

Troubled Families B Synthetic dataset

January 2026

Introduction

This file describes the synthetic data created using the output from the Gera (2020) thesis.

The experiment and creation of the synthetic data is described [here](#) in Little et al. (2025). There are 3797 records.

The synthetic data is designed so that it would solely reproduce the information found in the covariance matrix in Gera (2020). It is unlikely to be useful for any other analysis.

Data Dictionary

To indicate that the data is synthetic, all of the variable names are prefixed with “synthetic_”

Variable name	Description	Label	Meaning

References

Little, C., Elliot, M., Allmendinger, R. (2025). Producing synthetic teaching datasets using evolutionary algorithms. *UNECE Expert Meeting on Statistical Data Confidentiality*.

https://unece.org/sites/default/files/2025-12/SDC2025_Sd_UnivManchester_SyntheticTeachingDatasets_Elliot_D.pdf

Gera, K. (2020). Assessing the efficacy of time-to-event analysis as a predictive risk model for child protection. Master’s thesis, University of Manchester.