The Journal of Open Source Software

# AOUSDOHtools: An R Package for Social Determinants of Health Survey data in the All of Us Research Program

**Zhirui Deng** [1], **Theresa A. Koleck** [1], **Chen Zhang** [2], **Peter D. R. Higgins** [3], **and Caitlin Dreisbach** [2,4]

**1** School of Nursing, University of Pittsburgh, Pittsburgh, PA, USA **2** School of Nursing, University of Rochester, Rochester, NY, USA **3** School of Medicine, University of Michigan, Ann Arbor, MI, USA **4** Goergen Institute for Data Science, University of Rochester, Rochester, NY, USA

**DOI:** 10.21105/joss.09306

**Software**

- Review
- Repository
- Archive

**Editor:** Nick Golding

**Reviewers:**

- @smasongarrison
- @hantswilliams

**Submitted:** 13 May 2025
**Published:** 28 November 2025



## Summary

`AOUSDOHtools` is an R package that was created to support standardized and reproducible scoring of the Social Determinants of Health (SDOH) constructs from survey data collected as part of the *All of Us* Research Program. Developed in conjunction with a user guide (Koleck et al., 2024), the package provides functions to process raw SDOH Survey responses and compute 30 literature-informed construct-level scores across 14 SDOH constructs, such as Neighborhood Cohesion, Social Support, and Perceived Stress.

The package is designed for use within the *All of Us* Researcher Workbench, a secure cloud-based platform where the de-identified data are accessed and analyzed. The package is compatible with both Jupyter and RStudio environments hosted on the platform. `AOUSDOHtools` automates the data cleaning, recoding, scoring, and variable construction, which enables researchers to generate interpretable SDOH scores for downstream analysis.

The package is openly developed and maintained on GitHub and available through CRAN (cran.r-project.org/package=AOUSDOHtools) and GitHub (github.com/zhd52/AOUSDOHtools). It is intended to facilitate equitable and scalable research by making complex SDOH survey data accessible and analysis-ready for approved researchers working within the *All of Us* ecosystem.


**Figure 1:** `AOUSDOHtools` hex sticker. Created with R package `hexSticker` (Yu et al., 2020).

## Statement of Need

### *All of Us (AOU)*

The *All of Us* Research Program (allofus.nih.gov), led by the National Institutes of Health, aims to create a large and diverse health database by enrolling over one million participants across the United States. The program supports research focused on individualized prevention, diagnosis, and treatment (National Institutes of Health, 2025).

Participants provide data through self-reported surveys, electronic health records (EHR), physical measurements, wearable devices (e.g. FitBit), and biospecimen collection (e.g. urine and blood specimens), made available to approved researchers via the secure cloud-based Researcher Workbench (researchallofus.org) (All of Us Research Program Investigators, 2019). A major strength of *All of Us* is its explicit focus on health equity, particularly through the inclusion of historically underrepresented populations in biomedical research (National Institutes of Health, 2025).

Despite these strengths, the complexity of the available data, especially the survey components, can present analytic challenges. Standardized and scalable tools are needed to support consistent data processing and analysis, but such tools are not included in the platform by default (Grayson et al., 2022).

## Social Determinants of Health (SDOH) Survey

Social Determinants of Health (SDOH) refer to non-medical conditions like housing, discrimination, and education that significantly impact individual and population health outcomes (Office of Disease Prevention and Health Promotion, 2025). SDOH factors are strongly associated with health disparities and inequities across racial, geographic, and economic lines (Williams & Mohammed, 2009). The *All of Us* SDOH Survey (researchallofus.org/data-tools/survey-explorer) captures diverse life domains relevant to these determinants (All of Us Research Hub, 2025a). While rich, the data are disaggregated, complex, and not readily suitable for analysis without preprocessing and scoring.

Transforming survey responses into meaningful constructs for research, such as scoring neighborhood safety or perceived stress, benefits from structured tools. Without such tools, the reproducibility and consistency of analyses across studies are at risk.

## Scoring Social Determinants of Health (SDOH) Constructs

The `AOUSDOHtools` R Package (Deng, Koleck, Zhang, Higgins, & Dreisbach (2025); cran.r-project.org/package=AOUSDOHtools) was developed to address the absence of standardized tools for processing and scoring the *All of Us* SDOH Survey. It operationalizes literature-informed scoring logic for 14 SDOH constructs (***Figure 2***), such as Neighborhood Cohesion, Social Support, and Perceived Stress. Each construct includes well-defined item groupings, reverse-coding procedures, and aggregation rules defined in the *All of Us* SDOH User Guide (Koleck et al., 2024). The package includes 30 functions that automate scoring across these constructs, enabling reproducible and scalable analyses within the *All of Us* Researcher Workbench (All of Us Research Hub, 2025b). Six of the constructs have one scoring option. Eight of the constructs have multiple scoring options. `AOUSDOHtools` is intended for use exclusively within the *All of Us* Researcher Workbench (Jupyter or Rstudio), respecting privacy and data governance regulations.

Several R packages provide general scoring or psychometric utilities, such as `psych` (Revelle, 2025), `PROscorer` (Baser, 2023), and `scorecard` (Xie, 2025), but these are designed for use across a wide range of datasets and do not implement the unique structure of *All of Us* data. In contrast, `AOUSDOHtools` incorporates the *All of Us* variable naming conventions, metadata alignment, and privacy-preserving workflows required by the controlled Workbench environment.

This package builds directly upon the *All of Us* SDOH User Guide (Koleck et al., 2024), which provides the conceptual and methodological foundation for scoring SDOH constructs within the *All of Us* data ecosystem. The guide details the origin and purpose of each construct, specifies validated source instruments (e.g., Everyday Discrimination Scale, Perceived Stress Scale), defines item inclusion and reverse-coding logic, and outlines the computation of both subscale and composite scores. `AOUSDOHtools` translates these specifications into executable R functions that enable consistent, transparent, and reproducible SDOH scoring within the

secure *All of Us* Researcher Workbench. By automating data cleaning, recoding, and variable construction, `AOUSDOHtools` promotes reproducibility, reduces coding burden, and makes complex survey constructs more accessible for health equity research, thereby accelerating scalable, population-level analyses within the *All of Us* program.

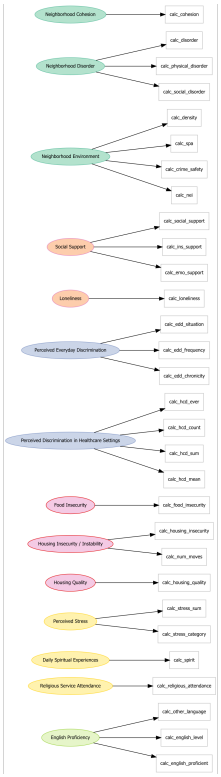


**Figure 2:** Overview of `AOUSDOHtools` functions linked to Social Determinant of Health constructs (Koleck et al., 2024). Created with R packages `DiagrammeR`, `DiagrammeRsvg`, and `rsvg` (Iannone, 2016; Iannone & Roy, 2024; Ooms & Bruggemann, 2025).

## Installation

The `AOUSDOHtools` package is available on CRAN (cran.r-project.org/package=AOUSDOHtools) and GitHub (github.com/zhd52/AOUSDOHtools). Users can install the stable release from CRAN or the development version from GitHub. Detailed installation instructions are provided in the package README and vignettes (R Core Team, 2025; Wickham, Hester, Chang, & Bryan, 2022).

This package is intended to use SDOH Survey data from the *All of Us* Research Program. As these data are only accessible within the secure *All of Us* Researcher Workbench (All of Us Research Hub, 2025b), the package needs to be installed and executed within that environment. Both the Jupyter and RStudio interfaces provided by the Researcher Workbench support the use of this package for in-platform analysis. The package uses the `tidyverse` framework for efficient data manipulation and visualization (Wickham et al., 2019; Wickham, 2023). The package also provides detailed documentation on how each score is derived, along with descriptions and value ranges for all supported constructs.

## Examples

After installation, users can apply `AOUSDOHtools` functions (***Figure 2***) directly to SDOH Survey data from the *All of Us* Research Program. In order to extract the SDOH data, a registered *All of Us* researcher would need to create a cohort using the cohort builder tool and select the premade concept set for the survey data. The concept set that includes the SDOH Survey data resides under "All Surveys" and then "Social Determinants of Health". By selecting the cohort and the concept set, a registered user can preveiw the dataset prior to launching the analytic platform.

Detailed examples are available in the package vignette (`vignette("AOUSDOHtools_examples")`). Example 1 demonstrates how to compute Neighborhood Cohesion scores using a synthetic dataset that mimics the expected structure of the *All of Us* survey data.

This workflow illustrates how raw concept-level responses from the *All of Us* data can be transformed into structured, construct-level scores. Functions for computing other SDOH constructs follow a similar structure. Users are encouraged to consult the package documentation for a complete list of scoring functions and usage details.

### Merging Resulting Scores

After computing individual construct scores using `AOUSDOHtools`, the resulting data frames can be combined into a single dataset for downstream analysis. Each scoring function returns a data frame indexed by `person_id`, which allows for merging using `purrr::reduce()` and `dplyr::full_join()` (Wickham, François, Henry, Muller, & Vaughan, 2023; Wickham & Henry, 2025).

Example 2 in the package vignette demonstrates how to combine all 14 constructs and their sub-scores. Full usage examples and additional workflows are also available in the package documentation on GitHub.

The resulting merged data frame provides a person-level summary of all available SDOH scores, ready for descriptive analysis or modeling. Participants missing responses are retained with NA values unless `na.rm = TRUE`, in which case incomplete rows are excluded.

## Development

`AOUSDOHtools` was created to support standardized and reproducible scoring of SDOH constructs derived from the *All of Us* Research Program. The package simplifies the transformation of raw survey responses into literature-informed, interpretable variables, allowing researchers to focus on data analysis and interpretation.

As the *All of Us* Research Program continues to evolve, with ongoing updates to the data model, survey content, and release versions, this package will be actively maintained to remain consistent with changes to the SDOH Survey. Substantial revisions to item content, concept identifiers, or scoring procedures may require corresponding updates to package functions.

The package is openly developed on GitHub, where users can report issues, request new features, and contribute to its development. Researchers are encouraged to review the package version and documentation to ensure compatibility with the specific release of the *All of Us* dataset they are using.

We welcome feedback, feature suggestions, and collaboration ideas. If you'd like to share your thoughts or contribute, please contact Zhirui Deng at zhd52@pitt.edu, or open an issue/submit a pull request on GitHub.

## Author Contributions

### Conceptualization

Theresa A. Koleck and Caitlin Dreisbach conceptualized the user guide. Theresa A. Koleck, Caitlin Dreisbach, Chen Zhang, and Peter D. R. Higgins contributed to its overall design.

### Software Development

Zhirui Deng developed the `AOUSDOHtools` R package, including scoring functions, documentation, unit tests, and package structure. Zhirui Deng also submitted and maintains the package on CRAN and GitHub.

### Function Development and Testing

Zhirui Deng, Chen Zhang, and Peter D. R. Higgins developed user guide functions. Zhirui Deng, Caitlin Dreisbach, Theresa A. Koleck, and Chen Zhang tested the package and contributed to updates to the scoring functions.

### Writing – Original Draft

Zhirui Deng drafted the manuscript.

### Writing – Review & Editing

All authors reviewed and revised the manuscript.

### Supervision

Theresa A. Koleck provided overall supervision.

### Licensing

Caitlin Dreisbach holds the MIT license for the version of the R package published on CRAN.

### Corresponding Author

Zhirui Deng (zhd52@pitt.edu)

## Acknowledgements

We thank the *All of Us* Research Program and its participants for making this work possible. Access to the Controlled Tier data through the *All of Us* Researcher Workbench enabled the development and testing of this package.

We also acknowledge all the authors of the published user guide (Koleck et al., 2024), whose work provided the conceptual and methodological foundation for scoring the Social Determinants of Health constructs implemented in `AOUSDOHtools`.

We are grateful to the Comprehensive R Archive Network (CRAN) team for maintaining the infrastructure that supports open source software distribution and reproducibility in R. We also acknowledge GitHub for providing the collaborative platform used to develop, maintain, and share the source code for this package.

## Conflict of Interest

The authors declare that they have no conflicts of interest related to this work.

## References


All of Us Research Hub. (2025b). All of us researcher workbench. Retrieved from https://www.researchallofus.org/

All of Us Research Hub. (2025a). Survey explorer – social determinants of health. Retrieved from https://www.researchallofus.org/data-tools/survey-explorer/

All of Us Research Program Investigators. (2019). The 'all of us' research program. *New England Journal of Medicine*, *381*(7), 668–676. doi:10.1056/NEJMsr1809937

Baser, R. (2023). *PROscorer: Functions to score commonly-used patient-reported outcome (PRO) measures and other psychometric instruments*. doi:10.32614/cran.package.proscorer

Deng, Z., Koleck, T. A., Zhang, C., Higgins, P. D. R., & Dreisbach, C. (2025). *AOUSDOHtools: Analyzing AOU SDOH survey data*. doi:10.32614/cran.package.aousdohtools

Grayson, S. C., Patzak, S. A., Dziewulski, G., Shen, L., Dreisbach, C., Lor, M., Conway, A., et al. (2022). Moving beyond table 1: A critical review of the literature addressing social determinants of health in chronic condition symptom cluster research. *Nursing Inquiry*, *30*(3), e12519. doi:10.1111/nin.12519

Iannone, R. (2016). *DiagrammeRsvg: Export DiagrammeR graphs as SVG*. doi:10.32614/cran.package.diagrammersvg

Iannone, R., & Roy, O. (2024). *DiagrammeR: Graph/network visualization*. doi:10.32614/cran.package.diagrammer

Koleck, T. A., Dreisbach, C., Zhang, C., Grayson, S., Lor, M., Deng, Z., Conway, A., et al. (2024). User guide for social determinants of health survey data in the all of us research program. *Journal of the American Medical Informatics Association*, *31*(12), 3032–3041. doi:10.1093/jamia/ocae214

National Institutes of Health. (2025). All of us research program. Retrieved from https://allofus.nih.gov/

Office of Disease Prevention and Health Promotion. (2025). Social determinants of health. Healthy people 2030. Retrieved from https://odphp.health.gov/healthypeople/priority-areas/social-determinants-health

Ooms, J., & Bruggemann, S. (2025). *Rsvg: Render SVG images into PDF, PNG, PostScript, or bitmap arrays*. doi:10.32614/cran.package.rsvg

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. doi:10.32614/r.manuals

Revelle, W. (2025). *Psych: Procedures for psychological, psychometric, and personality research*. doi:10.32614/cran.package.psych

Wickham, H. (2023). *Tidyverse: Easily install and load the "tidyverse"*. doi:10.32614/cran.package.tidyverse

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., & others. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. doi:10.21105/joss.01686

Wickham, H., François, R., Henry, L., Muller, K., & Vaughan, D. (2023). *Dplyr: A grammar of data manipulation*. doi:10.32614/cran.package.dplyr

Wickham, H., & Henry, L. (2025). *Purrr: Functional programming tools*. doi:10.32614/cran.package.purrr

Wickham, H., Hester, J., Chang, W., & Bryan, J. (2022). *Devtools: Tools to make developing r packages easier*. doi:10.32614/cran.package.devtools

Williams, D. R., & Mohammed, S. A. (2009). Discrimination and racial disparities in health: Evidence and needed research. *Journal of Behavioral Medicine*, *32*(1), 20–47. doi:10.1007/s10865-008-9185-0

Xie, S. (2025). *Scorecard: Credit risk scorecard*. doi:10.32614/cran.package.scorecard

Yu, G., Gatto, L., Rainer, J., Gibb, S., Corty, R., & Neitmann, T. (2020). *hexSticker: Create hexagon sticker in r*. doi:10.32614/cran.package.hexsticker