

# conmat: generate synthetic contact matrices for a given age-stratified population

Nicholas Tierney  <sup>1</sup>¶, Chitra Saraswati  <sup>1</sup>, Aarathy Babu  <sup>1</sup>, Michael Lydeamore  <sup>3</sup>, and Nick Golding  <sup>1,2</sup>

1 The Kids Research Institute Australia, WA, Australia 2 University of Western Australia, WA, Australia

3 Department of Econometrics and Business Statistics, Monash University, VIC, Australia ¶

Corresponding author

DOI: [10.21105/joss.08326](https://doi.org/10.21105/joss.08326)

## Software

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

---

Editor: Charlotte Soneson  <sup>1</sup>

Reviewers:

- [@anthbapt](#)
- [@Bisaloo](#)

Submitted: 10 March 2025

Published: 05 February 2026

## License

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

## Summary

Contact matrices describe the number of contacts between individuals. They are used to create models of infectious disease spread. `conmat` is an R package which generates synthetic contact matrices for arbitrary input demography, ready for use in infectious disease modelling.

The `conmat` package exposes model fitting and prediction separately to the user. Users can fit a model based on a contact survey, then predict from this model to their own demographic data. This means users can generate synthetic contact matrices for any region, with any contact survey.

We demonstrate a use-case for `conmat` by creating contact matrices for a sub-national entity (in this case, a state) in Australia.

For users who do not wish to run the entire `conmat` pipeline, we have pre-generated synthetic contact matrices for 200 countries, based on a list of countries from the United Nations, using a model fit to the POLYMOD contact survey. These resulting synthetic contact matrices, and the associated code, can be found in the `syncomat` analysis pipeline ([GitHub](#), [Zenodo](#)) ([Saraswati et al., 2024](#)).

## Statement of need

Infectious diseases like influenza and COVID-19 spread via social contact. If we can understand patterns of contact—which individuals are more likely be in contact with each other—then we will be able to create accurate models of how disease spreads. Epidemiologists and public policy makers can use these models to make decisions to keep a population safe and healthy.

Empirical estimates of social contact are provided by social contact surveys, which provide samples of the frequency and type of social contact across different settings (home, work, school, other). These surveys can be projected onto a given demographic structure, resulting in what are known as ‘synthetic’ contact matrices. A widely used approach by Prem et al. ([2017](#), [2021](#)) produced such matrices for 177 countries at ‘urban’ and ‘rural’ levels for each country using data from eight European countries by Mossong et al. ([2008](#)).

However, there were major limitations with the methods in Prem et al. ([2021](#)). First, not all countries were included in their analyses. Second, the contact matrices only covered broad population groups within entire countries. Third, the code provided by Prem et al. was not designed for replicability and easy modification with user-defined inputs.

The `conmat` package was developed to fill the specific need of creating contact matrices for arbitrary age categories and populations to inform infectious disease models. We developed

the method to output synthetic contact matrices. We also provided methods to create next-generation matrices.

## Example

We will generate a contact matrix for Tasmania, a state in Australia, using a model fitted from the POLYMOD contact survey. We can get the age-stratified population data for Tasmania from the Australian Bureau of Statistics (ABS) with the helper function `abs_age_state()`:

```
tasmania <- abs_age_state("TAS")
head(tasmania)

# A tibble: 6 × 4 (conmat_population)
#>   year state lower.age.limit population
#>   <dbl> <chr>      <dbl>      <dbl>
#> 1 2020 TAS            0      29267
#> 2 2020 TAS            5      31717
#> 3 2020 TAS           10      33318
#> 4 2020 TAS           15      31019
#> 5 2020 TAS           20      31641
#> 6 2020 TAS           25      34115
```

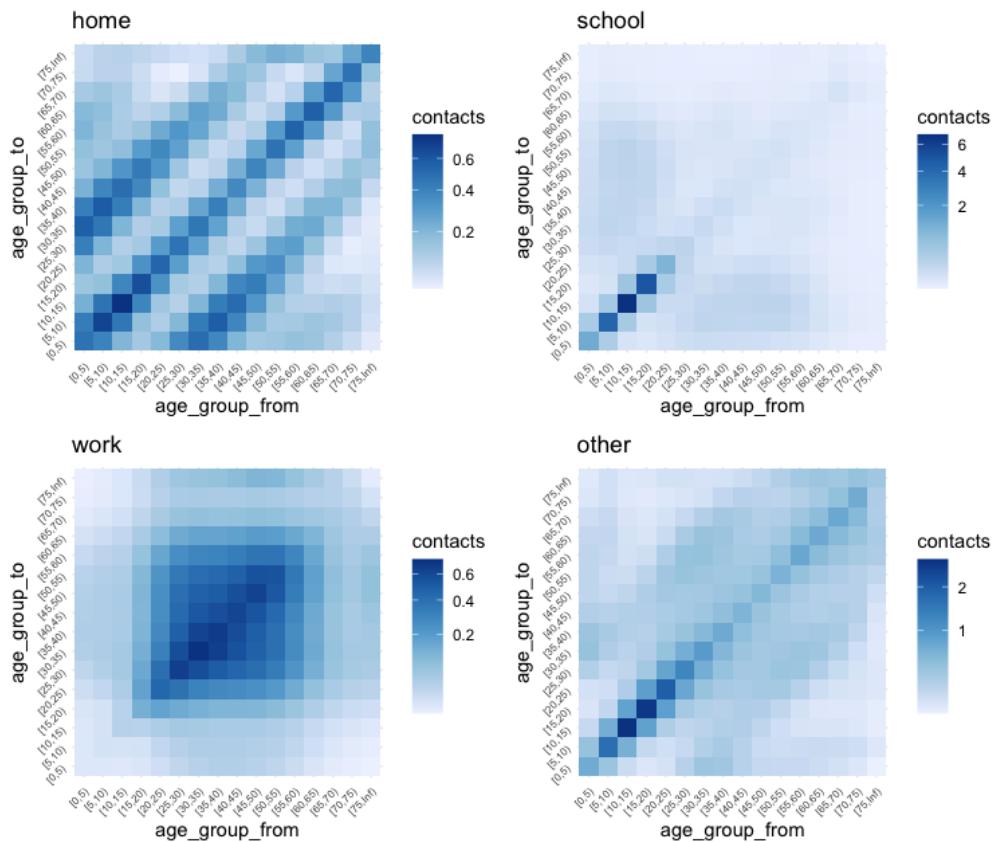
We can then generate a synthetic contact matrix for Tasmania, by extrapolating the contact patterns between age groups learned from the POLYMOD study, using `extrapolate_polymod()`.

```
tasmania_contact <- extrapolate_polymod(population = tasmania)
tasmania_contact
```

We can plot the resulting contact matrix for Tasmania with `autoplot`, shown in Figure 1.

```
autoplot(tasmania_contact)
```

Setting-specific synthetic contact matrices



**Figure 1:** Contact patterns between individuals for different age groups across four settings: home, work, school, and other. The x axis shows the age groups for focal individuals ('from'), and the y axis shows the age groups for people those individuals have contact with ('to'), coloured by the average number of contacts the individual has in that age group. We see different contact patterns in different settings, for example, diagonal with 'wings' for the home setting.

## Implementation

The overall approach of conmat has two parts:

- 1) fit a model to predict individual contact rates, using an existing contact survey;
- 2) predict a synthetic contact matrix using age population data.

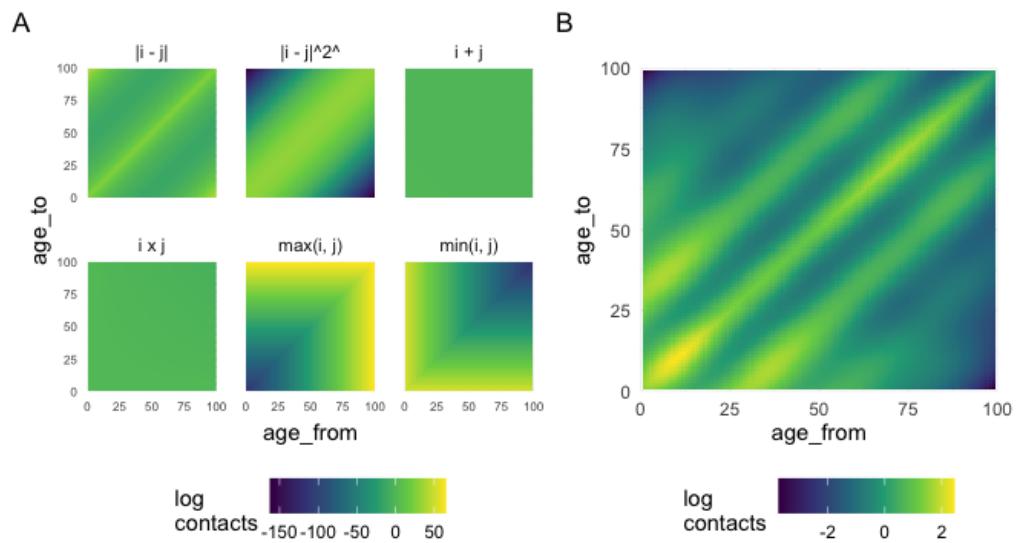
### Model fitting

conmat was built to predict at four settings: work, school, home, and other. One model is fitted for each setting. Each model fitted is a Poisson generalised additive model (GAM) with a log link function, which predicts the count of contacts, with an offset for the log of the number of participants. The model has six covariates to explain key features of the relationship between ages, and two optional covariates for attendance at school or work.

Each cell in the resulting contact matrix (after back-transformation of the link function),

indexed  $(i, j)$ , is the predicted number of people in age group  $j$  that a single individual in age group  $i$  will have contact with per day. The sum over all age groups  $j$  for a particular age group  $i$  is the predicted total number of contacts per day for each individual of age group  $i$ .

The model captures typical features of inter-person contact, where individuals primarily interact with people of similar age (the diagonals of the matrix), and with grandparents and/or children (the so-called ‘wings’ of the matrix). The key features of the relationship between the age groups are displayed in Figure 2 for the home setting.



**Figure 2:** Partial predictive plot (A) and overall synthetic contact matrix (B) for the Poisson GAM fitted to the POLYMOD contact survey in the home setting. The strong diagonal elements, and parents/grandparents interacting with children result in the classic ‘diagonal with wings’ shape.

Visualising the partial predictive plots for other settings (school, work and other) show patterns that correspond with real-life situations. A full visualisation pipeline is available at <https://idem-lab.github.io/conmat/dev/articles/visualising-conmat.html>.

## Conclusions and future directions

The conmat software provides a flexible interface to generating synthetic contact matrices using population data and contact surveys. These contact matrices can then be used in infectious disease modelling and surveillance. Public health decisions are based on models using these matrices as input data. This is the first piece of software that can provide context-appropriate contact matrices for a population, which means more accurate models of disease.

The main strength of conmat is its interface requiring only age population data to create a synthetic contact matrix. Current approaches provide only a selection of country-level contact matrices. This software can predict to arbitrary demography, such as sub-national or simulated populations.

We provide a trained model of contact rate that is fit to the POLYMOD survey for ease of use. The software also has an interface to train models to other contact surveys, such as Jarvis et al. (2024). This is important as POLYMOD represents contact patterns in 8 countries in Europe, and contact patterns are known to differ across nations and cultures.

The covariates used by conmat were designed to represent the key features that are typically present in a contact matrix for different settings (work, school, home, other). Including

other sources of information that may better describe these contact patterns, such as inter-generational mixing, or differences in school ages of a local demographic, may improve model performance.

The code underlying this software was used as a key input into several models for COVID-19 transmission and control in Australia and contributed to decisions around vaccination policy (Conway et al., 2023; McVernon et al., n.d.; Ryan et al., 2024).

Software is never finished, and the software in its current format has proven useful for infectious disease modelling. In time, we hope it can become more widely used and be useful for more applications in epidemiology and public health.

## Acknowledgements

The authors would like to thank Gerry Ryan for his contributions to *commat*, as well as the authors of “Projecting social contact matrices in 152 countries using contact surveys and demographic data”, Kiesha Prem, Alex Cook, and Mark Jit, for making their original data public. We would also like to thank reviewers Hugo Gruson and Anthony Baptista for their contributions.

## References

Conway, E., Walker, C. R., Baker, C., Lydeamore, M. J., Ryan, G. E., Campbell, T., Miller, J. C., Reboli, N., Yeung, M., Kabashima, G., Geard, N., Wood, J., McCaw, J. M., McVernon, J., Golding, N., Price, D. J., & Shearer, F. M. (2023). COVID-19 vaccine coverage targets to inform reopening plans in a low incidence setting. *Proceedings of the Royal Society B: Biological Sciences*, 290(2005), 20231437. <https://doi.org/10.1098/rspb.2023.1437>

Jarvis, C., Coletti, P., Backer, J., Munday, J., Faes, C., Beutels, P., Althaus, C., Low, N., Wallinga, J., Hens, N., & Edmunds, J. (2024). *CoMix data (last round in BE, CH, NL and UK)* [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.11154066>

McVernon, J., McCaw, J., Tierney, N., Miller, J., Lydeamore, M., Golding, N., Shearer, F., Geard, N., Cameron, Z., Baker, C., Walker, C., Ross, J., Wood, J., Conway, E., & Mueller, I. (n.d.). *Doherty institute - modelling*. Retrieved August 19, 2024, from <https://www.doherty.edu.au/our-work/institute-themes/viral-infectious-diseases/covid-19/covid-19-modelling/modelling>

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G. S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., & Edmunds, W. J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Medicine*, 5(3), e74. <https://doi.org/10.1371/journal.pmed.0050074>

Prem, K., Cook, A. R., & Jit, M. (2017). Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLOS Computational Biology*, 13(9), e1005697. <https://doi.org/10.1371/journal.pcbi.1005697>

Prem, K., van Zandvoort, K., Klepac, P., Eggo, R. M., Davies, N. G., Group, C. for the M. M. of I. D. C. W., Cook, A. R., & Jit, M. (2021). Projecting contact matrices in 177 geographical regions: An update and comparison with empirical data for the COVID-19 era. *PLOS Computational Biology*, 17(7), e1009098. <https://doi.org/10.1371/journal.pcbi.1009098>

Ryan, G. E., Shearer, F. M., McCaw, J. M., McVernon, J., & Golding, N. (2024). Estimating measures to reduce the transmission of SARS-CoV-2 in Australia to guide a “National Plan” to reopening. *Epidemics*, 47, 100763. <https://doi.org/10.1016/j.epidem.2024.100763>

Saraswati, C. M., Lydeamore, M., Golding, N., Babu, A., & Tierney, N. (2024). *syncomat: Synthetic Contact Matrices for 200 UN Countries* (Version v1.0.0) [Data set]. Zenodo.

<https://doi.org/10.5281/zenodo.11365943>