**Maciej Beręsewicz, PhD**
*Poznań University of Economics and Business*
*Department of Statistics*
*Poland*
*maciej.beresewicz@ue.poznan.pl*

**Cover Letter** January 30, 2025

Dear Editors of the Journal of Statistical Software,

we are writing to submit a revised manuscript of our article entitled ***singleRcapture**: An R Package for Single-Source Capture-Recapture Models* for review to the Journal of Statistical Software (previous submission number: 5641). We have considered all the comments on paper and software (see responses on next page).

Population size estimation is a major challenge in official statistics, social sciences, and natural sciences. The problem can be tackled by applying capture-recapture methods, which vary depending on the number of sources used, particularly on whether a single or multiple sources are involved. This paper focuses on the first group of methods and introduces the `singleRcapture` package in the R language (version 0.2.2 the same as on CRAN).

Our package serves to bridge a significant gap, as the SSCR methods are either not available at all or are only partially implemented in existing R packages and other open-source software. It offers state-of-the-art single-source capture-recapture (SSCR) models (e.g. zero-truncated one-inflated regression) together with new developments proposed by the authors, and provides a user-friendly application programming interface (API).

The novelty and our contribution can be summarised as follows:
- The package implements and extends existing SSCR methods together with new developments proposed by the authors.
- The package can be used to produce point estimates and calculate their variances; it also implements several bootstrap variance estimators.
- It implements diagnostics (e.g. rootograms) to assess quality and conduct sensitivity analysis (e.g. dfbetas).
- The package relies on the `S3Methods`, which can be easily applied by users who know how to run regression in R (e.g. `stats::glm`, `countreg`).
- The package enables flexible estimation of population size for user-specified strata and is fully integrated with the `sandwich` package.
- The `singleRcapture` package can be used to implement custom function families, which is an option advanced users will find particularly useful.

The `singleRcapture` package has been created for users interested in estimating the size of populations, particularly those that are difficult to reach or measure, for which information is available only from one source and dual/multiple system estimation is not applicable. The package has been developed since 2021 and the full history can be found at the Github repository `https://github.com/ncn-foreigners/singleRcapture`.

To the best of our knowledge, there is no open-source software that can be used to estimate population size with SSCR methods and includes variance estimators or diagnostics. That is why we believe that the paper will be of interest to the readership of the Journal of Statistical Software.

Thank you for your consideration of this manuscript.

Sincerely,
Maciej Beręsewicz & Piotr Chlebicki

**Replies to comments from the Editor (submission 5641)**

- The submission "singleRcapture: An R Package for Single-Source Capture-Recapture Models" presents an R package which implements single-source capture-recapture methods for population size estimation using zero-truncated, zero-one truncated and ero-truncated one-inflated Poisson, Geometric and Negative Binomial regression as well as Zelterman's and Chao's regression. This is the first resubmission of a previous version in which some issues on the implementation and article have been raised. Some of them have been solved in the current submission but some were insufficiently addressed. We decided not to send the current submission into review to let the authors correct remaining issues. We ask them to provide a point-by-point answer to detailed comments along their future submission.

  - We have corrected the code as suggested. Moreover we have included a html vignette, based on the submitted JSS paper, so the users will be given detailed information on the functionalities of the `singleRcapture` package.

- The output of `help(package = "singleRcapture")` indicates that man page titles are still not consistently in title style. Please correct.

  - We removed the `code` from titles and double checked the titles. Now everything should be in the right order.

- `plot(basicModel)` `Error in plot.singleRStaticCountData(basicModel) : Argument plotType must be provided as a character or integer or NULL.` Please, provide a default behavior (default value) for `plotType`. More specifically, if NULL is allowed, it seems natural that it is set as the default value of `plotType`.

  - We have changed the behaviour of the function and removed information about the `NULL`. If a user will not specify the plot or provide it as NULL then the `"qq"` is selected. We provide this information in the documentation

    ```
    #' @param plotType character parameter (default \code{"qq"})
    specifying type of plot to be made.
    ```

    as well as in the code

    ```
    if (missing(plotType) || is.null(plotType)) {
      plotType <- eval(formals()$plotType)[1]
    } else if (is.numeric(plotType) && length(plotType) == 1) {
      # Handle numeric index
      valid_plots <- eval(formals()$plotType)
      if (plotType > 0 && plotType <= length(valid_plots)) {
        plotType <- valid_plots[as.integer(plotType)]
      } else {
        stop("Numeric plotType must be between 1 and ", length(valid_plots))
      }
    } else if (is.character(plotType)) {
      # Use match.arg for character input validation against the options
      plotType <- match.arg(plotType)
    } else {
      stop("plotType must be a character string, or a numeric index")
    }
    ```

- `anova(basicModel)` `Error in anova.singleRStaticCountData(basicModel) : The anova method for singleRStaticCountData class doesn't work yet.` It would seem desirable that this method is implemented or that the authors explain why this is too much an effort to provide it.

  - *Godwin, R. T., & Böhning, D. (2017). Estimation of the population size by using the one-inflated positive Poisson model. Journal of the Royal Statistical Society Series C: Applied Statistics, 66(2), 425-448* recommends using LR tests to test for one-inflation (i.e. comparing models) and the standard Anova tests are not a recommended approach.
  - In general, the literature on SSCR methods suggests using the LR test, goodness-of-fit tests or plots to select between models. The final choice of estimator should be based on the appropriateness of the assumptions made and the experience of the researcher (and previous results, if available).
  - It should be noted that also the **countreg** does not implement the `anova` method as indicated in the paper the Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. Journal of statistical software, 27(8), 1-25. as they specify that: `No anova() method is provided, but the general coeftest(), waldtest() from lmtest, and linearHypothesis() from car can be used for Wald tests and lrtest() from lmtest for LR tests of nested models.` (p. 7)
  - On the other hand, the `VGAM` package implements the `anova` function for the `pospoisson` and other functions, so if the user is interested in evaluating the coefficients, this approach can be used. However, as noted above, the goal of SSCR is not the model parameters themselves, but the results in terms of population size. The models should be compared on this basis.
    Moreover the theory for `anova` for vector generalized linear models is more complicated than for generalized linear models. In particular is more than one way to generalized `anova` to this broader class of models cf. e.g. `?VGAM::anova.vglm`.
  - We have added more informative comment regarding the `anova` function.

```
#' @importFrom stats anova
#' @method anova singleRStaticCountData
#' @exportS3Method
anova.singleRStaticCountData <- function(object, ...) {
  stop("The custom anova method for singleRStaticCountData class is not yet implemented.
  If the goal is to compare models we recommend using `lmtest::lrtest` instead. ")
}
```

- The code still contains commented lines which usefulness is unclear.

  - We have removed most of the comments from the package. The only ones that remain in the code are the ones that contextualize the purpose of the code.