The Journal of Open Source Software

lifetimeRisk: An R package for lifetime risk analysis with competing risks

Matthew W. Segar [1¶], **Byron C. Jaeger** [2], **and Ambarish Pandey** [3]

1 Department of Cardiology, Texas Heart Institute, Houston, TX **2** Perisphere Real World Evidence LLC, Austin, Texas **3** Division of Caridology, Department of Internal Medicine, UT Southwestern Medical Center, Dallas, TX ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

Lifetime risk estimation is fundamental to understanding long-term disease burden and provides important insights for epidemiologists and clinicians assessing population health across diverse medical domains. The `lifetimeRisk` package provides tools for calculating lifetime risk of adverse outcomes while properly accounting for competing risks, implementing the methodology established by the SAS Practical Incidence Estimators (PIE) macro (Beiser et al., 2000). Originally developed for Alzheimer's disease research in the Framingham Study, the PIE macro provides a widely-used framework for lifetime risk calculations across epidemiological research, from cardiovascular disease and cancer to dementia and chronic diseases. The package enables person-year calculations, age-specific incidence rates, cumulative incidence with competing risks, and age-adjusted rates using the same statistical algorithms as the original PIE macro. All statistical computations have been validated against the reference SAS implementation to ensure methodological consistency and numerical equivalence for research applications.

Statement of need

Lifetime risk analysis extends traditional short-term risk assessment and has influenced medical research and clinical practice across multiple disciplines. The fundamental challenge addressed by lifetime risk methodology is that conventional 10-year risk models often underestimate long-term disease burden, particularly in younger populations who may face substantial cumulative risk despite appearing low-risk in short-term assessments. The statistical complexity lies in properly accounting for competing risks, where multiple potential outcomes can preclude the occurrence of the primary endpoint of interest. This is particularly relevant in aging populations where cardiovascular disease, cancer, and other-cause mortality compete as terminal events.

The impact of this methodology is evident in cardiovascular epidemiology, where lifetime risk studies have changed understanding of disease burden and clinical decision-making. Landmark studies by Lloyd-Jones et al. and Berry et al. demonstrated marked differences in lifetime cardiovascular disease risks across racial groups (Berry et al., 2012; Lloyd-Jones et al., 1999). These studies, along with subsequent publications demonstrating lifetime risks exceeding 30% even for individuals with optimal risk factors (Wilkins et al., 2012) and significant differences across sex-based groups (Pandey et al., 2018), have established lifetime risk as important for clinical decision-making and population health assessment.

Despite widespread use of these methods in high-impact research, accessible implementations have been limited. While R packages such as `survival`, `cmprsk`, and `etm` provide foundational competing risks methods, none offer the lifetime risk calculation framework with standardized output formats established by the PIE macro. The `lifetimeRisk` package fills this gap by providing an R implementation of the PIE macro methodology and enables researchers to

perform lifetime risk analyses without requiring SAS software.

Implementation

Statistical methodology

Lifetime risk estimation employs life table methodology to estimate the absolute probability that an individual will develop a disease if they live long enough, accounting for competing risks of death from other causes. This approach differs from Cox proportional hazards models by providing absolute risk estimates rather than relative hazards. The mathematical foundation begins with age-specific incidence rates and mortality rates from cohort data, calculating the probability of developing the disease of interest, dying from the disease, and dying from competing causes for each age interval.

The lifetime risk is computed as the cumulative probability of experiencing the event across all age intervals until a maximum age, typically 85 or 95 years. This requires estimation of age-specific hazard rates $\lambda_d(a)$ for the disease of interest and $\lambda_c(a)$ for competing risks at age a, where the cumulative incidence function $F(t) = \int_0^t \lambda_d(a) \times S(a), da$, and $S(a)$ represents the overall survival function incorporating both disease and competing mortality.

The PIE macro methodology extends basic life table approaches by implementing competing risks adjustments that properly account for interdependence between multiple potential outcomes (Fine & Gray, 1999). Unlike Kaplan-Meier survival curves that treat competing events as censored observations, the PIE approach explicitly models competing risks to provide unbiased estimates of cumulative incidence. This ensures that the sum of all cause-specific cumulative incidence functions equals one minus the overall survival probability.

Software architecture

The `lifetimeRisk` package implements this statistical methodology through a software architecture that prioritizes computational efficiency, numerical accuracy, and user accessibility. The core statistical engine translates the PIE macro algorithms into R code. Mathematical equivalence is maintained while leveraging the survival package for fundamental computations, `data.table` for efficient data management and transformation, and `ggplot2` for visualization.

Central to the package's reliability is its validation framework that ensures numerical accuracy and methodological consistency with the reference SAS implementation. All statistical computations have been tested against the original PIE macro through validation studies using both simulated and real-world datasets spanning different study designs, population characteristics, and event rates. The validation process encompasses comparison of age-specific incidence rates, cumulative incidence estimates, confidence intervals, and summary statistics.

Performance benchmarking demonstrates that the package scales efficiently with dataset size. Table 1 shows execution times and memory usage across datasets ranging from 1,000 to 1,000,000 observations. The package maintains consistent processing rates of approximately 50,000-60,000 observations per second for larger datasets, with memory usage scaling linearly with data size. These performance characteristics make the package suitable for large-scale epidemiological studies typical in cardiovascular and population health research.

Table 1. Performance benchmarks for `lifetimeRisk` package across different dataset sizes.

Dataset Size	Median Time (seconds)	Memory Usage (MB)	Observations per Second
1,000	0.04	0.0	23,882
5,000	0.10	0.2	50,570
10,000	0.18	0.4	56,204

Dataset Size	Median Time (seconds)	Memory Usage (MB)	Observations per Second
25,000	0.54	1.0	46,167
50,000	0.88	1.9	56,714
100,000	1.64	3.8	61,024
1,000,000	14.44	38.1	69,269

Note: Benchmarks performed using `microbenchmark` package with 5 replications per dataset size on standard computing hardware.

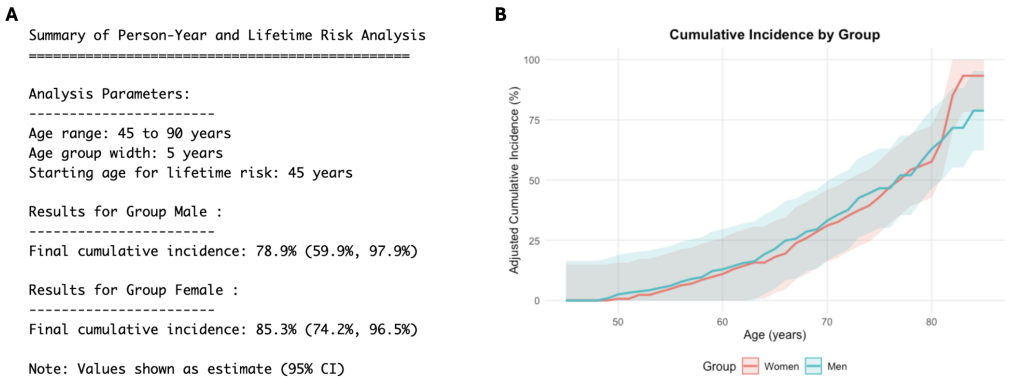

Figure 1: Example output from the 'lifetimeRisk' package demonstrating (A) summary statistics for sex-stratified lifetime risk analysis showing final cumulative incidence estimates, and (B) competing risks-adjusted cumulative incidence curves illustrating lifetime risk trajectories between men and women from age 45 to 90 years, with shaded confidence bands indicating 95% confidence intervals.

The package provides a function library designed for both novice and expert users. Key functions include `pie_analysis()` for complete lifetime risk analysis implementing the full PIE macro methodology, `calculate_age_specific_rates()` for detailed age-stratified incidence computations, `calculate_cumulative_incidence()` for competing risks-adjusted cumulative incidence calculation (**Figure 1A**), and specialized visualization and export functions. The `plot_lifetime_risk()` function generates publication-ready graphics with customizable confidence intervals (**Figure 1B**), while `create_lifetime_risk_table()` produces standardized summary tables suitable for manuscript inclusion.

Research applications and impact

The `lifetimeRisk` package addresses a need in epidemiological research for a freely accessible, open-source implementation of life time risk methodology, which has been widely cited in epidemiological literature and has influenced clinical practice guidelines. The package has been successfully applied in cardiovascular epidemiology research, including our recent multicohort study examining P-wave parameters and lifetime atrial fibrillation risk (Segar et al., 2025). In this analysis of 25,508 participants from 4 prospective cohort studies, we used the `lifetimeRisk` package to demonstrate that participants with multiple ECG abnormalities had lifetime AF risks reaching 35.7% compared to 22.9% for those with minimal abnormalities, with participants having 4+ ECG abnormalities living an average of 17.1 years free of AF compared to 21.7 years for those with none. This application demonstrates the package's capability to handle large-scale multicohort analyses and produce clinically meaningful lifetime risk estimates that inform patient care and risk stratification.

The broad applicability of lifetime risk methods extends across medical specialties. Cancer epidemiologists have used these approaches to develop global estimates of lifetime cancer risk, while neurological research has employed lifetime risk calculations to understand dementia patterns. Public health researchers have applied the methodology to examine health disparities and assess long-term health impacts of exposures. Recent applications demonstrate continued relevance, including population-specific lifetime risk tools and comprehensive models for diverse populations.

The `lifetimeRisk` package facilitates these research applications by providing a documented platform that removes technical barriers to implementing lifetime risk analyses. The package includes documentation, worked examples across different research domains, and flexible functions that accommodate various study designs and research questions, enabling researchers without extensive statistical programming backgrounds to apply these methods to their research.

Acknowledgements

This implementation builds upon the foundational statistical methodology established by Beiser et al. in the original SAS PIE macro and acknowledges the extensive validation and application work conducted in landmark epidemiological studies. We recognize the impact of this methodology on epidemiological research and clinical practice across medical disciplines.

References

Beiser, A., D'Agostino Sr, R. B., Seshadri, S., Sullivan, L. M., & Wolf, P. A. (2000). Computing estimates of incidence, including lifetime risk: Alzheimer's disease in the framingham study: The practical incidence estimators (PIE) macro. *Statistics in Medicine*, *19*(11), 1495–1522. https://doi.org/10.1002/(SICI)1097-0258(20000615)19:11%3C1495::AID-SIM441%3E3.0.CO;2-E

Berry, J. D., Dyer, A., Cai, X., Garside, D. B., Ning, H., Thomas, A., Greenland, P., Van Horn, L., Tracy, R. P., & Lloyd-Jones, D. M. (2012). Lifetime risks of cardiovascular disease. *New England Journal of Medicine*, *366*(4), 321–329. https://doi.org/10.1056/NEJMoa1012848

Fine, J. P., & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, *94*(446), 496–509. https://doi.org/10.1080/01621459.1999.10474144

Lloyd-Jones, D. M., Larson, M. G., Beiser, A., & Levy, D. (1999). Lifetime risk of developing coronary heart disease. *The Lancet*, *353*(9147), 89–92. https://doi.org/10.1016/S0140-6736(98)10279-9

Pandey, A., Omar, W., Ayers, C., LaMonte, M., Klein, L., Allen, N. B., Kuller, L. H., Greenland, P., Eaton, C. B., Gottdiener, J. S., & others. (2018). Sex and race differences in lifetime risk of heart failure with preserved ejection fraction and heart failure with reduced ejection fraction. *Circulation*, *137*(17), 1814–1823. https://doi.org/10.1161/CIRCULATIONAHA.117.031622

Segar, M. W., Lambeth, K., Rosenblatt, A., Razavi, M., Pandey, A., & Saeed, M. (2025). Electrocardiographic p-wave parameters and lifetime atrial fibrillation risk: A multicohort study. *Heart Rhythm*. https://doi.org/10.1016/j.hrthm.2025.02.308

Wilkins, J. T., Ning, H., Berry, J., Zhao, L., Dyer, A. R., & Lloyd-Jones, D. M. (2012). Lifetime risk and years lived free of total cardiovascular disease. *JAMA*, *308*(17), 1795–1801. https://doi.org/10.1001/jama.2012.14312