




BGmisc: An R Package for Extended Behavior Genetics Analysis

S. Mason Garrison ¹¶, Michael D. Hunter ², Xuanyu Lyu ^{1,3,4}, Jonathan D. Trattner ¹, and S. Alexandra Burt ⁵

¹ Department of Psychology, Wake Forest University, North Carolina, USA ² Department of Human Development and Family Studies, Pennsylvania State University, Pennsylvania, USA ³ Institute for Behavioral Genetics, University of Colorado at Boulder, Boulder, USA ⁴ Department of Psychology & Neuroscience, University of Colorado at Boulder, Boulder, USA ⁵ Department of Psychology, Michigan State University, Michigan, USA ¶ Corresponding author

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Open Journals](#) 

Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#))

Summary

Behavior genetics is a field that studies how our genes and environment contribute to differences in behavior and traits among individuals. Traditionally, twin studies have long been a cornerstone of this field, helping researchers understand how genetics influence behavior. Recently, the focus has expanded to include studies with more complex family structures, e.g., children of twins (CoT, [D'Onofrio et al., 2003](#)), mother-daughter-aunt-niece, (MDAN, [Rodgers, Bard, Johnson, D'Onofrio, & Miller, 2008](#)). These broader studies offer more detailed insights into how our genes and environment shape us, but they also make analyzing and organizing the data more complex. BGmisc simplifies the analysis of these complex data structures by offering a comprehensive suite of functions for accommodating families of any size, from twins to extensive pedigrees.

Statement of need

The move towards analyzing complex family structures in behavior genetics introduces challenges in data structuring and modeling. The data structures inherent in these more complicated family designs are orders of magnitude larger than traditional designs. For example, in the classical twin study, a family will consist of a single pair of twins (i.e., two people), whereas in the MDAN design, a family consists of two mother-daughter pairs (i.e., four people). This problem quickly becomes intractable when applied to extended family pedigrees, which can encompass up to hundreds of thousands of individuals in a single family (e.g., [Garrison et al., 2023](#)).

This shift towards extended family models underscores the limitations of existing genetic modeling software. Packages like OpenMx ([Neale et al., 2016](#)), EasyMx ([Hunter, 2023](#)), and kinship2 ([J. P. Sinnwell, Therneau, & Schaid, 2014](#); [J. Sinnwell & Therneau, 2022](#)) were developed with smaller, classical family designs in mind. In contrast, the BGmisc R package was specifically developed to structure and model extended family pedigree data.

Two widely-used R packages in genetic modeling are OpenMx ([Neale et al., 2016](#)) and kinship2 ([J. P. Sinnwell et al., 2014](#); [J. Sinnwell & Therneau, 2022](#)). The OpenMx ([Neale et al., 2016](#)) package is a general-purpose software for structural equation modeling that is popular among behavior geneticists ([Garrison, 2018](#)) for its unique features, like the `mxCheckIdentification()` function. This function checks whether a model is identified, determining if there is a unique solution to estimate the model's parameters based on the observed data. In addition, EasyMx ([Hunter, 2023](#)) is a more user-friendly package that streamlines the process of building and

estimating structural equation models. It seamlessly integrates with OpenMx's infrastructure. Its functionalities range from foundational matrix builders like `emxCholeskyVariance` and `emxGeneticFactorVariance` to more specialized functions like `emxTwinModel` designed for classical twin models. Despite their strengths, EasyMx and OpenMx have limitations when handling extended family data. Notably, they lack functions for handling modern molecular designs (Kirkpatrick, Pritikin, Hunter, & Neale, 2021), modeling complex genetic relationships, inferring relatedness, and simulating pedigrees.

Although not a staple in behavior genetics, the `kinship2` (J. P. Sinnwell et al., 2014) package provides core features to the broader statistical genetics scientific community, such as plotting pedigrees and computing genetic relatedness matrices. It uses the Lange algorithm (Lange, 2002) to compute relatedness coefficients. This recursive algorithm is discussed in great detail elsewhere, laying out several boundary conditions and recurrence rules. The `BGmisc` package extends the capabilities of `kinship2` by introducing an alternative algorithm to calculate the relatedness coefficient based on network models. By applying classic path-tracing rules to the entire network, this new method is computationally more efficient by eliminating the need for a multi-step recursive approach.

Features

The `BGmisc` package offers features tailored for extended behavior genetics analysis. These features are grouped under two main categories, mirroring the structure presented in our vignettes.

Modeling and Relatedness

- **Model Identification:** `BGmisc` evaluates whether a variance components model is identified and fits the model's estimated variance components to observed covariance data. The technical aspects related to model identification have been described by Hunter, Garrison, Burt, & Rodgers (2021).
- **Relatedness Coefficient Calculation:** Using path tracing rules first described by Wright (1922) and formalized by McArdle & McDonald (1984), `BGmisc` calculates the (sparse) relatedness coefficients between all pairs of individuals in extended pedigrees based solely on mother and father identifiers.
- **Relatedness Inference:** `BGmisc` infers the relatedness between two groups based on their observed total correlation, given additive genetic and shared environmental parameters.

Pedigree Analysis and Simulation

- **Pedigree Conversion:** `BGmisc` converts pedigrees into various relatedness matrices, including additive genetics, mitochondrial, common nuclear, and extended environmental relatedness matrices.
- **Pedigree Simulation:** `BGmisc` simulates pedigrees based on parameters including the number of children per mate, generations, sex ratio of newborns, and mating rate.

Collectively, these tools provide a valuable resource for behavior geneticists and others who work with extended family data. They were developed as part of a grant and have been used in several ongoing projects (Burt, 2023; Garrison et al., 2023; Hunter et al., 2023; Lyu et al., 2023) and theses (Lyu, 2023).

Availability

The `BGmisc` package is open-source and available on both GitHub at <https://github.com/R-Computing-Lab/BGmisc> and the Comprehensive R Archive Network (CRAN) at <https://>

86 [//cran.r-project.org/package=BGmisc](https://cran.r-project.org/package=BGmisc). It is licensed under the GNU General Public License.

87 Acknowledgments

88 The current research is supported by the National Institute on Aging (NIA), RF1-AG073189.
89 We want to acknowledge assistance from Carlos Santos.

90 References

- 91 Burt, S. A. (2023). Mom genes: Leveraging maternal lineage to estimate the contributions of
92 mitochondrial DNA. *Behavior Genetics*.
- 93 D'Onofrio, B. M., Turkheimer, E. N., Eaves, L. J., Corey, L. A., Berg, K., Solaas, M. H.,
94 & Emery, R. E. (2003). The role of the children of twins design in elucidating causal
95 relations between parent characteristics and child outcomes. *Journal of Child Psychology*
96 *and Psychiatry*, 44(8), 1130–1144. doi:[ff3dbm](https://doi.org/10.1111/j.1469-7610.2003.01385.x)
- 97 Garrison, S. M. (2018). Popular Structural Equation Modeling Programs for Behavior Ge-
98 netics. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(6), 972–977.
99 doi:[10.1080/10705511.2018.1493385](https://doi.org/10.1080/10705511.2018.1493385)
- 100 Garrison, S. M., Lyu, X., Hunter, M. D., Rodgers, J. L., Smith, K. R., Coon, H., & Burt, S.
101 A. (2023). Analyzing extended cousin similarity to unravel the mystery of mtDNA and
102 longevity. *Behavior Genetics*.
- 103 Hunter, M. D. (2023). *EasyMx: Easy model-builder functions for 'OpenMx'*. Retrieved from
104 <https://CRAN.R-project.org/package=EasyMx>
- 105 Hunter, M. D., Garrison, S. M., Burt, S. A., & Rodgers, J. L. (2021). The Analytic Identification
106 of Variance Component Models Common to Behavior Genetics. *Behavior Genetics*, 51(4),
107 425–437. doi:[10.1007/s10519-021-10055-x](https://doi.org/10.1007/s10519-021-10055-x)
- 108 Hunter, M. D., Lyu, X., Garrison, S. M., Rodgers, J. L., Smith, K., Coon, H., & Burt, S. A.
109 (2023). Modeling mtDNA effects from extended pedigrees in the utah population database.
110 *Behavior Genetics*.
- 111 Kirkpatrick, R. M., Pritikin, J. N., Hunter, M. D., & Neale, M. C. (2021). Combining Structural-
112 Equation Modeling with Genomic-Relatedness-Matrix Restricted Maximum Likelihood in
113 OpenMx. *Behavior Genetics*, 51(3), 331–342. doi:[10.1007/s10519-020-10037-5](https://doi.org/10.1007/s10519-020-10037-5)
- 114 Lange, K. (2002). Genetic Identity Coefficients. In K. Lange (Ed.), *Mathematical and*
115 *Statistical Methods for Genetic Analysis*, Statistics for Biology and Health (pp. 81–96).
116 New York, NY: Springer. doi:[10.1007/978-0-387-21750-5_5](https://doi.org/10.1007/978-0-387-21750-5_5)
- 117 Lyu, X. (2023, May). *Statistical power analysis on mtDNA effects estimation* (Master's thesis).
118 Wake Forest University.
- 119 Lyu, X., Hunter, M. D., Rodgers, J. L., Smith, K. R., Coon, H., Burt, S. A., & Garrison, S. M.
120 (2023). Statistical power analysis on mtDNA effects estimation. *Behavior Genetics*.
- 121 McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the reticular action
122 model for moment structures. *British Journal of Mathematical and Statistical Psychology*,
123 37, 234–251. doi:[10.1111/j.2044-8317.1984.tb00802.x](https://doi.org/10.1111/j.2044-8317.1984.tb00802.x)
- 124 Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M.,
125 Estabrook, R., et al. (2016). OpenMx 2.0: Extended Structural Equation and Statistical
126 Modeling. *Psychometrika*, 81(2), 535–549. doi:[f8rfrg](https://doi.org/10.1007/s11336-015-9498-4)

- 127 Rodgers, J. L., Bard, D. E., Johnson, A., D'Onofrio, B., & Miller, W. B. (2008). The Cross-
128 Generational Mother–Daughter–Aunt–Niece Design: Establishing Validity of the MDAN De-
129 sign with NLSY Fertility Variables. *Behavior Genetics*, 38(6), 567–578. doi:[10.1007/s10519-](https://doi.org/10.1007/s10519-008-9225-0)
130 [008-9225-0](https://doi.org/10.1007/s10519-008-9225-0)
- 131 Sinnwell, J. P., Therneau, T. M., & Schaid, D. J. (2014). The kinship2 r package for pedigree
132 data. *Human Heredity*, 78, 91–93. doi:[10.1159/000363105](https://doi.org/10.1159/000363105)
- 133 Sinnwell, J., & Therneau, T. (2022). *kinship2: Pedigree functions*. Retrieved from [https:](https://CRAN.R-project.org/package=kinship2)
134 [//CRAN.R-project.org/package=kinship2](https://CRAN.R-project.org/package=kinship2)
- 135 Wright, S. (1922). Coefficients of inbreeding and relationship. *The American Naturalist*,
136 56(645), 330–338. doi:[10.1086/279872](https://doi.org/10.1086/279872)

DRAFT