Elucidating the Dynamics and Impact of the Gut Microbiome on Maternal Nutritional Status During Pregnancy in Rural Pakistan. Study Protocol for a Prospective, Longitudinal Observational Study.

Authors:
Yaqub Wasan[1], Jo-Anna B. Baxter[2,3], Carolyn Spiegel-Feld[4], Kehkashan Begum[1], Arjumand Rizvi[1], Junaid Iqbal[1], Jessie M. Hulst[3,5,6], Robert Bandsma[3,5,6], Shazeen Suleman[7], Sajid B. Soofi[1], John Parkinson[6,8,9]* Zulfiqar A. Bhutta[1-3,10]*

Author affiliations:
[1]Centre of Excellence in Women and Child Health, Aga Khan University, Karachi, Pakistan
[2]Centre for Global Child Health, Hospital for Sick Children, Toronto, Canada
[3]Department of Nutritional Sciences, University of Toronto, Toronto, Canada
[4]Program in Molecular Medicine, Hospital for Sick Children, Toronto, Canada
[5]Division of Gastroenterology, Hepatology, and Nutrition at the Hospital for Sick Children
[6]Department of Pediatrics, Temerty Faculty of Medicine, University of Toronto
[7]Department of Pediatrics, Stanford University, Stanford, CA; Global Health Faculty Fellow, Center for Innovation in Global Health, Stanford University, Stanford, CA.
[8]Department of Biochemistry, University of Toronto, Toronto, Canada
[9]Department of Molecular Genetics, University of Toronto, Toronto, Canada
[10]Dalla Lana School of Public Health, University of Toronto, Toronto, Canada
*To whom correspondence should be addressed:
zulfiqar.bhutta@aku.edu, *jparkin@sickkids.ca*

Abstract
Background: Nutritional status critically contributes to maternal and child morbidities and mortality. Undernutrition during pregnancy and preconception is linked to adverse pregnancy and birth outcomes and has downstream effects on the growth and development of children. The gut microbiome has a profound influence on the nutritional status of the host. Dysbiosis has emerged as a major factor in the development of malnutrition. This phenomenon is understudied in settings with a high prevalence of undernutrition, and further investigation is warranted to better understand such interactions.
Methods: The study is a prospective, longitudinal observational study to investigate the impact and relationship between prokaryotic and eukaryotic microbes in the gut and their association with maternal health and birth outcomes among young women, 17-24 years of age in Matiari District, Pakistan. We aim to enroll 400 pregnant women with low and normal BMIs at the time of recruitment (<16 weeks of gestation). Data collection includes demographics, morbidities, medication use, empowerment, dietary recall, and anthropometrics. Stool and blood specimens are collected at several time points for different laboratory assays.
Discussion: Limited studies have examined the role of the microbiome in maternal health and birth outcomes where the prevalence of undernutrition is high. Here we aim to investigate the interplay of maternal and child undernutrition with the composition (diversity) and functionality of the gut microbiome and capture how it is influenced by environmental and genetic factors. Findings are expected to provide valuable insights into the understanding of undernutrition.
Study registration: The study is registered with ClinicalTrials.gov Identifier: NCT05108675.

Keywords: Maternal, Infant, BMI, Gut Microbiome, Dysbiosis, and Pregnancy Outcomes.

Background:

Undernutrition during pregnancy is associated with an increased risk of poor birth outcomes and intra-uterine growth restriction of the fetus [1] and globally, is a leading cause of death in under five children [2] as well as maternal morbidity and mortality [3, 4]. Moreover, women with low body mass index (BMI) are at increased risk for adverse pregnancy and birth outcomes [5]. Undernutrition before or during pregnancy may also have a long-term impact on the offspring [6-8]. Among women in low and middle-income countries (LMICs), several micronutrient deficiencies often co-exist due to insufficient dietary intake [1]. Pre-existing micronutrient deficiencies may be exacerbated during pregnancy as a result of the increased metabolic requirements [9]. In particular, adolescent pregnancies are associated with a 50% increased risk of stillbirths and neonatal deaths, and an increased risk of preterm birth, low birth weight (LBW), and asphyxia [10-12]. In Pakistan, The National Nutritional Survey (2018) revealed that more than 42% of women of reproductive age were anemic, and 54% and 27.3% were vitamin D and A deficient. Above 40% of children under 5 were stunted, and 17.7% had severe muscle wasting [13]. The AMANHI, a multicentre study showed stillbirths and neonatal mortality rates in district Matiari [14] higher than the national rates [15].

The intestinal microbiome has emerged as a key factor affecting nutritional status, with impaired maturation contributing to undernutrition [16-22]. During pregnancy, dramatic changes to the gut microbiota occur, with a decrease in individual (alpha) diversity but an increase in population (beta) diversity [23]. Relative to the first trimester, microbiota in the third-trimester exhibit higher abundances of *Proteobacteria*, typically associated with obesity in humans, and when transplanted to mice, result in increased adiposity and insulin insensitivity. Such adaptations may increase energy extraction from the diet to support pregnancy [24], raising interest in dietary supplements to improve pregnancy outcomes [25].

Cumulative pathogen exposures in children confer a high risk for poor growth [26, 27] and environmental enteric dysfunction (EED) [28-30]. EED is thought to be triggered by dysbiosis [31, 32]), initiated by nutrient deficiencies, antibiotic treatment, and/or pathogen exposure. This may further exacerbate pathogen colonization, impair development of the mucosal immune system, and disrupt, by as yet unknown mechanisms, metabolic processes that supply nutrients and energy for normal growth [19]. Maternal EED during pregnancy has also been shown to adversely impact birth outcomes[33].

In an attempt to understand the contribution of the gut microbiome on nutritional status during pregnancy, this study aims to systematically investigate how shifts in the gut microbial community impacts nutritional status during pregnancy in young women. Leveraging whole microbiome RNASeq (metatranscriptomics) and metabolomics, our aim is to further develop mechanistic insights into the relationships between host socio-environmental factors, nutritional status, and microbiome dynamics during pregnancy, and how they contribute to birth outcomes and infant growth during the first year of life.

Objectives:

The primary aim of this study is to assess if alterations of the microbiota in the maternal gut (dysbiosis) are associated with maternal gestational weight gain, and to determine the association between maternal microbiome dysbiosis during pregnancy and birth outcomes (i.e., birth weight, preterm births, small for gestational age, large for gestational age), infant growth, nutritional status, and health status in the first year of life. Several secondary outcomes are integrated to better understand the influence of maternal factors such as dietary intake, maternal BMI, exposure to pathogens, on the gut health and microbiome of infants. Finally, this study will explore how socio-economic factors; including gender, poverty, exclusion,

and empowerment influence the health of a mother's microbiome. Details are provided in supplementary material Table 1.

Study hypothesis:

The study hypothesises that the alterations of the microbiota in the maternal gut (dysbiosis) exacerbated by nutritional status or pathogen exposure during pregnancy, impact weight gain during pregnancy because of impaired nutrient absorption, leading to corresponding negative birth outcomes.

Methods:

Design:

This study employs a prospective, longitudinal observational design to be able to capture maternal and infant outcomes over time. The study aims to collect blood and stool samples from pregnancy to 1-year post-partum, to monitor the dynamic relationships between microbiome community structure and function with gut health and host nutritional status. The study is registered with ClinicalTrials.gov Identifier: NCT05108675 and approved by the national bioethics committee (NBC) in Pakistan (NBC Ref: No.4-87/NBC-700/21/820) and the institutional ethics review committee (ERC) at the Aga Khan University (AKU; ERC No.2021-6085-17561) and the research ethics board (REB) at the Hospital for Sick Children (SickKids; REB number: 1000076773).

Study status:

The first participant was enrolled on November 25, 2021. Data collection for the study is expected to be completed by the first quarter of 2024.

Study setting and population:

The study is being conducted in a rural district, Matiari, within Sindh province in Pakistan. Matiari is 200 kilometers away from Karachi and includes more than 1400 villages and a population of about 800,000. The study setting has a well-established community and health system liaison, basic demographic surveillance, and field centers for research. Furthermore, due to its involvement in several large cluster randomized trials, households are well characterized. This district is representative of typical conditions in Pakistan, and there is a close working relationship with the community, civic society leaders, and public health departments.

At the core of this study are two complementary cohorts of young women, one in Matiari, Pakistan and one in Toronto, Canada. Here we focus on the Pakistani cohort, which is expected to be at increased risk of undernutrition [34], as well as exposed to multiple pathogens and parasites associated with intestinal disease. We will recruit young, married women, including newlyweds, 17-24 years of age, living in Matiari District, Pakistan. We focused on this younger demographic due to our lack of knowledge on the microbiome of young women, and their increased vulnerability to undernutrition; two in every five young women (15-24) living in Matiari are underweight [35], many exhibit suboptimal dietary diversity [36], and more than 90% experience some form of micronutrient deficiency (38% iron deficiency anemia, 81.1% vitamin D deficiency and 31.8% vitamin A deficiency [34]. The cohort based in Toronto is focused on immigrant and refugee populations and will be described elsewhere.

Sample size:

We aim to recruit 400 young women 17-24 years of age into two groups based on their BMI at timing of pregnancy identification: group 1 included those with a 'normal ' BMI (i.e., BMI between 18.5 to 25 kg/m^2) and group 2 included those with an underweight BMI (i.e., <18.5 kg/m^2), as per World Health Organization's

(WHO) guideline definitions [37]). The adequacy of the sample size was verified using the 'pwr' package (version 1.2-2) in R (version 3.6.1). Calculations were based on the correlation between α-diversity (Shannon index) and weight gain during pregnancy. Assuming 400 participants are recruited into the study, and a type I error rate of 0.05, there will be 80% power to detect a correlation coefficient (*r*) >0.14. This is conventionally considered a small effect size [38]. Thus, we expect to be powered to reveal a significant association between weight gain during pregnancy and microbial diversity.

Identification of Pregnancies:

To identify pregnancies within the study area, we leveraged the network that we had established in a previous trial, the Matiari emPowerment and Preconception Supplementation (MaPPS) Trial [39, 40]. Using existing participant lists, women and families are initially approached by phone to get information on early pregnancies. Volunteers from within local villages and primary health care facilities area also encouraged to report new pregnancies. Research field staff also perform random checks of households and villages and meet with lady health workers and volunteers in the field to get information on pregnancies. Verbal consent is taken from the identified pregnant women to share their pregnancy status with the study team for further eligibility screening.

Eligibility Screening:

Pregnancy identification data is assembled at the field office on daily basis. The study team contacts the pregnant women for eligibility screening at their homes. The study staff obtains verbal consent to take the potential participant's height and weight to determine the anthropometric eligibility criteria (BMI <25 kg/m^2) and to conduct a pregnancy test to confirm pregnancy. Women 17-24 years of age who are in good general health, without known chronic diseases, and <16 weeks of gestation are invited to take part in the study. To participate, women will also be asked to provide written consent and agree to comply with the study procedure. Those with a BMI ≥25 kg/m^2 at the time of recruitment and/or, participating in another nutritional trial, and/or report taking antibiotics in the last three months, and/or screen for potential signs of COVID-19 are not eligible for study participation. Only one participant per household is enrolled in the study, given the potential for within household dietary and microbial similarities.

Participant recruitment:

Pregnant women who meet the eligibility criteria are formally invited for study participation following written, informed consent. Informed consent is administrated following AKU's well-established protocols for research ethics compliance. Participants are also explained the option to sign an additional section of the consent form for future genetic testing of biospecimens. Participants are given the option to discuss the consent form with their families, before agreeing to participate in the study.

Data and specimen collection:

All data collection is completed by trained female data collectors at the participant's home. This includes verbal data collection, anthropometric assessments, and blood and stool sample collection. Data collection at birth, which is collected either at a health facility or at home, depends on where the mother is available soon after delivery. Study research staff aim to complete birth anthropometrics within 72 hours post-delivery. Time points of data collection, biospecimen, and anthropometric assessments are provided in Table 1. For all structured interviews with the study participants, study personnel use tablet-based Research Electronic Data Capture (REDCap) applications to guide data collection customized to the visit [41, 42].

Demographics:

Household and personal demographics captured in this study include information regarding the participant's age, gender, sex, occupation, the language spoken in the home, religion, income, number of people this income supports, education, housing, defecation, hand washing, reproductive history and marital status. Data collection scales are adapted from the Pakistan Demographic and Health Survey (PDHS) [15].

Maternal empowerment and household food insecurity:

An empowerment questionnaire is deployed to collect data about self-efficacy using the Generalized Self-Efficacy scale, developed by Schwarzer and Jerusalem[43]. Perceived social support is measured using the Multi-dimensional Scale of Perceived Social Support (MSPSS), developed by Zimet and colleagues[44]. Perceived parental stress is measured using the Perceived Stress Scale (PSS-10)[45]. Lastly, food insecurity is assessed through the Household Food Insecurity Access Scale (HFIAS) [46].

24-hour food recall:

To link the microbiome to nutritional status and nutritional intake, with a focus on calories and macronutrients, an interactive semi-quantitative, 24-hour paper-based dietary recall [47] is administered by the research staff at the participants' homes.

Minimum dietary diversity score for women (MDD-W):

The MDD-W is a population-level indicator for dietary diversity for women aged 15-49, based on 10 food groups. [48] The MDD-W reflects what they have eaten over the previous 24 hours, and participants are asked at the end of the questionnaire whether this reflects their diet over the previous 3 months. The research team will calculate the MDD-W from the dietary recalls completed at baseline, 30-34 weeks post-conception, and at 12 months.

Infant Feeding:

Guidance developed by the WHO in 2010, in the "*Indicators for assessing infant and young child feeding practices (Part 2 Measurement)"* [49] is used to assess infant feeding. At the 12-month visit, research staff administer the NutricheQ questionnaire, a tool designed for toddlers aged 1 to 3 years of age, with a focus on markers for inadequate or excessive intake and dietary imbalances [50]. Two food insecurity questions, that assess maternal and infant's annual food insecurity are also included [46].

Birth history and pregnancy outcomes:

Birth and labor history questions administered at the post-conception visit (24-72 hours) gather additional information on: mode of delivery, gestational age, newborn anthropometrics i.e., weight, length, and head circumference, placental insufficiency, antibiotic use, among other birth characteristics.

Health and medicine use assessment:

Morbidity assessment captures mortality, morbidity, and medication usage of mother and infant at several timepoints (Table 1).

Blood sample collection:

Certified paramedics are trained by AKU faculty and senior management of Nutrition Research Laboratory (NRL) on blood collection using AKU's Standard Operating Procedures (SOPs). Venous blood specimens are collected from participating mothers (5 mL) and infants (3 mL) and distributed into two types of vacutainers. 0.6 mL of blood transferred to an SST tube (Yellow cap BD vacutainer: BD, PL6 7BP, UK) for

ferritin and c-reactive protein (CRP) analysis, while the rest will be transferred to an EDTA tube (Purple cap BD vacutainer: BD, 1 Becton Drive, Franklin Lakes, NJ 07417 USA) for HB and MCV analysis, as well as for future storage. These tubes will be transported to the field-based lab in a Coleman portable freezer maintained at 2-8°C. The extracted serum will then be flash-frozen and stored at -80°C until the point of analysis.

Stool sample collection:

Participating mothers are provided with sterile stool containers a day before collection to provide freshly passed stool samples to field staff, while infants' samples are collected in diapers and transferred to containers by the field team. If the stool is mixed with urine, collections are rescheduled. Collected samples are transported to the field-based lab in a Coleman portable freezer maintained at 2-8°C. At the lab, the sample is aliquoted in 4 cryovials and stored at -80°C until further processing.

Anthropometry:

Maternal height and weight are measured using a digital floor scale (seca 813, Seca, Hamburg, Germany) and stadiometer (Seca 213); mid upper arm circumference (MUAC) is taken through Seca 201 tapes. Triceps skinfold thickness (SFT) is measured using a skinfold caliper (Holtain CRYMYCH, UK). We use Seca scales for infant anthropometry, which includes measurement of weight (seca 354), length (seca 417), MUAC (seca 201), and head circumference (seca 212); these measurements are conducted using standardized procedures. All measurements are collected in duplicate, by two study personnel, using standardized procedures, as specified in the anthropometric data collection SOP adopted from the intergrowth study [51]. The average (mean) of acceptable paired measures is used in subsequent analyses.

Table 1: Overview of study visits, visit activities, data, and specimen collection.

Procedures	Pregnancy Surveillance	MV1: 10-14 Weeks Post-Conception	MV2: 30-34 Weeks Post-Conception	MV3: Maternal 24-72 Hours Post- Partum	IV1: Infant 24-72 Hours Post-delivery (same visit as of mom)	MV4: Maternal 3 months post-partum	IV2: Child 3-Months (same visit as of mom)	MV5: Maternal 6 Months Post-Partum	IV3: Infant 6 Months (same visit)	MV6: Maternal 12-Months Post-Partum	IV4: Child 12-Months (same visit as of mom)
Pregnancy Surveillance	X										
Screening and Confirmation of Pregnancy		X									
Informed Consent		X									
Household and individual demographics		X									
24-hour food recall		X	X							X	
Minimum Dietary Diversity						X					

Procedures	Pregnancy Surveillance	MV1: 10-14 Weeks Post-Conception	MV2: 30-34 Weeks Post-Conception	MV3: Maternal 24-72 Hours Post- Partum	IV1: Infant 24-72 Hours Post-delivery (same visit as of mom)	MV4: Maternal 3 months post-partum	IV2: Child 3-Months (same visit as of mom)	MV5: Maternal 6 Months Post-Partum	IV3: Infant 6 Months (same visit)	MV6: Maternal 12-Months Post-Partum	IV4: Child 12-Months (same visit as of mom)
Score for Women (MDD-W)											
Infant feeding practices							X		X		X
Anthropometric Measurements		X	X	X	X	X	X	X	X	X	X
Empowerment		X				X					
Birth and Labour History				X	X						
Health and Medication Use Assessment			X	X	X	X	X	X	X	X	X
Stool Sample		X	X			X	X			X	X
Blood Sample		X	X							X	X
Study Exit										X	X
MV= Maternal visit; IV= Infant visit											

Data management:

All tablets are synchronized daily to upload data from the REDCap data platform to a secure, web-based server hosted at the AKU campus. The tablets include built-in logic and range checks to ensure data quality. Paper-based study forms for the 24-hour food recall are checked by a study monitor for consistency and completeness. Dual entry of paper-based data is performed to reduce data entry errors. Data entry screens are developed using Visual FoxPro software (Microsoft). De-identified data is stored in a password-protected data base.

Outcome measures:

For the primary outcomes, to determine weight gain during pregnancy, maternal weight is measured at specified timepoints (Table 1). Gut dysbiosis/dynamics will be assessed using 16S and 18S rDNA surveys applied to the maternal stool samples to monitor the bacterial and eukaryotic components of the microbiome. WHO z-scores for weight, length, and head circumference at birth and during the first year of infant will be used to assess the association between maternal gut dysbiosis and birth outcome and infant health. Details are provided in Table 1 of supplementary material.

Stool analysis:

Maternal and infant stool samples from all time points undergoes DNA extraction at the NRL, AKU. Maternal stool RNA extraction is completed for selected participants (100 with the highest BMI and 100 with the

lowest BMI at time of enrollment) from stool collected at the two pregnancy visits. Stool samples of these participants are also analyzed for inflammatory markers (Table 2). The stool samples and extracted DNA and RNA samples are then batch shipped to the Hospital for Sick Children, Toronto, Canada to complete downstream analyses, including sequencing, metabolomics and biobanking.

Table 2: Stool assays, methods, instruments, and processing laboratories

Stool Analyte	Amount of stool (mg)	Assay method	Instrument	Processing Lab
Calprotectin	100	Sandwich Immunoassay	Liaison immune analyzer, Diasorin.	NRL, AKU, Pakistan
Lipocalin-2	100	Sandwich Immunoassay	Epoch 2 microplate reader, Biotek.	NRL, AKU, Pakistan
Claudin-15	300	Competitive Enzyme Immunoassay	Epoch 2 microplate reader, Biotek.	NRL, AKU, Pakistan
RNA Extraction	250	Column Extraction Technology	ZymoBIOMICS™ RNA Miniprep Kit (R2001)	NRL, AKU, Pakistan
DNA Extraction	200	Column Extraction Technology	ZymoBIOMICS™ DNA Miniprep Kit (D4300)	NRL, AKU, Pakistan
Stool Archiving for Shipment	2000	-	ThermoScientific TSX-Series -80°C Ultralow temperature freezer	NRL, AKU, Pakistan
Stool Metabolomics*	100	Microbiome Metabolism Assay	LC-MS	TMIC, Alberta, Canada
*SCFAs, amino acids, intermediates in glycolysis and nucleotide metabolism				

Blood analysis:

Blood samples are analyzed at AKU for hemoglobin, mean cell volume, ferritin, and CRP concentration, with additional aliquots shipped to SickKids for further analysis (Table 3). The Metabolomics Innovation Center (TMIC) will conduct metallomics analysis using the TMIC metallomics platform to investigate micronutrients (e.g., Zn, Mg and Se).

Table 3: Blood assays, methods, and processing laboratories

Blood Analyte	Amount of blood/serum sample (µL)	Assay method	Instrument Used	Processing Lab
Hemoglobin + mean cell volume	500	Photometric assay	Sysmex p100 hematology analyzer	Matiari Research Laboratory, Aga Khan University, Pakistan
Ferritin	200	Immunoturbidimetric assay	CobasC311Analyzer, Roche diagnostics	NRL, AKU, Pakistan
C-reactive protein	200	Immunoturbidimetric assay	CobasC311Analyzer, Roche diagnostics	NRL, AKU, Pakistan

Serum Archiving for Shipment	1000	-	ThermoScientific TSX-Series -80°C Ultralow temperature freezer	NRL, AKU, Pakistan
Serum Metabolomics study	50	Metallomics assay	ICP-MS	TMIC, Alberta, Canada

Profiling microbial community structure:

Microbial communities are analyzed through 16S and 18S rDNA surveys using established methods that target the V4 region of the 16S rRNA gene to capture bacterial taxa [52-54] and the V4V5 region of the 18S rRNA gene to capture eukaryotic taxa. DNA library preps include error-correcting barcodes [55] for multiplexing of samples. Sequencing is performed to generate ~50,000 2x150bp paired end reads per sample. To define taxonomic diversity, species profiles from 16S and 18S rDNA data will be clustered to identify differences in community structure across samples. We will utilize the QIIME2 platform [56], MOTHUR [57], multivariate approaches such as Permutation Multivariate Analysis of Variance (i.e. PERMANOVA-S a method that can associate microbiome changes with outcome measures while accounting for confounders [58]. Differences between groups in microbiome community structure are tested by analysis of similarities (ANOSIM) and co-occurrence analysis [59]. To better define bacterial pathogen burden, we apply TaqMan array card technology for the simultaneous detection of 19 common enteropathogens [60].

Profiling microbial community function:

After total RNA extraction and rRNA depletion (RiboZero Gold Kit, Illumina, San Diego, Ca, or equivalent), libraries will be constructed and Illumina-based sequencing will be performed to generate ~30 million 2x150bp paired-end reads per sample (our rarefaction analyses have previously shown such sequencing depth is sufficient to identify the vast majority of species and enzymes present in the samples [61]). Reads are processed for quality and contaminants using the MetaPro pipeline [62]. Reads are assembled using SPAdes [63] and subsequently annotated with taxonomic and functional assignments. Expression is normalized to Reads per Kilobase of transcript per million mapped reads (RPKM). Annotations are mapped onto biochemical pathways and complexes such as those defined by the Kyoto Encyclopedia of Genes and Genomes [64]. The output of these analyses is readouts of microbial gene expression detailing biochemical activities as well as the taxa responsible.

Statistical Analysis:

Normally distributed continuous data is shown as a mean and standard deviation, and median and Interquartile range (IQR) is calculated for non-normally distributed data. Categorical data is presented using proportions.

For the primary outcomes, maternal gut bacteria and eukaryotic profiles is used to calculate Bray-Curtis dissimilarity metrics between individual samples which is leveraged in principal co-ordinate analyses to determine the extent samples collected at the first or third trimester, exhibiting similar gestational weight gains, co-cluster. Permutational multivariate analysis of variance (PERMANOVA) tests will assess the degree of overlap between samples exhibiting low gestational weight gain versus samples exhibiting high gestational weight gain. Next, we attempt to correlate changes in the alpha diversity (as measured by the Shannon and Simpson indices) of the gut microbiome samples between the first and third trimester, with

gestational weight gain. To examine the influence of individual taxa on gestational weight gain, we perform bivariate analyses (Pearson, Spearman). The Benjamini-Hochberg procedure is applied to correct p-values while controlling for false-discovery rates.

To complement these analyses, we also undertake an integrative modeling strategy based on the Similarity Network Fusion framework [65] to analyze the contribution of each variable (clinical, microbiome, and gender-related, see Table 1 of supplementary material) on gestational weight gain. This allows the integration of all available datasets to uncover their global substructures that can be associated with gestational weight gain. In an alternative approach, we also employ Random Forests to identify combinations of variables that correlate with gestational weight gain. Approaches for secondary and exploratory analysis and details on future genetic testing are provided in the supplementary material.

For the rest of the outcomes, followed by general linear models i.e., PERMANOVA and the Bray-Curtis dissimilarity metric. DESeq2, a method for differential analysis of count data, is applied to investigate both associations of specific taxa with the clinical variables, together with the strength of those associations. These analyses reveal which clinical variables (including exposure to pathogens and parasites) correlate with the maternal microbiome from a taxonomic perspective. Additionally, microbiome structural and functional profiles are generated from the difference in 1) taxonomic abundances; 2) gene expression; and 3) metabolite concentrations, between the first and third trimesters. Profile differences are utilized in the PERMANOVA and DESeq2 approaches as described above to identify associations between clinical variables and changes in microbiome structure and function.

Discussion:

Nutritional status during pregnancy plays an important role in maternal health and birth outcomes [66, 67]. Maternal undernutrition during pregnancy often leads to fetal growth restriction, which increases the risk of neonatal deaths and childhood stunting by 2 years of age [2]. In Pakistan, large scale surveys and cohort studies have suggested multifaceted undernutrition and adverse pregnancy and birth outcomes [13-15]. Data from a cohort of young women (15-24 years) living in rural Pakistan revealed that more than 90 percent lives with minimum one micronutrient deficiency [34]. Dietary intake was limited to fewer types of foods, mainly staples [36]. Near 40 percent women were underweight [35].

The gut microbiome can have a profound influence on host's nutritional status, yet few studies of the dynamics between nutritional status and the gut microbiome during pregnancy have been conducted. Further, a few studies focusing on child undernutrition have revealed a key role for gut microbiota [16-22]. In particular, dysbiosis, or the loss of diversity/beneficial microbes and gain of pathobionts, has emerged as a major factor in the development of undernutrition. To date most studies of the gut microbiome have focused on bacterial components, typically neglecting the contribution of eukaryotic microbiota. Despite the fact that many such eukaryotes include parasites, such as *Giardia*, *Cryptosporidium,* and *Entamoeba*, each representing a significant burden on global healthcare with considerable implications for gut health [68-71]. Interestingly, not all parasitic infections cause disease; instead, many infections remain asymptomatic with disease emerging as a consequence of interactions between the eukaryotic and bacterial microbiome and the host immune system [72, 73]. With the emergence of new marker gene technology, based on the 18S/5S/28S locus, there is now the opportunity to profile eukaryotic communities and examine their impact within the context of the gut microbiome.

Thus, this study will inform the relationships between host nutritional status and microbiome dynamics during pregnancy, and how they contribute to gestational weight gain during pregnancy, in addition to

several other pregnancy and birth-related outcomes. Understanding the role of the microbiome on maternal health and birth outcomes, as well as the influence of enteric eukaryotic microbes, such as parasites, on the bacterial microbiome and host nutrition offers great potential in the identification of modifiable factors to improve health and nutrition outcomes.

To help establish causal relationships between microbiome dynamics, pathogen exposure, and nutritional status during pregnancy and to examine whether manipulation of the microbiome can improve nutritional status, future work is expected to leverage stool samples collected here, in fecal microbiome transplant studies using animal models.

Limitations

The study follows the STROBE guidelines; however, we expect controlling for all confounding variables may not be possible. For example, verbal data is collected through interviews, while some follow ups collect three to six months recall data on morbidity, medication use and care seeking which may impact the reliability of reporting mothers. Of particular note, the study population experienced an unprecedent flooding event in 2022 which disrupted daily life and dietary patterns. Further, the flooding event is also expected to have increased exposure of the study population to additional pathogens, with corresponding impacts on microbiome dynamics. Additionally, due to restricted hours of operation relative to the passing of stool by infants, there may be heterogeneity in the amount of time between passing of stool and collection by the field team. During analysis, we may find that we identify no significant differences between microbial diversity or composition in relation to our primary or secondary outcomes. Such findings would elevate the importance of the metatranscriptomic analyses to deliver more mechanistic investigations. It is possible that RNA quality and yields from stool is poor. In such events we will revert to performing whole microbiome DNA (which is more stable than RNA) sequencing (metagenomics) which also has the capacity to deliver functional insights. Finally, we acknowledge that by focusing on young women, 17-24 years of age, that findings may not be generalizable to younger or older demographics. At the same time, studies of the microbiome within this age group are lacking and hence this study is aimed at directly addressing this knowledge gap to deliver a wealth of information to better serve the health care needs of this important demographic.

Funding:

The study is funded by Canadian Institute of Health Research (CIHR)

Competing interests:

The authors declare that they have no competing interests.

Authors' contributions

Dr. Zulfiqar Ali Bhutta (PI) and JP envisioned the trial and got funding. CS, JBB, JH, RB, SS, SBS and YW drafted the protocol. KB and JI helped finalize laboratory work for Pakistan based cohort. AR assisted with data analysis plan. YW, JBB, CS developed questionnaires and field SOPs. YW produced initial draft of this manuscript with inputs from team mentioned above. All the listed authors reviewed and approved final draft for publication.

Abbreviations:

AMANHI	Alliance for Maternal and Newborn Health Improvement
ANOSIM	Analysis of Similarities
BMI	Body Mass Index
CRP	C-reactive protein
DNA	Deoxyribonucleic acid
EDTA	Ethylenediamine tetraacetic acid
EED	Environmental Enteric Dysfunction
ERC	Ethics Review Committee
HFIAS	Household Food Insecurity Access Scale
IV	Infant Visit
LBW	Low Birth Weight
LMICs	Low and Middle-income Countries
MAL-ED	Malnutrition and Enteric Disease
MaPPS	Matiari Empowerment and Preconception Supplementation
MCV	Mean Corpuscular Volume
MDD-W	Minimum dietary diversity score for women
MSPSS	Multi-Dimensional Scale of Perceived Social Support
MUAC	Mid Upper Arm Circumference
MV	Maternal visit
NBC	National Bioethics Committee
NRL	Nutrition Research Laboratory
PDHS	Pakistan Demographic and Health Survey
PERMANOVA	Permutational Multivariate Analysis of Variance
QIIME	Quantitative Insights into Microbial Ecology
RPKM	Reads per Kilobase of transcript per million mapped reads
REB	Research Ethics Board
REDCap	Redcap Research Electronic Data Capture
RPKM s	Reads Per Kilobase of Transcript Per Million Mapped
RNA	Ribonucleic acid
SCFA	Short-chain fatty acids
SFT	Triceps Skinfold Thickness
SOPs	Standard Operating Procedures
SST	Serum Separator Tube
STROBE	Strengthening the reporting of observational studies in epidemiology
TMIC	The Metabolomics Innovation Center
WHO	World Health Organization

References:

1. Ramakrishnan, U., et al., *Effect of women's nutrition before and during early pregnancy on maternal and infant outcomes: a systematic review.* Paediatr Perinat Epidemiol, 2012. **26 Suppl 1**: p. 285-301.
2. Black, R.E., et al., *Maternal and child undernutrition and overweight in low-income and middle-income countries.* Lancet, 2013. **382**(9890): p. 427-451.
3. Todd, C.S., et al., *Maternal nutrition intervention and maternal complications in 4 districts of Bangladesh: A nested cross-sectional study.* PLoS Medicine, 2019. **16**(10): p. e1002927.
4. Tomkins, A., *Nutrition and maternal morbidity and mortality.* British Journal of Nutrition, 2001. **85**(S2): p. S93-S99.
5. Katz, J., et al., *Mortality risk in preterm and small-for-gestational-age infants in low-income and middle-income countries: a pooled country analysis.* Lancet, 2013. **382**(9890): p. 417-425.
6. Harding, J.E., *The nutritional basis of the fetal origins of adult disease.* Int J Epidemiol, 2001. **30**(1): p. 15-23.
7. Barker, D.J., *Fetal origins of coronary heart disease.* BMJ, 1995. **311**(6998): p. 171-4.
8. Roseboom, T., S. de Rooij, and R. Painter, *The Dutch famine and its long-term consequences for adult health.* Early Hum Dev, 2006. **82**(8): p. 485-91.
9. Berti, C., et al., *Micronutrients in pregnancy: current knowledge and unresolved questions.* Clin Nutr, 2011. **30**(6): p. 689-701.
10. WHO, A.P., *Unmet needs, undone deeds.* A review of the literature and programmes, WHO, Geneva, Switzerland, 2007.
11. Haldre, K., et al., *Is a poor pregnancy outcome related to young maternal age? A study of teenagers in Estonia during the period of major socio-economic changes (from 1992 to 2002).* Eur J Obstet Gynecol Reprod Biol, 2007. **131**(1): p. 45-51.
12. Paranjothy, S., et al., *Teenage pregnancy: who suffers?* Arch Dis Child, 2009. **94**(3): p. 239-45.
13. UNICEF, *National Nutritional Survey 2018, Pakistan.* New York: UNICEF, 2019. **1**.
14. Ahmed, I., et al., *Population-based rates, timing, and causes of maternal deaths, stillbirths, and neonatal deaths in south Asia and sub-Saharan Africa: a multi-country prospective cohort study.* The Lancet Global Health, 2018. **6**(12): p. e1297-e1308.
15. Studies, N.I.o.P., *Pakistan demographic and health survey (PDHS) 2017–18*. 2018, Macro International Inc Maryland.
16. Tidjani Alou, M., et al., *Gut Bacteria Missing in Severe Acute Malnutrition, Can We Identify Potential Probiotics by Culturomics?* Frontiers in Microbiology, 2017. **8**(899).
17. Smith, M.I., et al., *Gut microbiomes of Malawian twin pairs discordant for kwashiorkor.* Science, 2013. **339**(6119): p. 548-54.
18. Subramanian, S., et al., *Persistent gut microbiota immaturity in malnourished Bangladeshi children.* Nature, 2014. **510**(7505): p. 417-421.
19. Blanton, L.V., et al., *Childhood undernutrition, the gut microbiota, and microbiota-directed therapeutics.* Science, 2016. **352**(6293): p. 1533-1533.
20. Velly, H., R.A. Britton, and G.A. Preidis, *Mechanisms of cross-talk between the diet, the intestinal microbiome, and the undernourished host.* Gut Microbes, 2017. **8**(2): p. 98-112.
21. Raoult, D., *Microbiota, obesity and malnutrition.* Microb Pathog, 2017. **106**: p. 1-2.
22. Reyes, A., et al., *Gut DNA viromes of Malawian twins discordant for severe acute malnutrition.* Proc Natl Acad Sci U S A, 2015. **112**(38): p. 11941-6.
23. Koren, O., et al., *Host remodeling of the gut microbiome and metabolic changes during pregnancy.* Cell, 2012. **150**(3): p. 470-80.

24. Astbury, S., et al., *Nutrient availability, the microbiome, and intestinal transport during pregnancy.* Appl Physiol Nutr Metab, 2015. **40**(11): p. 1100-6.
25. Bisanz, J.E., et al., *Microbiota at Multiple Body Sites during Pregnancy in a Rural Tanzanian Population and Effects of Moringa-Supplemented Probiotic Yogurt.* Applied and Environmental Microbiology, 2015. **81**(15): p. 4965-4975.
26. Rogawski, E.T., et al., *Determinants and Impact of Giardia Infection in the First 2 Years of Life in the MAL-ED Birth Cohort.* J Pediatric Infect Dis Soc, 2017. **6**(2): p. 153-160.
27. Platts-Mills, J.A., et al., *Pathogen-specific burdens of community diarrhoea in developing countries: a multisite birth cohort study (MAL-ED).* Lancet Glob Health, 2015. **3**(9): p. e564-75.
28. Semba, R.D., et al., *Environmental Enteric Dysfunction is Associated with Carnitine Deficiency and Altered Fatty Acid Oxidation.* EBioMedicine, 2017. **17**: p. 57-66.
29. Crane, R.J., K.D. Jones, and J.A. Berkley, *Environmental enteric dysfunction: an overview.* Food Nutr Bull, 2015. **36**(1 Suppl): p. S76-87.
30. Jones, K.D., et al., *Childhood malnutrition: toward an understanding of infections, inflammation, and antimicrobials.* Food Nutr Bull, 2014. **35**(2 Suppl): p. S64-70.
31. Levy, M., et al., *Dysbiosis and the immune system.* Nat Rev Immunol, 2017. **17**(4): p. 219-232.
32. Petersen, C. and J.L. Round, *Defining dysbiosis and its influence on host immunity and disease.* Cellular Microbiology, 2014. **16**(7): p. 1024-1033.
33. Lauer, J.M., et al., *Biomarkers of maternal environmental enteric dysfunction are associated with shorter gestation and reduced length in newborn infants in Uganda.* Am J Clin Nutr, 2018. **108**(4): p. 889-896.
34. Baxter, J.-A.B., et al., *Characterizing Micronutrient status and risk factors among late adolescent and young women in rural Pakistan: A cross-sectional assessment of the Mapps trial.* Nutrients, 2021. **13**(4): p. 1237.
35. Baxter, J.-A.B., et al., *Drivers of malnutrition among late adolescent and young women in rural Pakistan: a cross-sectional assessment of the MaPPS trial.* BMJ open, 2023. **13**(5): p. e063734.
36. Baxter, J.A.B., et al., *Dietary diversity and social determinants of nutrition among late adolescent girls in rural Pakistan.* Maternal & child nutrition, 2022. **18**(1): p. e13265.
37. WHO, *Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee.* , in *WHO Technical Report Series 854.* . 1995: Geneva.
38. Cohen, J., *Statistical power analysis for the behavioral sciences*. 1988, Hillsdale, N.J.: L. Erlbaum Associates.
39. Baxter, J.-A.B., et al., *Feasibility and effect of life skills building education and multiple micronutrient supplements versus the standard of care on anemia among non-pregnant adolescent and young Pakistani women (15–24 years): a prospective, population-based cluster-randomized trial.* Reproductive health, 2018. **15**(1): p. 103.
40. Baxter, J.-A.B., et al., *Effect of life skills building education and micronutrient supplements provided from preconception versus the standard of care on low birth weight births among adolescent and young Pakistani women (15–24 years): a prospective, population-based cluster-randomized trial.* Reproductive health, 2018. **15**(1): p. 104.
41. Harris, P.A., et al., *The REDCap consortium: building an international community of software platform partners.* Journal of biomedical informatics, 2019. **95**: p. 103208.
42. Harris, P.A., et al., *Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support.* Journal of biomedical informatics, 2009. **42**(2): p. 377-381.
43. Schwarzer, R. and M. Jerusalem, *The general self-efficacy scale (GSE).* Anxiety, Stress, and Coping, 2010. **12**(1): p. 329-345.

44. Zimet, G.D., et al., *The Multidimensional Scale of Perceived Social Support.* Journal of Personality Assessment, 1988. **52**(1): p. 30-41.
45. Cohen, S., T. Kamarck, and R. Mermelstein, *A Global Measure of Perceived Stress.* Journal of Health and Social Behavior, 1983. **24**(4): p. 385-396.
46. Coates, J., A. Swindale, and P. Bilinsky, *Household Food Insecurity Access Scale (HFIAS) for Measurement of Household Food Access: Indicator Guide (v. 3).* Academy for Educational Development, 2007. **3**.
47. Gibson, R. and E. Ferguson, *An Interactive 24-Hour Recall for Assessing the Adequacy of Iron and Zinc Intakes in Developing Countries.* 2008.
48. Project, I., *Data4Diets: building blocks for diet-related food security analysis.* USAID Advancing Nutrition, 2018.
49. Organization, W.H., *Indicators for assessing infant and young child feeding practices: part 2: measurement.* 2010.
50. Morino, G.S., et al., *NutricheQ Questionnaire assesses the risk of dietary imbalances in toddlers from 1 through 3 years of age.* Food Nutr Res, 2015. **59**: p. 29686.
51. Villar, J., et al., *International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). International standards for newborn weight, length, and head circumference by gestational age and sex: the Newborn Cross-Sectional Study of the INTERGROWTH-21st Project.* Lancet, 2014. **384**(9946): p. 857-68.
52. Coburn, B., et al., *Lung microbiota across age and disease stage in cystic fibrosis.* Sci Rep, 2015. **5**: p. 10241.
53. Copeland, J.K., et al., *Seasonal community succession of the phyllosphere microbiome.* Mol Plant Microbe Interact, 2015. **28**(3): p. 274-85.
54. Tyler, A.D., et al., *Characterization of the gut-associated microbiome in inflammatory pouch complications following ileal pouch-anal anastomosis.* PLoS One, 2013. **8**(9): p. e66934.
55. Hamady, M., et al., *Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex.* Nat Methods, 2008. **5**(3): p. 235-7.
56. Caporaso, J.G., et al., *QIIME allows analysis of high-throughput community sequencing data.* Nat Methods, 2010. **7**(5): p. 335-6.
57. Schloss, P.D., et al., *Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities.* Appl Environ Microbiol, 2009. **75**(23): p. 7537-41.
58. Tang, Z.-Z., G. Chen, and A.V. Alekseyenko, *PERMANOVA-S: association test for microbial community composition that accommodates confounders and multiple distances.* Bioinformatics, 2016. **32**(17): p. 2618-2625.
59. Weiss, S., et al., *Correlation detection strategies in microbial data sets vary widely in sensitivity and precision.* ISME J, 2016. **10**(7): p. 1669-1681.
60. Liu, J., et al., *A laboratory-developed TaqMan Array Card for simultaneous detection of 19 enteropathogens.* J Clin Microbiol, 2013. **51**(2): p. 472-80.
61. Xiong, X., et al., *Perilipin-2 modulates dietary fat-induced microbial global gene expression profiles in the mouse intestine.* Microbiome, 2017. **5**(1): p. 117.
62. Taj, B., et al., *MetaPro: a scalable and reproducible data processing and analysis pipeline for metatranscriptomic investigation of microbial communities.* Microbiome, 2023. **11**(1): p. 1-23.
63. The Metabolomics Innovation Centre. *Services*. 2019; Available from: *https://www.metabolomicscentre.ca/services*.
64. Kanehisa, M., et al., *From genomics to chemical genomics: new developments in KEGG.* Nucleic Acids Res, 2006. **34**(Database issue): p. D354-7.

65. Wang, B., et al., *Similarity network fusion for aggregating data types on a genomic scale.* Nat Methods, 2014. **11**(3): p. 333-7.
66. Vaivada, T., et al., *Evidence-based interventions for improvement of maternal and child nutrition in low-income settings: what's new?* Curr Opin Clin Nutr Metab Care, 2017. **20**(3): p. 204-210.
67. Bhutta, Z.A., et al., *Evidence-based interventions for improvement of maternal and child nutrition: what can be done and at what cost?* Lancet, 2013. **382**(9890): p. 452-477.
68. Kotloff, K.L., et al., *Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study.* The Lancet. **382**(9888): p. 209-222.
69. Striepen, B., *Parasitic infections: Time to tackle cryptosporidiosis.* Nature, 2013. **503**(7475): p. 189-91.
70. Leclerc, H., L. Schwartzbrod, and E. Dei-Cas, *Microbial agents associated with waterborne diseases.* Crit Rev Microbiol, 2002. **28**(4): p. 371-409.
71. Gilchrist, C.A., et al., *Role of the Gut Microbiota of Children in Diarrhea Due to the Protozoan Parasite Entamoeba histolytica.* J Infect Dis, 2016. **213**(10): p. 1579-85.
72. Bar, A.K., et al., *The Interplay of Host Microbiota and Parasitic Protozoans at Mucosal Interfaces: Implications for the Outcomes of Infections and Diseases.* PLoS Negl Trop Dis, 2015. **9**(12): p. e0004176.
73. Burgess, S.L. and W.A. Petri, Jr., *The Intestinal Bacterial Microbiome and E. histolytica Infection.* Curr Trop Med Rep, 2016. **3**: p. 71-74.