Validating the Performance of the Modified Early Obstetric Warning System Multivariable Model to Predict Maternal Intensive Care Unit Admission

Helen M. Ryan, MBBCh BAO, MRCPI;[1,2,3] Meghan A. Jones, MB BChir;[1,3] Beth A. Payne, PhD;[1,3] Sumedha Sharma, MPH;[1,3] Anna M. Hutfield, RN;[1,3] Tang Lee, MSc;[1,3] U. Vivian Ukah, MSc;[1,3] Keith R. Walley, MD, PhD;[4,5] Laura A. Magee, MD, MSc;[6,7] Peter von Dadelszen, MBChB, DPhil[6,7]

[1]Department of Obstetrics and Gynaecology, University of British Columbia, Vancouver, BC

[2]Department of Family Practice, University of British Columbia, Vancouver, BC

[3]Child and Family Research Institute, University of British Columbia, Vancouver, BC

[4]Department of Medicine, University of British Columbia, Vancouver, BC

[5]Centre for Heart Lung Innovation, St. Paul's Hospital, Vancouver, BC

[6]Institute of Cardiovascular and Cell Sciences, St. George's, University of London, London, UK

[7]Department of Obstetrics and Gynaecology, St. George's University Hospitals NHS Foundation Trust, London, UK

Abstract

Objectives: To evaluate the performance of the Modified Early Obstetric Warning System (MEOWS) to predict maternal ICU admission in an obstetric population.

Design: Case-control study.

Setting: Two maternity units in Vancouver, Canada, one with ICU facilities, between January 1, 2000, and December 31, 2011.

Patients: Pregnant or recently delivered ($\leq$6 weeks) women admitted to the hospital for $>$24 hours. Three control patients were randomly selected per case and matched for year of admission.

Measurements and Main Results: Retrospective, observational, case-control validation study investigating the physiologic predictors of admission in the 24-hour period preceding either ICU admission $>$24 hours (cases) or following admission (control patients). Model performance was assessed based on sensitivity, specificity, and predictive values. Forty-six women were admitted to the ICU for $>$24 hours (0.51/1000 deliveries); the study included 138 randomly selected control patients. There were no maternal deaths in the cohort. MEOWS had high sensitivity (0.96) but low specificity (0.54) for ICU admission $>$24 hours, whereas $\geq$1 one red trigger maintained sensitivity (0.96) and improved specificity (0.73).

Conclusion: Altering MEOWS trigger parameters may improve the accuracy of MEOWS in predicting ICU admission. Formal modelling of a MEOWS scoring system is required to support evidence-based care.

Key Words: Early warning systems, obstetric, Modified Early Obstetric Warning System, ICU admission

Corresponding Author: Dr. Peter von Dadelszen, St. George's, University of London, London, UK. pvd@sgul.ac.uk

Competing interests: See Acknowledgements.

Received on September 1, 2016

Accepted on January 30, 2017

Résumé

Objectifs : Évaluer le rendement du Modified Early Obstetric Warning System (MEOWS) quant à sa capacité de prédire les admissions maternelles aux soins intensifs dans une population obstétricale.

Modèle : Étude cas-témoin.

Milieu : Deux maternités de Vancouver (Canada), dont une pourvue d'un service de soins intensifs, entre le 1er janvier 2000 et le 31 décembre 2011.

Patientes : Femmes enceintes ou ayant accouché récemment (six semaines ou moins) hospitalisées pendant plus de 24 heures. Trois témoins ont été sélectionnées aléatoirement pour chaque cas et jumelées selon l'année d'admission.

Mesures et principaux résultats : Étude cas-témoin rétrospective observationnelle de validation examinant les facteurs physiologiques prédisant l'admission dans les 24 heures précédant l'admission aux soins intensifs pour plus de 24 heures (cas) ou suivant l'admission (témoins). Le rendement du modèle a été évalué selon la sensibilité, la spécificité et les valeurs prédictives. Quarante-six femmes ont été admises aux soins intensifs pendant plus de 24 heures (0,51/1 000 accouchements); l'étude comprenait 138 témoins sélectionnées aléatoirement. Il n'y a eu aucun décès maternel dans la cohorte. Le MEOWS avait une sensibilité élevée (0,96), mais une spécificité faible (0,54) pour les admissions aux soins intensifs durant plus de 24 heures, alors qu'au moins un

déclencheur rouge a maintenu la sensibilité (0,96) et amélioré la spécificité (0,73).

Conclusion : Modifier les paramètres des déclencheurs du MEOWS pourrait améliorer l'exactitude du système quant à sa capacité de prédire les admissions aux soins intensifs. La modélisation officielle d'un système de notation du MEOWS est requise pour appuyer les soins fondés sur des données probantes.

J Obstet Gynaecol Can 2017;39(9):728–733

https://doi.org/10.1016/j.jogc.2017.01.028

INTRODUCTION

Early warning systems—physiologic "track and trigger" systems—are clinical prediction models that involve serial clinical observations ("track") with criteria ("trigger") to identify patients at risk of complications.[1] The use of an EWS is recent in maternity care, with variations in measured parameters and threshold "trigger" values.[2]

Based on obstetric EWSs from nine U.K. hospitals, the Modified Early Obstetric Warning System tool was proposed by Confidential Enquiries into Maternal and Child Health in 2007[3] (Table 1). However, the MEOWS tool development was undertaken by highly informed experts without formal prediction model development methodology, resulting in the assignment of arbitrary cut-offs to continuous clinical variables.[4,5] The resulting CEMACH MEOWS tool has not been adequately externally validated.[4,5] The 12 MEOWS variables include maternal vital signs and other clinical observations and dipstick proteinuria (Table 1). All are routine clinical observations, although some are subjective measures (e.g., whether lochia is "heavy").

Used by doctors and midwives or nurses in settings such as accident and emergency departments to identify high-risk pregnant women, MEOWS is time-consuming, has been variably implemented, and appears to preferentially aid in the management of women with clearly established risk of morbidity.[2,6]

The aim of this study was to externally validate the current CEMACH MEOWS tool using Canadian data for pregnant and recently pregnant, hospitalised women and to assess for the possibility of developing a multivariable tool based on the current MEOWS tool.

ABBREVIATIONS

AUROC	area under the receiver operating characteristic
CEMACH	Confidential Enquiries into Maternal and Child Health
EWS	early warning systems
MEOWS	Modified Early Obstetric Warning System
NPV	negative predictive value
BP	systolic blood pressure

MATERIALS AND METHODS

This retrospective, observational, case-control study was performed in two tertiary obstetric units (British Columbia Women's Hospital and Health Centre and St. Paul's Hospital, Vancouver, BC). The University of British Columbia clinical research ethics board approved the project for both institutions (REB# H10-00654). The STARD checklist for this study is presented as an appendix.

Cases included pregnant or recently pregnant (<6 weeks after the end of the pregnancy, irrespective of gestational age at the end of the pregnancy) admitted women who subsequently required admission to the ICU for >24 hours (January 1, 2000, to December 31, 2011; a period of stable patterns of care). Because these were the limits of the data available, no formal sample size calculation was performed.

Control patients were the first three women identified from hospital databases who were either pregnant or recently pregnant and admitted to the hospital for >24 hours during that time who did not receive critical care, matching for year of admission but for neither antenatal nor postpartum state of cases at the time of ICU admission.

De-identified MEOWS variables (Table 1) plus six demographic variables (age, BMI, smoking status, gravidity, gestational age on admission, and birth weight) were extracted from the hospital case notes for the 24-hour period prior to ICU admission (cases) and the first 24 hours following admission (control patients) and entered into a REDCap database, Version 4.12.0.5 (Vanderbilt University, Nashville, TN); for multiple measurements, the most abnormal values were selected. We did not impute missing values. When MEOWS variables were missing, it was assumed that these observations were not recorded because they were not perceived to be abnormal and they would not have fallen into the "red" or "amber" trigger categories.

Neither the data collectors nor the statisticians were blinded to whether the woman was a case or control patient. "MEOWS activation" was defined as the occurrence of ≥1 "red" or ≥2 "amber" MEOWS triggers (Table 1).

Table 1. MEOWS variables: normal ranges and trigger criteria (CEMACH 2007)

Variable MEOWS vital signs components	Normal range for pregnancy	"Amber" (caution) trigger	"Red" (urgent) trigger
Respiratory rate (breaths/min)	10–20	21–30	<10 or >30
Temperature (°C)	37–38	35–36	<35 or >38
Heart rate (beats/min)	51–99	40–50 or 100–120	<40 or >120
Systolic blood pressure (mmHg)	101–149	90–100 or 150–160	<90 or >160
Diastolic blood pressure (mmHg)	<80	80–90	>90
Other components			
Oxygen saturation (%)	≥90	No trigger	<90
Looks unwell	No	Yes	No trigger
Neurologic response	Fully responsive	Responsive to voice	Responsive only to pain or unresponsive
Pain score[a]	0–1	2–3	No trigger
Amniotic fluid	Clear	No trigger	Green
Lochia	Light to moderate with no odour	No trigger	Heavy or foul
Passed urine (yes/no)	Yes	No trigger	No trigger
Dipstick proteinuria	Negative or trace	No trigger	>2+

[a]The "pain score" assesses pain on movement, deep breathing, or coughing as follows: 0 for no pain at rest or on movement, 1 for no pain at rest but slight pain on movement, 2 for intermittent pain at rest and moderate pain on movement, and 3 for intermittent pain at rest and moderate pain on movement.

The primary outcome was maternal ICU admission >24 hours.

Continuous data (median [interquartile range (IQR)], compared using Mann-Whitney U test); categorical data (Fisher exact test); and 95% CIs for sensitivity, specificity, positive predictive value, negative predictive value, and area under the receiver operating characteristic were calculated using SPSS, Version 22.0 (Armonk, NY); VassarStats (Poughkeepsie, NY); and GraphPad Prism, Version 6.01 (La Jolla, CA). A P value < 0.05 was used for statistical significance.

As a planned secondary analysis, we explored modifications to the existing CEMACH MEOWS model. Initially, a univariable analysis was performed to identify which of the clinical MEOWS chart measurements had the strongest association with the primary outcome. We included both highest and lowest values of temperature, heart rate, and blood pressure and the respiratory rate, resulting in nine variables in total. A cut-off significance value of $P < 0.05$ was used to select the variables that were entered into a logistic regression model with stepwise elimination of the variable with the least significant association with the primary outcome.

RESULTS

Of 90 714 deliveries at the two sites (2000-2011), 46 women (eight pregnant and 36 postpartum) required ICU admission (cases; ICU admission rate of 0.51/1000 deliveries), and 138 (109 pregnant and 29 postpartum) control patients were identified.

Non-vital signs MEOWS variables were poorly reported in the charts, precluding their assessment (Table 2). The model using the MEOWS components (either ≥1 red or ≥2 amber scores) had high sensitivity (0.96) but low specificity (0.54) for identifying women at risk of ICU admission >24 hours; ≥1 red trigger maintained sensitivity (0.96) and improved specificity (0.73); ≥2 red triggers lowered sensitivity (0.65) and increased specificity (0.89). No comparison could be made with a development data set because CEMACH MEOWS was developed by arbitrarily assigning cut-offs.

Following stepwise logistic regression, of eight significant variables identified using univariable analyses ($P < 0.05$; maximum temperature, maximum heart rate, minimum heart rate, maximum systolic blood pressure, minimum sBP, maximum diastolic blood pressure, minimum diastolic blood pressure, and respiratory rate), four variables were significantly associated with ICU admission >24 hours: maximum temperature (β co-efficient 1.14; $P = 0.005$); heart rate (β 0.06; $P < 0.001$); sBP (β 0.05; $P < 0.001$); and respiratory rate (β 0.22; $P = 0.001$). The AUROC for the four-variable equation was 0.91 (95% CI 0.83 to 0.95), with activation of ≥1 red or ≥2 amber triggers providing high sensitivity, specificity, and NPV (Table 3). When MEOWS activation was limited to ≥1 red trigger, sensitivity fell.

Table 2. Comparison of demographic variables, patient characteristics and MEOWS variables (n [%] women or median [IQR], as applicable)

Variable	ICU admitted women (n = 46)	Non-ICU admitted women (n = 138)	*P** value	AUROC
Demographics				
Age (years)	35.5 (30.8–39.0)	34.0 (30.75–37.0)	0.1	-
BMI (kg/m^2)	25.7 (15.9–30.5)	26.3 (23.5–30.0)	0.2	-
Missing	11 (23.9%)	49 (35.5%)		
Smoking (y/n)	5/46 (10.9%)	15/138 (10.9%)	1	-
Multiple pregnancy	3/46 (6.5%)	6/138 (4.3%)	0.5	-
Gravidity	3.0 (2.0–4.25)	3.0 (2.0–4.0)	0.4	-
Missing	16 (34.8%)	44 (31.9%)		
Gestational age	38+2 (35+0–39+4)	39+0 (38+0–40+1)	0.001	-
Birth weight (g)	3005 (2480–3448)	3315 (3005–3685)	0.003	-
Missing	6 (13.0%)	11 (8.0%)		
MEOWS vital signs components				
Respiratory rate (breaths/min)	20 (18–28)	19.5 (18–20)	0.0015	0.6516
Missing	1 (2.2%)	4 (2.9%)		
Temperature, maximum (°C)	37.4 (37.2–38.1)	37 (36.8–37.3)	<0.0001	0.7279
Temperature, minimum (°C)	36.4 (36–36.7)	36.3 (36–36.6)	0.9078	0.5058
Heart rate, maximum (beats/min)	125 (96–139.8)	90 (81–98)	<0.0001	0.8280

DISCUSSION

Main Findings

This is the first non-U.K. validation of MEOWS, particularly for prediction of ICU admission. Because MEOWS is a "screening test," these initial physiologic measurements taken up to 24 hours prior to ICU admission are most informative in identifying the severity of illness prior to further deterioration or implementation of resuscitation and stabilization.

Reassuringly, the original MEOWS model was highly sensitive for prediction of ICU admission, with a high NPV. However, with low specificity, many uncomplicated control pregnancies also triggered MEOWS. Conversely, using any single MEOWS red trigger more accurately predicted ICU admission in our cohort, with equivalent sensitivity and higher specificity and NPV, potentially reducing both workload and false-positive MEOWS responses.

From the secondary analysis and modelling, the performance of the four-variable model (Table 2) had high sensitivity, specificity, NPV, and AUROC for ICU admission, highlighting the importance of these four variables to identify pregnant and recently pregnant women at risk of critical illness.

Strengths and Limitations

MEOWS was originally implemented as a tool to improve recognition, treatment, and referral of women developing a critical illness.[2,3,6] Therefore, we chose ICU admission >24 hours as the ultimate outcome measure to validate MEOWS. Some may consider ICU admission as a process indicator rather than an outcome measure in itself. However, the decision to transport a woman from a standalone obstetric unit, such as British Columbia Women's Hospital and Health Centre, to a nearby full-service hospital's ICU is a significant and costly one, and one deemed worthy of avoidance in studies of obstetric high-dependency unit performance.[7,8] The choice of admission >24 hours as an element of the outcome was to address concerns that thresholds for admission might vary between the two sites, British Columbia Women's Hospital and Health Centre (without in-house ICU support) and St. Paul's (the hospital to which critically ill British Columbia Women's Hospital and Health Centre patients are referred), and to improve generalizability of our findings across health systems. Receiving critical care for >24 hours is an expensive undertaking, and our opinion is that reproductive age women who remain in ICU for at least that period are likely to consistently reach a similar threshold of illness wherever they are.

We assessed 46 pregnant and recently pregnant women receiving critical care, a rare outcome. Even though our two-site study includes a broad maternity population, both centres are referral centres (potential population bias), and there were no universally defined entry criteria for ICU admission of obstetric patients. However, all cases experienced either organ failure or shock that precipitated admission to ICU.

Table 3. Comparison of MEOWS chart and four significant MEOWS variables using different "trigger" combinations

	Fisher exact *P*[a]	RR (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	LR+ (95% CI)	LR− (95% CI)
MEOWS chart triggers								
≥1 red or ≥2 amber	<0.0001	15.83 (3.96–63.36)	0.96 (0.84–0.99)	0.54 (0.46–0.63)	0.41 (0.32–0.51)	0.97 (0.90–0.995)	2.10 (1.73–2.54)	0.08 (0.02–0.31)
≥1 red	<0.0001	13.35 (4.99–35.71)	0.91 (0.78–0.97)	0.72 (0.63–0.79)	0.52 (0.41–0.63)	0.96 (0.90–0.99)	3.23 (2.44–4.27)	0.12 (0.05–0.31)
≥2 red	<0.0001	5.79 (3.49–9.60)	0.65 (0.50–0.78)	0.89 (0.82–0.94)	0.67 (0.51–0.80)	0.89 (0.82–0.93)	6.00 (3.56–10.11)	0.39 (0.26–0.58)
Four significant MEOWS triggers								
≥1 red or ≥2 amber	<0.0001	11.14 (4.97–24.99)	0.87 (0.73–0.95)	0.84 (0.77–0.90)	0.65 (0.51–0.76)	0.95 (0.89–0.98)	5.5 (3.66–8.13)	0.16 (0.07–0.33)
≥1 red	<0.0001	9.14 (4.90–17.04)	0.78 (0.63–0.89)	0.88 (0.82–0.93)	0.69 (0.55–0.81)	0.92 (0.86–0.96)	6.75 (4.15–10.97)	0.25 (0.14–0.43)

RR, relative risk; PPV, positive predictive value; LR+, positive likelihood ratio; LR−, negative likelihood ratio.

[a]Comparing ICU-admitted with non-ICU-admitted pregnant women.

Our study is retrospective and case-control in nature. Three randomly selected control patients per case increased power to detect differences between groups. We matched for year of admission to ensure comparable standard of care. Minimal matching ensured control patients were representative of the obstetric population at risk of, but avoiding, critical care.[9] The two groups had similar baseline demographics.

In addition, many of the subjective variables, subject to user bias and interobserver variability, were either poorly or not routinely recorded (Table 2); hence, their utility was not evaluated. Therefore, we have not excluded the potential for subjective variables collected in MEOWS to have improved model performance.

Two national surveys in the United Kingdom have shown increasing use of obstetric EWSs from 19% (2007) to 100% (2014),[2,10] but direct evidence that they improve outcomes is lacking.[11] Variation in implementation of MEOWS equates to a potential loss of recognition of, and behavioural response to, clinical deterioration, whereas universal implementation increases workload and cost.[6] Non-obstetric EWSs tend to over-predict maternal morbidity and mortality in obstetric populations.[12]

In a single-unit validation of MEOWS (primary outcome: obstetric morbidity), the sensitivity was high (89%) but the positive predictive value was low and there were no ICU admissions or deaths,[13] whereas we found that MEOWS had a high sensitivity for predicting ICU admission. Carle et al.[14] developed and internally validated a new obstetric EWS predicting maternal death in women receiving critical care, which is more comparable with scoring systems like the Acute Physiology and Chronic Health Evaluation. We were unable to validate this EWS because fraction of inspired oxygen was not recorded and there were no maternal deaths in our cohort.

Previous MEOWS validations noted that temperature was not predictive of maternal morbidity,[13,14] differing from our cohort, in which highest temperature was strongly associated with ICU admission. High sBP as a predictor of ICU admission was consistent with the findings of Singh et al.[13] but differed from the ICU-based study.[14] Some measured variables in MEOWS, such as respiratory rate, are vulnerable to human error.[12] Oxygen saturation, as determined using either arterial or pulse oximetry measurement, is a valuable predictor of the need for ICU admission[15] and of death or serious complications in women with preeclampsia.[16–18] Although these findings support the use of pulse oximetry in an obstetric EWS, and

oxygen saturation showed significance with ICU admission ($P = 0.043$), it was too poorly recorded in our population for inclusion in the model. However, it is probable that oxygen saturation would contribute to an objectively derived MEOWS decision tool.

CONCLUSION

Our study has externally validated the performance of MEOWS in terms of prediction of ICU admission. Given the low specificity of MEOWS, the current threshold mandating a response may have been set too low; with improved specificity and NPV and maintained sensitivity, using any single red trigger activation. Prior to the development and full validation of an objectively derived MEOWS tool, we suggest that a modified MEOWS such as the one described here may better identify and focus early interventions on those women most at risk of requiring critical care. A statistically robust and externally validated MEOWS is an important and urgent need.

ACKNOWLEDGEMENTS

Funding support was provided by the Canadian Institute of Health Research catalyst grant: Maternal health: from Pre-conception to Empty Nest. The funders were not involved in the conduct of the study. P.v.D. has been a consultant to Alere International for work that is unrelated to this MEOWS study. The authors are very grateful for the data collection by the following University of British Columbia students: Joni Kooy, Nancy Liu, Brittany Tarras, Shannon Lockhart, Runshan Felar Yu, Chris Lim, Adrienne McKercher, Nelson Luk, Saba Marzara, Yuki Ueda, Annie Tran, Navdeep Dha, Joany Tang and Vivien Cao.

SUPPLEMENTARY DATA

Supplementary Data related to this article can be found at http://dx.doi.org/10.1016/j.jogc.2017.01.028.

REFERENCES

1. Gao H, McDonnell A, Harrison DA, et al. Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. Intensive Care Med 2007;33:667–9.
2. Swanton RD, Al-Rawi S, Wee MY. A national survey of obstetric early warning systems in the United Kingdom. Int J Obstet Anesth 2009;18:253–7.
3. Confidential Enquiry into Maternal and Child Health. Saving Mothers' Lives: Reviewing Maternal Deaths to Make Motherhood Safer - 2003-2005. London, UK: Confidential Enquiry into Maternal and Child Health; 2007. Available at: http://www.publichealth.hscni.net/sites/default/files/Saving%20Mothers'%20Lives%202003-05%20.pdf. Accessed on February 26, 2017.
4. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 2015;351:h5527.
5. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015;350:g7594.
6. Mackintosh N, Watson K, Rance S, et al. Value of modified early obstetric warning system (MEOWS) in managing maternal complications in the peripartum period: an ethnographic study. BMJ Qual Saf 2014;23:26–34.
7. Ryan M, Hamilton V, Bowen M, et al. The role of a high-dependency unit in a regional obstetric hospital. Anaesthesia 2000;55:1155–8.
8. Sultan P, Arulkumaran N, Rhodes A. Provision of critical care services for the obstetric population. Best Pract Res Clin Obstet Gynaecol 2013;27:803–9.
9. Grimes DA, Schulz KF. Compared to what? Finding controls for case-control studies. Lancet 2005;365:1429–33.
10. Isaacs RA, Wee MY, Bick DE, et al. A national survey of obstetric early warning systems in the United Kingdom: five years on. Anaesthesia 2014;69:687–92.
11. Knight M, Kenyon S, Brocklehurst (MBRRACE-UK) P, et al. Saving Lives, Improving Mothers' Care: Lessons Learned to Inform Future Maternity Care from the UK and Ireland Confidential Enquiries into Maternal Deaths and Morbidity 2009-2012. Oxford: National Perinatal Epidemiology Unit, University of Oxford; 2014. Available at: https://www.npeu.ox.ac.uk/downloads/files/mbrrace-uk/reports/Saving%20Lives%20Improving%20Mothers%20Care%20report%202014%20Full.pdf. Accessed on February 26, 2017.
12. Lappen JR, Keene M, Lore M, et al. Existing models fail to predict sepsis in an obstetric population with intrauterine infection. Am J Obstet Gynecol 2010;203:573.e1–5.
13. Singh S, McGlennan A, England A, et al. A validation study of the CEMACH recommended modified early obstetric warning system (MEOWS). Anaesthesia 2012;67:12–8.
14. Carle C, Alexander P, Columb M, et al. Design and internal validation of an obstetric early warning score: secondary analysis of the Intensive Care National Audit and Research Centre Case Mix Programme database. Anaesthesia 2013;68:354–67.
15. Cuthbertson BH, Boroujerdi M, McKie L, et al. Can physiological variables and early warning scoring systems allow early recognition of the deteriorating surgical patient? Crit Care Med 2007;35:402–9.
16. Millman AL, Payne B, Qu Z, et al. Oxygen saturation as a predictor of adverse maternal outcomes in women with preeclampsia. J Obstet Gynaecol Can 2011;33:705–14.
17. von Dadelszen P, Payne B, Li J, et al. Prediction of adverse maternal outcomes in pre-eclampsia: development and validation of the fullPIERS model. Lancet 2011;377:219–27.
18. Payne BA, Hutcheon JA, Dunsmuir D, et al. Assessing the incremental value of blood oxygen saturation (SpO2) in the miniPIERS (Pre-eclampsia Integrated Estimate of RiSk) risk prediction model. J Obstet Gynaecol Can 2015;37:16–24.

SUPPLEMENTARY DATA

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	1
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	3–4
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	5
	4	Study objectives and hypotheses	5
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	6
Participants	**6**	Eligibility criteria	6
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	6
	8	Where and when potentially eligible participants were identified (setting, location and dates)	6
	9	Whether participants formed a consecutive, random or convenience series	6
Test methods	**10a**	Index test, in sufficient detail to allow replication	6 & Figure 1
	10b	Reference standard, in sufficient detail to allow replication	7
	11	Rationale for choosing the reference standard (if alternatives exist)	7
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	6
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	6
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	6
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	6
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	1
	15	How indeterminate index test or reference standard results were handled	6
	16	How missing data on the index test and reference standard were handled	6
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	6–7
	18	Intended sample size and how it was determined	6
RESULTS			
Participants	**19**	Flow of participants, using a diagram	N/A
	20	Baseline demographic and clinical characteristics of participants	Table 1
	21a	Distribution of severity of disease in those with the target condition	Table 1
	21b	Distribution of alternative diagnoses in those without the target condition	N/A
	22	Time interval and any clinical interventions between index test and reference standard	6
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	8
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	8 & Tables 1 & 2
	25	Any adverse events from performing the index test or the reference standard	N/A

Continued

Continued

Section & Topic	No	Item	Reported on page #
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	9-10
	27	Implications for practice, including the intended use and clinical role of the index test	11
OTHER INFORMATION			
	28	Registration number and name of registry	N/A
	29	Where the full study protocol can be accessed	N/A
	30	Sources of funding and other support; role of funders	12

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.