

# Analyzing Intergenerational Cardiometabolic Phenotypes in OMOP CDM: A GUSTO Use-Case Study

Erwin Tantoso<sup>1</sup>, Cindy Ho<sup>1</sup>, Jing Yang Tan<sup>1</sup>, Karthik Seetharaman<sup>2</sup>, Peter Hoffman<sup>2</sup>, Mengling Feng<sup>3,4</sup>, Kee Yuan Ngiam<sup>5,6</sup>, Sebastian Maurer-Stroh<sup>1,7,8</sup>, Johan G Eriksson<sup>9,10,11,12</sup>, Mukkesh Kumar<sup>1,9</sup>

Institution(s) of origin:

<sup>1</sup> Bioinformatics Institute, Agency for Science Technology and Research, Singapore

<sup>2</sup> D4L data4life Asia Limited

<sup>3</sup> Saw Swee Hock School of Public Health, National University of Singapore, National University Health System, Singapore

<sup>4</sup> Institute of Data Science, National University of Singapore, Singapore

<sup>5</sup> Division of General Surgery (Thyroid & Endocrine Surgery), National University Hospital Singapore, Singapore

<sup>6</sup> Department of Biomedical Informatics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

<sup>7</sup> Yong Loo Lin School of Medicine, National University of Singapore, Singapore

<sup>8</sup> Department of Biological Sciences, National University of Singapore, Singapore

<sup>9</sup> Institute for Human Development and Potential, Agency for Science Technology and Research, Singapore

<sup>10</sup> Department of Obstetrics and Gynaecology and Human Potential Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

<sup>11</sup> Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland

<sup>12</sup> Folkhälsan Research Center, Helsinki, Finland

## Background

Intergenerational cardiometabolic risk is influenced by maternal metabolic health during pregnancy and subsequent child growth and metabolic development. The Developmental Origins of Health and Disease framework suggests that prenatal and early-life exposures may contribute to later obesity, glucose dysregulation, hypertension, and cardiovascular disease risk (1). Evidence from the Hyperglycemia and Adverse Pregnancy Outcome Follow-up Study showed that maternal glucose levels during pregnancy were associated with childhood adiposity and glucose metabolism at 10 - 14 years of age (2,3). GUSTO-based work has also linked maternal glycaemia with offspring abdominal adiposity in the neonatal and preschool periods (4).

GUSTO provides a multi-ethnic Singapore birth cohort for studying pregnancy and early-life determinants of growth and metabolic health (5). Representing GUSTO in OMOP CDM makes a novel contribution to maternal-child intergenerational family studies by enabling standardized mother-child linkage, reusable phenotype definitions, and scalable evidence generation using multi-generational data. The OMOP CDM standardizes measurements, observations, vocabularies, and person-level relationships, including linkage through the fact\_relationship table. The GUSTO OMOP Data Catalogue (6,7) showcases GUSTO data assets converted into OMOP CDM and has been positioned for contribution towards the OHDSI Evidence Network (8). This study links maternal Pregnancy Week 26 (PW26) cardiometabolic profiles with child trajectories from age 6 to 14 years. Eventually, these

profiles can be used for Data4Life Data2Evidence (D2E) platform cohort creation and National Electronic Health Records (NEHR)-linked cardiovascular outcome analysis.

## **Methods**

We conducted a proof-of-concept intergenerational cardiometabolic phenotyping study using GUSTO data represented in OMOP CDM. The analysis used OMOP person, measurement, observation, and fact\_relationship tables, together with curated concepts defining maternal and child variables. Mother-child dyads were constructed using the fact\_relationship table, with reciprocal relationship concept identifiers for “Natural mother” (4277283) and “Natural child” (4326600), enabling linkage of maternal pregnancy records to child follow-up records.

Eligible dyads required valid linkage, maternal PW26 cardiometabolic measurements, gestational age information, and child longitudinal growth or cardiometabolic measurements from approximately age 6 to 14 years. Gestational age from observation was used to align maternal measurements to PW26. Maternal profiles were derived from standardized features representing adiposity, fasting glycemia, post-load glycemia, blood pressure, and lipid-related metabolic risk. Feature-based clustering identified maternal PW26 profiles.

Child longitudinal profiles were modelled using multivariate latent class mixed modelling across repeated growth and cardiometabolic outcomes (9). Maternal and child profile memberships were linked through OMOP-derived dyads, and profile pairing was summarized using cross-tabulations and visualization as preliminary, hypothesis-generating analysis.

## **Results**

### Data Completeness

From the GUSTO OMOP dataset, 1,217 valid mother-child dyads were identified through the OMOP fact\_relationship table. Completeness was assessed before clustering to define analytic samples for maternal PW26 profiling and child trajectory modelling. Maternal PW26 variables showed varying completeness across adiposity, fasting glycemia, post-load glycemia, blood pressure, and lipid-related markers; 1,116 maternal subjects had more than 50% completeness. Child longitudinal availability varied across individuals and age windows. Trajectory analysis focused on repeated growth and cardiometabolic measurements from approximately age 6 to 14 years, with 947 children having valid measurements and more than 50% completeness.

### Maternal PW26 Cardiometabolic Clusters

Maternal PW26 cardiometabolic clustering identified a three-class solution (Figure 1) using standardized features for adiposity, fasting glycemia, post-load glycemia, blood pressure, and lipid-related metabolic risk. The classes were interpreted as adiposity-mild glycemic (cluster 1: 372 subjects), glycemic-lipid risk (cluster 2: 93 subjects), and favourable (cluster 3: 651 subjects) profiles. The favourable profile showed the lowest overall cardiometabolic risk, while the adiposity-mild glycemic profile reflected higher adiposity with modest glycemic variation. The glycemic-lipid risk profile captured a more adverse metabolic pattern involving glycemic and lipid-related risk markers.

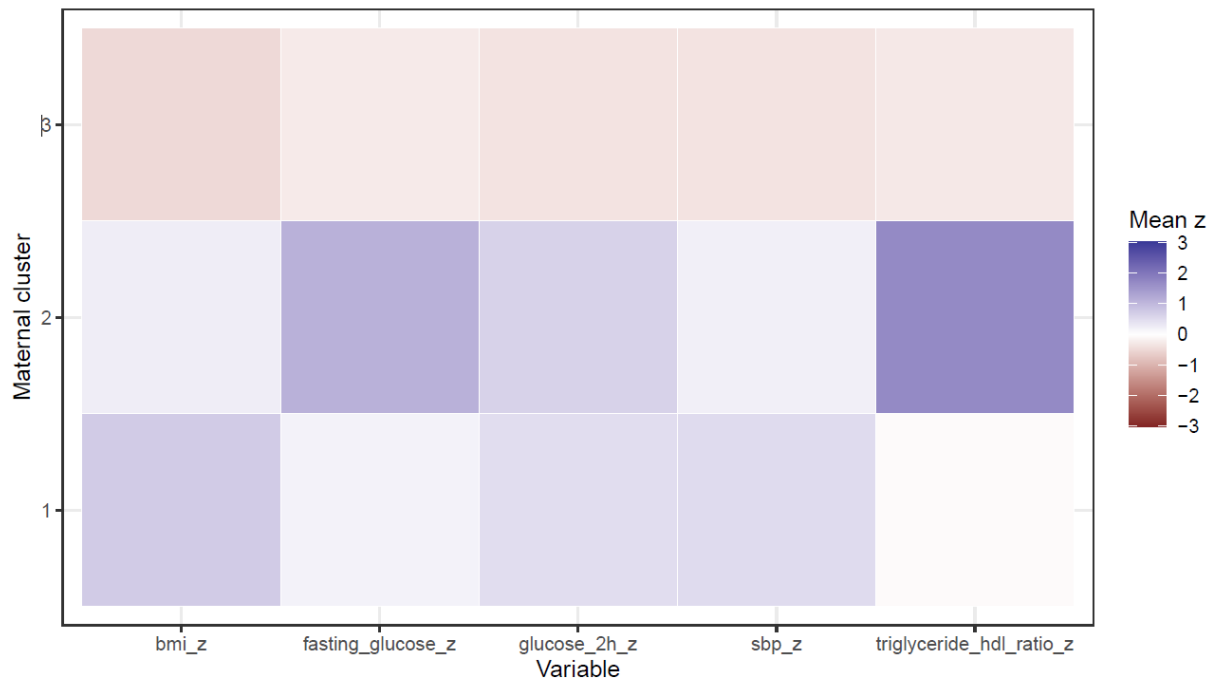


Figure 1: Maternal pregnancy week 26 (PW26) adiposity-glycemic heatmap. The measurements are standardized across all maternal subjects. Three distinct maternal subjects are shown with cluster 1 (adiposity – mild glycemic), cluster 2 (glycemic – lipid risk) and cluster 3 (favourable).

### Child Longitudinal Growth and Cardiometabolic Clusters

Child age-6+ longitudinal clustering was performed using commonly available cardiometabolic variables: child BMI, fasting glucose, systolic blood pressure, and triglyceride/HDL ratio. The three-class solution showed modest but interpretable separation across repeated measurements (Figure 2).

Feature-based child clustering using standardized mean values across the selected variables identified three reproducible child adiposity-cardiometabolic profiles (Figure 3): a lower adiposity-metabolic profile (cluster 1: 296 subjects), an average adiposity-metabolic profile (cluster 2: 574 subjects), and a higher adiposity-lipid profile (cluster 3: 77 subjects).

From a prevention perspective, the child profiles may provide a framework for risk-stratified health promotion. The lower adiposity-metabolic profile may be managed through universal maintenance strategies, while the average profile may benefit from early preventive management. The higher adiposity-lipid profile may warrant closer monitoring and structured family-based lifestyle intervention, particularly around physical activity, diet quality, adiposity, blood pressure, and lipid-related risk. These implications are exploratory and should be validated against future clinical outcomes, but they illustrate how OMOP-derived mother-child profiles could support future NEHR/D2E-enabled preventive cohort definitions.

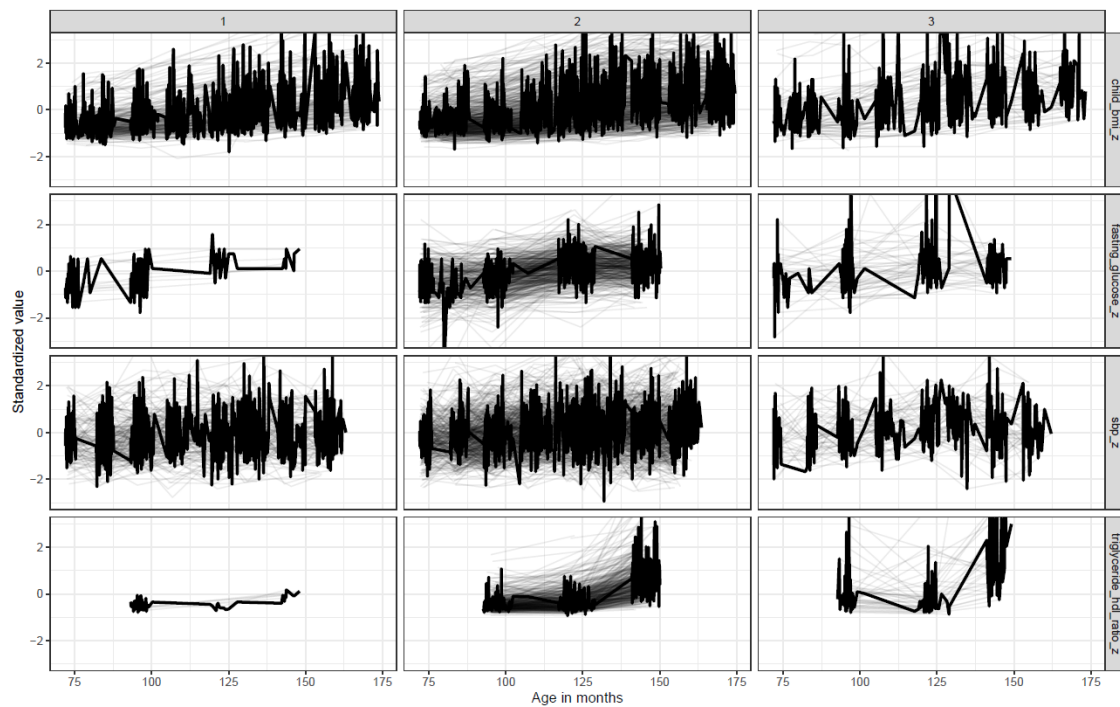


Figure 2: Child age-6+ standardized adiposity and cardiometabolic trajectories by child cluster. Individual grey lines represent child-level longitudinal measurements, and black lines represent the cluster-level average trajectory for each standardized feature. Profiles were derived using age-6+ measurements of child BMI, fasting glucose, systolic blood pressure, and triglyceride/HDL ratio. Values are shown as standardized z-scores, with the y-axis fixed from -3 to 3 for visualization. The three child clusters broadly represent a lower adiposity-metabolic profile (cluster 1: 296 subjects), an average adiposity-metabolic profile (cluster 2: 574 subjects), and a higher adiposity-lipid profile (cluster 3: 77 subjects).

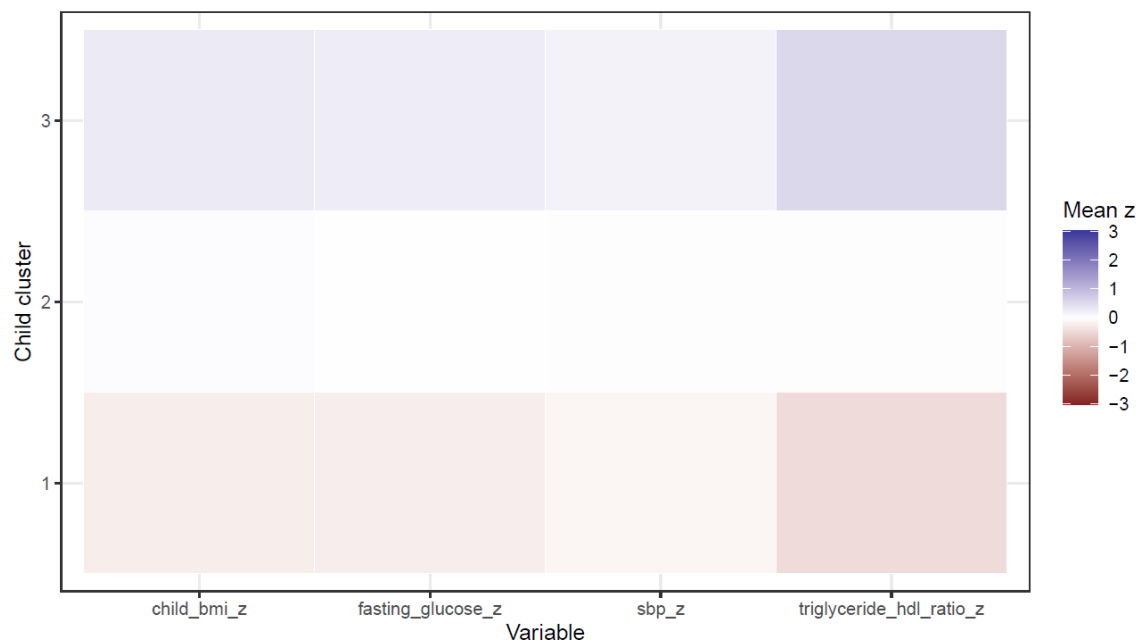


Figure 3: Child growth/adiposity heatmap. The standardized mean value is shown for each variable for each cluster. There are three distinct clusters, i.e. cluster 1 (lower adiposity – metabolic profile), cluster 2 (average adiposity – metabolic profile) and cluster 3 (higher adiposity – lipid profile)

### Cross-tabulation of Maternal-Child Clusters

Maternal PW26 clusters and child trajectory classes were linked through OMOP-derived mother-child dyads, yielding 868 linked pairs. Cross-tabulation (Table 1) showed that mothers in the glycemic-lipid risk profile had a higher proportion of children in the higher adiposity-lipid profile. To further examine this pattern, multinomial logistic regression was performed using child cluster 1 as the reference outcome and maternal cluster 3 as the reference maternal group, with adjustment for child sex, race, and gestational age. Compared with maternal cluster 3, maternal cluster 2 was associated with higher odds of child cluster 3 membership relative to child cluster 1 membership (OR 2.29, 95% CI 1.00–5.23,  $p = 0.049$ ), although this finding should be interpreted cautiously given the borderline statistical significance and wide confidence interval. Maternal cluster 1 was not significantly associated with child cluster membership. These findings are exploratory and do not establish causality but demonstrate that OMOP CDM can support linked intergenerational phenotype discovery. The generated profiles will guide future D2E cohort definition and NEHR-linked cardiovascular outcome analysis.

Table 1: Cross-tabulation of maternal PW26 cardiometabolic clusters and child cardiometabolic clusters. Values represent number and percentage of children within each maternal cluster. **Child cluster 1** was used as the reference outcome and **maternal cluster 3** as the reference exposure group in adjusted multinomial logistic regression

| Mother                                | Child Cluster 1 (lower adiposity) | Child Cluster 2 (average adiposity) | Child Cluster 3 (higher adiposity) |
|---------------------------------------|-----------------------------------|-------------------------------------|------------------------------------|
| Cluster 1 (adiposity – mild glycemic) | 80 (28.27%)                       | 177 (62.54%)                        | 26 (9.19%)                         |
| Cluster 2 (glycemic – lipid risk)     | 21 (30.43%)                       | 37 (53.62%)                         | 11 (15.94%) <sup>a</sup>           |
| Cluster 3 (favourable)                | 170 (32.95%)                      | 310 (60.08%)                        | 36 (6.98%)                         |

Note: <sup>a</sup> In adjusted multinomial logistic regression, maternal cluster 2 was associated with higher odds of child cluster 3 versus child cluster 1 compared with maternal cluster 3 (OR 2.29, 95% CI 1.00–5.23,  $p = 0.049$ ). Model adjusted for child sex, race, and gestational age.

### **Conclusion**

This use-case demonstrates the feasibility of using OMOP CDM for intergenerational cardiometabolic profile discovery in a birth cohort setting. The fact\_relationship table enabled reusable mother-child linkage, while OMOP measurement and observation tables supported standardized extraction of pregnancy, birth, and child follow-up variables. Maternal PW26 clustering identified three interpretable profiles: favourable, adiposity-mild glycemic, and glycemic-lipid risk. Child longitudinal profiling from age 6 to 14 years provided complementary growth and cardiometabolic trajectories, enabling exploratory mother-child phenotype pairing.

These preliminary profiles can guide future D2E-based cohort definition. Future work will formalize the profiles into computable cohorts, link GUSTO-derived profiles to NEHR-derived OMOP data and evaluate associations with later cardiovascular outcomes and potential intervention windows.

## References

1. Yeung EH, Robledo C, Boghossian N, Zhang C, Mendola P. Developmental Origins of Cardiovascular Disease. *Curr Epidemiol Rep*. 2014 Mar 1;1(1):9–16. doi:10.1007/s40471-014-0006-4 PubMed PMID: 25364653; PubMed Central PMCID: PMC4214855.
2. Lowe WL, Lowe LP, Kuang A, Catalano PM, Nodzenski M, Talbot O, et al. Maternal glucose levels during pregnancy and childhood adiposity in the Hyperglycemia and Adverse Pregnancy Outcome Follow-up Study. *Diabetologia*. 2019 Apr;62(4):598–610. doi:10.1007/s00125-018-4809-6 PubMed PMID: 30648193; PubMed Central PMCID: PMC6421132.
3. Scholtens DM, Kuang A, Lowe LP, Hamilton J, Lawrence JM, Lebenthal Y, et al. Hyperglycemia and Adverse Pregnancy Outcome Follow-up Study (HAPO FUS): Maternal Glycemia and Childhood Glucose Metabolism. *Diabetes Care*. 2019 Mar;42(3):381–92. doi:10.2337/dc18-2021 PubMed PMID: 30617141; PubMed Central PMCID: PMC6385697.
4. Tint MT, Sadananthan SA, Soh SE, Aris IM, Michael N, Tan KH, et al. Maternal glycemia during pregnancy and offspring abdominal adiposity measured by MRI in the neonatal period and preschool years: The Growing Up in Singapore Towards healthy Outcomes (GUSTO) prospective mother-offspring birth cohort study. *Am J Clin Nutr*. 2020 Jul 1;112(1):39–47. doi:10.1093/ajcn/nqaa055 PubMed PMID: 32219421; PubMed Central PMCID: PMC7351532.
5. Soh SE, Tint MT, Gluckman PD, Godfrey KM, Rifkin-Graboi A, Chan YH, et al. Cohort profile: Growing Up in Singapore Towards healthy Outcomes (GUSTO) birth cohort study. *Int J Epidemiol*. 2014 Oct;43(5):1401–9. doi:10.1093/ije/dyt125 PubMed PMID: 23912809.
6. GUSTO Data Vault: Laying the foundations for an open science system with OMOP Data Catalogue – OHDSI [Internet]. [cited 2026 Jun 4]. Available from: <https://www.ohdsi.org/2023showcase-407/>
7. GUSTO Data Vault [Internet]. [cited 2026 Jun 4]. Available from: <https://gustodatavault.sg/>
8. Data Partner List – OHDSI Evidence Network [Internet]. 2025 [cited 2026 Jan 15]. Available from: <https://ohdsi.github.io/EvidenceNetwork/dataPartners/dataPartnerList.html>
9. Proust-Lima C, Philipps V, Lique B. Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package lcmm. *J Stat Softw*. 2017;78(2). doi:10.18637/jss.v078.i02