The Journal of Open Source Software

individual: An R package for individual-based epidemiological models

Giovanni D. Charles[1] **and Sean L. Wu**[2]

1 MRC Centre for Global Infectious Disease Analysis, Abdul Latif Jameel Institute for Disease and Emergency Analytics (J-IDEA), Imperial College London, London, UK. **2** Division of Epidemiology and Biostatistics, School of Public Health, University of California, Berkeley, CA 94720, USA

DOI: 10.21105/joss.03539

Software

- Review
- Repository
- Archive

Editor: Mikkel Meyer Andersen

Reviewers:

- @seabbs
- @strengejacke
- @SteRoe

Submitted: 16 June 2021
Published: 05 October 2021

Summary

`individual` is an R package which provides users a set of useful primitive elements for specifying individual-based models (IBMs), also called agent-based models (ABMs), with special attention to models for infectious disease epidemiology. Users build models by specifying variables for each characteristic describing individuals in the simulated population using data structures from the package. `individual` provides efficient methods for finding subsets of individuals based on these variables, or cohorts. Cohorts can then be targeted for variable updates or scheduled for events. Variable updates queued during a time step are executed at the end of a discrete time step, and the code places no restrictions on how individuals are allowed to interact. These data structures are designed to provide an intuitive way for users to turn their conceptual model of a system into executable code, which is fast and memory efficient.

Statement of need

Complex stochastic models are crucial for many tasks in infectious disease epidemiology. Such models can formalize theory, generate synthetic data, evaluate counterfactual scenarios, forecast trends, and be used for statistical inference (Ganyani et al., 2021). IBMs are a way to design disaggregated simulation models, usually contrasted with mathematical models, which may model a density or concentration of individuals, or otherwise lump individuals with similar attributes together in some way (Shalizi, 2006). For modeling finite numbers of individuals with significant between-individual heterogeneity and complex dynamics, IBMs are a natural modeling choice when a representation using mathematical models would be cumbersome or impossible (Willem et al., 2017). Even if an aggregated representation were feasible, there are many reasons why an individual-based representation is to be preferred. Synthetic data may need to produce individual level outcomes, which aggregated models by their very nature are unable to provide (Tracy et al., 2018). Other complexities, such as when events occur after a random delay whose distribution differs from a Markovian one, mean even aggregated models will need to store individual completion times, necessitating more complex simulation algorithms and data structures; in such cases it is often more straightforward to adopt an individual-based representation from the start.

For practical use, individual-based models need to balance comprehensibility and speed. A fast model whose code is only understood by the author can be difficult to use as a basis for scientific exploration, which necessarily requires the development of multiple models to test hypotheses or explore sensitivity to certain assumptions. On the other hand a clear yet slow model can be practically unusable for tasks such as uncertainty quantification or statistical

inference. `individual` provides a toolkit for users to write models that is general enough to cover nearly all models of practical interest using simple, standardized code which is fast enough to be useful for computationally taxing applications.

State of the field

There are many software libraries for epidemiological simulation, both in R and other programming languages. However, based on our review of existing software, no other library exists in the R language which provides users with a set of primitive elements for defining epidemiological models without imposing strong restrictions upon the type of model that may be simulated (e.g.; compartmental, network, etc.), or limiting users to particular mathematical forms for model dynamics.

General R packages

Generic individual-based simulation packages in R include IBMPopSim (Giorgi et al., 2020), ibm (Oliveros-Ramos, 2016) and ibmcraftr (Tun, 2016). IBMPopSim provides sophisticated simulation algorithms, but requires users to input C++ code as a string which is then compiled, making it difficult to interface with the existing R ecosystem.

Epidemiological R packages

EpiModel (Jenness et al., 2018) allows the simulation of highly detailed discrete time models on networks, relying on the statnet (Krivitsky et al., 2003-2020) project for classes and algorithms. However due to its focus on directly transmitted diseases, `individual` may be more applicable to other epidemiological situations such as vector-borne diseases. In addition it does not offer an interface for compiled code.

hybridModels (Marques et al., 2020), similarly provides tools for simulating epidemics on dynamic networks. However, it is fully implemented in R, limiting the scope for scale and optimisation.

Other packages in R are more specialised or restrict the model's transmission dynamics to specific mathematical forms (e.g.; mass action). These include SimInf (Bauer et al., 2016), nosoi (Lequime et al., 2020), SPARSEMODr (Mihaljevic et al., 2021), EpiILMCT (Almutiry & Deardon, 2020) and EpiILM (Warriyar et al., 2020).

Design principles

Because in many epidemiological models the most important representation of state is a finite set of mutually exclusive values, such as the Susceptible, Infectious, Recovered classes from the well-known SIR model (Allen, 2017), `individual` uses a bitset to store these data. At the R level users can call set operations (union, intersection, complement, symmetric difference, set difference) which are implemented as bitwise operations in the C++ source. This lets users write clear, highly efficient code for updating their model, fully in R.

In contrast to other individual-based modeling software, where users focus on defining a type for simulated individuals, in `individual` users instead define variables, one for each characteristic of the simulated population. Individual agents are defined by their their position in each bitset giving membership in a variable, or element in a vector of integers or floats. This design is similar to a component system, a design pattern to help decouple complicated types (Nystrom, 2014). Because of this disaggregated representation of state, performing

operations to find and schedule cohorts of individuals benefits from fast bitwise operators. This state representation is (to our knowledge), novel for epidemiological simulation. While Rizzi et al. (2018) proposed using a bitset to represent the state of each simulated individual, the population was still stored as types in an array.

`individual` uses `Rcpp` (Eddelbuettel & François, 2011) to link to C++ source code, which underlies the data structures exposed to the user. The API for `individual` uses `R6` (Chang, 2020) classes at the R level which users call to create, update, and query variables. `individual` also provides a C++ header-only interface which advanced users can link to from their R package. Users can then write their own C++ code or benefit from other packages with a compiled interface, significantly enhancing the extensibility of `individual`'s API, and documentation on interacting with `individual`'s C++ API is available in the package documentation.

After a user has specified all the variables in their model, dynamics are specified by processes which run each time step, and events which can be scheduled to update specific cohorts in the future. The simulation loop then executes processes, fires events and updates state on each discrete time step.

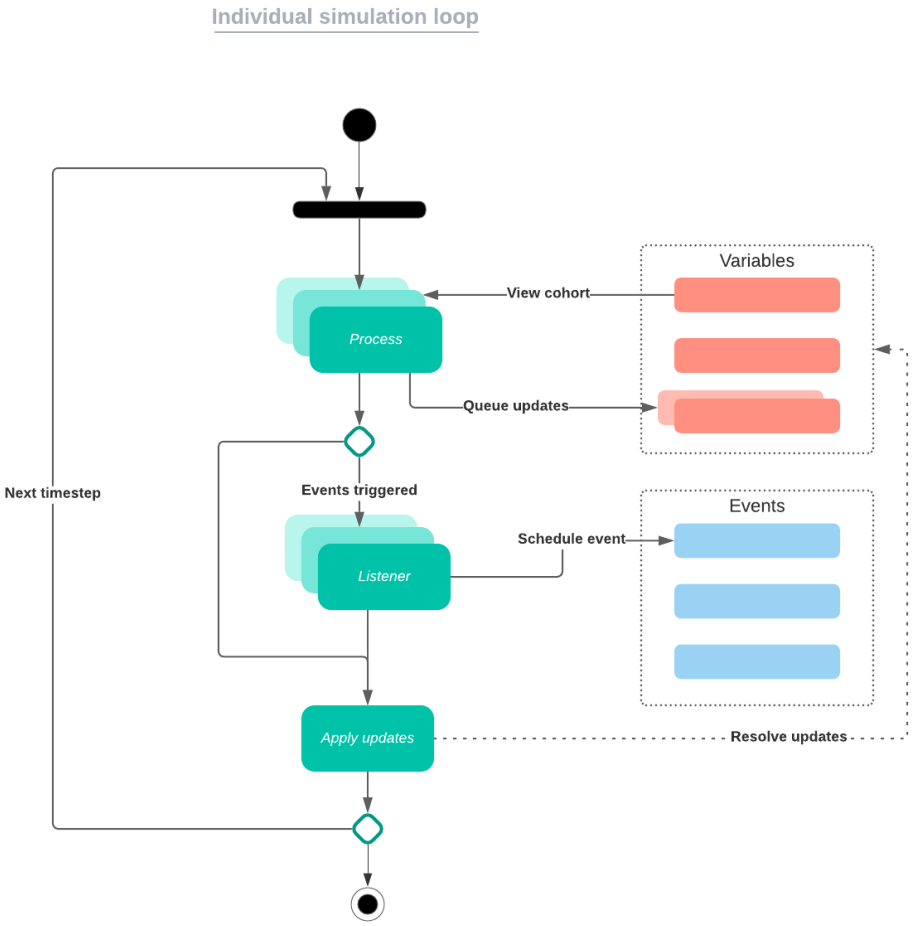

Figure 1: A flow diagram for the simulation loop

Licensing and availability

`individual` is licensed under the MIT License, with all source code stored at GitHub. Requests, suggestions, and bug reports are encouraged via filing an issue. A general guide on how to contribute to `individual` is available at the package's website. The automated test coverage can be found at codecov.io. Example code can be found in the tutorial section of the package documentation.

Acknowledgements

We would like to thank Dr. Pete Winskill and Dr. Oliver Watson for their encouragement, testing and contributions to the repository. And Dr. Richard Fitzjohn for his early technical feedback and expert advice on R package development.

References

Allen, L. J. (2017). A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling*, *2*(2), 128–142. https://doi.org/10.1016/j.idm.2017.03.001

Almutiry, W., & Deardon, R. (2020). *EpiILMCT: Continuous time distance-based and network-based individual level models for epidemics*. https://CRAN.R-project.org/package=EpiILMCT

Bauer, P., Engblom, S., & Widgren, S. (2016). Fast event-based epidemiological simulations on national scales. *The International Journal of High Performance Computing Applications*, *30*(4), 438–453. https://doi.org/10.1177/1094342016635723

Chang, W. (2020). *R6: Encapsulated classes with reference semantics*. https://CRAN.R-project.org/package=R6

Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, *40*(8), 1–18. https://doi.org/10.18637/jss.v040.i08

Ganyani, T., Faes, C., & Hens, N. (2021). Simulation and analysis methods for stochastic compartmental epidemic models. *Annual Review of Statistics and Its Application*, *8*(1), 69–88. https://doi.org/10.1146/annurev-statistics-061120-034438

Giorgi, D., Kaakai, S., & Lemaire, V. (2020). *IBMPopSim: Individual based model population simulation*. https://CRAN.R-project.org/package=IBMPopSim

Jenness, S. M., Goodreau, S. M., & Morris, M. (2018). EpiModel: An r package for mathematical modeling of infectious disease over networks. *Journal of Statistical Software*, *84*. https://doi.org/10.18637/jss.v084.i08

Krivitsky, P. N., Handcock, M. S., Hunter, D. R., Butts, C. T., Klumb, C., Goodreau, S. M., & Morris, M. (2003-2020). *Statnet: Software tools for the statistical modeling of network data*. Statnet Development Team. http://statnet.org

Lequime, S., Bastide, P., Dellicour, S., Lemey, P., & Baele, G. (2020). Nosoi: A stochastic agent-based transmission chain simulation framework in r. *Methods in Ecology and Evolution*, *11*(8), 1002–1007. https://doi.org/10.1111/2041-210X.13422

Marques, F. S., Grisi-Filho, J. H. H., Silva, J. C. R., Almeida, E. C., & Júnior, J. L. S. (2020). hybridModels: An R package for the stochastic simulation of disease spreading

in dynamic networks. *Journal of Statistical Software*, *94*(6), 1–32. https://doi.org/10.18637/jss.v094.i06

Mihaljevic, J. R., Borkovec, S., Ratnavale, S., Hocking, T. D., Banister, K. E., Eppinger, J. E., Hepp, C. M., & Doerry, E. (2021). SPARSEMODr: Rapid simulations of spatially explicit and stochastic models of infectious disease. *medRxiv*. https://doi.org/10.1101/2021.05.13.21256216

Nystrom, R. (2014). *Game programming patterns*. Genever Benning.

Oliveros-Ramos, R. (2016). *Ibm: Individual based models in r*. https://CRAN.R-project.org/package=ibm

Rizzi, R. L., Kaizer, W. L., Rizzi, C. B., Galante, G., & Coelho, F. C. (2018). Modeling direct transmission diseases using parallel bitstring agent-based models. *IEEE Transactions on Computational Social Systems*, *5*(4), 1109–1120. https://doi.org/10.1109/tcss.2018.2871625

Shalizi, C. R. (2006). Methods and techniques of complex systems science: An overview. *Complex Systems Science in Biomedicine*, 33–114. https://doi.org/10.1007/978-0-387-33532-2_2

Tracy, M., Cerdá, M., & Keyes, K. M. (2018). Agent-based modeling in public health: Current applications and future directions. *Annual Review of Public Health*, *39*(1), 77–94. https://doi.org/10.1146/annurev-publhealth-040617-014317

Tun, S. T. T. (2016). *Ibmcraftr: Toolkits to develop individual-based models in infectious disease*. https://CRAN.R-project.org/package=ibmcraftr

Warriyar, V. K. V., Almutiry, W., & Deardon, R. (2020). *EpiILM: Spatial and network based individual level models for epidemics*. https://CRAN.R-project.org/package=EpiILM

Willem, L., Verelst, F., Bilcke, J., Hens, N., & Beutels, P. (2017). Lessons from a decade of individual-based models for infectious disease transmission: A systematic review (2006-2015). *BMC Infectious Diseases*, *17*(1), 1–16. https://doi.org/10.1186/s12879-017-2699-8