervals both in terms of coverage probability and expected length.
2. I constructed a new confidence interval for the risk difference and risk ratio taking into account the Se and Sp of a diagnostic test that have proven to be easy to use and computationally feasible with better performance compared to other unadjusted methods.
3. I proposed a new sample size procedure for one- and two-sample binomial tests considering misclassification and potential drop-out rate to prevent researchers from a considerable power loss in their studies.
4. The adjustments I proposed for the adaptive “promising zone design” allows clinical trial boards to have an unbiased estimate of the treatment effect and accurate conditional power calculation at the interim analysis.
5. I extended the logistic model for the case when the probability of misclassification is covariate-dependent. This extension has major advantages over existing methods for the analysis of sensitive survey questions when respondents are reluctant to answer honestly and the degree of honesty is supposed to depend on certain covariates.

4. Publications related to the thesis research

4.1 Peer-reviewed journals

Hársfalvi, P., & Reiczigel, J. (2023). Profile likelihood confidence interval for the prevalence assessed by an imperfect diagnostic test. *Preventive Veterinary Medicine*, 214, 105886.

Hársfalvi, P., & Reiczigel, J. (2024). The effect of misclassification on sample size for one and two-sample tests with binary endpoints. *Journal of Biopharmaceutical Statistics*, 1-12.

4.2 Referred journals

Hársfalvi, P., & Singer, J. (2023). Confidence limits for risk differences and risk ratios adjusted for estimated sensitivity and specificity. *Biostatistics & Epidemiology*, 7(1), 2183614.

Hársfalvi, P., & Reiczigel, J. (2023). Profil-likelihood konfidenciaintervallum alkalmazása betegség prevalenciájára, ha a diagnosztikai teszt hibával terhelt= Profile likelihood confidence interval for the prevalence assessed by an imperfect diagnostic test. *STATISZTIKAI SZEMLE*, 101(11), 1033-1043.

4.3 Conference presentations

1. 9th Statistics and Biopharmacy Conference – Paris

Applying the promising zone method for sample size re-estimation in clinical trials when the binomial endpoint is based on a diagnostic test

2. 44th Annual Conference of the International Society for Clinical Biostatistics – Milan

Profile likelihood confidence interval for the prevalence assessed by an imperfect diagnostic test