# Comparing machine learning and statistical models in capture-recapture estimators: a simulation study on bias and precision

Research Report

**Luca Y. Kogelheide (8159556)**

Supervisors: Dr. Joep Burger (CBS) and Dr. Jonas Klingwort (CBS)

Methodology and Statistics for the Behavioural, Biomedical and Social Sciences - Utrecht University

Date: 22.12.2024

Word count: 2498

# 1 Introduction

The primary objective of national statistical institutes is to produce high-quality statistics that inform the public and guide policymakers (European Commission 2018). Traditionally, such statistics are generated using probability surveys. However, in recent decades, several challenges associated with survey-based data collection have emerged, including declining response rates (Leeuw et al. 2018; Meyer et al. 2015) and an increased response burden (Hu et al. 2017; Richardson et al. 1996). These issues arise particularly in diary surveys, which are time-intensive and impose a substantial response burden on respondents.

As a result, participants may report inaccurately or do not respond at all, potentially resulting in significant underreporting (Ashley et al. 2009; Schmidt 2014). Within the Total Survey Error Framework (Biemer 2010), such shortcomings are understood as both measurement error and nonresponse error, respectively, leading to biased estimates if not adequately addressed. While nonresponse error can often be mitigated through statistical weighting techniques, addressing measurement error remains challenging due to the difficulty of validating the amount. Typically, a second independent data source is needed to quantify the measurement error before it can be corrected.

One potential approach to quantifying measurement error is to link survey data to a second data source, thereby combining multiple data sources (Lohr and Raghunathan 2017). Klingwort et al. (2019) proposed using capture-recapture (CRC) techniques to quantify measurement error by linking sur-

vey data with sensor data as a secondary data source. Although weighting the survey estimates can account for selective nonresponse, the CRC estimators offer the additional advantage of correcting for measurement error. By comparing the counts and characteristics of the units in both data sources, CRC estimators estimate the portion of the population that is misclassified or unobserved in either data source, thus correcting for measurement errors such as underreporting. The difference between the basic survey estimate and the CRC estimate is the estimated amount of measurement error in the survey. An in-depth study comparing multi-source model-based CRC estimates with design-based survey estimates identified underreporting, likely due to the high (perceived) response burden, as the cause for the difference between survey and CRC estimates (Klingwort et al. 2021).

# 2 Research Background

Recent advances in machine learning (ML) have expanded its applications within the field of capture-recapture methodologies. For example, in ecological research, ML is predominantly applied to automate labor-intensive and time-consuming tasks such as labeling images produced by camera traps (Green et al. 2020; Whytock et al. 2021) or identifying animals through data from bioacoustic sensors (Wang 2024). However, one topic that has received limited attention so far is the application of ML algorithms in the capture-recapture estimation process. Model-based CRC estimators use additional information to improve the accuracy of the estimate. Usually, this is done by applying regression models to map these covariates to the target variable.

Since ML algorithms are highly flexible in capturing complex, non-linear relationships (Boulesteix and Schmid 2014), they might provide more accurate predictions than statistical models and thus provide better estimates.

In a recent study, Walraad et al. (2024) explored the application of gradient boosting algorithms, specifically XGBoost, in the context of model-based CRC estimators. By applying both XGBoost and generalized linear models (GLM) to map auxiliary variables to the target variable, the study demonstrated that ML-based approaches outperform statistical models in terms of predictive accuracy. However, improvements in prediction performance had little impact on the resulting CRC point estimates, which showed little variation across modeling approaches. This raises questions about the extent to which advances in predictive modeling translate into improvements in CRC estimation. Moreover, since Walraad et al. (2024) conducted their analysis using real-world data, it remains unclear which model-estimator combination yields less biased or more accurate estimates.

Building on this work, the current study aims to investigate whether machine learning algorithms produce less biased and more precise capture-recapture estimators compared to traditional statistical models. To address this research question, a simulation study with known true population size is conducted, comparing different model-estimator combinations regarding bias and precision. This study not only extends previous research, but also lays the groundwork for integrating ML techniques into CRC estimation practices more reliably.

# 3 Simulation

This simulation study builds on the work of Klingwort et al. (2019) and Walraad et al. (2024) and follows a similar framework. In these studies, the goal was to quantify underreporting in a survey estimating the total number of days transport vehicles were on the road, using Weigh-in-Motion road sensor data as a second capture occasion and auxiliary variables form register data from Statistics Netherlands. For readers interested in the original study design, we refer to the study by Klingwort et al. (2019). In our study, we aim to estimate the same total using capture-recapture estimators, applied to data from two capture occasions drawn from the same population.

Data simulation and analysis were performed using R version 4.4.0 (R Core Team 2024). For data wrangling, tidyverse version 2.0.0 was used (Wickham et al. 2019). All plots were created using ggplot2 version 3.5.1 (Wickham 2016). We use the ADEMP (aim, data generating mechanism, estimands, methods, performance metrics) framework to report the simulation study (Morris et al. 2019). In this section, we describe the methods of the simulation study before describing the statistical methods in Section 4. The code and data will be publicly available at a later stage.

## 3.1 Aim

The aim of this study is to compare different combinations of (statistical or machine learning) models and CRC estimators with respect to bias and precision, as described in Section 3.4.

## 3.2 Data generating mechanism

We simulate 48 different scenarios by varying the following simulation parameters in a fully factorial manner: sample size of the first capture occasion $n_1 \in \{100, 500\}$, sample size of the second capture occasion $n_2 \in \{1000, 2500\}$, response rate $rr \in \{1, 0.8, 0.5\}$, and the amount of underreporting $fnr \in \{0, 0.1, 0.3, 0.5\}$. For each scenario, we draw $n_{sim} = 1000$ samples for each capture occasion, to minimize the Monte Carlo uncertainty. In a later stage, we will determine $n_{sim}$ based on the Monte Carlo standard error for each of the performance metrics (see Section 3.5). The input seed is '1337'.

### Population

We simulate a population of size 10,000. The binary target variable, which indicates whether a vehicle was 'on the road' (1) or 'not on the road' (0), is simulated from a Bernoulli distribution independently for each day in a given quarter, spanning 91 days (7 days a week for 13 weeks):

$$Y_{i,j} \sim Bern(p_i)$$

The parameter $p_i$ is the probability of a given unit $i$ being on the road. To ensure that this probability is constrained between 0 and 1, it is the output of a sigmoid function $\frac{1}{1+exp(-f)}$, with $f$ being an arbitrary linear predictor of covariates:

$$f = \pi x_1 + sin(x_2) + exp(x_3) + 20 + x_4^2$$

where $\pi$ is a constant, and $x_1, x_2, x_3$, and $x_4$ are the following covariates:

- $x_1$ is a binary variable drawn from a Bernoulli distribution: $x_1 \sim Bern(0.5)$,
- $x_2$ is a categorical variable with 3 categories of equal size: $x_2 \in (1, 2, 3)$,
- $x_3$ is a categorical variable with 8 categories of equal size: $x_3 \in (1, 2, 3, 4, 5, 6, 7, 8)$,
- $x_4$ is a continuous variable drawn from a normal distribution with mean 5 and standard deviation 1.5 : $x_4 \sim \mathcal{N}(5, 1.5^2)$.

**Capture Occasions**

Once the binary target variable is generated for each unit across all days, the first capture occasion is simulated by drawing a simple random sample of size $n_1$ from the population. Since in the original survey, the survey period is one week, we randomly assign the numbers 1 to 13 to each unit in the sample to indicate in which quarter week the sampled unit is included for the first capture occasion. This indicator variable is only included for the first (survey) capture occasion. For the second capture occasion, we draw a sample of size $n_2$ and the units are recorded for each week of the quarter.

To reflect the collection of real-word survey data, we include unit nonresponse by introducing the response rate ($rr$) as a hyperparameter. Similarly to unit nonresponse, we include missing sensor data which can be thought of as a sensor failing to capture a passing vehicle. This idea of a sensor capture rate is treated similarly to the response rate in all aspects and takes on the same values as the response rate parameters indicated above.

To account for underreporting, we introduce a false negative rate, denoted as *fnr*, which represents the probability of observing a 0 for the target variable while the true value is 1. Specifically, for each unit where $Y_{i,j} = 1$, there is a probability *fnr* that the reported/observed values will be incorrectly recorded as 0, reflecting underreporting in the data.

## 3.3 Estimand

The estimand $\theta$ is the total number of days vehicles have been on the road during a given quarter. Specifically, the ture population total refers to the sum of the binary target variable $\delta_{i,j}$ (1 = on the road, 0 otherwise) for all vehicles across the 91 days in a quarter:

$$\theta = \sum_{i=1}^{N} \sum_{j=1}^{91} \delta_{i,j},$$

where $N$ is the total number of vehicles in the population.

Furthermore, the estimated amount of underreporting is the difference between the survey estimate $\hat{\theta}^{\text{svy}}$ and each given estimator.

## 3.4 Methods

We will apply 8 estimators to each simulated data set: the survey estimator $\hat{\theta}^{\text{svy}}$; the Lincoln-Peterson estimator $\hat{\theta}^{\text{LP}}$; the Huggins estimator using both logistic regression, i.e., a generalized linear model (GLM) $\hat{\theta}^{\text{HUG}^{\text{GLM}}}$, and XGBoost (XGB) $\hat{\theta}^{\text{HUG}^{\text{XGB}}}$; and the Log-linear estimator using both, a Poisson regression, i.e., a GLM $\hat{\theta}^{\text{LL}^{\text{GLM}}}$, and XGBoost $\hat{\theta}^{\text{HUG}^{\text{XGB}}}$. The Log-linear estimator can be constructed in two ways: either as a purely model-based

estimator by replacing the observed counts with predicted counts $\hat{\theta}^{\text{LL}_{repl}}$, or by supplementing the observed counts with predicted counts $\hat{\theta}^{\text{LL}_{supp}}$. Applying these two methods for the GLM and XGBoost approaches, we obtain $\hat{\theta}^{\text{LL}^{\text{GLM}}} \in \{\ hat\theta^{\text{LL}^{\text{GLM}}_{repl}}, \hat{\theta}^{\text{LL}^{\text{GLM}}_{supp}}\}$ and $\hat{\theta}^{\text{LL}^{\text{XGB}}} \in \{\hat{\theta}^{\text{LL}^{\text{XGB}}_{repl}}, \hat{\theta}^{\text{LL}^{\text{XGB}}_{supp}}\}$. For a more detailed explanation of each estimator, see Section 4.3.

## 3.5 Performance metrics

To assess the performance of each estimator, we evaluate both bias and precision using three performance metrics: relative bias, the coefficient of variation, and relative root mean squared error.

Relative bias measures the accuracy of an estimator:

$$\text{Relativ Bias} = \frac{\hat{\theta} - \theta}{\theta},$$

where $\hat{\theta}$ are the estimated values and $\theta$ is the true value.

The coefficient of variation (CV) quantifies the precision of the estimator by measuring the variability of the estimates relative to their mean:

$$\text{CV} = \frac{SD(\hat{\theta})}{\bar{\theta}},$$

where $\bar{\theta}$ is the mean of the estimated values across all iterations.

The relative root mean squared error combines both bias and precision:

$$\text{Relative RMSE} = \frac{\sqrt{\frac{1}{n_{sim}} \sum_{i=1}^{n_s im} (\theta_i - \theta)^2}}{\theta},$$

where $\theta_i$ is the estimated value for each iteration and $n_{sim}$ is the number of iterations.

Please note that Monte Carlo uncertainty estimates for each of the three performance metrics will follow at a later stage.

# 4 Methods

## 4.1 Capture-Recapture

Capture-recapture techniques are widely applied to estimate demographic parameters. Initially developed to meet the needs of ecologists in estimating population sizes in wild animal populations (Seber 1982), these methods are now frequently used in the social and medical sciences for applications such as estimating homelessness (Coumans et al. 2017), census undercounts, disease incidence, or criminal activity (Böhning et al. 2018; Pollock 1991).

The CRC estimators applied here rely on data collected on two independent capture occasions. First, a sample of $n_1$ units is drawn from the population and marked. Second, an independent sample of $n_2$ units is drawn. The overlap between the two samples, represented by the number of units marked already $m$, provides the basis for estimating the total population size. CRC estimators considered in this work are subject to the following assumptions (Otis et al. 1978):

- The population is closed (that is, no changes occur due to births, deaths, or migration during the study period).
- The inclusion probability of the first capture occasion is independent

of the inclusion probability of the second capture occasion.

- All units in the population have a nonzero probability of being included in each capture.
- For at least one of the capture occasions, the inclusion probabilities are homogeneous across units.
- Marked units can be accurately identified and perfectly linked across capture occasions.
- All units in the list belong to the population.
- There are no duplicate entries in the lists.

## 4.2 Definitions and notations

We define the indicator variable $\delta_{i,j}^{\text{svy}}$ for the first capture occasion (the survey) as follows: It takes the value 1 if vehicle $i$ was reported to be on the road on day $j$, and 0 otherwise. Similarly, we define the indicator variable $\delta_{i,j}^{\text{sen}}$ for the second capture occasion (sensor data) as 1 if the vehicle $i$ was recorded at least once by a sensor on day $j$, and 0 otherwise.

## 4.3 Estimators

### Survey Estimator

The survey estimator is based on the first capture occasion and represents the total number of days vehicles have been on the road during a quarter. In practice, this statistic is a yearly estimate. However, for this simulation, a

quarterly estimate of the population total is sufficient. The survey estimator is being calculated by:

$$\hat{\theta}^{\text{svy}} = \sum_{i=1}^{r} \left( w_i \sum_{j=1}^{7} \delta_{i,j}^{\text{svy}} \right),$$

where $r$ denotes the number of respondents in the survey, and $w_i$ is the survey weight for vehicle $i$. It is calculated as:

$$w_i = 13 \cdot \frac{N}{r},$$

where $N$ is the size of the population, $r$ is the number of respondents, and factor 13 extrapolates from a weekly estimate to a quarterly estimate. For this simulation, I assume not stratification.

**Lincoln-Peterson Estimator**

The Lincoln-Peterson estimator (Lincoln 1930; Petersen 1896) is the most basic of the CRC estimators and uses of the previously derived quantities $n_1, n_2$, and $m$ from two capture occasions, without any auxiliary information. It can be seen as a ratio estimator for population totals, where we equate the proportion of marked units in the second sample ($n_2/m$) with the proportion of marked units in the entire population ($n_1/N$). Rearranging this proportion yields the LP estimator for the total population size:

$$\hat{\theta}^{\text{LP}} = \frac{n_1 n_2}{m},$$

where the quantities are calculated as follows:

$$n_1 = \sum_{i=1}^{r} \left( w_i \sum_{j=1}^{7} \delta_{i,j}^{\text{svy}} \right) = \hat{\theta}^{\text{svy}},$$
$$n_2 = \sum_{i=1}^{r} \left( w_i \sum_{j=1}^{7} \delta_{i,j}^{\text{sen}} \right),$$
$$m = \sum_{i=1}^{r} \left( w_i \sum_{j=1}^{7} \delta_{i,j}^{\text{svy}} \cdot \delta_{i,j}^{\text{sen}} \right).$$

**Huggins Estimator**

The estimator proposed by Huggins (1989) and Alho (1990) considers heterogeneity in the capture probabilities. It is the Horvitz and Thompson (1952) estimator, where model-based estimates of the inclusion probabilities replace the design-based inclusion probabilities:

$$\hat{\theta}^{HUG} = \sum_{i=1}^{r} w_i \sum_{j} \frac{1}{\hat{\psi}_{ij}},$$

where $\hat{\psi}_{ij}$ is the probability that vehicle $i$ on day $j$ is either reported as used in the survey, recorded by a sensor, or both.

To model the inclusion probabilities, we need to calculate the conditional probabilities $p_{i,j}^{\text{svy}}$ $\text{P}(\delta_{i,j}^{\text{svy}} = 1 | \delta_{i,j}^{\text{sen}} = 1)$ and $p_{i,j}^{\text{sen}}$ $\text{P}(\delta_{i,j}^{\text{sen}} = 1 | \delta_{i,j}^{\text{svy}} = 1)$. We can either apply a GLM, specifically a logistic regression, assuming $\delta_{i,j}$ follows a Bernoulli distribution with probability $p_{i,j}$ and the logit of $p_{i,j}$ is a linear combination of auxiliary variables. Another approach is to apply machine learning algorithms for classification. Here, we use XGBoost (XGB) from the R package *caret* version 7.0-1.

Therefore, we compare two different approaches to estimate the capture

probabilities for the Huggins estimator:

$$\hat{\theta}^{\text{HUG}} \in \{\hat{\theta}^{\text{HUG}^{\text{GLM}}}, \hat{\theta}^{\text{HUG}^{\text{XGB}}}\}.$$

**Log-linear Estimator**

In contrast to the Huggins estimator, the Log-linear (LL) estimator (Fienberg 1972) is an unconditional likelihood estimator and displays CRC problems as 2 by 2 contingency tables. Adding auxiliary information, the number of strata increases by $4 \prod_{g=1}^{G} C_g$, where $C_g$ is the number of categories of the covariate g and G the number of categorical covariates used to define the strata. Fitting a statistical or machine learning model to the observed counts, we can use this model to predict all counts. There are two approaches to construct the LL estimator. We can either replace the observed counts with the purely model-based predicted counts and sum over all strata:

$$\hat{\theta}^{\text{LL}^{\text{repl}}} = \sum_{h=1}^{H} \bar{y}_h.$$

Or we can supplement the sum of observed counts (set $\mathcal{S}$) with predicted counts for the cells where $\delta_h^{\text{svy}} = \delta_h^{\text{svy}} = 0$ (set $\mathcal{R}$):

$$\hat{\theta}^{\text{LL}^{\text{supp}}} = \sum_{h \in \mathcal{S}} y_h + \sum_{h \in \mathcal{R}} \hat{y}_h.$$

Similar to the Huggins estimator, there are to approaches two estimating $y_h$. We can either apply a GLM, specifically a Poisson regression, assuming $y_h$ follows a Poisson distribution with mean and variance $\lambda_h$ and the log of

$\lambda_h$ is a linear combination of covariates. Otherwise, we can apply machine learning, specifically XGBoost (XGB). Therefore, we compare four versions of log-linear estimators, namely by replacing or supplementing observed counts (repl or supp) and the way of estimating $y_h$ (GLM or XGB), leading to:

$$\hat{\theta}^{LL} \in \{\hat{\theta}^{\mathrm{LL}^{\mathrm{GLM}}_{repl}}, \hat{\theta}^{\mathrm{LL}^{\mathrm{GLM}}_{supp}}, \hat{\theta}^{\mathrm{LL}^{\mathrm{XGB}}_{repl}}, \hat{\theta}^{\mathrm{LL}^{\mathrm{XGB}}_{supp}}\}.$$

Please note that the sections on the Huggins estimator and on the Log-linear estimator are mainly based on the paper of Walraad et al. (2024), as I did not yet find the time to conduct an in-depth review of the literature on these estimators.

# 5 Results

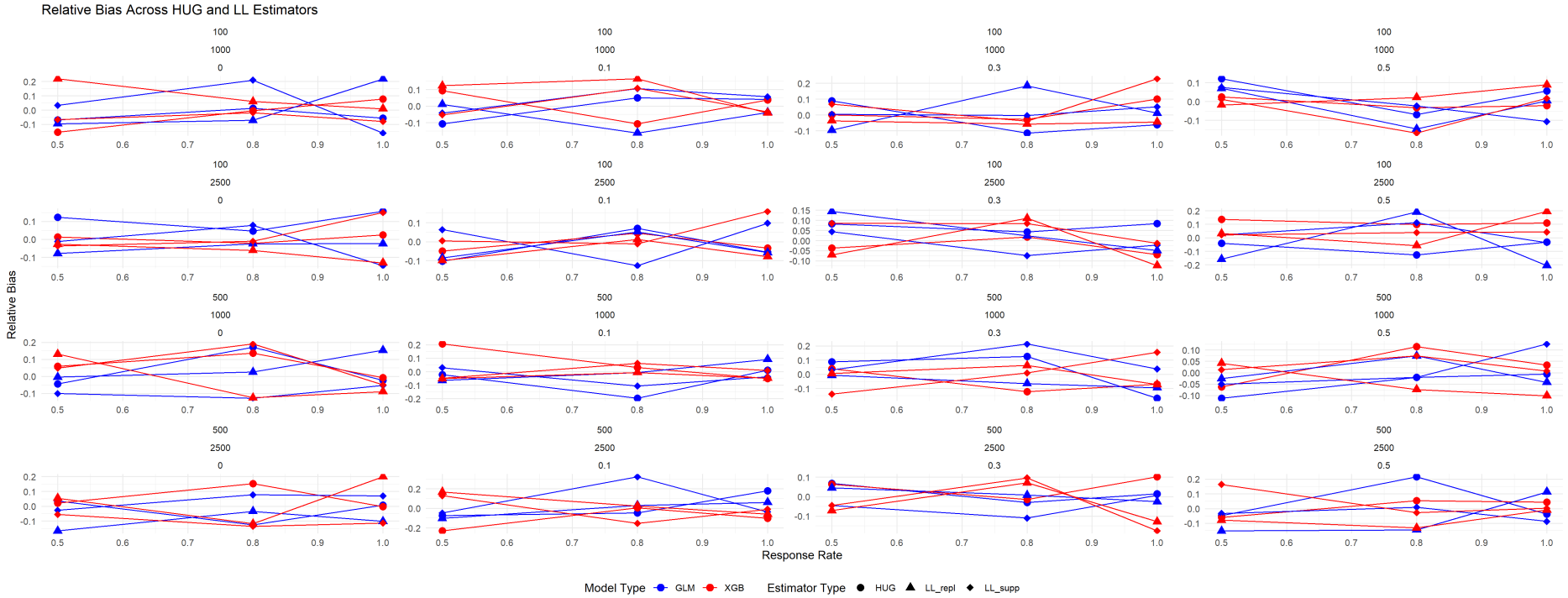
Table 1 and Figure 1 show possible ways to display information for each model-estimator combination across all simulation scenarios for one specific performance metric.

Table 1: Simulation Results: Relative Bias of Different Model-Estimator Combinations

| Scenarios: $(n_1, n_2, rr, fnf)$ | $\hat{\theta}^{\text{svy}}$ | $\hat{\theta}^{\text{LP}}$ | $\hat{\theta}^{\text{HUG}^{\text{GLM}}}$ | $\hat{\theta}^{\text{HUG}^{\text{XGB}}}$ | $\hat{\theta}^{\text{LL}^{\text{GLM}}_{\text{repl}}}$ | $\hat{\theta}^{\text{LL}^{\text{GLM}}_{\text{supp}}}$ | $\hat{\theta}^{\text{LL}^{\text{XGB}}_{\text{repl}}}$ | $\hat{\theta}^{\text{LL}^{\text{XGB}}_{\text{supp}}}$ |
|---|---|---|---|---|---|---|---|---|
| (100, 1000, 1, 0) | value1 | value2 | value3 | value4 | value5 | value6 | value7 | value8 |
| (500, 1000, 1, 0) | value1 | value2 | value3 | value4 | value5 | value6 | value7 | value8 |
| (...) | ... | ... | ... | ... | ... | ... | ... | ... |
| (100, 2500, 0.5, 0.5) | value1 | value2 | value3 | value4 | value5 | value6 | value7 | value8 |
| (500, 2500, 0.5, 0.5) | value1 | value2 | value3 | value4 | value5 | value6 | value7 | value8 |

**Note.** This is an examplary table of displaying results for each simulation scenarios and each model-estimator combination for the performance metrics relative bias. Due to its length, this would probably be include in the appendix, and most results will be shown using figures.

Figure 1: Relative Bias Across HUG and LL Estimators for each Simulation Scenario




**Note.** This is simulated data and does not correspond to the expected data at all. This is merely an exemplary figure showing what a possible figure for the real results could look like.

# References


Alho, Juha M (1990). "Logistic regression in capture-recapture models". In: *Biometrics*, pages 623–635.

Ashley, D, T Richardson, and D Young (2009). "Recent information on the under-reporting of trips in household travel surveys". In: *Australasian Transport Research Forum (ATRF), 32nd, 2009, Auckland, New Zealand*. Volume 32.

Biemer, Paul P (2010). "Total survey error: Design, implementation, and evaluation". In: *Public opinion quarterly* 74.5, pages 817–848. DOI: 10.1093/poq/nfq058.

Böhning, Dankmar, John Bunge, and Peter GM Heijden (2018). *Capture-recapture methods for the social and medical sciences*. CRC Press Boca Raton. DOI: 10.4324/9781315151939.

Boulesteix, Anne-Laure and Matthias Schmid (2014). "Machine learning versus statistical modeling". In: *Biometrical Journal* 56.4, pages 588–593. DOI: 10.1002/bimj.201300226.

Coumans, AM, MJLF Cruyff, Peter GM Van der Heijden, JRLM Wolf, and HJSIR Schmeets (2017). "Estimating homelessness in the Netherlands using a capture-recapture approach". In: *Social Indicators Research* 130, pages 189–212. DOI: 10.1007/s11205-015-1171-7.

European Commission, Eurostat (2018). *European statistics code of practice – For the national statistical authorities and Eurostat (EU statistical authority)*. Publications Office of the European Union. DOI: 10.2785/798269.

Fienberg, Stephen E (1972). "The multiple recapture census for closed populations and incomplete 2k contingency tables". In: *Biometrika* 59.3, pages 591–603.

Green, Austin M, Mark W Chynoweth, and Çağan Hakkı Şekercioğlu (2020). "Spatially explicit capture-recapture through camera trapping: a review of benchmark analyses for wildlife density estimation". In: *Frontiers in Ecology and Evolution* 8, page 563477. DOI: 10.3389/fevo.2020.563477.

Horvitz, Daniel G and Donovan J Thompson (1952). "A generalization of sampling without replacement from a finite universe". In: *Journal of the American statistical Association* 47.260, pages 663–685.

Hu, Mengyao, Garrett W Gremel, John A Kirlin, and Brady T West (2017). "Nonresponse and underreporting errors increase over the data collection week based on paradata from the National Household Food Acquisition and Purchase Survey". In: *The Journal of Nutrition* 147.5, pages 964–975. DOI: 10.3945/jn.116.240697.

Huggins, RM991431 (1989). "On the statistical analysis of capture experiments". In: *Biometrika* 76.1, pages 133–140.

Klingwort, Jonas, Bart Buelens, and Rainer Schnell (2019). "Capture–Recapture Techniques for Transport Survey Estimate Adjustment Using Permanently Installed Highway-Sensors". In: *Social Science Computer Review* 39.4, pages 527–542. DOI: 10.1177/0894439319874684.

Klingwort, Jonas, Joep Burger, Bart Buelens, and Rainer Schnell (2021). "Transition from survey to sensor-enhanced official statistics: Road freight transport as an example". In: *Statistical Journal of the IAOS* 37.4, pages 1289–1299. DOI: 10.3233/SJI-210821.

Leeuw, Edith de, Joop Hox, and Annemieke Luiten (2018). "International nonresponse trends across countries and years: An analysis of 36 years of labour force survey data". In: *Survey methods: insights from the field*, pages 1–11. DOI: 10.13094/SMIF-2018-00008.

Lincoln, Frederick Charles (1930). *Calculating waterfowl abundance on the basis of banding returns*. 118. US Department of Agriculture.

Lohr, Sharon L. and Trivellore E. Raghunathan (2017). "Combining Survey Data with Other Data Sources". In: *Statistical Science* 32.2, pages 293–312. DOI: 10.1214/16-STS584.

Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan (Nov. 2015). "Household Surveys in Crisis". In: *Journal of Economic Perspectives* 29.4, pages 199–226. DOI: 10.1257/jep.29.4.199.

Morris, Tim P, Ian R White, and Michael J Crowther (2019). "Using simulation studies to evaluate statistical methods". In: *Statistics in medicine* 38.11, pages 2074–2102. DOI: 10.1002/sim.8086.

Otis, David L, Kenneth P Burnham, Gary C White, and David R Anderson (1978). "Statistical inference from capture data on closed animal populations". In: *Wildlife monographs* 62, pages 3–135.

Petersen, Carl Georg Johannes (1896). "The yearly immigration of young plaice in the Limfjord from the German sea". In: *Rept. Danish Biol. Sta.* 6, pages 1–48.

Pollock, Kenneth H (1991). "Review papers: modeling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present, and future". In: *Journal of the American Statistical Association* 86.413, pages 225–238.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Richardson, Anthony J, Elizabeth S Ampt, and Arnim H Meyburg (1996). "Nonresponse issues in household travel surveys". In: *Conference proceedings*. Volume 10, pages 79–114.

Schmidt, Tobias (2014). "Consumers' recording behaviour in payment diaries–empirical evidence from Germany". In: *Survey Methods: Insights from the Field*, page 16. DOI: 10.13094/SMIF-2014-00008.

Seber, George Arthur Frederick (1982). *The estimation of animal abundance and related parameters (2nd ed)*. London: Charles Griffin.

Walraad, Maaike, Jonas Klingwort, and Joep Burger (2024). "Incorporating Machine Learning in Capture-Recapture Estimation of Survey Measurement Error". In: *Survey Research Methods*. Volume 18. 2, pages 99–112. DOI: 10.18148/srm/2024.v18i2.8307.

Wang, Yuheng (2024). "Statistical inference in acoustic spatial capture-recapture: integrating machine learning for species identification". PhD thesis. The University of St Andrews. DOI: 10.17630/sta/1168.

Whytock, Robin C, Jedrzej Świeżewski, Joeri A Zwerts, Tadeusz Bara-Słupski, Aurélie Flore Koumba Pambo, Marek Rogala, Laila Bahaa-el-din, Kelly Boekee, Stephanie Brittain, Anabelle W Cardoso, et al. (2021). "Robust ecological analysis of camera trap data labelled by a machine learning model". In: *Methods in Ecology and Evolution* 12.6, pages 1080–1092. DOI: 10.1111/2041-210X.13576.

Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4. URL: https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani (2019). "Welcome to the tidyverse". In: *Journal of Open Source Software* 4.43, page 1686. DOI: 10.21105/joss.01686.