

Troubled Families B Synthetic dataset

January 2026

Introduction

This file describes the synthetic data created using the output from the Gera (2020) thesis. The thesis centred on the Troubled Families Programme (TFP) and used Time-to-event analysis to assess the probability of children receiving child protection interventions. Note that there were two datasets described in the thesis, we used the YOB (year of birth) dataset which captured 3797 children belonging to 3060 families. This contains only children who were born between 2014 and 2018.

The experiment and creation of the synthetic data is described [here](#) in Little et al. (2025), and further details about the original data are contained in Gera (2020). There are 3797 records in the synthetic dataset, and each represents a child born between 2014 and 2018.

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
synthetic_yob	Year of birth	1	2014
		2	2015
		3	2016
		4	2017
		5	2018
synthetic_num_adults	Number of adults in child's family	Range: 1-8	
synthetic_single_par	Whether child's family is headed by a single parent	0	No
		1	Yes
synthetic_deprivation	Aggregate deprivation index for family	Range 1-8	
synthetic_num_criteria	Number of TFP criteria that child's family met	Range 2-6	
synthetic_outcome	Whether a child received a Child Protection Intervention	0	No
		1	Yes
synthetic_duration_to_outcome	The duration in years from a child's birth / start of a TFP intervention to the point at which they were either censored or received a child protection intervention	Range 1-5	

synthetic_fam_cp	Whether a family member had received a child protection intervention prior to the birth event	0	No
		1	Yes
synthetic_fam_ftex	Whether a family member had received a fixed term exclusion from school prior to the birth event	0	No
		1	Yes
synthetic_fam_mhda	Whether a family member had been recorded as having mental health, drug or alcohol issues	0	No
		1	Yes
synthetic_num_siblings	The number of siblings that a child had at the time of birth	Range 0-9	
synthetic_age_tfp	Child's age at time their family began receiving TFP support	Range 0-9	
synthetic_adults_x_siblings	Interaction term of number of adults and number of siblings	Range 0-32	

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.