

Supplementary Table - Proposed/projected variables, time points and data source.

Dataset (Approx. number of variables)	Data format and source	Collection period	Examples of types of exposures ascertained*	Examples of types of outcomes ascertained*
Currently already in progress				
Short questionnaire (200 variables)	Self-administered via online portal using a tablet. Source: Mother	Pregnancy (ideally first trimester before 12 weeks, leeway: before delivery) Between 2017-2021	<ul style="list-style-type: none"> • Parity • Gravidity • Psychosocial details (social support, number of household members) • Age of menarche • Contraceptives • Education • Employment 	<ul style="list-style-type: none"> • Previous pregnancy complications such as miscarriages and stillbirth • Congenital malformations • Current pregnancy issues including anxiety • Happiness levels
Long questionnaire (330 variables)	Self-administered via online portal using a tablet. Source: Mother	Pregnancy (ideally third trimester. However, any time before delivery) Between 2018-2022	<ul style="list-style-type: none"> • Ethnicity of paternal and maternal family • Infertility • Oral health • Menstrual cycle • Pregnancy planning • Previous Caesarean sections • Smoking • Alcohol consumption 	<ul style="list-style-type: none"> • Mental health including depression (EPDS), anxiety (STAI) and social support (MOS-SS) • Health Literacy (BRIEF) • Work absenteeism • Husband's health
Medical record extraction exercise 1 (600 variables)	Extracted from multiple portals according to hospital location by trained staff. Extracted onto offline app. Source: Hospital	Pregnancy and delivery collection period: Between 2018 – 2023	<ul style="list-style-type: none"> • Length of gestation • Consanguinity • Complete blood count • Gestational age at presentation at hospital 	<ul style="list-style-type: none"> • Length of labor • Complications during pregnancy • Down's syndrome • Pregnancy outcome (live birth, still birth) • Gestational diabetes • Gestational hypertension • Anemia

Currently being processed or will be processed in the future; All details resemble data collected by predecessors and are prone to changes.				
Food frequency questionnaire for mother (approx. 300 variables)	Self-administered via online portal using a tablet. Source: Mother	At any time during pregnancy Collection period: 2019 – 2021	Dietary habits	Food related illness, allergies and consumption habits
Infant Questionnaire 1 and 2 (in progress) (approx. 1000 variables)	Self-administered via online portal using a tablet. Source: Mother	At 6 months and 12 months of child's life Collection period: 2020 – 2022	<ul style="list-style-type: none"> • Baby anthropometrics • Baby feeding practices • Sleeping patterns of the baby • Pregnancy intervals of future babies 	<ul style="list-style-type: none"> • Onset of illness, if any • Milestones • Post-partum issues with respect to the mother
Medical record extraction exercises 2 and 3 (approx. 600 variables)	Extracted from multiple portals according to hospital location by trained staff. Extracted into offline app. Source: Hospital	Birth till 12 months of life Collection period: 2020-2022	<ul style="list-style-type: none"> • Baby birth details including gender, head and abdominal circumference • Other baby anthropometrics • Laboratory and imaging results • Discharge details 	<ul style="list-style-type: none"> • Birth complications • Admission to NICU • APGAR score • Hospital admissions
Child Questionnaires (in progress) (approx. 1500 variables)	Self-administered via online portal using a tablet. Source: Mother and child	At 18, 36 months and 5 and 7 years of child's life Collection period: 2021-2028	<ul style="list-style-type: none"> • Child anthropometrics • Diet and lifestyle • Immunization • Sleeping patterns 	<ul style="list-style-type: none"> • Onset of illness, if any • Milestones • Motor and behavioral developments • Relationships with family • Emotions and behavior (ASEBA; BASC-2; CRS-R; CSI-4; DMSD; PSC; RBPC; SDQ)
Medical record extraction exercise 4, 5, 6 and 7 (approx. 1500 variables)	Extracted from multiple portals according to hospital location by trained staff. Extracted into offline app. Source: Hospital	From 18 months till 7 years (at least 4) Collection period: 2021-2028	<ul style="list-style-type: none"> • Immunizations • Exposure to allergens • Genetic testing • Lab and imaging results (if any) 	<ul style="list-style-type: none"> • Onset of illness, if any (if missed out by mother) • Hospital admissions • Anthropometry including BMI • Allergies

Adolescent Questionnaires (approx. 1000 variables)	Self-administered via online portal using a tablet. Source: Mother and child	At 11, 14 and 16 years Collection period: 2027-2038	<ul style="list-style-type: none"> • Diet • Physical activity • Education • Lifestyle 	<ul style="list-style-type: none"> • Health issues • Maltreatment • Psychological behavior • Emotions and behavior (ASEBA; BASC-2; CRS-R; CSI-4; DMSD; PSC; RBPC; SDQ)
Medical record extraction exercise 8, 9, 10 (approx. 1000 variables)	Extracted from multiple portals according to hospital location by trained staff. Extracted into offline app. Source: Hospital	From 11 years to 16 years Collection period: 2027-2038	<ul style="list-style-type: none"> • Allergies • Psychological behavior • Personality traits • Anthropometry 	<ul style="list-style-type: none"> • Onset of illness, if any (if missed out in questionnaires) • Hospital admissions • Obesity • Reproductive health (puberty)

* Please note that due to the longitudinal design of the Mutaba'ah study many outcomes during pregnancy will become exposures for outcomes during infancy, childhood, and adolescence. For example, we are interested in the lifestyle behaviours (exposures) associated with gestational diabetes mellitus (GDM; outcome during pregnancy) and then GDM becomes an exposure for specific outcomes late in the pregnancy (e.g. preeclampsia, delivery complications), birth (e.g. macrosomia), infancy (e.g. growth and development), and childhood (e.g. obesity), and adolescence (e.g. non-alcoholic fatty liver disease). APGAR denotes Appearance, Pulse, Grimace, Activity, and Respiration Score; ASEBA denotes Achenbach System of Empirically Based Assessment; BASC-2 denotes Behavior Assessment System for Children, Second Edition; BMI denotes Body Mass Index; BRIEF denotes the Brief Health Literacy Screening Tool; CSI-4 denotes the Child Symptom Inventory-4; CRS-R denotes Conners' Rating Scale – Revised; DMSD denotes the Devereux Scales of Mental Disorders; EPDS denotes Edinburgh Peri/Post-Natal Depression Scale; MOS-SS denotes Medical Outcomes Study Social Support Scale; NICU denotes Neonatal Intensive Care Unit; PSC denotes Pediatric Symptom Checklist; RBPC denotes Revised Behavior Problem Checklist; SDQ denotes Strengths and Difficulties Questionnaire; STAI denotes State-Trait Anxiety Inventory.