pprof: An R Package for Provider Profiling

Xiaohan Liu[1], **Lingfeng Luo**[1], **Xiangeng Fang**[1], **Yubo Shao**[1], **Lin Yin Guo**[1], **Tao Xu**[1], **and Kevin He**[1¶]

1 Department of Biostatistics, School of Public Health, University of Michigan ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Sehrish Kanwal

Reviewers:

- @hknd23

Submitted: 03 July 2025
Published: unpublished

Summary

The pprof package is an open-source R software specifically developed for provider profiling, enabling the evaluation and comparison of healthcare provider performance. It offers both linear models for continuous outcomes and logistic regression for binary outcomes, seamlessly incorporating both fixed and random effects. In the field of provider profiling, datasets are growing in both sample size and the number of providers, leading to highly clustered structures and imposing significant computational burdens on existing R functions. pprof addresses these challenges by implementing computationally efficient algorithms tailored to different outcome types. For linear fixed-effect models, it employs a profile likelihood method (Hsiao, 2022) while for logistic fixed-effect models, it utilizes the serial blockwise inversion Newton algorithm (Wu et al., 2022). Beyond model fitting, pprof provides comprehensive post-modeling functionalities, including the calculation of standardized measures, hypothesis testing, and confidence interval estimation. Moreover, pprof enables statistical inference for individual provider effects, which lack in many existing packages. Currently under active development, **pprof** will soon include models for count outcomes and time-to-event outcomes, further expanding its applicability. This comprehensive suite of tools makes pprof a useful tool for robust and efficient provider performance evaluation in large and clustered healthcare datasets.

Statement of Need

Provider profiling plays a critical role in the healthcare industry by enabling the assessment and comparison of performance of healthcare providers. Accurate profiling facilitates the identification of low-performing providers, promotes accountability, and supports quality improvement initiatives, ultimately leading to enhanced patient outcomes and more efficient healthcare delivery systems (He et al., 2013; Horwitz et al., 2013; Jones & Spiegelhalter, 2011; Kalbfleisch & Wolfe, 2013; Normand et al., 1997; Spiegelhalter et al., 2012; Wu et al., 2022). As healthcare data become increasingly large-scale, particularly with highly clustered structures inherent in provider-specific datasets, the demand for powerful analytical tools capable of efficiently processing and analyzing such data has increased.

However, existing functions often fall short in addressing the computational challenges posed by large and highly clustered datasets in the field of provider profiling. Specifically, in linear fixed-effect models, conventional statistical software like R's `lm` and `glm` functions rely on a dummy variable approach to represent provider effects. This method imposes a substantial computational burden as the number of providers increases. Similarly, for binary outcomes, traditional estimation algorithms such as Newton-Raphson and Fisher scoring become computationally infeasible when dealing with thousands of providers and extensive sample sizes. The computational cost of inverting the information matrix escalates dramatically, imposing a significant burden even on high-performance workstations (He et al., 2013). Addressing these limitations, the pprof package offers advanced algorithms tailored for both binary and continuous outcomes, including the serial blockwise inversion Newton (SerBIN) algorithm for binary fixed-effect models (Wu et al., 2022) and a profile likelihood method for linear fixed-effect models (Hsiao, 2022). These innovations significantly reduce computational costs and enhance scalability.

Additionally, many existing functions lack comprehensive statistical inference capabilities for individual provider effect parameters, which are essential for accurately assessing and comparing provider performance. The `pprof` package addresses this gap by enabling statistical testing of individual provider effects and detecting both high- and low- performing outliers. In particular, for binary outcomes, where existing tools lack suitable inferential approaches for identifying providers with outlier performance, especially in cases involving small providers with extreme outcomes, where estimates of provider effects are often unstable. As a result, traditional Wald tests, which rely not only on large-sample approximations but also on point estimates of parameters, tend to yield inaccurate results in these settings due to the instability of the estimates and the poor approximation of small-sample properties. To address this inferential gap for binary outcomes, the `pprof` package implements score and exact tests for provider effects. The exact tests leverage finite-sample distributions, with the Poisson-binomial distribution representing a special case (Wu et al., 2022). These features provide more accurate and reliable assessment of provider performance for binary outcomes, overcoming both computational and inferential challenges.

Furthermore, in the field of provider profiling, standardized measures play a crucial role, enabling meaningful comparisons across providers, thereby facilitating the identification of outliers and areas for improvement (He et al., 2013). These measures adjust for varying patient populations and case mixes, ensuring that performance evaluations are fair and accurate. The `pprof` package addresses this need by providing both indirect and direct standardized measures and outputting both the expected and observed outcomes seamlessly.

In summary, `pprof` offers a variety of risk-adjusted model development and comprehensive post-modeling functionalities, including the calculation of standardized measures, confidence interval estimation, statistical inference and visualizations commonly used in the provider profiling domain, thereby delivering a robust and efficient solution for provider performance evaluation in large and highly clustered healthcare datasets. The functions and workflow of `pprof` are summarized in the flowchart in Figure 1.

Package Overview

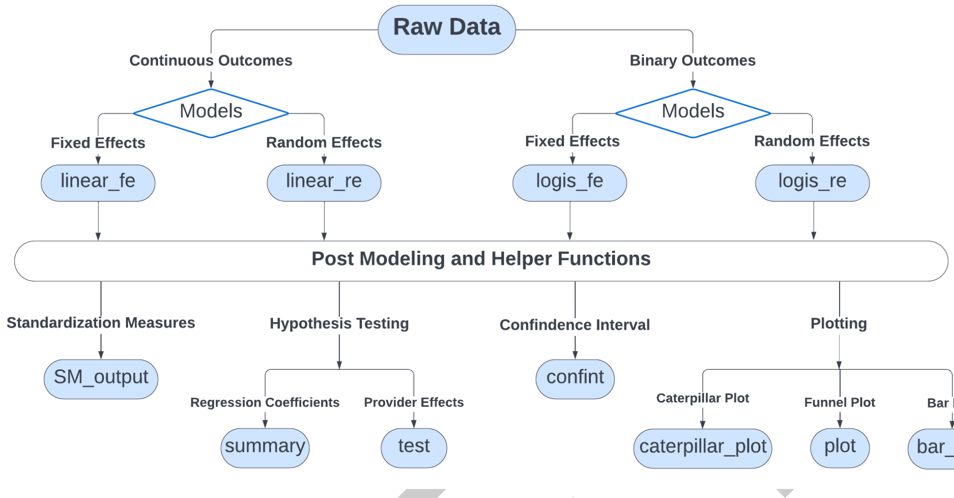

Figure 1: Flowchart for functions in the `pprof` package: This flowchart outlines the primary functions of the `pprof` package, including model fitting (`linear_fe`, `linear_re`, `logis_fe`, `logis_re`), standardized measures (`SM_output`), hypothesis testing for provider effect (`test`), confidence interval estimation for provider effect and standardized measures (`confint`), summary statistics for covariate estimates (`summary`), and visualization (`caterpillar_plot`, `funnel_plot`, `bar_plot`).

The `pprof` package is designed to facilitate robust provider profiling through its comprehensive suite of modeling, standardization, inference, and visualization tools. At its core, the package currently offers four primary modeling functions tailored to different types of outcomes.

For continuous outcomes, `pprof` provides both fixed-effect and random-effect linear models. The `linear_fe` function implements a fixed-effect linear model utilizing the profile likelihood method, which involves transforming the observed variables by subtracting the appropriate provider means (Hsiao, 2022). This transformation enables the application of the least squares method to the adjusted data, thereby allowing for the estimation of both regression parameters and provider effect parameters. Complementing this, the `linear_re` function offers a random-effect linear model by extending the `lmer` function from the widely acclaimed **lme4** package (Bates et al., 2015). This function defaults to using Restricted Maximum Likelihood (REML) for parameter estimation, ensuring reliable and efficient modeling of random effects.

For binary outcomes, `pprof` fits a fixed logistic model that leverages the Serial Blockwise Inversion Newton (SerBIN) algorithm (Wu et al., 2022). This advanced algorithm enhances the computational efficiency and scalability of logistic regression models in the context of large and highly clustered healthcare datasets, addressing the limitations of traditional generalized linear model (GLM) approaches. Moreover, `pprof` also provides the capability to fit a random-effect logistic model by extending the `glmer` function from `lme4` (Bates et al., 2015), thereby increasing the package's comprehensiveness and offering users the flexibility to choose the most appropriate modeling approach based on their data characteristics. Each of these model functions outputs essential diagnostic and summary statistics, including parameter estimates, variance components, residual standard errors, and fitted values. This comprehensive output ensures that users have access to all necessary information for thorough model evaluation and interpretation.

Beyond model fitting, `pprof` encompasses a suite of generic functions that extend its analytical capabilities. The `SM_output` function generates standardized measures for each fitted model, facilitating meaningful comparisons across providers by adjusting for varying patient populations and case mixes. Additionally, the `confint` function computes confidence intervals for both provider effect parameters and standardized measures, providing users with critical inferential statistics necessary for robust performance evaluation. The `test` function conducts hypothesis testing of provider effects, enabling the identification of significantly high or low-performing providers through rigorous statistical testing. Furthermore, the `summary` function delivers comprehensive summary statistics for regression parameters, offering users a clear and concise overview of model estimates.

Additionally, `pprof` includes comprehensive visualization capabilities essential for interpreting and presenting provider performance data. The package offers caterpillar plots, funnel plots, and bar plots, each serving distinct purposes in visualizing provider performance from various perspectives. Caterpillar plots visualize the standardized measures alongside their confidence intervals for each provider, enabling the clear identification of providers performing above or below expectations. Funnel plots, generated by the `plot` function, aim to identify healthcare providers with unusual performance by plotting standardized measures against a precision parameter, with control limits forming a funnel shape around the target. Providers that lie beyond these control limits are considered out of control and warrant further investigation (Spiegelhalter, 2005; Wu et al., 2023). `bar_plot` generates bar charts displaying the percentage of flagged results based on provider sizes, facilitating straightforward comparisons and helping to identify patterns related to provider scale. Together, these visualization tools provide comprehensive insights from multiple perspectives, enhancing the interpretability and actionable understanding of provider performance metrics.

Data Example

To illustrate the effectiveness and practical application of the `pprof` package, we conducted an analysis using the Early Childhood Longitudinal Study (ECLS) data (Tourangeau et al., 2015). This publicly available dataset tracks over 18,000 children from kindergarten through fifth grade, providing a comprehensive collection of student-level information. For our demonstration, we utilized the fifth-grade cross-sectional data, focusing on mathematical assessment as the primary outcome measure. The mathematical assessment encompassed 18 topics, including data analysis, statistics, and probability, which collectively evaluated each student's competency in conceptual knowledge, procedural knowledge, and problem-solving. These competencies were consolidated into a single math score, where lower scores indicated lower proficiency, higher scores denoted higher proficiency.

The primary predictors of interest in our analysis were parent-reported annual household income and the gender of the children. Household income was categorized into 18 ordinal ranges, ranging from the lowest category of $5,000 or less (designated as level 1) to the highest category of $200,000 or more (designated as level 18). For the purposes of this analysis, income was treated as a continuous predictor, while gender was treated as a categorical variable. To ensure the robustness of our analysis, we retained only complete cases by excluding observations with missing values. In this dataset, each child was nested within a specific school, which served as the clustering variable. The final sample comprised 9,101 individuals from 2275 schools.

```r
install.packages('pprof')
library(pprof)
data(ecls_data)
```

Given that the outcome variable was continuous, we employed a fixed-effect linear model as an example to demonstrate the application of `pprof`.

```r
formula.fe <- as.formula("Math_Score ~ Income + id(School_ID) + Child_Sex")
fit.fe <- linear_fe(formula = formula.fe, data = ecls_data)
```

Figure 2 displays the estimated standardized measures along with their 95% confidence intervals for each provider. Due to the inclusion of schools with a small number of students, we developed the model using the entire dataset; however, the visualization only includes schools with more

The Journal of Open Source Software

than five students to ensure the reliability of the estimates. In the context of continuous outcomes, the results of indirect and direct standardizied measures are identical; therefore, both caterpillar plots are the same.

```
school_counts <- table(model_dat$School_ID)
schools_with_morethan5 <- as.numeric(names(school_counts[school_counts > 5]))
CI_fe <- confint(fe_pl, stdz = c("indirect", "direct"), parm = schools_with_morethan5)
caterpillar_fe_indirect <- caterpillar_plot(CI_fe$CI.indirect, use_flag = T,
errorbar_width = 0.5)
caterpillar_fe_direct <- caterpillar_plot(CI_fe$CI.direct, use_flag = T,
errorbar_width = 0.5)
```

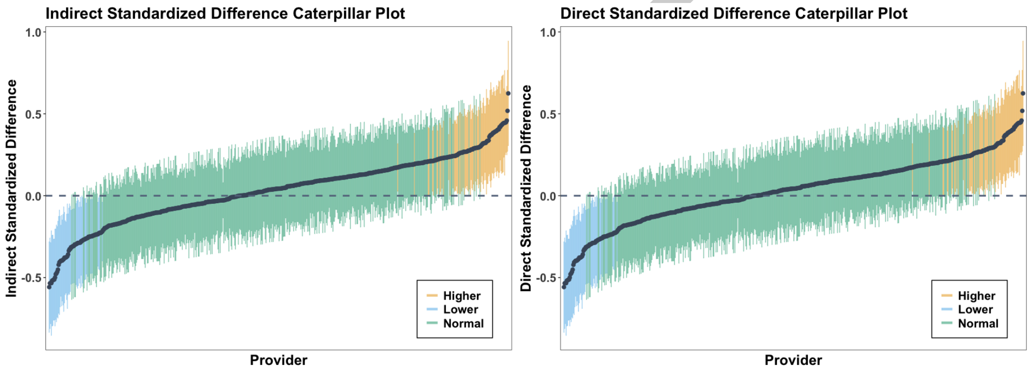

Figure 2: Caterpillar plots of standardized measures from fixed-effect linear model using ECLS data: Caterpillar plots display standardized measures (black dots) with 95% confidence intervals (vertical bars) for each school. The dashed horizontal line represents the reference level, corresponding to a standardized difference of 0. The left plot shows indirect standardizated difference, and the right plot shows direct standardizated difference. In the case of fixed-effect linear models, both standardization methods yield identical results, resulting in identical plots. Schools are flagged as "higher than expected," "lower than expected," or "as expected" based on whether their confidence intervals lie entirely above, below, or include the reference value (0).

Availability

Stable releases of the pprof package is already available via the Comprehensive R Archive Network. Alternatively, the pprof package is available on GitHub (https://github.com/UM-KevinHe/pprof).

Funding

The work is partially supported by the National Institutes of Health under the award numbers R01 DK129539.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

He, K., Kalbfleisch, J. D., Li, Y., & Li, Y. (2013). Evaluating hospital readmission rates in dialysis facilities; adjusting for hospital effects. *Lifetime Data Analysis*, *19*(4), 490–512.

Horwitz, L., Partovain, C., Lin, Z., Herrin, J., Grady, J., Conover, M., Montague, J., Dillaway, C., Bartcazk, K., Suter, L., & others. (2013). *Hospital-wide all-cause risk-standardized* *readmission measure: Measure methodology report. Yale new heaven health services corporation/center for outcomes research and evaluation (YNHHSC/CORE).*

Hsiao, C. (2022). *Analysis of panel data* (4th edition). Cambridge University Press.

Jones, H. E., & Spiegelhalter, D. J. (2011). The identification of "unusual" health-care providers from a hierarchical model. *The American Statistician*, *65*(3), 154–163.

Kalbfleisch, J. D., & Wolfe, R. A. (2013). On monitoring outcomes of medical providers. *Statistics in Biosciences*, *5*, 286–302.

Normand, S.-L. T., Glickman, M. E., & Gatsonis, C. A. (1997). Statistical methods for profiling providers of medical care: Issues and applications. *Journal of the American Statistical Association*, *92*(439), 803–814.

Spiegelhalter, D. J. (2005). Funnel plots for comparing institutional performance. *Statistics in Medicine*, *24*(8), 1185–1202.

Spiegelhalter, D. J., Sherlaw-Johnson, C., Bardsley, M., Blunt, I., Wood, C., & Grigg, O. (2012). Statistical methods for healthcare regulation: Rating, screening and surveillance. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *175*(1), 1–47.

Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., Hagedorn, M. C., Daly, P., & Najarian, M. (2015). Early childhood longitudinal study, kindergarten class of 2010-11 (ECLS-k: 2011). User's manual for the ECLS-k: 2011 kindergarten data file and electronic codebook, public version. NCES 2015-074. *National Center for Education Statistics*.

Wu, W., Kuriakose, J. P., Weng, W., Burney, R. E., & He, K. (2023). Test-specific funnel plots for healthcare provider profiling leveraging individual-and summary-level information. *Health Services and Outcomes Research Methodology*, *23*(1), 45–58.

Wu, W., Yang, Y., Kang, J., & He, K. (2022). Improving large-scale estimation and inference for profiling health care providers. *Statistics in Medicine*, *41*(15), 2840–2853. https://doi.org/10.1002/sim.9387