medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Clinical and economic utility of a preterm birth predictor derived from an analysis of a large and diverse pregnancy cohort

Julja Burchard[1], Glenn R. Markenson[2], George R. Saade[3], Louise C. Laurent[4], Kent D. Heyborne[5], Dean V. Coonrod[6], Corina N. Schoen[7], Jason K. Baxter[8], David M. Haas[9], Sherri Longo[10], Scott A. Sullivan[11], Carol A. Major[12], Sarahn M. Wheeler[13], Leonardo M. Pereira[14], Emily J. Su[15], Kim A. Boggess[16], Angela F. Hawk[17], Amy H. Crockett[18], Thomas J. Garite[1], J. Jay Boniface[1], Tracey C. Fleischer[1]*, Angela C. Fox[1], Ashoka D. Polpitiya[1], Gregory C. Critchfield[1], Paul E. Kearney[19]

[1]Sera Prognostics, Inc., [2]Boston Medical Center, [3]University of Texas Medical Branch, [4]University of California, San Diego, [5]Denver Health and Hospital Authority, [6]Maricopa Integrated Health System, [7]University of Massachusetts Baystate, [8]Thomas Jefferson University Hospital, [9]Indiana University, [10]Ochsner Baptist Medical Center, [11]Medical University of South Carolina, [12]University of California, Irvine, [13]Duke University, [14]Oregon Health and Science University, [15]University of Colorado School of Medicine, [16]University of North Carolina, [17]Regional Obstetrical Consultants, [18]Greenville Health System, [19]Data Incites, Inc.

*Corresponding author: 2749 East Parleys Way, Suite 200, Salt Lake City, UT 84109; (801) 990-0597

Abstract

Objective

Evaluate clinical utility and cost effectiveness of identifying pregnancies at increased risk of preterm birth using a validated proteomic biomarker risk predictor to enable proactive intervention

Study Design

Pregnancies at elevated risk (≥15%) of preterm birth were identified in a cohort from TREETOP (NCT02787213), a study independent of biomarker development. In the screening arm, higher-risk subjects received simulated interventions based on published efficacy of multimodal treatment or care-management alone. Subjects in the non-screening arm received no interventions. Neonatal and maternal length of stay, neonatal mortality and morbidity and neonatal costs were compared between arms.

Results

Multimodal/care-management modeled treatments predicted reductions in neonatal (30%/22%) and maternal (9.2%/8.5%) hospital stays, neonatal morbidity and mortality (41%/29%), and neonatal costs (34%/16%) for the screening vs. non-screening arm.

Conclusion

Modeled interventions applied to pregnancies identified as higher-risk by a proteomic biomarker risk predictor demonstrate clinically and economically meaningful improvements in neonatal and maternal outcomes.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

Preterm birth (PTB) occurs in approximately 10% of all births and is the leading cause of neonatal death in the United States (1, 2). In addition to neonatal morbidity and mortality, the economic impact of PTB on the healthcare system is enormous and estimated to be greater than 25 billion U.S. dollars annually (3). Direct medical care accounts for the majority of these costs which are significantly higher for individuals who are delivered at earlier gestational ages (GA). The effective application of interventions such as progesterone, cervical cerclage and intensive case management to ameliorate PTB and its sequelae requires tools to identify pregnancies at risk. Clinical factors used to assess the risk of PTB are poorly predictive and are present in only a small minority of pregnancies. A history of previous PTB is a traditional predictor of recurrent PTB but is seen in only approximately 4% of all pregnancies and predicts only 11% of all spontaneous PTBs (sPTBs) (4, 5). Similarly, a short cervical length measured by transvaginal ultrasound is a widely used predictor of sPTB, but accounts for only 2% of all pregnancies and an additional 6% of all sPTBs beyond those who are identified by a history of prior preterm delivery (6, 7). Biomarker-based risk stratification tools offer an important solution to predict and potentially ameliorate PTB and its sequelae. Recent years have seen the onset of precision diagnostic tools that enable the targeting of interventions. Of particular interest here are tools that predict PTB in women with no apparent risk well in advance of delivering in an effort to improve outcomes (8). Saade and colleagues described and clinically validated a novel serum proteomic biomarker risk predictor for sPTB based on a combination of the expression ratio of insulin-like growth factor-binding protein 4 (IBP4, gene symbol IGFBP4)

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

to sex hormone-binding globulin (SHBG) and clinical variables using samples from the Proteomic Assessment of Preterm Risk (PAPR) study (NCT01371019) (9). In a follow up study, the proteomic biomarker risk predictor was observed to be predictive of medically indicated PTB (10). Additionally, the mass spectrometry assay measuring the protein ratio IBP4/SHBG in pregnant serum has been analytically validated (11) and a threshold at which the probability of sPTB <37 weeks of gestation is at least twice the baseline rate has been established (9).

In addition to clinical and analytical validation, establishing the clinical and economic utility of a test is important to determine its value in clinical practice for all healthcare stakeholders including patients, physicians, healthcare workers and insurers (12). Studies of clinical utility are frequently classified as those establishing *efficacy* or those establishing comparative *effectiveness* (13-16). Studies of efficacy establish a causal link between a novel protocol and positive outcomes, typically in a highly controlled and less generalizable setting or population. Randomized control trials (RCTs) are the gold standard for establishing efficacy. Comparative effectiveness studies, in contrast, establish a correlative, but not causal, link between the test and clinical outcomes in more generalizable settings or populations, and so, provide 'real world' evidence of clinical utility. Ideally, a predictive test has evidence of both efficacy and comparative effectiveness. Recent years have seen much work on the evolution of clinical utility thinking and application (14-16).

The purpose of the current analysis, named ACCORDANT (Analyses aCross COngruent studies ReDucing Adverse pregNancy ouTcomes), is to determine the clinical and economic utility of the proteomic biomarker risk predictor when used as a screening test for a low-risk pregnant

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

population in improving neonatal outcomes when coupled with established interventions (17-24), such as progesterone and care-management. The ACCORDANT study utilized a cohort of pregnancies from the TREETOP study (25), which is independent from sources of biomarker development and intervention efficacy measurement. TREETOP was an observational study including only standard care and not previously utilized in modeling studies. In contrast to pure modeling studies, the TREETOP subjects were all assigned to higher or lower risk of preterm delivery < 37 weeks of gestation risk based on actual proteomic biomarker mass spectrometry measurements performed as samples were collected. Modeled interventions were applied to higher-risk pregnancies in the ACCORDANT screening arm. The effectiveness of ACCORDANT's test-and-treat approach was quantified in terms of reductions in neonatal hospital length of stay, neonatal morbidity and mortality, maternal hospital length of stay, and neonatal costs. Furthermore, the relative impact on neonatal outcomes of two interventions, namely multimodal intervention and care-management alone, is assessed.

Methods

Study design overview

An overview of the ACCORDANT simulation study is depicted in Figure 1. Subjects originate from the Multicenter Assessment of a Spontaneous Preterm Birth Risk Predictor (TREETOP) study (NCT02787213) (25). Subjects are randomized into screening and non-screening arms. Patients in the screening arm identified as higher-risk have proteomic biomarker risk predictor results, determined in the TREETOP study, that are at or above a pre-defined 15% risk of sPTB < 37 weeks of gestation. Higher-risk patients receive simulated treatment with either of two

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

intervention models (described in detail below) resulting in shifts in gestational age at birth. For higher-risk patients with shifted GA at birth, outcomes are calculated according to their new GA at birth. Unscreened and lower risk screened patients retained their respective original GA at birth observed in the TREETOP study. The two arms of the study are then compared to assess differences in neonatal and maternal length of hospital stay, neonatal morbidity and mortality index and associated neonatal costs. To estimate variability in outcomes, the study was repeated 500 times, applying variation to GA shifts in each iteration. These sources of variation and all assumptions for implementation of the study are detailed in Table 1. The routing of all subjects to each arm allows for a demographically unbiased comparison of the two arms.

Table 1 Summary of ACCORDANT assumptions.

Input	Type	Center	Spread	Distribution	References
Size of population	demographic	as observed	as observed	n/a	Markenson (25)
Composition of population	demographic	as observed	as observed	n/a	Markenson (25)
Test performance	clinical	as observed	as observed	n/a	Markenson (25)
Untreated GA at birth	clinical	as observed	as observed	n/a	Markenson (25)
Untreated neonatal length of stay	clinical	as observed	as observed	n/a	Markenson (25)
Untreated maternal length of stay	clinical	as observed	as observed	n/a	Markenson (25)
Untreated neonatal morbidity and mortality	clinical	as observed	as observed	n/a	Markenson (25)
Multimodal intervention	clinical	*max_m*=9.5 wk, *mid_m*=30.75 wk, *rate_m*=-1/2.3 *	*max_SD*=3.09 wk, *mid_SD*=30.13 wk, *rate_SD*=-1/2.9 *	3-parameter sigmoid	Branch (17)
Care-management intervention	clinical	*max_m*=5.25 wk, *mid_m*=30.75 wk, *rate_m*= -1/2.1	*max_SD*=3.09 wk, *mid_SD*=30.13 wk, *rate_SD*=-1/2.9	3-parameter sigmoid	Stankaitis (26), Newman (18), Newnham (27)

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Treated neonatal length of stay for GA<35 weeks	clinical	GA-dependent as published; 80% of mean applied to gamma, 20% to lognormal deviates	GA-dependent as published; 60% of variance applied to gamma, 40% to lognormal deviates	method of moments replicating parametric and non-parametric summary statistics	Phibbs (28)
Treated neonatal length of stay for GA ≥35 weeks	clinical	GA-dependent as published; 10% of mean applied to gamma, 90% to lognormal deviates	GA-dependent as published; 90% of variance applied to gamma, 10% to lognormal deviates	method of moments replicating parametric and non-parametric summary statistics	Phibbs (28)
Treated neonatal length of stay for GA<35 weeks	clinical	GA-dependent as published; 85% of mean applied to gamma, 15% to lognormal deviates	GA-dependent as published; 60% of variance applied to gamma, 40% to lognormal deviates	method of moments replicating parametric and non-parametric summary statistics	Beam (29)
Treated neonatal length of stay for GA ≥35 weeks	clinical	GA-dependent as published; 15% or mean applied to gamma, 85% to lognormal deviates	GA-dependent as published; 90% of variance applied to gamma, 10% to lognormal deviates	method of moments replicating parametric and non-parametric summary statistics	Beam (29)
Treated maternal length of stay for GA<35 weeks	clinical	GA-dependent as published; 20% of mean applied to gamma, 80% to lognormal deviates	GA-dependent as published; 60% of variance applied to gamma, 40% to lognormal deviates	method of moments replicating parametric and non-parametric summary statistics	Phibbs (28)
Treated maternal length of stay for GA≥35 weeks	clinical	GA-dependent as published; 20% of mean applied to gamma, 80% to lognormal deviates	GA-dependent as published; 80% of variance applied to gamma, 20% to lognormal deviates	method of moments replicating parametric and non-parametric summary statistics	Phibbs (28)
Treated neonatal morbidity and mortality	clinical	See Supplemental Table 1		multinomial	Markenson (25)

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Treated or untreated neonatal length of stay for cases of perinatal mortality	clinical	1 day longer than maximum observed	0	n/a	Branch (17)
Neonate cost for cases of perinatal mortality	economic	$89,069	$0	point estimate: average observed	Phibbs (28)
Neonate cost	economic	GA-dependent cost per day as published applied to neonatal length of stay	as for neonatal length of stay	method of moments: blend of gamma and lognormal	Phibbs (28)
Neonate cost	economic	GA-dependent cost per day as published applied to neonatal length of stay	as for neonatal length of stay	method of moments: blend of gamma and lognormal	Beam (29)
Progesterone cost	economic	$20	$0	point estimate	Fried (30), Cohen (31), Cahill (32), Kowalski (33)
Care-management cost	economic	$53	$0	point estimate	Estimated from CMS physician fee schedule for PREVENT protocol care

**max_m, max_SD* are the upper asymptotes; *mid_m, mid_SD* are the inflection points; and *rate_m, rate_SD* are the slopes of the three parameter sigmoid functions, respectively. Here *m* signifies mean, *SD* signifies standard deviation and wk represents 7 days.

The ACCORDANT analysis thus has the following components, each of which are detailed below: TREETOP study subjects randomized into two arms; application of the proteomic biomarker risk predictor to TREETOP subjects to identify those with preterm birth risk ≥15%; and estimation of outcomes/costs resulting from the above steps in the two simulated study arms (Figure 1).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

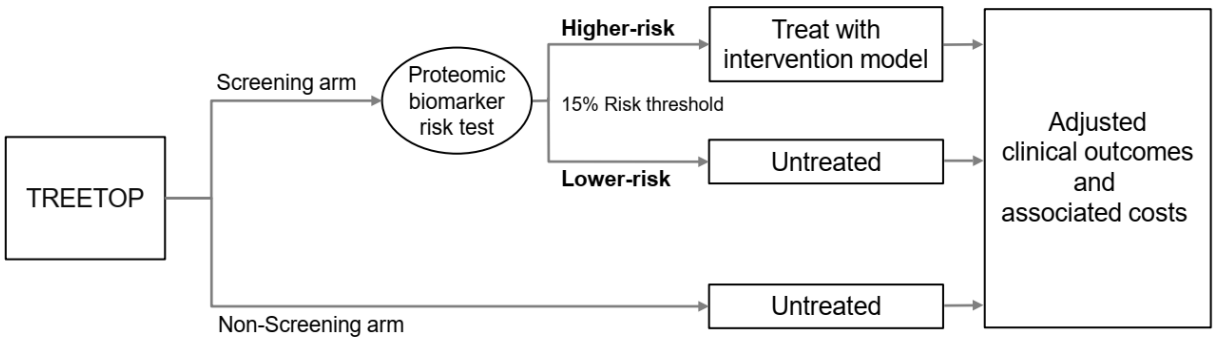

Figure 1 Pictorial overview of the ACCORDANT study. TREETOP subjects were screened prospectively using the proteomic biomarker risk test. Subjects were routed to both screening and non-screening arms. Those in the screening arm were stratified into higher and lower risk groups based on their proteomic biomarker risk scores. The GAs at birth of higher-risk subjects were shifted according to published intervention efficacies. Associated outcomes were calculated for the treated group and for subjects with missing or truncated data in both arms. Outcomes were compared between the screening and non-screening arms, as a whole, and between the most severely affected 10% of each arm by outcome.

TREETOP Study subjects consist of a randomly selected subset (14 sites, n=847) of 34% of all TREETOP subjects who underwent serum sample collection in the protein ratio's validated GA window (19 $^{1/7}$ – 20 $^{6/7}$ weeks) as reported in Markenson, et al. (25). These subjects were randomly selected by a third-party statistician for analysis and reflect the clinical and demographic factors of the overall study. We note that the non-selected subjects in TREETOP remain blinded for future studies. All TREETOP clinical data were recorded electronically, monitored centrally on site and were subject to source document verification as pre-specified. Neonatal outcomes were collected through 28 days of life.

Application of the proteomic biomarker risk predictor

The proteomic biomarker measurements were assayed on all TREETOP serum samples as they were collected, using a standardized lab process previously validated and documented (11),

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

consistent with clinical intended use. A previously validated probability threshold (34) was used to assign higher and lower preterm risk. As per Saade et al. (9), proteomic measurements were translated into risk predictions for each subject and a threshold corresponding to 15% risk of sPTB < 37 weeks of gestation was employed to stratify higher vs. lower risk in each pregnancy. Higher-risk subjects were routed for intervention (see intervention models below).

The intervention models

Intervention models provided the means and standard deviations of shifts in GA at birth with treatment relative to a subject's untreated GA at birth. Specifically, for each treated subject in TREETOP, her original GA at birth is shifted by an amount selected randomly from the normal distribution defined by the mean and standard deviation of the change for the GA at birth with treatment that is expected for pregnancies with the untreated GA at birth observed for that subject in TREETOP. Two intervention models are assessed in this analysis: multimodal (17) and care-management (18, 26, 27).

The multimodal intervention model is based on the observed effect size of the combination of 17-α-hydroxyprogesterone caproate and care-management applied to higher-risk pregnancies as determined by the proteomic biomarker risk predictor in the Prediction and Prevention of Preterm Birth (PREVENT-PTB, NCT03530332) randomized control trial (17).

The care-management intervention model is based on the combination of three controlled studies (18, 26, 27) of the effect of care-management applied to higher-risk pregnancies as determined by clinical and demographic data. The shifts in GA at birth for the care-

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

management intervention model, as indexed by untreated GA at birth, were derived using the metafor R package (35).

Intervention effects were formalized by the following two functions defining means and standard deviations, respectively, of shift in GA at birth. For each observation x, where x is the untreated GA at birth in weeks, calculation using the functions shown establishes sigmoid curves relating untreated GA at birth to the magnitude (mean) and variability (standard deviation) of shift in GA at birth with intervention:

$$shift_in_GA_at_birth_mean(x) = \frac{max_m}{1 + e^{-(mid_m - x) * rate_m}}$$

and

$$shift_in_GA_at_birth_stdev(x) = \frac{max_SD}{1 + e^{-(mid_SD - x) * rate_SD}} \quad ,$$

where max_m, max_SD are the upper asymptotes; mid_m, mid_SD are the inflection points; and $rate_m$, $rate_SD$ are the slopes of the three parameter sigmoid functions, respectively. Here m signifies mean, SD signifies standard deviation. Table 1 specifies the numerical values for coefficients used in the two functions above for calculating the means and standard deviations of shifts in GA at birth representing the multimodal and care-management intervention models. For illustration, magnitudes of shift in GA at birth with intervention are presented in Table 2 for weeks 28 to 36.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Table 2 Magnitude of shift in gestational age (GA) at birth with modeled interventions for each GA week.

	GA at birth (week)								
Intervention model	28	29	30	31	32	33	34	35	36
Multimodal intervention: magnitude of shift in GA at birth (weeks)	7.3	6.5	5.5	4.5	3.5	2.6	1.9	1.3	0.9
Care-management intervention: magnitude of shift in GA at birth (weeks)	4.1	3.7	3.1	2.5	1.9	1.3	0.9	0.6	0.4

Derivation of outcomes and costs

We describe the derivation of neonatal and maternal length of stay, neonatal morbidity and mortality, and cost for each subject in the screening and non-screening arm of this analysis. For all subjects, these four outcomes are either within error of the values observed in the TREETOP study if unaffected by intervention, mortality or missingness; or are assigned as a function of GA at birth.

Derivation of neonatal and maternal length of hospital stay

Information from two datasets, Beam (29) and Phibbs (28), was used as the source of post-intervention lengths of stay, providing generalizability as these datasets include over 750,000 and 1,000,000 pregnancies, respectively. In addition, because TREETOP truncated collection of neonatal length of stay information for each subject at 28 days with information missing for some subjects, missing or truncated TREETOP neonatal length of stay information was

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

augmented using information from Beam or Phibbs. Missing/truncated values and values altered by interventions were calculated by randomized sampling from these datasets based on distributions defined by the published mean and variance estimates using the method of moments and confirmed by nonparametric estimates also supplied and by distributions in TREETOP. The published neonatal length of stay data and thus also the derived distributions were indexed by GA at birth, allowing for unbiased assignment of neonatal length of stay to each treated and untreated subject in the study based on subject GA at birth. Based on the observation in Beam that late preterm distributions are differently shaped than early preterm distributions, we applied distinct distributions to early and late preterm GAs at birth (see Table 1). Length of stay error with respect to TREETOP observations was set at ± 1 day. Derivation of maternal length of stay was conducted in the same manner except only using the Phibbs dataset (maternal length of stay was not available in the Beam dataset).

Derivation of neonatal morbidity and mortality

The assessment of neonatal morbidity and mortality was based on an established scoring system (6). Affected infants were assigned a morbidity and mortality index score that increases from 1 (mild) to 2 (moderate) to 3 (severe) for each additional diagnosis of respiratory distress syndrome, bronchopulmonary dysplasia, intraventricular hemorrhage grade III or IV, all stages of necrotizing enterocolitis, periventricular leukomalacia or proven severe sepsis; with a score of 4 assigned to perinatal mortality. The scale uses hospital stays to determine index scores if the length of stay gives a higher score than concomitant diagnoses: 1-4 days give a score of 1, 5-20 days a score of 2 and >20 days a score of 3. For example, level 1 includes those neonates

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

with at most one adverse outcome and up to four days of hospital stay. While the published scale accounts for neonatal intensive care unit stays, here total length of neonatal hospital stay is used instead to leverage the universality of hospital admission and greater simplicity of calculation of hospital length of stay vs. level of care. Subjects in this analysis are assigned an initial neonatal morbidity and mortality index based on TREETOP observations. If the subject's GA at birth changes, the neonatal morbidity and mortality score for that subject is updated probabilistically in accordance with multinomial distributions indexed by GA at birth, shown in Supplemental Table 1. As the neonatal morbidity and mortality index is an ordinal scale without numeric error, modeled neonatal morbidity and mortality was restricted to the original observation in untreated subjects and to values no greater than the original observation in treated subjects.

Estimation of neonate cost

Neonate costs were not collected in the TREETOP study and are wholly derived from the Phibbs and Beam cost datasets for each subject in this analysis in either arm of the study. Costs in these datasets are indexed by GA at birth, as is neonatal length of stay, and so can be derived from neonatal length of stay as a mean cost per day of stay. To maintain the known tight association of cost and length of stay and to utilize the variance already incorporated into the neonatal length of stay calculation, the neonate cost is obtained by first determining the neonatal length of stay for that subject (as defined above) and then determining the cost from the conversion factors indexed by GA at birth in the Phibbs and Beam datasets, respectively.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Those subjects who underwent intervention were also assigned an additional cost from weeks 22 to 35 for progesterone treatment and/or care-management (see Table 1 for details).

Comparison of study arms and statistical methods

Statistical analyses were performed in R (3.5.1 or higher, MRAN). As this study is exploratory as per ISPOR (Professional Society for Health Economics and Outcomes Research) classification, p-values <0.05 were considered significant.

Changes in neonatal and maternal lengths of stay due to treatment were assessed by time-to-event analysis. While multiple approaches give similar results, we report Cox proportional hazards regression p-values for the main effect of treatment arm, either in the study as a whole or in the top 10% of subjects per arm.

Significance of changes in the number of subjects in each neonatal morbidity and mortality level was assessed using a one-sided Fisher's Exact test pre-specified for reduction in higher levels of neonatal morbidity and mortality.

As cost data are not normally distributed (cost exponentially increases as GA at birth decreases) and no changes are expected in the large majority of babies who are born healthy and at term, changes in cost with treatment were assessed by a bootstrap test of total costs across the 500 iterations, either in each arm as a whole or in the top 10% of neonates per arm by cost.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Results

The sub-cohort of the TREETOP population used in this study is described in Markenson et al. (25), and consisted of 847 subjects enrolled from 14 sites in the United States. Of the 847 subjects, 36% were identified as higher-risk (≥15%) of sPTB < 37 weeks of gestation by the proteomic biomarker risk predictor. Here we summarize the impact of two modeled interventions (multimodal, care-management) on important clinical and economic outcomes on this cohort of 847 pregnancies as selected by the proteomic biomarker risk predictor. Application of the intervention models to the higher-risk subjects in the screening arm resulted in an average prolongation of gestation in the arm as a whole of under a day (0.8 multimodal / 0.3 care-management). However, as this average prolongation of gestation is dominated by term deliveries, we also examined the average prolongation of gestation modeled for the cohort of pregnancies in the bottom 10% of GA at birth in each arm. For this cohort the average prolongation of gestation is 8 days with the multimodal intervention and 3 days with the care-management intervention. We then examined the impact of these GA shifts on neonatal length of hospital stay, maternal length of hospital stay, neonatal morbidity and mortality, and neonate costs in the TREETOP population.

Tables 3 and 4 detail the impact of the two intervention models on the above outcomes for the TREETOP population. Across all subjects, using the Phibbs and Beam datasets, neonatal length of stay was reduced by 26% and 30% (multimodal) and by 19% and 22% (care-management), respectively. For subjects in the top 10% of neonatal length of hospital stay, neonatal length of stay was reduced by 46% and 47% (multimodal) and by 33% and 34% (care-management),

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

respectively for the two datasets. In all cases, regardless of intervention model and outcome database used, the reductions were significant (Table 3).

Table 3 Intervention impact on neonatal length of stay, maternal length of stay and cost.

Outcome	Intervention model	Population	Mean control	Mean screened	% Reduction	Mean difference	p-value
Phibbs neonatal length of stay	multimodal	all	3.9	2.9	26%	1.0	**0.0018**
Beam neonatal length of stay	multimodal	all	3.6	2.5	30%	1.1	**<0.001**
Phibbs maternal length of stay	multimodal	all	2.7	2.5	9.2%	0.25	**<0.001**
Phibbs cost	multimodal	all	$5,300	$3,500	34%	$1,800	**0.018**
Beam cost	multimodal	all	$12,000	$8,800	26%	$3,200	**0.0030**
Phibbs neonatal length of stay	care-management	all	3.9	3.2	19%	0.74	**0.029**
Beam neonatal length of stay	care-management	all	3.6	2.8	22%	0.79	**0.0070**
Phibbs maternal length of stay	care-management	all	2.7	2.5	8.5%	0.23	**0.0010**
Phibbs cost	care-management	all	$5,300	$4,400	16%	$880	0.20
Beam cost	care-management	all	$12,000	$10, 000	16%	$1,900	0.098
Phibbs neonatal length of stay	multimodal	top 10%	20	11	46%	9.0	**<0.001**
Beam neonatal length of stay	multimodal	top 10%	19	10	47%	9.0	**<0.001**
Phibbs maternal length of stay	multimodal	top 10%	5.6	4.6	17%	0.97	**0.0011**
Phibbs cost	multimodal	top 10%	$38,000	$18,000	52%	$20,000	**0.0040**
Beam cost	multimodal	top 10%	$66,000	$36,000	44%	$29,000	**0.0010**
Phibbs neonatal length of stay	care-management	top 10%	20	13	33%	6.5	**0.024**
Beam neonatal length of stay	care-management	top 10%	19	13	34%	6.6	**0.020**

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Phibbs maternal length of stay	care-management	top 10%	5.6	4.7	16%	0.88	**0.0041**
Phibbs cost	care-management	top 10%	$38,000	$28,000	27%	$10,000	0.13
Beam cost	care-management	top 10%	$66,000	$49,000	25%	$17,000	0.089

'Phibbs' and 'Beam' denote the outcome/cost database used for each outcome. The population assessed is either all of TREETOP ('all') or those in the top 10% of each measure ('top 10%'). Statistical tests used for neonatal length of stay, maternal length of stay and cost are, respectively, Cox proportional hazards, Cox proportional hazards and bootstrap, with p-values <0.05 in bold.

Across all subjects, maternal length of stay was reduced 9% (multimodal and care-management) using the Phibbs outcome database. For subjects in the top 10% of neonatal length of hospital stay, maternal length of stay was reduced 17% (multimodal) and 16% (care-management). In all cases, regardless of intervention model, the reductions were significant (Table 3).

Across all subjects, neonatal costs were reduced by 34% and 26% with savings of $1,800 and $3,200 per pregnancy screened (multimodal) and 16% with a savings of $880 and $1,900 per pregnancy screened (care-management), respectively for the Phibbs and Beam databases. The two models differed in cost reduction with the Beam (commercial payer data) cost reductions being approximately double the Phibbs cost reduction. Reductions of neonatal costs by the multimodal intervention were significant (Table 3).

Change in the neonatal morbidity and mortality index upon intervention was significant in reduction of the proportion of neonates with scores at or above level 2 (multiple negative outcomes and/or extended hospital stay) versus level 1 (at most 1 negative outcome and up to 4 days hospital stay), for both intervention models. Reductions in the proportion of neonatal

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

morbidity and mortality level 2 or higher were 41% (multimodal) and 29% (care-management) respectively (Table 4).

Table 4 Intervention impact on neonatal morbidity and mortality.

Intervention model	Subset	Level 1 (mild)	Level 2 (moderate)	Level 3 (severe)	Level 4 (death)	p-value (1 vs. >1)	p-value (1-2 vs. >2)	p-value (1-3 vs. 4)	Proportion moderate+	% Reduction in moderate+
Control	Control	90.6%	6.5%	2.5%	0.5%				9.4%	
Multimodal	Screening	94.5%	4.1%	1.4%	0.0%	**0.0015**	**0.035**	0.19	5.5%	41%
Care-management	Screening	93.3%	4.7%	1.9%	0.1%	**0.025**	0.14	0.19	6.7%	29%

Comparisons are made among the four levels of the neonatal morbidity and mortality index severity between the control arm and the screening arm where the intervention model is applied. Fisher's Exact test is used to generate a significance estimate, with p-values <0.05 in bold. The percent reduction in levels 2 or higher is determined for each intervention.

Discussion

Preterm birth remains the most compelling issue in obstetrics and neonatology, creating adverse short and long-term outcomes for newborns. In efforts at identification and treatment of at-risk mothers-to-be with early interventions (progesterone, cerclage, and others) and acute therapies in symptomatic patients (tocolytics, corticosteroids, magnesium sulfate for neuroprotection and antibiotics), acute maternal administration of antenatal corticosteroids

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

and magnesium sulfate have been demonstrated to improve outcomes (36). The overall frequency of premature delivery has not been materially reduced in the US. The strategy most used currently for lowering the rate of premature delivery with resultant decreases in serious neonatal morbidities is identification of specific risk factors or specific higher-risk groups and application to these women of approaches that make sense for the specific etiology associated with their particular risk factor, e.g., cerclage for cervical shortening or premature cervical dilations. For small numbers of patients with specific risk factors an approach such as progesterone therapy for previous preterm delivery or short cervix on endovaginal ultrasound are further examples (20, 21). However, this approach is limited by the fact that such risk factors only capture a small percentage of the patients who will ultimately deliver prematurely. This approach of risk identification therefore would hold promise if we could identify a much larger proportion of those destined to deliver prematurely and apply an intervention or combination of interventions appropriate for the group chosen. There is substantial reason to believe that a benefit would be seen.

In this paper we assessed the clinical effectiveness and economic benefit of using a proteomic biomarker risk predictor for identifying higher-risk patients in the TREETOP study for treatment with established interventions for preterm birth including the use of medications and proactive care-management. Our innovative analytical approach combines real-world observational data with simulation of prolonged gestation based on published treatment efficacy to provide real-world effectiveness estimates. As reduction in adverse neonatal outcomes is in reality the compelling reason for delaying the GA at which the baby is born, we chose several critical metrics to determine whether the modeled studies resulted in improved neonatal outcomes.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

The results projected impressive potential improvements in neonatal morbidity and mortality, neonatal length of stay and in neonatal cost of care, which is also a surrogate for decreased morbidities requiring extended neonatal intensive care hospitalization. Encouragingly, neonatal improvements occurred without increasing, but rather decreasing maternal length of stay.

While one cannot necessarily assume efficacy of progesterone treatment in a broader population, and especially in patients who are delivered prematurely for medical or obstetrical indications, as it has in reality only been demonstrated to be effective in women with previous preterm deliveries by some studies in those with a short cervix on ultrasound, a stronger assertion can be made of the potential benefit for care-management. Several studies in diverse populations have demonstrated reductions in prematurity, shortened durations of NICU stay and/or improvements in composite neonatal morbidity with such programs (18, 19, 26, 27). And since all of these were in were in populations with diverse reasons for identification of increased risk for prematurity, their combined findings lend support that a marker which is not specific for any particular etiology or particular risk of prematurity may be similarly effective on all.

The results of this exercise are important, demonstrating not only the potential benefits for mothers and babies but also the utility of effectiveness assessments combining real-world data and published evidence. Furthermore, this study provides confidence that future studies and real-world use of test-and-treat strategies are worthwhile, ethically justified for pregnant women and are likely to demonstrate positive results. As well, this modeling exercise provides a compelling reason to proceed with well-designed and controlled prospective studies of progesterone and/or care-management in patients identified as higher-risk with such a well

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

demonstrated, sensitive protein marker of risk for prematurity. Finally, these results are based on interventions that are currently used in practice, known to be safe and generally found to be acceptable to pregnant women (24, 26, 27, 36). Thus, there is a high potential for the proteomic biomarker risk predictor to be a clinically important component of risk stratification for pregnant women that could lead to tangible gains in reducing the impact of preterm birth.

Conflict of Interest

AF, AP, TF, GC, JB, JJB, TG and PK are stockholders, employees, or contract employees of Sera Prognostics. All other authors report no conflict of interest.

Availability of Data and Materials

Data supporting the results presented here can be requested at data-sharing@seraprognostics.com. Data will not be made publicly available, or in any format, that may violate a subject's right to privacy. For example, dating information or identifiers that would allow data to be integrated, thereby enabling the potential identification of study subjects, are protected.

Funding

Sponsorship (funding, study design and execution) of this work was provided by Sera Prognostics, Inc.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Author Contributions

J.B., G.C.C., J.J.B, and P.E.K conceptualized the study. G.R.M., G.R.S., L.C.L, K.D.H., D.V.C, C.N.S, J.K.B, D.M.H, S.L., S.A.S., C.A.M, S.M.W., L.M.P., E.J.S., K.A.B., A.F.H., and A.H.C provided subjects, clinical data, and serum samples for analyses. A.C.F curated data. Formal analyses were done by J.B. (including validation), A.D.P, and P.E.K., with methodology designed by J.B, J.J.B, and P.E.K. Visualization/data presentation and/or manuscript draft preparation was done by J.B., T.C.F., T.J.G. G.C.C., J.J.B., and P.E.K. The study was supervised by J.J.B. and P.E.K., with P.E.K. serving as project administrator. All authors reviewed, edited, and approved the manuscript.

Acknowledgements

We would like to acknowledge the study coordinators and research personnel at the study sites. We would also like to acknowledge the Sera clinical laboratory and clinical operations teams.

References

1. Blencowe H, Vos T, Lee AC, Philips R, Lozano R, Alvarado MR, et al. Estimates of neonatal morbidities and disabilities at regional and global levels for 2010: introduction, methods overview, and relevant findings from the Global Burden of Disease study. Pediatr Res. 2013;74 Suppl 1:4-16.
2. Martin JA, Hamilton BE, Osterman MJK, Driscoll AK, Drake P. Births: Final Data for 2017. Natl Vital Stat Rep. 2018;67(8):1-50.
3. Waitzman NJ, Jalali A, Grosse SD. Preterm birth lifetime costs in the United States in 2016: An update. Semin Perinatol. 2021:151390.
4. Laughon SK, Albert PS, Leishear K, Mendola P. The NICHD Consecutive Pregnancies Study: recurrent preterm delivery by subtype. Am J Obstet Gynecol. 2014;210(2):131 e1-8.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

5. Petrini JR, Callaghan WM, Klebanoff M, Green NS, Lackritz EM, Howse JL, et al. Estimated effect of 17 alpha-hydroxyprogesterone caproate on preterm birth in the United States. Obstet Gynecol. 2005;105(2):267-72.
6. Hassan SS, Romero R, Vidyadhari D, Fusey S, Baxter JK, Khandelwal M, et al. Vaginal progesterone reduces the rate of preterm birth in women with a sonographic short cervix: a multicenter, randomized, double-blind, placebo-controlled trial. Ultrasound Obstet Gynecol. 2011;38(1):18-31.
7. McIntosh J, Feltovich H, Berghella V, Manuck T. The role of routine cervical length screening in selected high- and low-risk women for preterm birth prevention. Am J Obstet Gynecol. 2016;215(3):B2-7.
8. Gyamfi-Bannerman C, Ananth CV. Trends in spontaneous and indicated preterm delivery among singleton gestations in the United States, 2005-2012. Obstet Gynecol. 2014;124(6):1069-74.
9. Saade GR, Boggess KA, Sullivan SA, Markenson GR, Iams JD, Coonrod DV, et al. Development and validation of a spontaneous preterm delivery predictor in asymptomatic women. Am J Obstet Gynecol. 2016;214(5):633.e1 - .e24.
10. Boggess K, Saade GR, Sullivan SA, Markenson GR, Iams JD, Coonrod DV, et al. 337: Use of a second trimester serum-based proteomic risk classifier for prediction of spontaneous and medically-indicated preterm birth. American Journal of Obstetrics & Gynecology. 2017;216(1):S204-S5.
11. Bradford C, Severinsen R, Pugmire T, Rasmussen M, Stoddard K, Uemura Y, et al. Analytical validation of protein biomarkers for risk of spontaneous preterm birth. Clinical Mass Spectrometry. 2017;3:25-38.
12. Micheel C, Nass SJ, Omenn GS, Institute of Medicine (U.S.). Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials. Evolution of translational omics : lessons learned and the path forward. Washington, D.C.: National Academies Press,; 2012. Available from: National Academies Press http://www.nap.edu/catalog.php?record_id=13297.
13. Institute of Medicine. Knowing What Works in Health Care: A Roadmap for the Nation. Eden J, Wheatley B, McNeil B, Sox H, editors. Washington, DC: The National Academies Press; 2008. 278 p.
14. Price D, Bateman ED, Chisholm A, Papadopoulos NG, Bosnic-Anticevich S, Pizzichini E, et al. Complementing the randomized controlled trial evidence base. Evolution not revolution. Annals of the American Thoracic Society. 2014;11 Suppl 2:S92-8.
15. Duddy A, Franke K, Whicher DM, Moloney RM, Karkare SU. Making Informed Decisions: Assessing the Strengths and Weaknesses of Study Designs and Analytic Methods for Comparative Effectiveness Research. A Briefing Document for Stakeholders.

Center for Medical Technology Policy (CMTP); 2012 February 2012.
16. Sun F, Bruening W, Uhl S, Ballard R, Tipton K, Schoelles K. AHRQ Technology Assessments. Quality, Regulation and Clinical Utility of Laboratory-developed Molecular Tests. Rockville (MD): Agency for Healthcare Research and Quality (US); 2010.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

17. Branch DW, VanBuren JM, Porter TF, Holmgren C, Holubkov R, Page K, et al. Prediction and Prevention of Preterm Birth: A Prospective, Randomized Intervention Trial. Am J Perinatol. 2021(EFirst).
18. Newman RB, Sullivan SA, Menard MK, Rittenberg CS, Rowland AK, Korte JE, et al. South Carolina Partners for Preterm Birth Prevention: a regional perinatal initiative for the reduction of premature birth in a Medicaid population. Am J Obstet Gynecol. 2008;199(4):393 e1-8.
19. Manuck TA, Henry E, Gibson J, Varner MW, Porter TF, Jackson GM, et al. Pregnancy outcomes in a recurrent preterm birth prevention clinic. Am J Obstet Gynecol. 2011;204(4):320 e1-6.
20. Romero R, Conde-Agudelo A, Da Fonseca E, O'Brien JM, Cetingoz E, Creasy GW, et al. Vaginal progesterone for preventing preterm birth and adverse perinatal outcomes in singleton gestations with a short cervix: a meta-analysis of individual patient data. Am J Obstet Gynecol. 2018;218(2):161-80.
21. Meis PJ, Klebanoff M, Thom E, Dombrowski MP, Sibai B, Moawad AH, et al. Prevention of recurrent preterm delivery by 17 alpha-hydroxyprogesterone caproate. N Engl J Med. 2003;348(24):2379-85.
22. Grobman WA, Thom EA, Spong CY, Iams JD, Saade GR, Mercer BM, et al. 17 alpha-hydroxyprogesterone caproate to prevent prematurity in nulliparas with cervical length less than 30 mm. Am J Obstet Gynecol. 2012;207(5):390 e1-8.
23. Blackwell SC, Gyamfi-Bannerman C, Biggio JR, Jr., Chauhan SP, Hughes BL, Louis JM, et al. 17-OHPC to Prevent Recurrent Preterm Birth in Singleton Gestations (PROLONG Study): A Multicenter, International, Randomized Double-Blind Trial. Am J Perinatol. 2020;37(2):127-36.
24. Hillemeier MM, Domino ME, Wells R, Goyal RK, Kum HC, Cilenti D, et al. Effects of maternity care coordination on pregnancy outcomes: propensity-weighted analyses. Matern Child Health J. 2015;19(1):121-7.
25. Markenson GR, Saade GR, Laurent LC, Heyborne KD, Coonrod DV, Schoen CN, et al. Performance of a proteomic preterm delivery predictor in a large independent prospective cohort. American journal of obstetrics & gynecology MFM. 2020;2(3):100140.
26. Stankaitis JA, Brill HR, Walker DM. Reduction in neonatal intensive care unit admission rates in a Medicaid managed care program. Am J Manag Care. 2005;11(3):166-72.
27. Newnham JP, White SW, Meharry S, Lee HS, Pedretti MK, Arrese CA, et al. Reducing preterm birth by a statewide multifaceted program: an implementation study. Am J Obstet Gynecol. 2017;216(5):434-42.
28. Phibbs CS, Schmitt SK, Cooper M, Gould JB, Lee HC, Profit J, et al. Birth Hospitalization Costs and Days of Care for Mothers and Neonates in California, 2009-2011. J Pediatr. 2019;204:118-25.e14.
29. Beam AL, Fried I, Palmer N, Agniel D, Brat G, Fox K, et al. Estimates of healthcare spending for preterm and low-birthweight infants in a commercially insured population: 2008-2016. J Perinatol. 2020;40(7):1091-9.
30. Fried I, Beam AL, Kohane IS, Palmer NP. Utilization, Cost, and Outcome of Branded vs Compounded 17-Alpha Hydroxyprogesterone Caproate in Prevention of Preterm Birth. JAMA Intern Med. 2017;177(11):1689-90.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.08.21262940; this version posted September 14, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

31. Cohen AW, Copel JA, Macones GA, Menard MK, Riley L, Saade GR. Unjustified increase in cost of care resulting from U.S. Food and Drug Administration approval of Makena (17alpha-hydroxyprogesterone caproate). Obstet Gynecol. 2011;117(6):1408-12.
32. Cahill AG, Odibo AO, Caughey AB, Stamilio DM, Hassan SS, Macones GA, et al. Universal cervical length screening and treatment with vaginal progesterone to prevent preterm birth: a decision and economic analysis. Am J Obstet Gynecol. 2010;202(6):548 e1-8.
33. Kowalski JT. A Guide to the Cost of Progesterone for Prevention of Preterm Labor. Proc Obstet Gynecol. 2011;1(3):Article 9 [4 p.].
34. Burchard J, Polpitiya AD, Fox AC, Randolph T, Fleischer TC, Dufford M, et al. Clinical Validation of a Proteomic Biomarker Threshold for Increased Risk of Spontaneous Preterm Birth. medRxiv. 2021:2021.01.23.21249902.
35. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. 2010. 2010;36(3):48.
36. Garite TJ, Combs CA. Obstetric interventions beneficial to prematurely delivering newborn babies: antenatal corticostetroids, progesterone, magnesium sulfate. Clin Perinatol. 2012;39(1):33-45.