

# Thom Benjamin Volker

Utrecht, The Netherlands

 [thomvolker.github.io](https://github.com/thomvolker)

 [t.b.volker@uu.nl](mailto:t.b.volker@uu.nl)

 [GitHub](#)

 [LinkedIn](#)

## Education and experience

---

**PhD Candidate Synthetic Data** **2022 - Current**

*Utrecht University, Statistics Netherlands*

- Supervisors: S. van Buuren, E.-J. van Kesteren, P.-P. de Wolf

**MSc Methodology and Statistics**

**2019 - 2022**

*Utrecht University*

*Valedictorian, Cum laude (GPA 8.9/10)*

- Internship missing data with Gerko Vink and Stef van Buuren
- Various student-assistant jobs: data visualization, Shiny app development, contributions to [ASReview](#) and teaching.

**MSc Sociology and Social Research**

**2019 - 2022**

*Utrecht University*

*Cum laude (GPA 8.7/10)*

**BA Liberal Arts & Sciences**

**2016 - 2019**

*Utrecht University*

*(GPA 7.7/10)*

## Teaching

---

### ◦ Post graduate-level courses

- [Solving Missing Data Problems in R](#) [2019 - 2023]: *Developing materials, supervising practicals.*
- [Machine learning with python](#) [2024]: *Supervising practicals.*
- [Statistical Programming with R](#) [2021]: *Supervising practicals.*
- [Survey Research: Statistical Analysis and Estimation](#) [2021]: *Supervising practicals.*

### ◦ Master level courses

- Data wrangling [2021 - Current]: *Supervising practicals and grading exams.*
- Battling the curse of dimensionality [2022 - 2025]: *Developing course materials, supervising practical sessions and grading exams.*
- Network Analysis [2020 - 2021]: *Developing course materials, supervising practicals and grading exams.*
- Methodological and Statistical Aspects of Social Science Research [2019 - 2020]: *Supervising practicals.*

### ◦ Bachelor level courses

- Theory Construction and Statistical Modeling [2019- 2020]: *Supervising practicals*
- Various undergraduate courses on standard statistical methods [2018-2021].

## Supervision

---

- Ludo Mels (Master's thesis, co-supervised with [Dr. Sara van Erp](#), 2026-2027).
- Heleen Brügger (Master's thesis, co-supervised with [Hanne Oberman](#) & [Gerko Vink](#), 2023-2024).
- Mirthe Hendriks (Master's thesis, co-supervised with [Gerko Vink](#), 2021).
- Stijn van den Broek (Master's thesis, co-supervised with [Gerko Vink](#), 2021).
- Romain Rey (Bachelor's thesis, co-supervised with [Irene Klugkist](#), 2020).

## Miscellanea

---

**IOPS PhD Representative [2022 - 2024]:** Represent PhD candidates within the Interuniversity Graduate School of Psychometrics and Sociometrics.

**Social committee [2022 - Current]:** Organising frequent social activities for the department.

**Debuut [2019 - 2020]:** Participate in buddy programme at Utrecht University to match potential first generation university students (~ 11 year olds) with actual students to get a flavour of what studying at a university entails.

## Received funding

---

### SynthEval

2023-2024

#### Project lead

- An R package for evaluating and improving the quality of synthetic data sets. For €15.000.

## Peer-reviewed publications

---

- Klugkist, I. & **Volker, T. B.** (2023). Bayesian Evidence Synthesis for Informative Hypotheses: An introduction. *Psychological Methods, Advance online publication*. [Paper](#).
- Van Leeuwen, A., Bagheri, A., **Volker, T. B.** & Van Brakel, Charlotte (2023). Verkenning van de inzet van topic modelling bij het analyseren van schrijfoopdrachten in de context van de flipped classroom. [Exploring the use of topic modelling for the analysis of written assignments in a flipped classroom context.] *Tijdschrift voor Hoger Onderwijs, 41* (2), 84-98. [Paper](#).
- **Volker, T. B.** & Vink, G. (2021). Anonymiced Shareable Data: Using mice to Create and Analyze Multiply Imputed Synthetic Datasets. *psych, 3*, 703-716. [Paper](#).
- **Volker, T. B.**, Buskens, V. & Raub, W. (Forthcoming). How Networks Facilitate Trust: A Synthesis on Control Effects. In *Handbook of Trust & Social Networks*.

## Software development

---

- `densityratio` [aut] - Developed R-package for directly estimating the ratio between two sample densities.
- `mice` [ctb] - Various additions to the multiple imputation software `mice`.
- `jaspMissingData` [ctb] - A JASP module for multiple imputation with the R-package `mice`.
- `ggmice` [ctb] - Adjusted functionality to allow for visualization of synthetic data.

## Preprints, vignettes, blogs and other writings

---

- **Volker, T.B.**, De Wolf, P.-P. & Van Kesteren, E. J. (Under review). A density ratio framework for evaluating the utility of synthetic data. [Paper](#).
- **Volker, T.B.**. Generating and evaluating synthetic data in R. [Commissioned tutorial for LMU Open Science Center](#).
- Poses, C. & **Volker, T.B.** (2025). `densityratio`: An R-package for distribution comparison through density ratio estimation. [Software vignette](#).
- Poses, C. & **Volker, T.B.** (2025). Covariate shift adjustment with the `densityratio` R-package. [Software vignette](#).
- **Volker, T.B.** (2025). High dimensional two-sample testing with the `densityratio` R-package. [Software vignette](#).
- **Volker, T.B.** (2025). Illustrative GANs for simply synthetic data with `keras` in R. [Blog post](#).
- **Volker, T.B.** (2025). Different ways of calculating OLS regression coefficients (in R). [Blog post](#).

- **Volker, T.B.** (2025). Statistical power of the Bayesian one-sample t-test. [Blog post](#).
- Research Data Management Support (2023). Data Privacy Handbook. [Online book](#).
- **Volker, T.B.** & Klugkist, I. (2023). Combining support for hypotheses over heterogeneous studies with Bayesian Evidence Synthesis: A simulation study. [arXiv](#).
- **Volker, T.B.** & Vink, G. (2022). futuremice: The future starts today. [https://www.gerkovink.com/futuremice/Vignette\\_futuremice.html](https://www.gerkovink.com/futuremice/Vignette_futuremice.html) Software vignette.

## Invited workshops, talks and conference presentations

---

- Synthetic data: How fake data can be useful. Guest lecture for elective course *Mimicking & Masking*, Utrecht University, 8 November 2026. [Slides](#).
- **densityratio**: An R-package for distribution comparison. Talk for ETH Zürich, Switzerland, 30 April 2026, . [Slides](#).
- Synthetic data for Enhancing open and reproducible social science. Invited talk for UTH Zürich, Switzerland, 23 April 2026. [Slides](#).
- A practical introduction to multiple imputation of missing data with the R-package **mice**. Invited workshop for Workshops for Ukraine, 22 January 2026. [Slides](#).
- Statistical disclosure control for medical output. Invited talk for VAC4EU Consortium, 11 November 2025. [Slides](#).
- Missing data in medical research: From practical problems to flexible solutions [with Hanne Oberman], Invited half day course at St. Antonius Hospital, 9 November 2025. [Course materials](#).
- Prediction intervals with missing data [with Florian van Leeuwen and Stef van Buuren], Conference presentation at ODISSEI, 4 November 2025. [Slides](#).
- Synthetic data: A solution for data access problems [with Peter Gerbrands, in Dutch], Invited talk at the Dutch Chamber of Commerce, 20 November 2025. [Slides](#).
- How networks facilitate trust: A synthesis on control effects [with Werner Raub and Vincent Buskens]. Conference presentation at Trust and Social Networks conference, 23 July 2025. [Slides](#).
- Measuring utility of synthetic data [in Dutch]. Invited talk for Dutch Knowledge Network Synthetic Data, 19 December 2023. [Slides](#).
- Assessing the utility of synthetic data: A density ratio perspective [with Peter-Paul de Wolf & Erik-Jan van Kesteren], UNECE Expert Meeting on Statistical Data Confidentiality, 26-28 September 2023, Wiesbaden, Germany. [Slides](#).
- Fake it 'till you make it: Generating synthetic data with high utility in R [with Erik-Jan van Kesteren], Tübingen Open Science Winter School, 13 February 2023, Tübingen, Germany. [Course materials](#).
- Creating synthetic data with **mice** in R [with Gerko Vink], UMCU, 4 November 2022, Utrecht, the Netherlands. [Slides](#).