CPRD Synthetic dataset codebook

January 2026

Introduction

This file describes the synthetic dataset created as part of the EDASIDA project. This used the analytical output from the paper by Trajano et al. (2025) which looked at the use of the drug Metformin to mitigate weight gain in patients with severe mental illness (SMI) who take second generation antipsychotic (SGA) medication. This study used data from the Clinical Practice Research Datalink (CPRD).

The paper used the CPRD which includes two anonymised databases: GOLD and AURUM. The databases include historical and current primary care records for more than 62 million patients. Information about the data is contained within the text, and in tables and figures. Specifically, this synthetic data creation experiment used the information in Table 1, Figure 2 and Supplemental Table 2.

The total number of records in the synthetic dataset is 22581, with each row representing an individual patient. The variables are categorical or continuous.

The synthetic data is designed so that it would solely reproduce the information found in Table1, Figure 2 and Supplemental Table 2 of the Trajano et al. (2025) paper - 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_gender	Gender	0	Male
		1	Female
synthetic_age_index	Age at index date	Continuous	
synthetic_age_pres	Age at first SGA prescription		
synthetic_age_diag	Age at first SMI diagnosis		
synthetic_ethnicity	Ethnicity	0	Asian
		1	Black
		2	Mixed/Other
		3	White
		4	Unknown
synthetic_sga_type	SGA initiated at index date	0	Aripiprazole
		1	Olanzapine
		2	Risperidone
		3	Quetiapine
synthetic_smi_diag	SMI diagnosis	0	Bipolar disorder

		1	Other non-organic psychoses
		2	Schizophrenia
synthetic_region	Region	0	East Midlands
		1	East of England
		2	London
		3	North East
		4	North West
		5	Northern Ireland
		6	Scotland
		7	South East
		8	South West
		9	Wales
		10	West Midlands
		11	Yorkshire & the Humber
synthetic_imd	2019 English IMD quintile	0	Unknown
		1	1 (Least deprived)
		2	2
		3	3
		4	4
		5	5 (Most deprived)
synthetic_weight_base	Baseline weight (kg)	Continuous	
synthetic_weight_6m	Weight at 6 months	Continuous	
synthetic_weight_1y	Weight at 1 year	Continuous	
synthetic_weight_2y	Weight at 2 years	Continuous	
synthetic_bmi_base	Baseline BMI	0	Underweight
		1	Healthy
		2	Overweight
		3	Obese
		4	Unknown
Comorbidities:			
synthetic_cerebrovascular	Cerebrovascular disease	0	No
		1	Yes
synthetic_myocardial	Myocardial infarction	0	No
		1	Yes
synthetic_liver	Liver disease	0	No
		1	Yes
synthetic_renal	Renal disease	0	No
		1	Yes
synthetic_hypertension	Hypertension	0	No
		1	Yes
synthetic_dyslipidaemia	Dyslipidaemia	0	No
		1	Yes
synthetic_diabetes	Diabetes	0	No
		1	Yes
synthetic_polycistic	Polycystic ovarian syndrome	0	No

		1	Yes
Other medications prescribed:			
synthetic_pres_antid	Antidepressants	0	No
		1	Yes
synthetic_pres_lipid	Lipid-regulating medications	0	No
		1	Yes
synthetic_pres_insulin	Insulin	0	No
		1	Yes
Prior exposure to substances:			
synthetic_alcohol	Alcohol misuse	0	No
		1	Yes
synthetic_substance	Substance misuse	0	No
		1	Yes
synthetic_ex_smoker	Ex-smoker	0	No
		1	Yes
synthetic_smoker	Current smoker	0	No
		1	Yes
Biochemical parameters:			
synthetic_hba1c	HbA1c	Continuous	
synthetic_glucose	Glucose	Continuous	
synthetic_diab_pcos	Diagnosed with diabetes and/or polycystic ovarian syndrome	0	No
		1	Yes
synthetic_group	Belonging to which group	0	SGA only
		1	SGA+Metformin

References

Farache Trajano L, Hayes JF, Launders N, Davies NM, Osborn DPJ, Richards-Belle A. Co-prescription of metformin and antipsychotics in severe mental illness: a UK primary care cohort study. *BMJ Mental Health*. 2025;28:e301505. https://doi.org/10.1136/bmjment-2024-301505