PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A metabolic clock as noninvasive blood tests of preterm birth and for gestational age assessment: a two-center retrospective study in the US
AUTHORS	Sylvester, Karl; Hao, Shiying; You, Jin; Zheng, Le; Tian, Lu; Yao, Xiaoming; Mo, Lihong; Ladella, Subhashini; Wong, Ronald; Shaw, Gary M.; Stevenson, David; Cohen, Harvet; Whitin, John; McElhinney, Doff; Ling, Xuefeng

VERSION 1 – REVIEW

REVIEWER	Lina Youssef BCNatal | Fetal Medicine Research Center (Hospital Clínic and Hospital Sant Joan de Déu), Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain.
REVIEW RETURNED	11-Jul-2020

GENERAL COMMENTS	Comments to the authors: The study aimed to identify a model based on metabolic pathway to determine gestational age and pick up pregnancies at high risk for preterm birth. The analysis is adequate and the manuscript is well written. The developed model has been validated in a distinct cohort. The main limitation of the study is its retrospective design with a modest number of patients and the collection of maternal blood samples at different timepoints of gestation in addition to non-considering the gestational age and other baseline characteristics in the prediction model of preterm birth. Abstract: -Lines 46: the results of both the development and the validation model should be stated adding also the specificity of the test. Introduction: -The authors mentioned in lines 70-72: “First trimester US imaging is the gold standard for GA determination, however there can be frequent discordance between US dating and a mother’s last known menstrual period (LMP). In these cases, follow-up testing by US is utilized to more accurately estimate GA.” Could they provide the reference for serial imaging in case of discordance? Most centers rely on first trimester US and omit LMP. -The authors mentioned in lines 72-73: “US measurements are not currently used to determine risk of premature birth (PTB)”. Some centers started to apply universal screening for preterm birth by measuring cervical length at the second trimester anatomy scan (18-24 weeks). Could the authors comment on this with an appropriate reference.

	-Line 85: The authors mentioned that metabolites were not used previously to determine GA. Please check this recent reference: Liang et al. Cell 181,71 Pages 1680-1692.e15 (https://doi.org/10.1016/j.cell.2020.05.002) -Lines 93-101 are better to be moved to the discussion section. Methods: -Lina 105: "PTB was defined by delivery at < 35 weeks GA." From the clinical point of view, PTB is defined as delivery <37 w, why did authors choose this definition? May be to exclude late preterm deliveries since they tend to be milder cases. -Line 109: the gold standard was the US measurement. Was it the first trimester US based on the crown-rump length? Please specify with an appropriate reference. -Line 116: Please include the IRB approval numbers. -Lines 146-148: "Results were compared with the insulin-like growth factor-binding protein 4 (IBP4)/sex hormone-binding globulin (SHBG) signature that is commercially available as a metabolic test for determining risk of PTB". This analysis has been done on another cohort according to Ext A.2. Please mention this in the main text with a brief description of the ELISA methods. Results: -Table 1: there is an error in the black race percentage in UAB cohort (it is now 6.9% □ it should be 76.9%) -There are differences in the baseline characteristics between term and PTB pregnancies. Have the authors adjusted the results for these differences? (if not, this should be added in the limitations section). -Line 201: "Samples collected before 35 weeks' GA were used to develop a model that differentiated PTB pregnancies from those full-term". Since the metabolites profile is correlated with GA, why do the authors choose this wide range of gestational weeks? Have they developed different algorithms for each pregnancy trimester for instance or a multivariate model including the GA and other variables like race, BMI…? Discussion: -Line 255: "There is a need to develop a more robust method than LMP and US that captures pregnancy progression". First trimester US is a very reliable method to determine GA and is the standard of care whenever it is available. This statement should be modified to (or something similar): "There is a need to develop a more robust method than LMP and an alternative to first trimester US that captures pregnancy progression" -What are the pathways differentially expressed in PTB? Could the authors share their insights regarding these pathways to enhance the understanding of PTB pathophysiology?

REVIEWER	Jeffrey Murray Dept of Pediatrics University of Iowa USA I have collaborated with, published with and managed grant support for this group previously but not related to this specific manuscript.
REVIEW RETURNED	18-Sep-2020

GENERAL COMMENTS	This is a first proof of principle publication establishing the feasibility of using metabolic testing of maternal serum during

	pregnancy to both establish gestational age or identify a risk profile for likely preterm birth. The study is timely, well conceived and the report largely well written and discussed. It has the potential to be a first step in both developing a better understanding of the underlying risks for PTB and to have clinical utility (potentially) in both high and low resource settings. There are a few concerns: In table 1 the nos. in () seem to refer to percentages but dont make sense for the numbe inr the UAB column in some places (10 preterm is 6% and 2 is 15% for ex). It might also be useful in table 1 to have a col of p values for difference between the two sites (things like primipara status, age of mo, hx of ptb, race seem quite different in addition the demographic data and descriptions of the populations is underreported. Medical conditions of the mother, were all PTB spontaneous, were twins excluded etc should all be noted either in text or table and some discussion of these potentially confounding factors included. efining “preterm” as before 35 weeks is non standard and while it does have clinical relevance will require substantial explanation if used for population/surveillance estimates and likely affects the outcomes reported here. In addition, the absence of the 35 and 36 week pregnancies and the very distorted numbers of PTB pregnancies compared to population numbers will need to be address in future population based studies. IBP4 and SHBG are referred to as “metabolic” markers but were proteomic from the report and it was a little unclear as to the utility of the comparison of their PTB predictor to that of the SHBG/IBP4 outside the narrow range for which the latter is recommended as it has not been validated outside a narrow window (18 to 20 weeks as noted). The prevalence corrected PPV values used the national PTB prevalence of 9.7% but its unclear why this would be better than using the local population value which might also be substantially different between Stanford and UAB and how might this affect results.Also as a side point all the positive predictive values might also be viewed in the context that just guessing that a random pregnancy will be term will be correct 90% of the time. In the replication study there were a substantial number of women with a prior PTB. Is it known if they received progesterone as a preventative as has been common practice in some centers and if so did this have any effect on the metabolic profiles? similarly were any of the women diabetic, on antihypertensive medications or with other complications of pregnancy and if so could they see any effect on values there (recognizing that the number would be quite small). They might also support their investigation with the extensive literature showing that newborn metabolic profiles using targeted metabolites are also highly predictive of GA and PTB. They state the data is available upon request. rather once published the data should be freely available publicly. Finally there is a missed opportunity to go a bit deeper in the discussion as to how this might be used both scientifically (to better understand the drivers of developmental changes that occur

	in pregnancy) and more importantly to outline better how this might be used clinically. For example they briefly propose that this could be used in places where US is unavailable but US is becoming far more widespread and has the added utility of identifying many pregnancy complications of immediate clinical concern (twins, placenta praevia, breech etc) and the cost/processing of the maternal serum collection could also be a challenge. Similarly might there be alternatives such as maternal urine that could also work and be easier to collect. So some mention of these issues to guide future thinking about their findings could help move the field quickly both scientifically and practically. All these concerns are relatively minor for this initial publication which is an important contribution to developing this a a method with substantial promise. The authors have provided substantial detail in the appendix, done a thorough job of noting not only its many strengths but also its limitations and it is written in a clear and accessible style.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Lina Youssef

Institution and Country: BCNatal | Fetal Medicine Research Center (Hospital Clínic and Hospital Sant Joan de Déu), Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain.

Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below

Comments to the authors:

The study aimed to identify a model based on metabolic pathway to determine gestational age and pick up pregnancies at high risk for preterm birth. The analysis is adequate and the manuscript is well written. The developed model has been validated in a distinct cohort. The main limitation of the study is its retrospective design with a modest number of patients and the collection of maternal blood samples at different timepoints of gestation in addition to non-considering the gestational age and other baseline characteristics in the prediction model of preterm birth.

Abstract:

-Lines 46: the results of both the development and the validation model should be stated adding also the specificity of the test.

Response: The abstract was revised by adding the development results and the specificities.

Introduction:

-The authors mentioned in lines 70-72: “First trimester US imaging is the gold standard for GA determination, however there can be frequent discordance between US dating and a mother’s last known menstrual period (LMP). In these cases, follow-up testing by US is utilized to more accurately estimate GA.” Could they provide the reference for serial imaging in case of discordance? Most centers rely on first trimester US and omit LMP.

Response: The text was revised to clarify to reflect current specialty society opinions. We have replaced lines 70-72 with the following sentences:

“According to the policy statement of the Committee on Obstetric Practice, the American Institute of Ultrasound in Medicine, and the Society for Maternal-Fetal Medicine, a pregnancy is considered optimally dated through a combination of last menstrual period (LMP) and an accurate US obtained prior to 22 0/7 weeks. Accordingly, LMP is dependent on maternal recall and many pregnancies do not present for a first prenatal US evaluation until the second or third trimester. Thus, there is a need for a molecular method that would complement the potential shortcomings of LMP recall and US dating outside the first trimester. Moreover, it is possible that molecular pregnancy dating will provide greater resolution to pregnancy risk then current information based on calendar dating (LMP) and anthropometrics (US).

-The authors mentioned in lines 72-73: “US measurements are not currently used to determine risk of premature birth (PTB)”. Some centers started to apply universal screening for preterm birth by measuring cervical length at the second trimester anatomy scan (18-24 weeks). Could the authors comment on this with an appropriate reference.

Response: We have revised lines 72-73 as follows: “Although experience is accumulating with the use of second and third trimester US for an estimation of risk of preterm birth, to date these measures have not been widely adopted, are subject to user experience and have reported variable performance characteristics.” References were added.

-Line 85: The authors mentioned that metabolites were not used previously to determine GA. Please check this recent reference: Liang et al. Cell 181,71 Pages 1680-1692.e15 (https://doi.org/10.1016/j.cell.2020.05.002)

Response: Our manuscript was submitted before the paper of Liang et al was in press. We revised the main text in recognition of their findings and added the citation.

“Attempts at estimating GA using molecular adaptations have included modeling of RNA, protein, immune cell changes, and most recently metabolites in maternal blood.”

-Lines 93-101 are better to be moved to the discussion section.
Response: We moved this section to the discussion section.

Methods:

-Line 105: “PTB was defined by delivery at < 35 weeks GA.” From the clinical point of view, PTB is defined as delivery <37 w, why did authors choose this definition? May be to exclude late preterm deliveries since they tend to be milder cases.

Response: Please refer to the 2nd reviewer same point comment.

-Line 109: the gold standard was the US measurement. Was it the first trimester US based on the crown-rump length? Please specify with an appropriate reference.

Response: It is the first trimester US based on crown-rump length. A reference was added.

-Line 116: Please include the IRB approval numbers.

Response: The IRB number was included.

-Lines 146-148: "Results were compared with the insulin-like growth factor-binding protein 4 (IBP4)/sex hormone-binding globulin (SHBG) signature that is commercially available as a metabolic test for determining risk of PTB". This analysis has been done on another cohort according to Ext A.2. Please mention this in the main text with a brief description of the ELISA methods.

Response: We clarified the IBP4/SHBG analysis in the main text: "ELISA tests were conducted on the SU and UAB cohorts to evaluate the insulin-like growth factor-binding protein 4 (IBP4)/sex hormone-binding globulin (SHBG) signature, a predictor that was validated in a prospective study as a predictor of spontaneous PTB[19]. Serum concentrations were measured using commercial kits Human IGFBP4 ELISA Kit (Abcam, Burlingame, CA, USA) and Human SHBG Quantikine ELISA Kit (R&D System Inc.). Results were compared with our metabolic model."

Results:

-Table 1: there is an error in the black race percentage in UAB cohort (it is now 6.9% ◊ it should be 76.9%)

Response: We revised to 76.9%.

-There are differences in the baseline characteristics between term and PTB pregnancies. Have the authors adjusted the results for these differences? (if not, this should be added in the limitations section).

Response: We mentioned in the limitation section: "baseline characteristics of patients were not included in the analysis."

-Line 201: "Samples collected before 35 weeks' GA were used to develop a model that differentiated PTB pregnancies from those full-term". Since the metabolites profile is correlated with GA, why do the authors choose this wide range of gestational weeks? Have they developed different algorithms for each pregnancy trimester for instance or a multivariate model including the GA and other variables like race, BMI…?

Response: To identify metabolites that might be able to discriminate a preterm birth delivery from a full-term delivery applicable for each trimester, we performed unblinded analysis, with samples collected before 35 weeks' GA, to identify metabolites could separate with statistical significance. We did not develop different algorithms for each pregnancy trimester. Neither did we develop multivariate models including the GA and other variables like race, BMI…

Discussion:

-Line 255: “There is a need to develop a more robust method than LMP and US that captures pregnancy progression”. First trimester US is a very reliable method to determine GA and is the standard of care whenever it is available. This statement should be modified to (or something similar): “There is a need to develop a more robust method than LMP and an alternative to first trimester US that captures pregnancy progression”

Response: We have modified the sentence to be more precise according to the reviewer’s suggestion. The full sentence now reads as follows: “There is a need to develop a more robust method than LMP and an alternative to first trimester US that captures pregnancy progression, a complex relationship of fetal and placental growth, development, and function.”

-What are the pathways differentially expressed in PTB? Could the authors share their insights regarding these pathways to enhance the understanding of PTB pathophysiology?
Response: To clarify the differentially expressed pathways in PTB, a paragraph was added in discussion section to explore the underlying biology of the contributing pathways for preterm birth.

Reviewer: 2

Reviewer Name: Jeffrey Murray

Institution and Country: Dept of Pediatrics, University of Iowa, USA

Please state any competing interests or state ‘None declared’: I have collaborated with, published with and managed grant support for this group previously but not related to this specific manuscript.

Please leave your comments for the authors below

This is a first proof of principle publication establishing the feasibility of using metabolic testing of maternal serum during pregnancy to both establish gestational age or identify a risk profile for likely preterm birth. The study is timely, well conceived and the report largely well written and discussed. It has the potential to be a first step in both developing a better understanding of the underlying risks for PTB and to have clinical utility (potentially) in both high and low resource settings. There are a few concerns:

In table 1 the nos. in () seem to refer to percentages but don’t make sense for the number in the UAB column in some places (10 preterm is 6% and 2 is 15% for ex). It might also be useful in table 1 to have a col of p values for difference between the two sites (things like primipara status, age of mo, hx of ptb, race seem quite different in addition the demographic data and descriptions of the populations is underreported. Medical conditions of the mother, were all PTB spontaneous, were twins excluded etc should all be noted either in text or table and some discussion of these potentially confounding factors included.
Response: The numbers in the UAB column were clarified. A column of p values measuring the differences between the two sites was added to Table 1.

In results we added: “Our SU and UAB cohorts were assembled: no complications of pregnancy were included; all deliveries were singleton; and all PTB were spontaneous.”

In limitation we added: "baseline characteristics of patients were not included in the analysis."

Defining "preterm" as before 35 weeks is non standard and while it does have clinical relevance will require substantial explanation if used for population/surveillance estimates and likely affects the outcomes reported here. In addition, the absence of the 35 and 36 week pregnancies and the very *distorted numbers of PTB* pregnancies compared to population numbers will need to be address in future population based studies.
Response: We agree with the reviewer that our "preterm" definition, albeit not standard, but does have clinical relevance. We consider subjects who are delivered at 35 or 36 weeks of GA are close to term subjects with delivery at 37 weeks or later. The late preterm deliveries were not considered since they tend to be milder cases. In order to identify the metabolic signature uniquely associated with earlier pregnancies when asymptomatic, we defined PTB subjects as delivery at < 35 weeks GA. In addition, the definition of PTB as <35 weeks in this study is related to sample density and availability. Given the sparse data points we were able to obtain for this POC study was limited for pregnancies delivering between 35-37 weeks, led us to define the cohorts as stated.

In definition section, we added: "PTB was defined by delivery at < 35 weeks GA in order to make a complete separation from the full-term subjects." We took the absence of the 35 and 36 week pregnancies as a limitation and added to the limitation section that "pregnancies with delivery at 35 or 36 weeks were not included in the study."

We also agree that follow-on population based studies are needed to further define the clinical utility of the methods and panels being developed herein. We mentioned at the end of the limitation: "A larger prospective cohort study with a reasonable ratio of full-term to preterm is necessary before applying the estimates and prediction to a broader population for clinical utility."

IBP4 and SHBG are referred to as "metabolic" markers but were proteomic from the report and it was a little unclear as to the utility of the comparison of their PTB predictor to that of the SHBG/IBP4 outside the narrow range for which the latter is recommended as it has not been validated outside a narrow window (18 to 20 weeks as noted).
Response: We deleted "metabolic". We made the comparison because 1) the IBP4 and SHBG ratio marker has been validated in a multi-site study and made into a commercial kit to predict preterm birth; 2) similar to our method, the IBP4 and SHBG ratio marker was measured from maternal blood; 3) the comparison has demonstrated the strength of our method – our preterm prediction model works in a wider GA window and isn't constrained by pre-pregnancy BMI.

The prevalence corrected PPV values used the national PTB prevalence of 9.7% but its unclear why this would be better than using the local population value which might also be substantially different between Stanford and UAB and how might this affect results. Also as a side point all the positive predictive values might also be viewed in the context that just guessing that a random pregnancy will be term will be correct 90% of the time.
Response: The prevalence corrected PPV in Alabama was recalculated according to the local PTB prevalence (12.5% in 2018). The updated result is 70.4%. Figure A.6 was updated accordingly.

The 12.5% PTB prevalence gives a correct rate of 87.5% in randomly classifying a term pregnancy, and our model has improved this value to 97.5% (i.e., the prevalence corrected NPV). We agree with the reviewer that the negative predictive rate (NPV) is "in the context that just guessing that a random pregnancy will be term will be correct 90% of the time**". The aim of the study, however, is to identify preterm birth cases, so the PPV is paid with more attention. A**

random preterm pregnancy will be correct 12.5% of the time. The prevalence corrected PPV of our model is 70.4%, which is 5.6 times higher than the value.

In the replication study there were a substantial number of women with a prior PTB. Is it known if they received progesterone as a preventative as has been common practice in some centers and if so did this have any effect on the metabolic profiles? Similarly were any of the women diabetic, on antihypertensive medications or with other complications of pregnancy and if so could they see any effect on values there (recognizing that the number would be quite small).
Response: Women in both SU and UAB cohorts had no record of complications of pregnancy. Two women in SU cohort had received progesterone and delivered at preterm. No woman had antihypertensive medications during pregnancy.

They might also support their investigation with the extensive literature showing that newborn metabolic profiles using targeted metabolites are also highly predictive of GA and PTB.
Response: References were added to the end of the 2nd paragraph in introduction.

They state the data is available upon request. Rather once published the data should be freely available publicly.
Response: Once published, data will be uploaded at the laboratory website (http://translationalmedicine.stanford.edu) and shared with National March Of Dimes Database.

Finally there is a missed opportunity to go a bit deeper in the discussion as to how this might be used both scientifically (to better understand the drivers of developmental changes that occur in pregnancy) and more importantly to outline better how this might be used clinically. For example they briefly propose that this could be used in places where US is unavailable but US is becoming far more widespread and has the added utility of identifying many pregnancy complications of immediate clinical concern (twins, placenta praevia, breech etc) and the cost/processing of the maternal serum collection could also be a challenge. Similarly might there be alternatives such as maternal urine that could also work and be easier to collect. So some mention of these issues to guide future thinking about their findings could help move the field quickly both scientifically and practically.
Response: We thank the reviewer for these comments regarding scientific insights and practical clinical applications. In the Discussion, we have added several sentences regarding the scientific insights afforded through the leading metabolic pathways contributing to the respective models of pregnancy progression, specific mention is made of glycerophospholipids signaling and several others. In addition, we now provide several sentences on the practicality and potential scalability of these measures as follows:

"Taken together, the analysis of the leading pathways found to significantly contribute to the metabolic pregnancy modeling herein provide ample insights to deepen our understanding of pregnancy progression and may facilitate the identification and interpretation of potential therapeutic targets. Further, we speculate that the platform and approaches outlined herein may be extended to the interrogation of additional conditions of pregnancy including abnormalities of placentation, gestational diabetes and fetal growth disturbances among others."

All these concerns are relatively minor for this initial publication which is an important contribution to developing this a method with substantial promise. The authors have provided substantial detail in the appendix, done a thorough job of noting not only its many strengths but also its limitations and it is written in a clear and accessible style.

VERSION 2 – REVIEW

REVIEWER	Lina Youssef BCNatal | Fetal Medicine Research Center (Hospital Clínic and Hospital Sant Joan de Déu), Institut d'Investigacions Biomèdiques August Pi I Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain.
REVIEW RETURNED	23-Oct-2020

GENERAL COMMENTS	The manuscript has improved and the authors have considered all the reviewers comments. I still have minor typo comments: Introduction: The final part “Our findings suggest that composite metabolic panel modeling may serve as a reproducible and precision approach to GA dating of pregnancy and prediction of PTB.” Should be like: Our aim is to investigate if composite metabolic panel modeling may serve as a reproducible and precision approach to GA dating of pregnancy and prediction of PTB. Discussion: “Plasma level of arginine and citrulline was significantly lowered in preterm babies” should be lower rather than lowered.