THE JOURNAL OF MATERNAL-FETAL & NEONATAL MEDICINE
2022, VOL. 35, NO. 25, 8878–8886
https://doi.org/10.1080/14767058.2021.2005573

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

Performance of a validated spontaneous preterm delivery predictor in South Asian and Sub-Saharan African women: a nested case control study

Rasheda Khanam[a*], Tracey C. Fleischer[b*], Nansi S. Boghossian[c*], Imran Nisar[d], Usha Dhingra[e], Sayedur Rahman[f], Angela C. Fox[b], Muhammad Ilyas[d], Arup Dutta[e], Nurun Naher[f], Ashoka D. Polpitiya[b], Usma Mehmood[d], Saikat Deb[e,g], Aziz Ahmed Choudhury[f], Md. Bahadur Badsha[b], Karim Muhammad[d], Said Mohammed Ali[g], Salahuddin Ahmed[f], Durlin E. Hickok[b], Najeeha Iqbal[d], Mohammed Hamad Juma[g], Md. Abdul Quaiyum[h], J. Jay Boniface[b], Sachiyo Yoshida[i], Alexandar Manu[i], Rajiv Bahl[i], Fyezah Jehan[d], Sunil Sazawal[e,g], Julja Burchard[b†‡] and Abdullah H. Baqui[a†‡]

[a]Department of International Health, Johns Hopkins University Bloomberg School of Public Health, Baltimore, United States; [b]Sera Prognostics, Inc., Salt Lake City, United States; [c]Department of Epidemiology and Biostatistics, University of South Carolina, Arnold School of Public Health, Columbia, United States; [d]Department of Paediatrics and Child Health, Aga Khan University, Karachi, Pakistan; [e]Global Division, Center for Public Health Kinetics, New Delhi, India; [f]Projahnmo Research Foundation, Dhaka, Bangladesh; [g]Public Health Laboratory-IDC, Pemba, Tanzania; [h]International Centre for Diarrheal Disease Research, Dhaka, Bangladesh; [i]World Health Organization (MCA/MRD), Geneva, Switzerland

ABSTRACT

Objectives: To address the disproportionate burden of preterm birth (PTB) in low- and middle-income countries, this study aimed to (1) verify the performance of the United States-validated spontaneous PTB (sPTB) predictor, comprised of the IBP4/SHBG protein ratio, in subjects from Bangladesh, Pakistan and Tanzania enrolled in the Alliance for Maternal and Newborn Health Improvement (AMANHI) biorepository study, and (2) discover biomarkers that improve performance of IBP4/SHBG in the AMANHI cohort.

Study design: The performance of the IBP4/SHBG biomarker was first evaluated in a nested case control validation study, then utilized in a follow-on discovery study performed on the same samples. Levels of serum proteins were measured by targeted mass spectrometry. Differences between the AMANHI and U.S. cohorts were adjusted using body mass index (BMI) and gestational age (GA) at blood draw as covariates. Prediction of sPTB $<$ 37 weeks and $<$ 34 weeks was assessed by area under the receiver operator curve (AUC). In the discovery phase, an artificial intelligence method selected additional protein biomarkers complementary to IBP4/SHBG in the AMANHI cohort.

Results: The IBP4/SHBG biomarker significantly predicted sPTB $<$ 37 weeks ($n = 88$ vs. 171 terms $\geq$ 37 weeks) after adjusting for BMI and GA at blood draw (AUC= 0.64, 95% CI: 0.57–0.71, $p < .001$). Performance was similar for sPTB $<$ 34 weeks ($n = 17$ vs. 184 $\geq$ 34 weeks): AUC = 0.66, 95% CI: 0.51–0.82, $p = .012$. The discovery phase of the study showed that the addition of endoglin, prolactin, and tetranectin to the above model resulted in the prediction of sPTB $<$ 37 with an AUC= 0.72 (95% CI: 0.66–0.79, p-value $< .001$) and prediction of sPTB $<$ 34 with an AUC of 0.78 (95% CI: 0.67–0.90, $p < .001$).

Conclusion: A protein biomarker pair developed in the U.S. may have broader application in diverse non-U.S. populations.

ARTICLE HISTORY
Received 1 September 2021
Revised 15 October 2021
Accepted 9 November 2021

KEYWORDS
Preterm birth; biomarkers; low- and middle-income countries; IBP4; SHBG

Introduction

Preterm birth (PTB) affects approximately 15 million infants annually, about 11% of all live births worldwide [1]. Globally, PTB and related complications are the leading causes of neonatal deaths (35%) [2,3] and of deaths in children under five

CONTACT Julja Burchard jburchard@seraprognostics.com Sera Prognostics, Inc, 2749 East Parleys Way, Suite 200, Salt Lake City, 84109, UT, United States; Abdullah H. Baqui abaqui@jhu.edu Johns Hopkins University Bloomberg School of Public Health, Room E-8153, 615 N. Wolfe Street, Baltimore, 21205, MD, USA
*joint first author
‡joint senior author
†joint corresponding author

Supplemental data for this article can be accessed here.

years [4]. Surviving preterm infants may experience significant morbidities such as chronic lung disease, hearing and visual impairments, neurodevelopmental disabilities [5], and chronic disease in adulthood [6,7]. The familial and economic burden of PTB is substantial [8].

The incidence of PTB ranges from approximately 5% in some European countries to 18% in certain African countries [9]. Worldwide, more than 60% of all PTBs occur in sub-Saharan Africa and South Asia [9]. The majority of studies identifying biomarkers predictive of spontaneous PTB (sPTB) may lack sufficient clinical performance, suffer from small subject numbers, or both [10], and have been conducted in high-income countries [11,12]. However, the underlying risk factors and causes of sPTB may differ in low-and middle-income countries (LMICs). Nevertheless, despite different upstream causes, there may be downstream pathway convergence, which may enable predictive tests developed in high-income countries to perform well in LMICs.

The World Health Organization coordinated the Alliance for Maternal and Newborn Health Improvement (AMANHI) study, involving about ten thousand pregnant women in three sites of South Asia and sub-Saharan Africa (Sylhet, Bangladesh; Karachi, Pakistan; and Pemba Island, Tanzania) [13]. The objectives of the AMANHI study were to establish a biobank and to identify biomarkers of adverse pregnancy outcomes in developing countries [13]. The study plan included steps to evaluate candidate biomarkers identified in high-income countries and to conduct novel discovery studies [13].

A maternal serum-based proteomics predictor of sPTB $<$ 37 weeks was validated in a United States (U.S.) cohort with maximal performance in serum collected $19^{1/7}$–$20^{6/7}$ weeks gestation [12]. The predictor was comprised of a ratio of two proteins, insulin-like growth factor-binding protein 4 (IBP4) and sex hormone-binding globulin (SHBG) and demonstrated better performance in a stratified BMI range of $>$ 22 – $\leq$ 37 [12]. Recently, these biomarkers were shown to predict very early PTB ($<$ 32 weeks) from spontaneous and medically indicated causes [14]. IBP4/SHBG also predicted neonatal morbidity and length of neonatal hospital stay, suggesting sensitivity to determinates of neonatal outcomes [14].

The objectives of this study were to: (1) verify the performance of the IBP4/SHBG biomarker in the AMANHI study cohort, and, (2) discover novel classifier proteins that improve the performance of IBP4/SHBG in the AMANHI study cohort. Exploratory analyses were conducted to identify additional novel proteomic and clinical variable biomarkers of sPTB across all three geographies combined, with validation planned utilizing future cohorts.

Materials and methods

Study design, settings, and participants

Between 2014 and 2018, the AMANHI biobanking study prospectively enrolled 10,001 pregnant women, identified through population-based surveillance in three countries: Bangladesh, Pakistan, and Tanzania [13]. Trained community health workers visited all women of child-bearing age in the study areas every two to three months to identify pregnancies and obtain informed consent. GA was determined using ultrasound before 20 weeks of gestation using standardized measurements [13]. Community health workers made four antepartum (8–19 weeks, 24–28 weeks, 32–36 weeks, and 38+ weeks of gestation) and two postpartum home visits to collect background characteristics, previous medical history, risk factors, exposures, outcomes, and morbidity for the index pregnancy. BMI was calculated from maternal height and weight measured at the enrollment visit. Maternal blood was collected and processed using a standardized protocol and serum samples were stored at -80°C. De-identified samples were shipped to the U.S. *via* courier in a liquid nitrogen dry shipper.

Selection of cases and controls

In developing the protocol for this study, a power and sample size analysis determined that 32 cases and 64 controls per site achieved 91% power to distinguish an AUC = 0.7 from AUC = 0.5 (random performance). Combining three sites, the case-control study comprised 300 subjects (100 sPTB cases $<$ 37 weeks of gestation and 200 control term deliveries $\geq$ 37 weeks) enrolled in 2014–2016. Inclusion criteria included the ability to consent, singleton pregnancy, and serum collection within $17^{0/7}$ and $19^{6/7}$ weeks. Exclusion criteria included signs/symptoms of preterm labor at the time of specimen collection, major fetal anomaly, blood transfusion during the current pregnancy, use of progesterone after $12^{6/7}$ weeks gestation, use of heparin, or serum hemolysis $>$ 100 mg/dl. Two-term births per case matched by the gestational week of blood draw and site were selected randomly from qualifying and available samples: Bangladesh (36 sPTB / 72 term), Pakistan (23 sPTB / 46 term), and Tanzania (40 sPTB / 80 term). One case and one control from Bangladesh

were excluded from analyses because the case sample did not show pregnancy-specific proteins, and the control sample was drawn in week 16.

Laboratory methods

De-identified samples were received blinded, randomized, and processed in a CLIA-certified laboratory according to a standard operating procedure [12,15]. Briefly, serum samples were depleted of high abundance proteins, trypsinized, fortified with stable isotope-labeled internal standard (SIS) peptides, desalted, and analyzed by coupled liquid chromatography-multiple reaction monitoring mass spectrometry (LC-MRM-MS) measuring 122 proteins associated with pregnancy, of placental origin, or for quality control. Peptides were quantified as the response ratio between endogenous and SIS peak area counts. Quality was assessed for each batch [12,15] and overall.

Statistical analyses

Significant differences ($p < .05$) in demographics and clinical variables between the U.S. validation and the AMANHI cohorts were determined using a *t*-test (means) or a Wilcoxon test (medians) for continuous variables and the Fisher's Exact test for categorical variables, with missing values excluded from analyses [16,17].

IBP4/SHBG biomarker scores were calculated as described [12,14]. As prespecified in the data analysis plan, because subjects were largely outside of the intended use in geography, anthropometrics and GA at blood draw, emphasis was on confirmation of the IBP4/SHBG biomarker after controlling for these differences. For validation in the AMANHI cohort, we tested the prediction of sPTB using logistic regression with models comprised of the IBP4/SHBG biomarker with and without adjustment for GA at blood draw and BMI. The appropriateness of the assumption of linear relationships was assessed by calculating average marginal effects [18]. Predictive performance was reported by the area under the receiver operator curve (AUC) with prespecified direction (cases > controls) and 95% confidence intervals (CI) calculated by DeLong's method [19]. A Wilcoxon one-sided test was used to calculate *p*-values. Subjects with missing BMI values were omitted, although imputation of missing BMI values using Multivariate Imputation by Chained Equations [20] yielded similar results.

To improve the performance of the baseline predictor (IBP4/SHBG + GA at blood draw * BMI), causal inference network analysis [21,22] was used to select additional proteins (log-transformed response ratios) and clinical variables as nodes directly causal of sPTB (Supplemental). Candidate proteomic and clinical variables were combined with the baseline predictor in logistic regression models as above. Significant classification performance improvement was defined as an AUC greater than the upper 95% confidence bound of the AUC of the baseline predictor (base R, stats package $\geq$ 4.0.3, [23]).

Prediction of early sPTB was assessed without exclusion or overrepresentation of late sPTBs or early-term births [24]. For example, to predict sPTB $<$ 34 weeks, subjects delivering $\geq$ 34 weeks were defined as controls and adjusted to their natural rate in the AMANHI population. To minimize bias, adjustment was repeated 100 times and median AUCs were reported.

For Kaplan-Meier analysis, subjects were divided into lower and higher risk groups based on percentile thresholds from 5th to 95th, in 5% increments. GA at birth was used as the time variable, and significance was assessed by the log-rank test.

Ethics approval

The study protocol was approved by the following ethics committees: WHO Institutional Review Board (IRB), International Centre for Diarrhoeal Disease Research Bangladesh (icddr,b) in Bangladesh, Aga Khan University in Pakistan, Zanzibar Medical Research and Ethics Committee in Tanzania, and the IRB of Johns Hopkins Bloomberg School of Public Health, U.S.A.

Results

Comparison of AMANHI and U.S. Validation cohorts

Clinical characteristics of the AMANHI cohort were compared to the original U.S. validation study cohort [12] (Table 1). The mean GA at blood draw was significantly different between the two studies (128 vs. 140 d, $p < .001$). The optimal blood draw window for the IBP4/SHBG biomarker was $19^{1/7}$–$20^{6/7}$ weeks gestation, while the AMANHI samples spanned weeks $17^{0/7}$–$19^{6/7}$. The mean BMI of the AMANHI cohort was significantly lower than the U.S. cohort (21.8 kg/m^2 vs. 27.7 kg/m^2, $p < .001$), with 155 of 259 subjects with recorded BMI falling below the optimal U.S. BMI

Table 1. Comparison of AMANHI and U.S. validation cohorts.

	AMANHI ($N = 298$)	U.S. Validation Cohort [12] ($N = 54$)	p-value
Maternal age (years)			
Mean (SD)	26.1 (5.97)	26.2 (6.25)	.90
Median [Q1, Q3]	25.0 [21.0, 30.0]	24.5 [21.3, 30.0]	.94
Body mass index (kg/m^2)			
Mean (SD)*	21.8 (4.19)	27.7 (6.22)	<.001
Median [Q1, Q3]*	20.8 [18.6, 24.0]	26.5 [22.3, 31.3]	<.001
Gravida			
Multigravida	235 (78.9%)	41 (75.9%)	.60
Primigravida	63 (21.1%)	13 (24.1%)	
Prior PTB$^+$ (Multigravida)			
No	225 (95.7%)	34 (82.9%)	.006
Yes	10 (4.2%)	7 (17.0%)	
Gestational age at blood draw (days)			
Mean (SD)	128 (6.39)	140 (4.20)	<.001
Median [Q1, Q3]	127 [122, 134]	139 [135, 144]	<.001
Gestational age at birth (days)			
Mean (SD)	267 (18.4)	265 (24.2)	.44
Median [Q1, Q3]	271 [256, 281]	273 [256, 281]	.90

*Missing BMI values omitted from calculations: AMANHI cohort, $n = 39$; U.S. cohort, $n = 1$.
$^+$Prior PTB is restricted to prior sPTB in the U.S. validation cohort. Prior PTB is not limited by the timing of initiation of delivery in the AMANHI cohort.

Table 2. Comparison of three sPTB predictive models containing IBP4/SHBG.

	sPTB < 37					sPTB < 34				
Model	AUC	95% CI	p-value	Cases (n)	Controls (n)	AUC	95% CI	p-value	Cases (n)	Controls (n)
U.S. validated predictor: **IBP4/SHBG**	0.55	0.48–0.62	.069	99	199	0.62	0.49–0.75	.044	19	214
Adjusted for demographic differences: **IBP4/SHBG + GABD*BMI**	0.64	0.57–0.71	<.001	88	171	0.66	0.51–0.82	.012	17	184
Discovery phase model: **IBP4/SHBG + GABD*BMI + (EGLN x PRL) / TETN**	0.73	0.66–0.79	<.001	88	171	0.78	0.67–0.90	<.001	17	184

GABD – gestational age at blood draw; GABD*BMI denotes the addition of the main effects of GABD and BMI plus their interaction (product): GABD + BMI + GABD:BMI.

(> 22 to ≤ 37 kg/m^2) [12]. The proportion of AMANHI subjects with a prior PTB was lower than in the U.S. cohort (Table 1). However, because AMANHI data collection for prior preterm birth was based on recall, the prevalence may be underestimated. There were no significant differences in maternal age, gravidity, or GA at delivery between the cohorts (Table 1).

Performance of the validated IBP4/SHBG sPTB predictor in the AMANHI cohort

IBP4/SHBG is influenced by GA and BMI [12], and SHBG blood levels are associated with BMI [25]. Thus, because the U.S. and AMANHI populations had significantly different blood draw windows and BMI, without a population adjustment for these variables the IBP4/SHBG biomarker score did not reach statistical significance ($p = .069$, Table 2). However, with adjustment for GA at blood draw and BMI, IBP4/SHBG significantly classified sPTB subjects (AUC = 0.64, 95% CI: 0.57–0.71, $p < .001$, Table 2).

To test performance in classifying early sPTB, subjects with sPTB < 34 weeks were compared to control births (≥ 34 weeks). Importantly, the baseline predictor (IBP4/SHBG, GA at blood draw and BMI) also significantly classified early (< 34 weeks) sPTB (AUC = 0.66, 95% CI: 0.51–0.82, $p = .012$).

Discovery of novel predictors for the AMANHI cohort

To discover improved sPTB prediction in AMANHI geographies, we used artificial intelligence network techniques to select new features. This conditional correlation network analysis identified direct antecedents of sPTB: primigravida, prior PTB, and twelve proteins, in addition to IBP4/SHBG, GA at blood draw, and BMI (Supplemental Figure). These top features were added to the baseline IBP4/SHBG predictor individually, in pairs, or in triplets. AUC was significantly improved over the baseline predictor only with the addition of three proteins. Extra clinical variables did not significantly improve performance.

The top predictor ranked by AUC included three new analytes: endoglin (EGLN) and prolactin (PRL) were positively associated with sPTB, and tetranectin

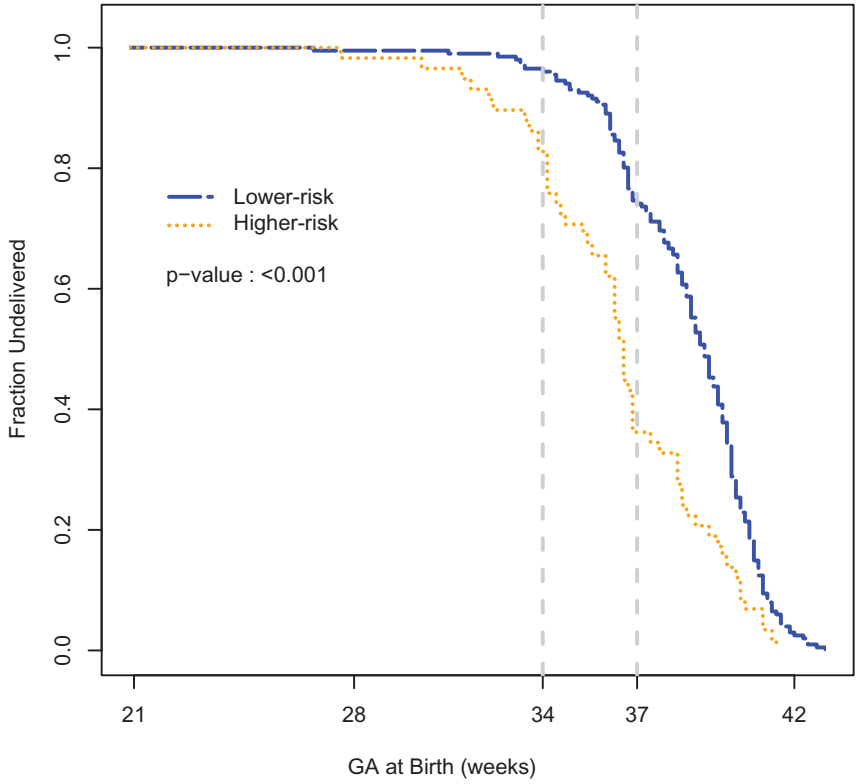

Figure 1. Kaplan-Meier analysis of GA at delivery for lower- and higher-risk groups. Time-to-event analysis shows the rate of births for the top 15% of scores (higher risk) in the discovery phase predictor (IBP4/SHBG + GA at blood draw * BMI + [EGLNxPRL]/TETN) vs. low scores (lower risk). Vertical lines indicate delivery at 34 and 37 weeks.

(TETN) was negatively associated, forming a second ratio in the algorithm, (EGLN x PRL)/TETN (Table 2). Performance for predicting sPTB < 37 or sPTB < 34 in this discovery phase predictor was AUC $= 0.72$, $p < .001$ and AUC $= 0.78$, $p < .001$, respectively (Table 2).

The subjects were then stratified into low- and high-risk groups at an 85th percentile threshold, where 15% of subjects would be deemed higher risk. A Kaplan-Meier analysis indicated that subjects in the high-risk group (the top 15%) delivered significantly ($p < .001$) earlier than those in the lower-risk group (Figure 1). Significant separation was also seen from 95th (5% at higher risk) to 15th (85% higher risk) percentile thresholds ($p < .001$–$.028$).

Discussion

We confirmed that a U.S.-validated proteomics predictor of sPTB could be applied in LMICs, after adjusting for expected demographic differences between the populations. In the subsequent discovery phase of this study, a predictor of sPTB < 37 weeks including three new proteins showed improved predictive performance.

The AMANHI cohort is different from the U.S. cohort. The mean BMI for AMANHI subjects was below the optimal BMI range identified in the U.S. cohort [12]. AMANHI blood samples were drawn in weeks 17–19 of gestation, whereas the U.S. test was validated for weeks 19–20 gestation and demonstrated dependence on the blood draw period [12]. Nevertheless, the IBP4/SHBG biomarker, when adjusted for these differences, significantly identified sPTB subjects. Importantly, the adjusted predictor performed well for the prediction of sPTB < 37 and < 34 weeks. As adverse outcomes are inversely related to GA [26], it is critical that a sPTB predictor be able to identify those patients destined for early delivery.

An AMANHI discovery phase predictor comprised of IBP4/SHBG, GA at blood draw, BMI, EGLN, PRL, and TETN demonstrated improved classification performance for both sPTB < 37 and < 34 weeks. The improved predictor significantly stratified patients by GA at delivery over a wide range of percentiles thresholds, demonstrating potential flexibility in implementation.

Optimal timing of prognostic test administration requires balancing the need for timely intervention with the ability to access women seeking care. Administration of a second-trimester test may appropriately address this balance in LMICs, where few women seek prenatal care in the first trimester [27–29]. Together, flexibility in stratification and timing of blood draw demonstrates that these biomarkers may be suitable for the development of a clinically useful sPTB diagnostic test applicable across LMIC geographies.

The biological plausibility of the IBP4/SHBG biomarker has been discussed elsewhere [12]. Briefly, SHBG's decreased abundance in the second trimester in women who subsequently develop sPTB [12] or preeclampsia [30], a major medical indication for PTB, may result from pro-inflammatory signals [31]. Decreased SHBG levels would be predicted to result in increased levels of free estrogens that oppose progesterone and deliver pro-labor signals [12]. Insulin-like growth factor (IGF) signaling pathways have been implicated as key regulators of placental development and fetal nutrient programming [32]. Higher maternal serum levels of IBP4, a key regulator of IGF2 bioavailability in the placenta bed [33], are associated with growth-restricted fetuses [34] and sPTB [12].

In the AMANHI cohort, we observed that median levels of both EGLN and PRL were elevated in sPTB relative to term, whereas TETN levels were decreased. Endoglin, a transmembrane coreceptor for transforming growth factor-beta (TGFβ) [35], regulates differentiation, cell migration, and angiogenesis [36,37]. EGLN is expressed in the placenta [38], where it inhibits trophoblast migration and invasion [39]. Elevated circulating levels of soluble EGLN (sEGLN) are associated with preeclampsia (PE) [40,41] and may serve as a biomarker to predict PE, particularly in combinations with other angiogenesis factors [42] and uterine artery doppler lowest pulsatility index [43,44]. Soluble EGLN levels were also elevated in women delivering infants who are small for gestational age [44–46], preterm [46,47], or in amniotic fluid from pregnancies complicated by intraamniotic infection [48]. However, the significance of sEGLN to predict sPTB, as opposed to PE, from serum at 17–19 weeks gestation is an unexpected finding of this study.

Prolactin, a pituitary growth factor responsible for the development of mammary glands and milk production, increases 10–20-fold in pregnancy [49]. Its expression in decidua during pregnancy [50], and reported pleiotropic activities including immunemodulation [51], regulation of insulin resistance by facilitating the transport of glucose and other nutrients across the placenta [52], and placental angiogenesis [53], suggest important roles in pregnancy health. Circulating and urine PRL levels (full length and anti-angiogenic fragment) were higher in severe vs mild PE, and predicted adverse maternal and fetal outcomes such as small for gestational age [54]. Cervicovaginal fluid PRL was more detectable in women symptomatic of preterm delivery than in asymptomatic women [55]. A systematic review and meta-analysis of potential sPTB biomarkers found that cervicovaginal PRL was one of three biomarkers out of 30 meeting inclusion criteria with a high ($>$ 10) positive likelihood ratio [10]; however, its utility to predict sPTB as a blood-based biomarker has not been reported to our knowledge.

TETN has been implicated in extracellular matrix remodeling and fibrinolysis *via* interactions with plasminogen [56] and fibrin [57]. Lower levels of serum/plasma TETN are associated with various disease states including cancer, particularly metastatic disease [58], arthritis [59,60], heart failure [61], and PE [62]. Exosomes derived from tumor cells over-expressing TETN reduced VEGF secretion and inhibited angiogenesis [63]. In both amniotic fluid and fetal serum, the correlation between TETN levels and gestational age was seen, suggesting a role in fetal maturation [64]. TETN is reported to be negatively regulated by TGFβ [65], a pathway of importance in decidualization and placentation [66–68]. By extension, TETN may be involved in trophoblast invasion. However, direct evidence for the role of TETN in preterm birth is lacking.

Studies examining the proteomics of sPTB in LMICs with a high burden of prematurity are limited. A recent study by Jehan et al. utilized a multi-omics approach to discover plasma and urine biomarkers of sPTB from AMANHI and Global Alliance to Prevent Prematurity and Stillbirth (GAPPS) studies in a combined cohort of 81 subjects [69]. Interestingly, even though some of the geographies differed between the studies, both highlight the prognostic potential of proteomics and the importance of inflammatory and glucose homeostasis pathways.

Our study had many strengths. The AMANHI study design included a large population-based cohort with early gestational dating conducted by trained professionals. Additionally, serum, sociodemographic, and pregnancy characteristics were collected in a harmonized manner across all three sites. In the completely independent AMANHI cohort, following a pre-specified process, we tested a previously discovered and

validated proteomics predictor and established its validity in this LMIC population.

Importantly, the analyses of prediction of early sPTB < 34 weeks should be interpreted with caution, due to small subject numbers. As well, the discovery phase predictor encompassing novel proteins requires further validation. This study was not sufficiently powered to allow for subset analyses. Future studies include exploration of the pathways leading to sPTB in different geographies.

Conclusions

We demonstrated that a serum protein predictor that was discovered, verified, and validated in the United States can predict sPTB in LMICs. Patient characteristics and timing of blood draw may be useful considerations when developing and applying a predictive test to a new geography.

Geolocation

The locations in this study included: Sylhet, Bangladesh (24.89904° N, 91.87198° E); Karachi, Pakistan (24.8607° N, 67.0011° E); and Pemba Island, Tanzania (5.0319° S, 39.7756° E).

Acknowledgments

We acknowledge the study participants for their essential contributions of time, care, samples and data; the dedicated field and data teams for implementing the AMANHI study; the Sera Clinical Laboratory for data generation; Max T. Dufford, BS, for data management; ChienTing Hsu, MS, Ryan M. Treacy, BS, and Jeff S. Flick, PhD, for preliminary analyses; and Todd L. Randolph, MD and Gregory C. Critchfield, MD, MS for critical review.

Disclosure statement

JJB, JB, TCF, ACF, MBB, ADP, and DEH are stockholders and employees or consultants of Sera Prognostics. JJB, JB, ADP, TCF, MBB and DEH have patents issued and pending related to this work. All other authors report no conflict of interest.

Funding

This work was funded by the Bill and Melinda Gates Foundation, including a grant to Sera Prognostics (Contract ID OPP1127876) and coordinated by the World Health Organization. The funders had no roles in designing of the biobank study, data collection, and data analysis.

ORCID

Rasheda Khanam http://orcid.org/0000-0002-9365-8594
Tracey C. Fleischer http://orcid.org/0000-0002-7141-4265
Nansi S. Boghossian http://orcid.org/0000-0001-7424-5148
Imran Nisar http://orcid.org/0000-0002-2378-4720
Sayedur Rahman http://orcid.org/0000-0002-4445-3699
Muhammad Ilyas http://orcid.org/0000-0001-9696-2084
Nurun Naher http://orcid.org/0000-0001-9687-1449
Ashoka D. Polpitiya http://orcid.org/0000-0001-8698-2188
Md. Bahadur Badsha http://orcid.org/0000-0002-6379-1554
Salahuddin Ahmed http://orcid.org/0000-0001-6771-0638
Durlin E. Hickok http://orcid.org/0000-0002-3421-5881
Najeeha Iqbal http://orcid.org/0000-0003-4026-5655
J. Jay Boniface http://orcid.org/0000-0002-5273-1089
Sachiyo Yoshida http://orcid.org/0000-0002-1101-8535
Rajiv Bahl http://orcid.org/0000-0001-7936-6985
Fyezah Jehan http://orcid.org/0000-0002-5874-4358
Sunil Sazawal http://orcid.org/0000-0002-8035-9085
Julja Burchard http://orcid.org/0000-0001-9446-5180
Abdullah H. Baqui http://orcid.org/0000-0001-8350-1983

Data availability statement

Deidentified individual participant data will be made available upon approval of the manuscript, including data dictionaries and data that underlie the results reported in this article. Data will be shared with researchers who contact either of the corresponding authors requesting use of the data for research on preterm birth in LMICs. Requestors will need to sign a data access agreement.

References

[1] Blencowe H, Cousens S, Chou D, et al. Born too soon: the global epidemiology of 15 million preterm births. Reprod Health. 2013;10(Suppl 1):S2.
[2] Harrison MS, Goldenberg RL. Global burden of prematurity. Semin Fetal Neonatal Med. 2016;21(2):74–79.
[3] Liu L, Oza S, Hogan D, et al. Global, regional, and national causes of child mortality in 2000-13, with projections to inform post-2015 priorities: an updated systematic analysis. Lancet. 2015;385(9966):430–440.
[4] March of Dimes, PMNCH, Save the Children, WHO. Born too soon: the global action report on preterm birth. In: Howson C, Kinney M, Lawn J, editors. Geneva: World Health Organization; 2012.
[5] Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. Lancet. 2008;371(9608):261–269.
[6] Hovi P, Andersson S, Eriksson JG, et al. Glucose regulation in young adults with very low birth weight. N Engl J Med. 2007;356(20):2053–2063.
[7] Rotteveel J, van Weissenbruch MM, Twisk JW, et al. Infant and childhood growth patterns, insulin sensitivity, and blood pressure in prematurely born young adults. Pediatrics. 2008;122(2):313–321.
[8] Frey HA, Klebanoff MA. The epidemiology, etiology, and costs of preterm birth. Semin Fetal Neonatal Med. 2016;21(2):68–73.

[9] Blencowe H, Cousens S, Oestergaard MZ, et al. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. Lancet. 2012;379(9832): 2162–2172.
[10] Conde-Agudelo A, Papageorghiou AT, Kennedy SH, et al. Novel biomarkers for the prediction of the spontaneous preterm birth phenotype: a systematic review and meta-analysis. BJOG. 2011;118(9):1042–1054.
[11] Kacerovsky M, Lenco J, Musilova I, et al. Proteomic biomarkers for spontaneous preterm birth: a systematic review of the literature. Reprod Sci. 2014;21(3): 283–295.
[12] Saade GR, Boggess KA, Sullivan SA, et al. Development and validation of a spontaneous preterm delivery predictor in asymptomatic women. Am J Obstet Gynecol. 2016;214(5):633.e1–633–e24.
[13] Baqui AH, Khanam R, Rahman MS, et al. Understanding biological mechanisms underlying adverse birth outcomes in developing countries: protocol for a prospective cohort (AMANHI biobanking) study. J Glob Health. 2017;7(2):021202.
[14] Markenson GR, Saade GR, Laurent LC, et al. Performance of a proteomic preterm delivery predictor in a large independent prospective cohort. Am J Obstetrics Gynecol MFM. 2020;2(3):100140.
[15] Bradford C, Severinsen R, Pugmire T, et al. Analytical validation of protein biomarkers for risk of spontaneous preterm birth. Clin Mass Spectrom. 2017;3: 25–38.
[16] Rich B. table1: tables of Descriptive Statistics in HTML. R package version 1.2. 2020.
[17] Robinson L. demoGraphic: Providing Demographic Table with the P-Value, Standardized Mean Difference Value. R package version 0.1.0; 2019.
[18] Leeper T. margins: Marginal Effects for Model Objects; 2021.
[19] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837–845.
[20] van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. J Stat Soft. 2011;45(3):1–67.
[21] Badsha MB, Fu AQ. Learning causal biological networks with the principle of mendelian randomization. Front Genet. 2019;10(460):460.
[22] Badsha MB, Martin EA, Fu AQ. MRPC: An R package for accurate inference of causal graphs. 2021.
[23] R Core Team, R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.
[24] Boniface JJ, Burchard J, Saade GR. Effects of selective exclusion of patients on preterm birth test performance. Obstet Gynecol. 2019;134(6):1333–1338.
[25] Xargay-Torrent S, Carreras-Badosa G, Borrat-Padrosa S, et al. Circulating sex hormone binding globulin: an integrating biomarker for an adverse cardio-metabolic profile in obese pregnant women. PLOS One. 2018; 13(10):e0205592.
[26] Glass HC, Costarino AT, Stayer SA, et al. Outcomes for extremely premature infants. Anesth Analg. 2015; 120(6):1337–1351.
[27] Saad-Haddad G, DeJong J, Terreri N, et al. Patterns and determinants of antenatal care utilization: analysis of national survey data in seven countdown countries. J Glob Health. 2016;6(1): 010404–010404.
[28] Agha S, Tappis H. The timing of antenatal care initiation and the content of care in Sindh, Pakistan. BMC Pregnancy Childbirth. 2016;16(1):190.
[29] Jiwani SS, Amouzou-Aguirre A, Carvajal L, et al. Timing and number of antenatal care contacts in low and middle-income countries: analysis in the countdown to 2030 priority countries. J Glob Health. 2020; 10(1):010502.
[30] Yu CK, Papageorghiou AT, Bindra R, et al. Second-trimester sex hormone-binding globulin and subsequent development of pre-eclampsia. J Matern Fetal Neonatal Med. 2004;16(3):158–162.
[31] Simo R, Saez-Lopez C, Barbosa-Desongles A, et al. Novel insights in SHBG regulation and clinical implications. Trends Endocrinol Metab. 2015;26(7):376–383.
[32] Forbes K, Westwood M. Maternal growth factor regulation of human placental development and fetal growth. J Endocrinol. 2010;207(1):1–16.
[33] Giudice LC, Conover CA, Bale L, et al. Identification and regulation of the IGFBP-4 protease and its physiological inhibitor in human trophoblasts and endometrial stroma: evidence for paracrine regulation of IGF-II bioavailability in the placental bed during human implantation. J Clin Endocrinol Metab. 2002; 87(5):2359–2366.
[34] Qiu Q, Bell M, Lu X, et al. Significance of IGFBP-4 in the development of fetal growth restriction. J Clin Endocrinol Metab. 2012;97(8):E1429–39.
[35] Cheifetz S, Bellón T, Calés C, et al. Endoglin is a component of the transforming growth factor-beta receptor system in human endothelial cells. J Biol Chem. 1992;267(27):19027–19030.
[36] Gregory AL, Xu G, Sotov V, et al. Review: the enigmatic role of endoglin in the placenta. Placenta. 2014; 35 Suppl (Suppl):S93–S9.
[37] Duff SE, Li C, Garland JM, et al. CD105 is important for angiogenesis: evidence and potential applications. Faseb J. 2003;17(9):984–992.
[38] Gougos A, St Jacques S, Greaves A, et al. Identification of distinct epitopes of endoglin, an RGD-containing glycoprotein of endothelial cells, leukemic cells, and syncytiotrophoblasts. Int Immunol. 1992;4(1):83–92.
[39] Caniggia I, Taylor CV, Ritchie JWK, et al. Endoglin regulates trophoblast differentiation along the invasive pathway in human placental villous explants. Endocrinology. 1997;138(11):4977–4988.
[40] Margioula-Siarkou G, Margioula-Siarkou C, Petousis S, et al. Soluble endoglin concentration in maternal blood as a diagnostic biomarker of preeclampsia: a systematic review and meta-analysis. Eur J Obstet Gynecol Reprod Biol. 2021;258:366–381.
[41] Jeyabalan A, McGonigal S, Gilmour C, et al. Circulating and placental endoglin concentrations in