JAMA | US Preventive Services Task Force | EVIDENCE REPORT

Preeclampsia Screening
Evidence Report and Systematic Review for the US Preventive Services Task Force

Jillian T. Henderson, PhD; Jamie H. Thompson, MPH; Brittany U. Burda, DHSc, MPH; Amy Cantor, MD, MPH

IMPORTANCE Preeclampsia is a complex disease of pregnancy with sometimes serious effects on maternal and infant morbidity and mortality. It is defined by hypertension after 20 weeks' gestation and proteinuria or other evidence of multisystem involvement.

OBJECTIVE To systematically review the benefits and harms of preeclampsia screening and risk assessment for the US Preventive Services Task Force.

DATA SOURCES MEDLINE, PubMed, and Cochrane Central Register of Controlled Trials databases from 1990 through September 1, 2015. Surveillance for new evidence in targeted publications was conducted through October 5, 2016.

STUDY SELECTION English-language trials and observational studies, including externally validated prediction models, of screening effectiveness, benefits, and harms from routine preeclampsia screening during pregnancy.

DATA EXTRACTION AND SYNTHESIS Independent dual review of article abstracts and full texts against a priori inclusion criteria. Meta-analysis was not performed because of clinical and statistical heterogeneity of included studies.

MAIN OUTCOMES AND MEASURES Maternal and infant health outcomes, including eclampsia, stroke, stillbirth, preterm birth, and low birthweight; screening and risk prediction test performance; harms of screening and risk assessment.

RESULTS Twenty-one studies (13 982 participants) were included. No studies directly compared the effectiveness of preeclampsia screening in a screened population vs an unscreened population; 1 US trial (n = 2764) found no difference in benefits or harms with fewer prenatal visits but was underpowered for rare, serious outcomes. For harms, a before-after comparison cohort noninferiority study of urine protein screening for specific indications compared with routine screening (n = 1952) did not identify harms with fewer urine screening tests. Four studies (n = 7123) reported external validation performance of 16 risk prediction models, 5 of which had good or better discrimination (*c* statistic >0.80) for prediction of preeclampsia, and positive predictive values of 4% in the largest, most applicable validation cohorts. Calibration was not reported despite being a key model performance measure. There were no studies of urine screening test performance conducted in asymptomatic primary care populations; 14 studies of protein urine test performance among women being evaluated for suspected preeclampsia (n = 1888) had wide-ranging test accuracy (sensitivity, 22%-100%; specificity, 36%-100%) and high statistical and clinical heterogeneity in tests used, eligibility criteria, and proteinuria prevalence (8.7%-93.8%).

CONCLUSIONS AND RELEVANCE Evidence to estimate benefits and harms of preeclampsia screening and the test performance of different screening approaches over the course of pregnancy was limited. Externally validated risk prediction models had limited applicability and lacked calibration and clinical implementation data needed to support routine use. Further research is needed to better inform risk-based screening approaches and improve screening strategies, given the complex pathophysiology and clinical unpredictability of preeclampsia.

JAMA. 2017;317(16):1668-1683. doi:10.1001/jama.2016.18315

Editorial page 1629

Related article page 1661 and JAMA Patient Page page 1700

Supplemental content

Related articles at jamacardiology.com jamainternalmedicine.com

Author Affiliations: Kaiser Permanente Research Affiliates Evidence-based Practice Center, Center for Health Research, Kaiser Permanente, Portland, Oregon (Henderson, Thompson, Burda); Oregon Health & Science University, Portland (Cantor).

Corresponding Author: Jillian T. Henderson, PhD, Kaiser Permanente Research Affiliates Evidence-based Practice Center, Center for Health Research, Kaiser Permanente Northwest, 3800 N Interstate Ave, Portland, OR 97227 (jillian.t.henderson@kpchr.org).

Approximately 2% to 8% of pregnancies are affected by preeclampsia, defined by the development of hypertension and proteinuria after 20 weeks' gestation. In the absence of proteinuria, additional features contribute to diagnosis (ie, thrombocytopenia, renal insufficiency, impaired liver function, pulmonary edema, cerebral or visual symptoms).[1] Preeclampsia is the second leading cause of maternal mortality worldwide.[2,3] In the United States the rate of preeclampsia increased from 3.4% in 1980 to 3.8% in 2010,[4] with the proportion of severe cases also increasing. Identifying women at higher risk for preeclampsia early in pregnancy, based on medical history and routine tests, could inform risk-based prevention and screening. Once preeclampsia is diagnosed, evidence-based interventions may reduce the risk or severity of maternal and infant health outcomes of preeclampsia; these include treatment of high blood pressure, administration of magnesium sulfate to prevent eclampsia, and induced delivery.[1,5-12]

In 1996, the US Preventive Services Task Force (USPSTF) recommended screening for preeclampsia using office-based blood pressure measurement for all pregnant women at the first prenatal visit and periodically throughout the remainder of the pregnancy (B recommendation).[13] The current review was commissioned to systematically review and update evidence on screening for preeclampsia.

Methods

Scope of Review

Detailed methods are available in the full evidence report at https://www.uspreventiveservicestaskforce.org/Page/Document/final-evidence-review153/preeclampsia-screening1. The analytic framework and key questions (KQs) guiding this review are shown in **Figure 1**.[15]

Data Sources and Searches

After an initial search for existing systematic reviews and guidelines, a comprehensive search was performed for primary literature in the MEDLINE, PubMed, and Cochrane Central Register of Controlled Trials databases from 1990 through September 1, 2015 (eMethods in the Supplement). Studies published before 1990 were excluded because of changes in diagnostic criteria and treatments in the past 25 years, limiting applicability of earlier evidence.[7,11,12] Reference lists of prior reports and publications were also searched. Since September 2015, we continued to conduct ongoing surveillance through article alerts and targeted searches of high-impact journals to identify major studies published in the interim that may affect the conclusions or understanding of the evidence and therefore the related USPSTF recommendation. The last surveillance was conducted on October 5, 2016, and identified no relevant new studies.

Study Selection

Two investigators independently reviewed 10 082 titles and abstracts and 378 full-text articles against prespecified inclusion criteria (**Figure 2**). Discrepancies were resolved through consensus discussions. English-language, fair- and good-quality studies of pregnant women and adolescents without a diagnosis of preeclampsia and asymptomatic for the condition were included. Studies among women with chronic hypertension, diabetes mellitus, or elevated risk for preeclampsia were also included. Studies were excluded if they solely focused on women seeking high-risk obstetric care, receiving infertility treatment, receiving inpatient care, or if they were conducted in countries not having a high development index designation according to the 2014 United Nations Development Programme.[16] Any standard diagnostic criterion for preeclampsia was allowed.[1,17,18]

Screening interventions included point-of-care tests and clinical tools routinely used in prenatal care to screen for preeclampsia, such as blood pressure measurements using manual or automated devices and point-of-care urine tests for proteinuria with qualitative, quantitative, visual, or automated readings. Only studies using the 24-hour urine test as the reference standard to calculate the diagnostic accuracy of urine protein tests were included. Secondary evaluations and tests used to assess preeclampsia severity or to confirm diagnosis were not included. Evidence on the benefits and harms (KQ1, KQ5) of preeclampsia screening was from randomized clinical trials (RCTs) and observational studies that reported on maternal and infant mortality, morbidity from eclampsia, HELLP (hemolysis, elevated liver enzyme levels, low platelet counts) syndrome, organ damage or failure, fetal growth restriction, preterm delivery, low birth weight, stillbirth, and placental abruption. Evidence was sought on the screening test performance of clinical blood pressure measurement, urinalysis, or both for identifying women with preeclampsia at the time of screening (KQ4), to compare the effectiveness of different screening protocols (eg, instruments, test procedures, timing of tests, rescreen intervals) (KQ4a), to assess the diagnostic accuracy of point-of-care tests for detecting proteinuria (KQ4b), and to evaluate risk-based screening protocols, compared with general screening (KQ4c).

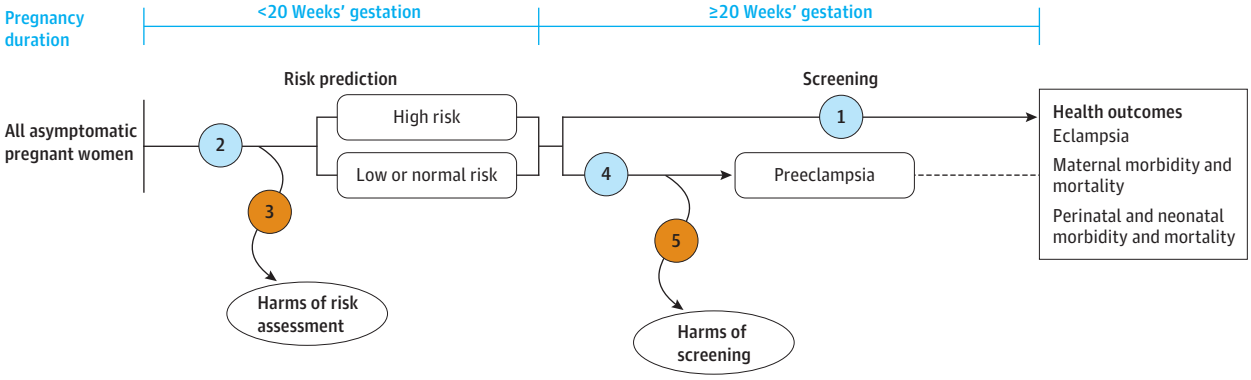
For assessment of preeclampsia risk (KQ2, KQ3), studies evaluating prediction models for use in the first 20 weeks of pregnancy were included to inform and differentiate screening and preventive interventions (eg, aspirin prophylaxis) before preeclampsia develops.[19-22] These were externally validated (ie, models tested in another population than the derivation study, assessing either performance or effect) multivariable risk prediction models using patient history and routinely collected clinical measures (eg, body mass index, weight, blood pressure) as well as serum markers and Doppler ultrasound measures (eg, uterine artery pulsatility index).

Quality Assessment and Data Extraction

Two investigators independently assessed the quality of all included studies using criteria predefined by the USPSTF[14] and supplemented them with other criteria from the Quality Assessment of Diagnostic Accuracy II for diagnostic accuracy studies (KQ4a)[23] and from the Newcastle-Ottawa Scale and Before-After Quality Assessment Tool[24] for observational studies (KQ3 and KQ5) (eTable 1 in the Supplement). Each included study received a final quality rating of good, fair, or poor; discrepancies were resolved through discussion. Poor-quality studies (ie, attrition >40%, differential attrition >20%, or other fatal flaws or cumulative effects of multiple minor flaws or missing information significant enough to limit confidence in the validity of results) were excluded. Good-quality studies met all or most of the assessment criteria; fair-quality studies met only some of the assessment criteria.

One investigator abstracted data from all included studies into an Access database (Microsoft Corp). A second investigator checked the data for accuracy.

Figure 1. Analytic Framework



Key questions

1 How effectively does screening for preeclampsia reduce maternal and perinatal morbidity and mortality?

a. Does effectiveness differ by screening protocol (eg, tests used, timing of tests, rescreen intervals) or preeclampsia risk status?

2 What is the effectiveness of risk assessment in early pregnancy for identifying women at high risk for preeclampsia?

3 What are the harms of preeclampsia risk assessment?

4 How effectively do screening tests (eg, blood pressure, proteinuria) identify women with preeclampsia?

a. How accurate are different screening tests for proteinuria?

b. How effective are different screening protocols (eg, instruments, test procedures, timing of tests, rescreen intervals) for identifying women with preeclampsia?

c. How should women at high risk for preeclampsia be screened differently from women at low or average risk?

5 What are the harms of screening for preeclampsia and do they differ by risk status or screening protocol?

Evidence reviews for the US Preventive Services Task Force (USPSTF) use an analytic framework to visually display the key questions that the review will address to allow the USPSTF to evaluate the effectiveness and safety of a preventive service. The questions are depicted by linkages that relate interventions and outcomes. A dashed line indicates health outcomes that follow an intermediate outcome. Further details are available from the USPSTF procedure manual.[14]

Data Synthesis and Analysis

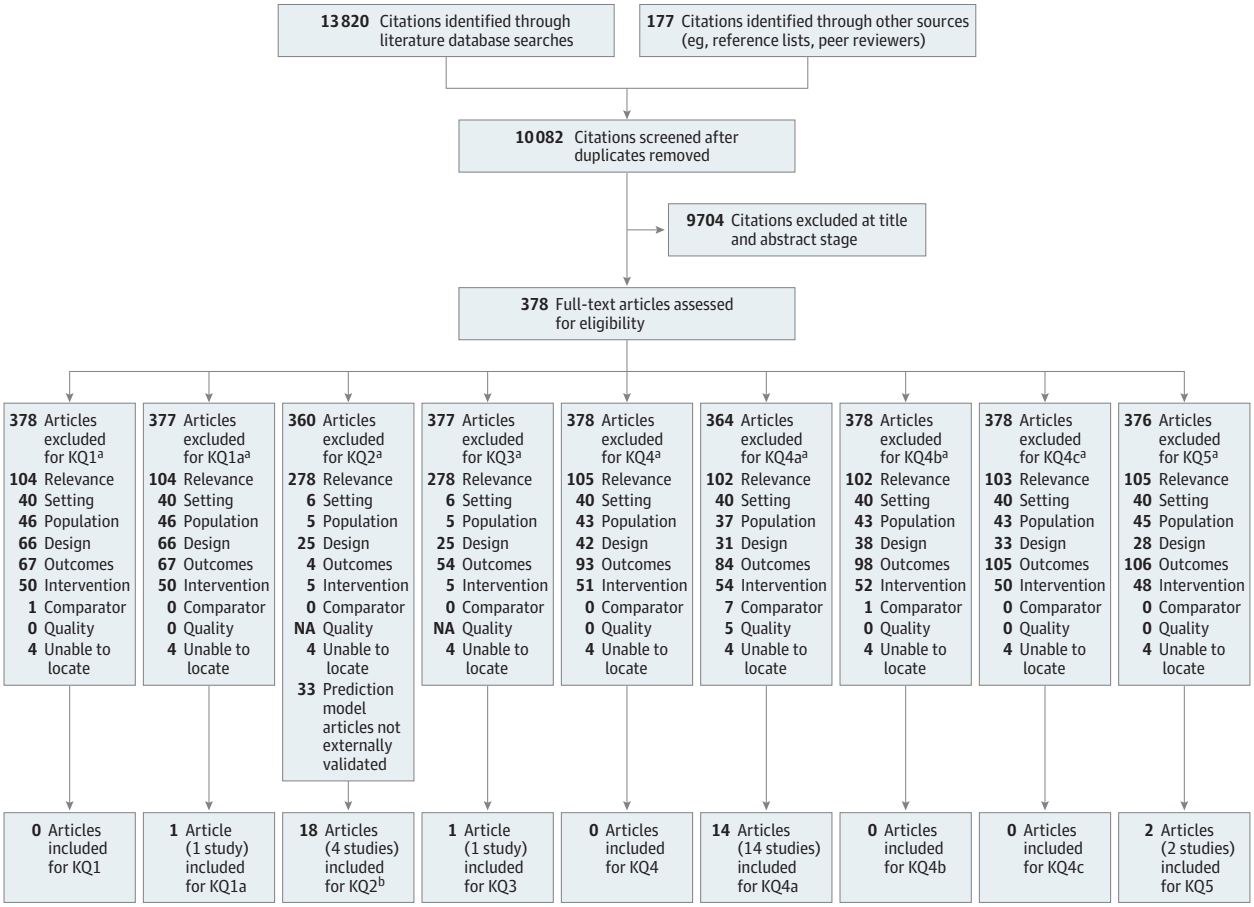
Summary evidence tables for each of the key questions include study population characteristics, study design features, and findings. Statistical pooling of results with meta-analysis was not possible for any outcomes because of statistical and clinical heterogeneity due to different study designs, interventions, reference standards, and populations.

Synthesis of included prediction models was informed by methodologic guidance for evaluating performance of multivariable risk prediction models.[25-29] Model performance was evaluated based on commonly recognized metrics. These include discrimination (*c* statistic), or area under a receiver operating characteristic curve plot, representing the probability that a case will have a higher risk score than a noncase. Sensitivity, specificity, positive predictive values (PPVs), and negative predictive values also measure discrimination. A priori risk-level cutpoints are optimal, but in the preeclampsia prediction literature "detection rates," analogous to sensitivity, were commonly reported, with risk cutpoints corresponding to a 10% false-positive rate (90% specificity).[27] Calibration reflects the extent to which the model predictions match the observed outcomes for individuals across different risk levels; goodness-of-fit tests (eg, Hosmer-Lemeshow test) are sometimes reported, but calibration plots that graphically depict the observed outcome frequencies against predicted probabilities are more informative.[28] Discrimination and calibration are both necessary for evaluating model performance in validation studies.[28] The models we identified with good or better discrimination based on the *c* statistic (≥0.80)[30] are described in this review. Models were classified as to whether they aimed to predict preeclampsia requiring early delivery (<34 weeks' gestation) or a later-onset diagnosis (≥34 weeks' gestation).

Results

Twenty-one included studies comprising 13 982 participants and reported in 35 publications were identified (Figure 2).[31-65] No studies directly compared the effectiveness of preeclampsia screening in a screened population vs an unscreened population (KQ1). One RCT[31] (n = 2764) on the benefits and harms of a reduced prenatal visit schedule (KQ1a, KQ5) and 1 observational before-after study (n = 1952) for potential harms of an indicated rather than routine

Figure 2. Literature Flow Diagram



NA indicates not applicable.

[a] Details about reasons for exclusion are as follows. Relevance: Study aim not relevant. Setting: Study was not conducted in a setting or country relevant to US primary care. Population: Study was not conducted in women and adolescents without a diagnosis of preeclampsia and asymptomatic for the condition. Design: Study did not use an included design. Outcomes: Study did not have relevant outcomes or had incomplete outcomes. Intervention: Study used an excluded intervention/screening approach. Comparator: Study lacked a comparison group. Quality: Study did not meet criteria for fair or good quality. Unable to locate: Library services could not locate article in which study was published.

[b] Included 4 studies (7 articles) on externally validated risk prediction models. Eleven articles were also identified that represent the model development studies related to external validation studies.

protein urine screening protocol (KQ5)[50] were included. Four external validation studies (n = 7123) evaluating 1 or more multivariable models for predicting preeclampsia risk (16 models total, 5 with good or better discrimination) were included (KQ2).[32-35] A single observational study (n = 255) evaluating the harms of risk prediction (KQ3) was included.[36] No studies evaluated the overall, protocol-specific, or risk-based effectiveness of screening tests for identifying women with preeclampsia (KQ4, KQ4b, KQ4c). Although no studies of the test accuracy of proteinuria screening were found among general prenatal care populations (KQ4a), 14 studies (n = 1888) examining the diagnostic accuracy of urine tests for proteinuria among women being evaluated for suspected preeclampsia were included to approximate test performance.[37-49,54]

Screening Effectiveness

Key Question 1a. Does preeclampsia screening effectiveness differ by screening protocol (eg, tests used, timing of tests, rescreen intervals) or preeclampsia risk status?

One fair-quality RCT conducted from 1992 to 1994 in a large managed care organization randomized 2764 pregnant women ages 18 to 39 presenting for prenatal care to a routine number of prenatal care visits (14 visits) or a schedule of fewer visits (9 visits) (eTable 2 in the Supplement).[31] A total of 2328 women completed the study: 1163 in the control group and 1165 in the intervention group. The study enrolled women at low risk for preeclampsia presenting for prenatal care before 13 weeks' gestation. Routine prenatal care consisted of visits that included screening for preeclampsia with blood pressure measurement and point-of-care proteinuria testing every 4 weeks between 8 and 28 weeks' gestation, then every 2 weeks until 36 weeks' gestation, then weekly until delivery for a total of 14 prenatal care visits. For the intervention, the study aimed to reduce the number of visits to 9 (at 8, 12, 16, 24, 28, 32, 36, 38, and 40 weeks of gestation). At baseline, there were no statistically significant differences between groups on maternal characteristics. During pregnancy, women in the control group had more health care visits in total ($P < .001$) than women in the intervention group, but

Table 1. Differences in Health Outcomes During Pregnancy, at Time of Delivery, or 6 Weeks Postpartum (Key Question 1a and Key Question 5)

Source, Study Design[a]	Outcome Category	Outcome	No. With Outcome/Sample Size (%)[b]		RR (95% CI)	*P* Value
			Intervention	Control		
McDuffie et al,[31] 1996 RCT[c]	Preeclampsia	Mild preeclampsia	59/1165 (5.1)	66/1163 (5.7)	0.94 (0.78-1.14)	.74
		Severe preeclampsia	10/1165 (0.9)	9/1163 (0.8)	1.05 (0.68-1.62)	.41
	Preterm birth	Preterm delivery <32 wk	10/1165 (0.9)	8/1163 (0.7)	1.11 (0.73-1.68)	.32
		Preterm delivery <37 wk	73/1165 (6.3)	63/1163 (5.4)	1.08 (0.92-1.27)	.19
	Delivery complications	Placenta abruptio	17/1165 (1.5)	11/1163 (0.9)	1.21 (0.90-1.64)	.13
		Apgar score at 5 min <7	18/1175 (1.6)	29/1176 (2.5)	0.77 (0.53-1.10)	.95
		Chorioamnionitis	9/1165 (0.8)	11/1163 (0.9)	0.90 (0.55-1.46)	.68
		Placenta previa	7/1165 (0.6)	9/1163 (0.8)	0.87 (0.50-1.52)	.70
		Postpartum hemorrhage with cesarean delivery	2/1165 (1.3)	3/1163 (2.2)	0.77 (0.26-2.27)	.77
		Postpartum hemorrhage with vaginal delivery	32/1165 (3.2)	33/1163 (3.2)	0.98 (0.77-1.27)	.47
		Preterm labor	79/1165 (6.8)	77/1163 (6.6)	1.01 (0.86-1.18)	.44
		Preterm premature rupture of membranes	38/1165 (3.3)	38/1163 (3.3)	1.00 (0.80-1.25)	.50
	Cesarean delivery	Cesarean delivery, overall	15/1165 (13.0)	140/1163 (12.0)	1.04 (0.93-1.17)	.25
	Perinatal and neonatal mortality	Stillbirth	5/1175 (0.4)	5/1176 (0.4)	1.00 (0.54-1.86)	.50
	Birthweight	Birthweight, mean (SD), g	3286/1175 (520)	3295/1176 (536)	NR	.66
		Very low birthweight (<1500 g), No. (%)	7/1175 (0.3)	6/1176 (0.3)	1.08 (0.65-1.79)	.39
		Low birthweight (<2500 g), No. (%)	64/1175 (5.4)	72/1176 (6.1)	0.94 (0.78-1.12)	.76
		Small for gestational age, No. (%)	36/1175 (3.1)	28/1176 (2.4)	1.13 (0.91-1.41)	.16
	Health care use during pregnancy	Total No. of visits, mean (SD)	12.0/1165 (4.2)	14.7/1163 (4.2)	NR	<.001
	Satisfaction with prenatal care at 6 wk postpartum	No. of prenatal visits, just right	494/589 (89.2)	473/600 (82.8)	NR	.002
		No. of prenatal visits, too few	49/589 (8.8)	6/600 (1.1)	NR	NR
		No. of prenatal visits, too many	11/589 (2.0)	92/600 (16.1)	NR	NR
		Quality of prenatal care, excellent or good	574/589 (97.5)	587/600 (97.8)	NR	.67
Rhode et al,[50] 2007 Before-after[d,e]	Preeclampsia	Preeclampsia or eclampsia	23/1019 (2.3)	36/933 (3.8)	NR	.001
	Preterm birth	Preterm delivery	50/1019 (4.9)	72/933 (7.7)	NR	.14
	Cesarean delivery	Cesarean delivery	181/1019 (17.8)	173/933 (18.5)	NR	.03
	Other maternal morbidity	Cystitis	33/1019 (3.3)	15/933 (1.7)	NR	<.001
		Gestational diabetes	42/998 (4.2)	81/874 (9.3)	NR	.82
		Gestational hypertension	58/1019 (5.7)	38/933 (4.1)	NR	<.001
		High blood pressure	81/1019 (8.0)	74/933 (7.9)	NR	<.001
		Pyelonephritis	4/1019 (0.40)	4/933 (0.40)	NR	<.001
		Asymptomatic bacteriuria	67/1019 (6.8)	79/933 (8.7)	NR	.05
		Urinary tract infection	141/933 (14.2)	140/907 (15.4)	NR	.04

Abbreviations: NR, not reported; RCT, randomized clinical trial; RR, relative risk.

[a] Both studies were of fair quality. For McDuffie et al,[31] quality was assessed using criteria developed by the US Preventive Services Task Force.[14] For Rhode et al,[50] quality was assessed using the Before-After Quality Assessment Tool.[24]

[b] Unless otherwise noted.

[c] Compared a reduced number of perinatal visits (9 prenatal visits) to usual care (14 prenatal visits).

[d] Rhode et al[50] used statistical tests for noninferiority. $P < .05$ indicates that rates are statistically equivalent (no greater than .04 in 1 direction).

[e] Compared women who were enrolled and delivered prior to August 15, 2002, who underwent routine urine screening, with women who were enrolled and delivered on or after August 15, 2002, who underwent indicated urine testing.

the mean difference in the number of visits between the 2 study groups was smaller than intended (12.0 [SD, 4.2] vs 14.7 [SD, 4.2]; $P < .001$) (**Table 1**). There were no statistically significant differences between groups on maternal health outcomes (eg, gestational diabetes, preeclampsia), delivery complications (eg, preterm delivery, cesarean delivery, postpartum hemorrhage), or neonatal outcomes (eg, birthweight, gestational age, stillbirth).

Screening Harms

Key Question 5. What are the harms of preeclampsia screening, and do they differ by risk status or screening protocol?

The same fair-quality trial (n = 2764) included for KQ1a found no difference in birth outcomes (eg, low birth weight, preterm birth, number of cesarean deliveries) with an intended reduction in the number of prenatal care visits (eTable 2 in the Supplement).[31] Power

was insufficient to detect differences for rare outcomes related to preeclampsia, particularly serious adverse maternal events such as progression to eclampsia, organ failure, stroke, and death.

An additional fair-quality retrospective before-and-after comparison cohort study (n = 1952) evaluated differences in health outcomes after a change in practice at a hospital-based nurse midwifery practice that primarily served low-income Hispanic women (74% of eligible study participants). The practice change was from routine prenatal dipstick urine testing to "clinically indicated" urine testing (eTable 5 in the Supplement).[50] All women in the study received urine tests at their first prenatal visit; those delivered before August 15, 2002 (n = 933), received routine urine screening with chemical reagent strips testing for bacteria or protein at all subsequent visits, whereas those delivered after August 15, 2002 (n = 1019), had subsequent urine screening only for certain conditions. Indications for urine testing were symptoms of a urinary tract infection, severe vomiting, weight loss of 0.9 kg or more since the previous visit, systolic blood pressure 140 mm Hg or higher, diastolic blood pressure 90 mm Hg or higher, or a health condition requiring periodic urine testing (eg, chronic hypertension, renal disease). Women in the routine urine testing group had used an average of 7.8 (range, 0-19) tests, whereas women in the indicated testing group had used an average of 1.4 (range, 0-16). Among the indicated testing group, the reasons for urine testing were urinary tract infection or vaginitis symptoms (31.5%) and elevated blood pressure or significant preeclampsia-related symptoms (35.6%).

The purpose of the study was to evaluate whether changes in the urine screening approach were safe; thus, statistical tests were designed to evaluate noninferiority—statistically significant *P* values indicated no difference between the 2 groups (Table 1). The study reported equivalence in the rates of diagnosis for preeclampsia/eclampsia, high blood pressure, and cesarean deliveries. Preterm delivery was not equivalent between groups, but the rate was higher in the routine testing group, supporting the noninferiority of indicated testing.

Similar to the evidence on benefits, for harms the absence of adequately powered studies, conducted more recently in broader prenatal care populations, limits the conclusions that can be drawn to evaluate preeclampsia screening protocols.

Accuracy of Screening Tests

Key Question 4a. How accurate are different point-of-care screening tests for proteinuria (a diagnostic criterion for preeclampsia)?

Fourteen studies (n = 1888) examined the diagnostic accuracy of urine tests for proteinuria among women being evaluated for suspected preeclampsia on the basis of positive point-of-care urine test results, high blood pressure, symptoms, or for undefined reasons (eTable 4 in the Supplement).[37-49,54] Six studies were conducted in the United States,[38,39,43,44,48,49] 4 in the United Kingdom,[37,42,46,47] 1 in New Zealand,[40] 1 in Canada,[41] 1 in Chile,[54] and 1 in the Netherlands.[45]

Twelve of the studies evaluated the accuracy of urine tests for protein to creatinine ratio (**Table 2**) in 1516 pregnant women.[37-45,48,49,54] The test sensitivities ranged from 65% (95% CI not calculable)[49] to 96% (95% CI, 88%-99%),[45] with most falling above 81% (**Figure 3**). Limited information on the specific protein to creatinine ratio index test used, differing test thresholds, and diverse study enrollment criteria, along with the dispersion of study data points, account for considerable clinical and statistical heterogeneity for diagnostic accuracy. Summary conclusions about overall performance could not be drawn.

Two studies evaluated the accuracy of urine tests for albumin to creatinine ratio using the DCA 2000 point-of-care system (Bayer Healthcare) in 321 pregnant women[40,47] (Table 2). The sensitivities were high (>90%), but specificities differed (Figure 3). In 1 study with high proteinuria prevalence[47] (45%), specificity remained high (>90%), but in the study with lower proteinuria prevalence[40] (8.7%), specificity was lower (<70%).

Four studies evaluated the accuracy of protein urine dipsticks in 634 pregnant women with mixed test performance characteristics.[39,40,46,47] The studies used dipsticks of different makes and models (Table 2), but all studies used the same reference standard. Sensitivities ranged from 22% to 100% and specificities from 36% to 100% (Figure 3). One dipstick test in a good-quality, high-proteinuria prevalence study[47] had both sensitivity and specificity above 80% for automated reading (with Clinitek 50) of the Multistix 8SG dipstick.[47] All other studies had very high sensitivity and low specificity or vice versa.

Likely owing to diversity of index tests used, study eligibility criteria, and proteinuria prevalence, there was considerable variation in performance of urine screening tests for protein. No evidence was found to estimate the accuracy of urine protein screening tests among healthy prenatal populations.

Risk Prediction

Key Question 2. What is the effectiveness of risk prediction in early pregnancy for identifying women at high risk for preeclampsia?

Four external validation studies (n = 7123) reported on the performance of 16 distinct risk prediction models in 7 articles (**Table 3**).[32-35,51-53] The outcomes differed: 6 models were developed for prediction of preeclampsia requiring delivery before 34 weeks of gestation,[55,57,59-62] 1 before 37 weeks of gestation,[58] 7 after 34 weeks of gestation,[55,56,60,62-65] and 2 predicting any preeclampsia.[58,63] Preterm preeclampsia is rare, so outcome prevalence was for models predicting later or any preeclampsia. An additional 11 articles reported on the model development studies related to these external validations.[55-65] Five of the externally validated models had *c* statistics indicating good or better discrimination (≥0.80) (**Table 4**).[55-58,66] The models were labeled with lead authors of the model development studies cited in external validation. Most of the included models were developed in the United Kingdom, with overlap in the investigators and funding source.

The models were validated with prospective cohort data collected in the United States by Oliveira et al (n = 2962),[33] Australia by Park et al (n = 3014),[34] Italy by Farina et al (n = 554),[32] and Norway by Skråstad et al (n = 541).[35] The timing of risk calculation occurred before 20 weeks' gestation for all models but varied depending on the gestation at which women presented and on the availability of variables needed for the model. The validation studies by Farina et al and Park et al enrolled women with singleton pregnancies presenting for aneuploidy screening; the validation study by Oliveira et al enrolled women with singleton pregnancies presenting for prenatal care in the first trimester; and the validation study by Skråstad et al enrolled nulliparous women, resulting in a slightly younger cohort (mean age, 26 years). All of the cohorts were enrolled sometime between 2007 and 2012.

The validation cohort study[33] most applicable to US primary care settings enrolled average-risk women presenting for prenatal care in the first trimester at 4 health centers in Baltimore, Maryland, and was used to evaluate a model by Poon et al (early preeclampsia)[55]

Table 2. Index Tests and Reference Standard Characteristics of Included Diagnostic Accuracy Studies (Key Question 4a) Sorted by Index Test and Threshold

		Index Test				Reference Standard		
Source	Quality[a]	Make, Manufacturer	Sampling Methods	Assay (Machine)	Operator and Reader	24-h Collection Method	Assay (Machine)	Operator and Reader
Protein Creatinine Spot Tests Only								
Young et al,[49] 1996	Fair	Protein to creatinine ratio spot test (NR)	One sample taken prior to initiation of 24-h collection (in 66% of samples) or immediately of the 24-h collection (34% of samples); none were first morning void	Standard spectrophotometric technique (Beckman analyzer)	NR	Avoided the first void sample	NR (NR)	NR
Durnwald and Mercer,[38] 2003	Fair	Protein to creatinine ratio spot ratio (NR)	One sample taken prior to initiation of 24-h collection	Biuret reaction; modified Jaffe reaction (NR)	NR	Outpatients collected urine in container; urine from select inpatients collected by Foley catheter	NR (NR)	NR
Wheeler et al,[48] 2007	Fair	Protein to creatinine ratio spot test (NR)	One sample taken prior to initiation of 24-h collection	Johnson & Biuret method (protein); 2-point rate method (creatinine) (Johnson Vitros 250)	NR	NR	NR (NR)	NR
Sethuram et al,[42] 2011	Fair	Protein to creatinine ratio spot test (NR)	One 10-mL sample taken prior to initiation of 24-h collection; avoided the first void sample	Benzethonium chloride turbidometric method (protein); Jaffe method (creatinine) (Abbott Diagnostics analyzer)	NR	Avoided the first void sample	Benzethonium chloride turbidometric method (protein) (Abbott Diagnostics analyzer)	NR
Tun et al,[44] 2012	Fair	Protein to creatinine ratio spot test (NR)	One sample from 24-h timed specimen or immediately after timed collection	NR (NR)	NR, sent to Health Network Laboratories	Started at time of admission; collected in 2 consecutive 12-h collections	ADVIA total protein urine assay (modified Fujita method) (NR)	NR, sent to Health Network Laboratory
Stout et al,[43] 2013	Fair	Protein to creatinine ratio spot (NR)	One sample taken prior to initiation of 24-h collection	Enzymatic creatinase (NR)	NR	Used first 24-h urine collection	Benzethonium chloride (NR)	NR
Lamontagne et al,[41] 2014	Good	Protein to creatinine ratio spot ratio (NR)	One sample taken prior to initiation of 24-h collection	Colorimetric method using pyrogallol red-molybdate (protein); Jaffe method (creatinine) (Beckman Coulter multianalyzer with Synchron LX system)	NR	Not collected with catheter	Colorimetric method using pyrogallol red-molybdate (protein); Jaffe method (creatinine) (Beckman Coulter multianalyzer with Synchron LX system)	NR
Verdonk et al,[45] 2014	Good	Protein to creatinine ratio spot test (Albustix, Siemens Healthcare Diagnostics)	5-mL samples at 8 AM, 12 PM (noon), and 5 PM	Enzymatic assay (creatinine), colorimetric assay (protein) (CREA plus, Roche Diagnostics)	NR	Began at midnight, restarted at midnight the next day if collection errors occurred	NR (NR)	NR
Bhide et al,[37] 2015	Fair	Protein to creatinine ratio spot test (NR)	One sample taken prior to initiation of 24-h collection	Pyragallol red (protein), Jaffe kinetic method (creatinine) (NR)	NR	NR	Pyragallol red (protein) (NR	NR
Valdés et al,[54] 2015	Fair	Protein to creatinine ratio spot test (NR)	One sample (15-20 mL)	NR (NR)	NR	Initiated on admission	NR (NR)	NR

(continued)

Table 2. Index Tests and Reference Standard Characteristics of Included Diagnostic Accuracy Studies (Key Question 4a) Sorted by Index Test and Threshold (continued)

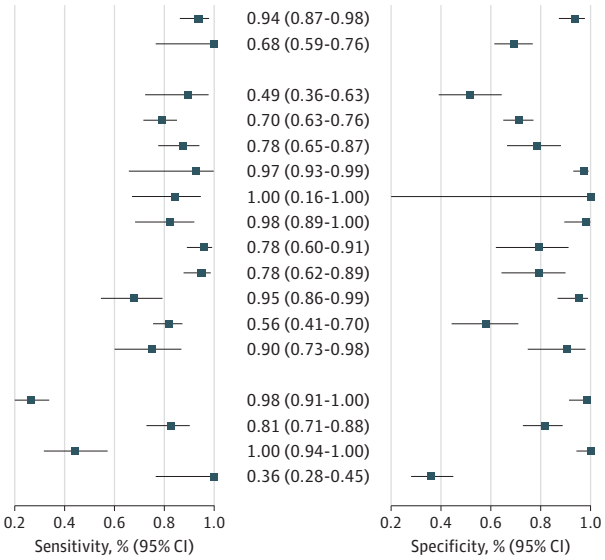
		Index Test				Reference Standard		
Source	Quality[a]	Make, Manufacturer	Sampling Methods	Assay (Machine)	Operator and Reader	24-h Collection Method	Assay (Machine)	Operator and Reader
Multiple Spot Urine Tests (Protein to Creatinine or Albumin to Creatinine Ratio) or Dipsticks								
Waugh et al,[47] 2005	Good	Dipstick-Clinitek Microalbumin (automated) (Clinitek Microalbumin, Bayer)	One 10-mL sample from early morning void before final 24-h specimen was added to the collection	Two semiquantitative immunoassays for albumin and creatinine (Clinitek 50)	NR	On waking, the first void discarded; 24-h collection started with the second urine specimen; the final specimen was the first void the following day	Benzethonium chloride assay (NR)	NR
		Dipstick-Microalbustix (visual) (Microalbustix, Bayer)		Two semiquantitative immunoassays for albumin and creatinine (NA)	Two observers			
		DCA 2000-point-of-care test (DCA 2000, Bayer)		Immunoturbidometric assay (albumin), colorimetric assay (creatinine) (NR)	NR			
		Dipstick-Multistix 8SG (visual) (Multistix 8SG, Bayer)		NR (NR)	Two observers			
		Dipstick-Multistix 8SG (automated) (Multistix 8SG, Bayer)		NR (Clinitek 50)	NR			
Dwyer et al,[39] 2008	Good	Protein to creatinine ratio spot test (NR)	One sample taken prior to initiation of 24-h collection either via clean catch or by catheter	Pyrogallol red/molybdate (protein); Jaffe rate (creatinine) (Synchron LX Systems)	Laboratory technician	24-h collection via either clean catch or catheter	NR (NR)	Laboratory technician
		Protein to creatinine ratio automated dipstick (Iris test strips, IRIS Inc or Arcray Inc)		3′3″5′5″ tetrachlorophenol-3,4,5,6-tetrabromophenolphthalein (protein error of pH indicator) (Autoanalyzers)				
Kyle et al,[40] 2008	Fair	Protein to creatinine ratio spot test (NR)	One sample taken prior to initiation of 24-h collection	NR (Abbott Ci8200 Analyzer)	Research midwife	First void discarded	Benzethonium chloride assay (NR)	NR
		Albumin/creatinine spot test (DCA 2000, Bayer Healthcare LLC)		NR (NR)				
		Dipstick (NR)		NR (NR)				
Protein Dipsticks Only								
Waugh et al,[46] 2001	Fair	Dipstick (BM-Test-5L, Boehringer Mannheim UK)	Two 10-mL aliquots of thoroughly mixed urine from 24-h collection	NR (NR)	Observer	Initiated at 8 AM	Benzethonium chloride or Bradford assay (ExcelGel with silver staining kit)	NR

Abbreviation: NR, not reported.

[a] Quality assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool.[23]

Figure 3. Diagnostic Accuracy of Point-of-Care Tests for Proteinuria (Key Question 4a)

Source	Threshold	Participants, No.	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Albumin to protein ratio spot test				
Waugh et al,[47] 2005	2.0 mg/mmol	171	0.94 (0.85-0.98)	0.94 (0.87-0.98)
Kyle et al,[40] 2008	≥2.0 mg/mmol	150	1.00 (0.75-1.00)	0.68 (0.59-0.76)
Protein to creatinine ratio spot test				
Tun et al,[44] 2012	17.0 mg/mmol	86	0.89 (0.71-0.98)	0.49 (0.36-0.63)
Stout et al,[43] 2013	>21.5 mg/mmol	356	0.78 (0.70-0.84)	0.70 (0.63-0.76)
Wheeler et al,[48] 2007	23.7 mg/mmol	126	0.87 (0.76-0.94)	0.78 (0.65-0.87)
Kyle et al,[40] 2008	≥30 mg/mmol	150	0.92 (0.64-1.00)	0.97 (0.93-0.99)
Sethuram et al,[42] 2011	>30 mg/mmol	32	0.83 (0.65-0.94)	1.00 (0.16-1.00)
Lamontagne et al,[41] 2014	≥30 mg/mmol	91	0.81 (0.67-0.92)	0.98 (0.89-1.00)
Verdonk et al,[45] 2014	≥30 mg/mmol	105	0.96 (0.88-0.99)	0.78 (0.60-0.91)
Bhide et al,[37] 2015	≥30 mg/mmol	117	0.95 (0.87-0.99)	0.78 (0.62-0.89)
Dwyer et al,[39] 2008	≥31.6 mg/mmol	116	0.66 (0.52-0.78)	0.95 (0.86-0.99)
Dumwald and Mercer,[38] 2003	≥33.9 mg/mmol	220	0.81 (0.74-0.87)	0.56 (0.41-0.70)
Valdés et al,[54] 2015	≥40.7 mg/mmol	72	0.74 (0.58-0.86)	0.90 (0.73-0.98)
Urine dipstick				
Waugh et al,[46] 2001	1+	197	0.22 (0.16-0.30)	0.98 (0.91-1.00)
Waugh et al,[47] 2005	1+	171	0.82 (0.71-0.90)	0.81 (0.71-0.88)
Dwyer et al,[39] 2008	1+	116	0.41 (0.28-0.55)	1.00 (0.94-1.00)
Kyle et al,[40] 2008	1+	150	1.00 (0.75-1.00)	0.36 (0.28-0.45)



One study[49] is not plotted, as it did not provide enough information to construct a 2 × 2 table.

Table 3. Study Characteristics of Prospective Cohort Studies Used for External Validation of Preeclampsia Risk Prediction Models (Key Question 2)

Source	Location	Location of Model Development	Study Population	Study Period	Sample Size, No.	Outcome Prevalence, No. (%)[a,b]	Funding
Farina et al,[32] 2011	Bologna, Italy	London, United Kingdom[55,56]	Women with singleton pregnancies enrolled at screening visit for early diagnosis of chromosomal and other fetal abnormalities, and delivery in tertiary care center	December 2007-April 2010	554	Late preeclampsia: 7.0 (39 cases)	Ricerca Fondamentale Orientata
Park et al,[34] 2013	Sydney, Australia	London, United Kingdom[55]	Women with singleton pregnancies presenting for aneuploidy screening	April 2010-March 2012	3066	Early preeclampsia: 0.4 (12 cases)	NR
Oliveira et al,[33] 2014	Baltimore, Maryland	London, United Kingdom[55]; St Louis, Missouri[57]	Women with singleton pregnancies	2007-2010	871-2962[c]	Early preeclampsia: 1.0-1.2 (10-30 cases) Late preeclampsia: 4.1-5.0 (78-116 cases)	Diagnostic Technologies Limited
Skråstad et al,[35] 2014	Trondheim, Norway	London and Gillingham, United Kingdom[58]	Nulliparous women	September 2010-March 2012	541	Any preeclampsia: 3.9 (21 cases) Preterm preeclampsia requiring delivery (<37 wk): 0.9 (5 cases)	Norwegian University of Science and Technology; National Center for Fetal Medicine

Abbreviation: NR, not reported.

[a] Early preeclampsia is defined as preeclampsia that occurs before 34 weeks' gestation.

[b] Late preeclampsia is defined as preeclampsia that occurs later in pregnancy (34 weeks' gestation or later).

[c] Model N depended on availability of variable needed for each predictive model.

and the model by Odibo et al.[57] The model by Poon et al[55] was the only one externally validated in more than 1 setting including the United States.[33,34] In the US validation study (n = 2833),[33] discrimination was moderate (*c* statistic, 0.80 [95% CI, 0.71-0.89]), and detection (52%) and PPV (4.2) were low, based on 29 cases (1% incidence); in the Australian validation cohort of women with singleton pregnancies attending aneuploidy screening (n = 3014),[34] discrimination was high (*c* statistic, 0.93 [95% CI, 0.92-0.94]), as was detection (91.7% [95% CI, 61.5-98.6]), but the PPV was low (3.6), based on only 12 cases.

The model by Odibo et al[57] had better discrimination and detection in the US validation cohort and had been developed in a US population. The model used clinical history, placental protein 13, pregnancy-associated plasma protein A, and the mean artery pulsatility index to predict preeclampsia-required delivery before 34 weeks' gestation (*c* statistic, 0.86 [95% CI, 0.73-0.99]). The model was validated with a smaller subset of the US cohort (n = 871; 29% of the 2969 women in the external validation cohort[33]), because not all women had data on a serum marker needed for the model.

Table 4. External Validation Performance of 5 Preeclampsia Risk Prediction Models With Good or Better Discrimination (*c* Statistic >0.80) (Key Question 2)

	Preeclampsia Requiring Early Delivery				Late Preeclampsia Diagnosis	
Descriptor	Poon et al,[55] 2010[a,b]	Poon et al,[55] 2010[a,b]	Odibo et al,[57] 2011	Akolekar et al,[58] 2013[a]	Onwudiwe et al,[56] 2008[a]	Poon et al,[55] 2010[a,b]
Model variables	Race, chronic hypertension history, conception mode, parity, MAP, PAPP-A, Doppler ultrasound UtA-PI	Race, chronic hypertension history, conception mode, parity, MAP, PAPP-A, Doppler ultrasound UtA-PI	Chronic hypertension, PAPP-A, PP-13, Doppler ultrasound UtA-PI	Age, weight, height, race/ethnicity, personal preeclampsia history, mother's preeclampsia history, parity, mode of conception, chronic health conditions, MAP, PAPP-A, PIGF, Doppler ultrasound UtA-PI	BMI, race/ethnicity, parity, personal preeclampsia history, MAP, Doppler ultrasound UtA-PI	Age, BMI, race, mother preeclampsia history, parity, MAP, Doppler ultrasound UtA-PI
External validation study	Oliveira et al,[33] 2014	Park et al,[34] 2013	Oliveira et al,[33] 2014	Skråstad et al,[35] 2014	Farina et al,[32] 2011	Farina et al,[32] 2011
No. of participants eligible for model validation cohort[c]	2833	3014	871	541	554	554
With preeclampsia outcome, No. (%)	29 (1.0)	12 (0.4)	10 (1.1)	5 (0.9)	39 (7.0)	39 (7.0)
Preeclampsia timing, wk[d]	<34	<34	<34	<37	>34	>34
c Statistic (95% CI)[e]	0.80 (0.71-0.89)	0.93 (0.92-0.94)	0.86 (0.73-0.99)	0.94 (0.86-1.00)	0.85 (0.78-0.93)	0.93 (0.88-0.98)
Calibration[f]	NR	NR	NR	NR	NR	NR
Detection, % (95% CI)[g]	52 (CI NR)	91.7 (61.5-98.6)	80 (CI NR)	80.0 (28.4-99.5)	74.4 (60.7-88.1)	84.6 (73.3-95.9)
PPV (95% CI)[h]	4.2 (2.6-6.5)	3.6 (2.0-7.0)	11.3 (5.3-21.5)	6.8 (1.9-16.5)	36.3	39.3
NPV (95% CI)[i]	99.6 (99.0-100.0)	99.9 (99.7-99.9)	99.8 (99.0-100.0)	99.8 (98.8-100.0)	97.9	98.7

Abbreviations: BMI, body mass index; MAP, mean arterial pressure; NPV, negative predictive value; NR, not reported; PAPP-A, pregnancy-associated plasma protein A; PIGF, placental growth factor; PPV, positive predictive value; UtA-PI, uterine artery pulsatility index.

[a] Model development supported by investigators and funds from the Fetal Medicine Foundation. Additional descriptive information about the model available in Wright et al.[66]

[b] Clinical history algorithm described in Poon et al.[55,59,67]

[c] Model N depended on availability of variable needed for each predictive model.

[d] Preeclampsia defined as requiring delivery, with the exception of the Farina external validation, which defined the outcome as the diagnosis of preeclampsia.

[e] A test performance statistic (equivalent to the area under the curve) used to assess discrimination, a model performance measure that refers to how well a model differentiates between persons with and without the outcome.[27]

[f] A model performance measure that refers to how well predicted risks compare with observed outcomes, preferably evaluated graphically by calibration plots and supplemented by a formal statistical test (Hosmer-Lemeshow test for logistic regression and its equivalent for Cox regression).[27]

[g] Analogous to sensitivity. The percentage of cases correctly classified based on a predefined false-positive risk threshold.[27] Detection for preeclampsia in this table was based on a fixed 10% false-positive rate (risk cutpoint for 90% specificity), which was the most commonly reported.

[h] A test performance statistic used to measure what proportion of patients who test positive have the disease. Not reported in the Farina external validation study; calculated by hand.

[i] A test performance statistic used to measure what proportion of patients who test negative do not have the disease. Not reported in the Farina external validation study; calculated by hand.

The model developed by Akolekar et al[58,66] and validated by Skråstad et al[35] was used to predict any preeclampsia requiring delivery before 37 weeks' gestation (*c* statistic, 0.94 [95% CI, 0.86-1.00]). There were 5 cases of early preeclampsia requiring delivery (incidence, 0.9%). Detection was 80%, and the PPV was 6.8. Two additional models used clinical history and uterine Doppler measures to detect later-onset preeclampsia, a more common outcome. These models, by Onwudiwe et al[56] and Poon et al (late preeclampsia),[55] had good or better discrimination, detection of 85% and 74%, and PPVs of 39.3 and 36.3, respectively, when validated in a small Italian cohort study (n = 554).[32]

Information on model calibration was not provided in any of the external validation studies, precluding a complete assessment of model performance. There were no randomized impact studies evaluating the health benefits or harms of risk assessment using multivariable prediction models compared with standard care.

Key Question 3. What are the harms of preeclampsia risk prediction?

One fair-quality, prospective cohort study (n = 255) conducted in Spain examined whether first-trimester risk prediction and clinical care based on risk status increased anxiety in pregnant women (eTable 3 in the Supplement).[36] Risk for early preeclampsia requiring delivery before 34 weeks was assessed using a model developed in Spain and with modest performance in the US-based validation study.[33] Pregnant women screened as high risk were recruited and matched with the next visiting low-risk screened woman in a first trimester screening unit (135 low risk, 120 high risk). After risk prediction, women received counseling on potential risks of preeclampsia. Women at high risk were followed up with

a protocol that included recommended daily intake of aspirin (150 mg) from the day of screening until 36 weeks' gestation and second-trimester ultrasonography at 20 to 22 weeks.[36] Low- and high-risk women did not differ in anxiety, measured with the Spielberg State-Trait Anxiety Inventory, at baseline or after risk prediction and counseling.[36] First-trimester preeclampsia risk prediction and counseling was not associated with greater maternal anxiety in the immediate short term when coupled with counseling and potentially preventive medication. The risk assessment model used and the content of the counseling were not well described, and the results may have limited generalizability.

Discussion

This review identified only 1 RCT, conducted more than 20 years ago, that compared different screening strategies and found no difference in benefits or harms from slightly fewer preeclampsia screening visits compared with the standard of care at the time.[31] The applicability of these findings to current practice settings and populations is limited, given changes to screening, diagnosis, and management practices, as well as population health, over the past 2 decades (**Table 5**). An observational study published in 2007 found no harms associated with indicated rather than routine urine testing for preeclampsia screening, but study design and setting limit applicability. Both of these studies were underpowered to assess very rare adverse events associated with preeclampsia, particularly serious maternal risks from eclampsia and stroke.

No studies directly evaluated the individual or combined test accuracy of blood pressure screening and urine protein screening for detecting the presence or absence of preeclampsia at a single point in time or cumulatively across pregnancy. Evidence to estimate the frequency of false-positive and false-negative readings for elevated blood pressure and proteinuria was not found. Understanding the optimal use of low-resource and relatively noninvasive screening and confirmatory tests (eg, additional blood pressure measurements, repeat point-of-care urine testing, diagnostic urine testing) has likely been a lower priority research question than those concerned with etiology and treatment of preeclampsia. Studies of current prenatal care populations, with a higher prevalence of obesity and other preeclampsia risk factors, are needed to derive more complete evidence-based approaches to preeclampsia screening. This is particularly important in the context of recent changes to diagnostic criteria that support additional tests for women with hypertension in the absence of proteinuria.[68]

This review did not include evidence on associations of blood pressure and proteinuria levels and the likelihood of developing preeclampsia later in the pregnancy. Such studies are important for establishing diagnostic criteria and identifying candidate markers for risk prediction but do not address the question of screening effectiveness, which this review aimed to evaluate. Chronic hypertension is associated with increased probability of developing preeclampsia,[69] and high blood pressure occurring for the first time during pregnancy is one of the diagnostic criteria for preeclampsia.[1] Thus, the current clinical practice of repeat measurement at clinical visits remains important for all pregnant women. The accuracy of individual blood pressure readings is optimized if conducted in accordance with guidance on clinical blood pressure measurement in general[70] and during pregnancy.[71,72]

Associations between proteinuria levels and adverse preeclampsia outcomes are less consistent.[73,74] Recognizing this, updated American College of Obstetricians and Gynecologists (ACOG) guidelines address other signs and symptoms that may be used to diagnose preeclampsia in the absence of proteinuria, including cerebral or visual symptoms, impaired liver or renal function, low platelet count, and pulmonary edema.[1] These changes to diagnostic criteria could increase the number of women identified with preeclampsia, require different approaches to diagnostic confirmation, and may lead to the development of new approaches to screening. Regardless of these changes, the presence of significant proteinuria remains a key diagnostic criterion for preeclampsia. Due to historical precedent and low resource requirements, urine dipstick testing is likely to continue, despite its recognized poor and variable test performance[75] (particularly with visual rather than automated readings[47,76]).

This review and others[74,77,78] identified a larger body of evidence on the performance of protein to creatinine ratio spot tests for detecting significant proteinuria. The high variability in the populations, tests evaluated, and accuracy limits conclusions that can be drawn regarding the optimal clinical application of these tests. Guidance from ACOG suggests that diagnosis can be based on the 24-hour protein test (>300 mg) or protein to creatinine ratio (>0.30). Guidance on the most accurate proteinuria screening approach is not provided by ACOG. Current evidence is not from a general screening population; rather, it is from among women with suspected preeclampsia and aims to determine whether protein to creatinine ratio spot tests are accurate enough to substitute for the more resource-intensive 24-hour collections, the reference standard for diagnosis of proteinuria in pregnancy.[74,77] The applicability of these results to point of care screening is limited. Urine tests for protein are conducted throughout pregnancy but practices may vary, as this is not a clinical standard per se but rather a long-standing practice tradition. False-positive results lead to further confirmatory testing and heightened surveillance. Maximizing single-test performance for a relatively inexpensive and noninvasive test may have limited value. Evidence on test performance in general prenatal care populations with repeat testing, and comparative studies of different approaches to screening, could better define the optimal role for point-of-care urine testing in routine preeclampsia screening or diagnostic evaluation.

This review sought multivariable risk prediction models that could be used for risk-based screening or for targeting other clinical preventive services, such as aspirin chemoprophylaxis for preeclampsia prevention.[22] Five of the 16 externally validated multivariable risk prediction models identified had good or better discrimination and PPVs ranging from 4% to 39%. Information on model calibration was not provided for any of the models, limiting evaluation of the likely performance or effect of clinical application. Moreover, serum markers and Doppler measures may not be routinely collected, limiting their feasibility for routine primary care risk assessment. Recent systematic reviews,[79,80] several methodological critiques,[26,81-83] and recent guidance from ACOG[84] support the findings of this review on the evidence limitations and absence of a well-supported model to be used in routine prenatal care for prediction of preeclampsia risk.

Table 5. Overall Summary of Evidence by Key Question

Key Question Topic	No. of Studies (Study Design), Sample Size	Quality	Limitations[a]	Consistency	US Primary Care Applicability	Summary of Findings[b]
KQ1: Preeclampsia screening effects on health outcomes	0	NA	NA	NA	NA	NA
KQ1a: Differences in effects of preeclampsia screening on health outcomes by protocol (eg, type of test; timing) or risk status	1 (RCT) n = 2764	Fair	Differences in number of visits between control and intervention groups not as pronounced as study design aim (12.0 vs 14.7 visits) Insufficient power for rare adverse outcomes (eg, very low birthweight, eclampsia, stillbirth)	NA	Low: Low-risk women seeking prenatal care in first trimester in a large US health maintenance organization Study nearly 20 y old; changes to clinical practice; insured women presenting for prenatal care early in pregnancy may differ from those presenting later	Fewer prenatal care visits did not have beneficial or harmful effects on health outcomes or diagnoses of preeclampsia Study has limited relevance to current clinical practice or population
KQ2: Preeclampsia multivariable risk prediction	Four external validation studies (prospective cohort); 16 externally validated risk prediction models; 5 externally validated models with good to excellent discrimination (*c* statistic ≥0.80) n = 541-3066	Review limited to models with external validation in prospective cohort studies[c]	Calibration not reported for any models; risk cutpoints not established a priori Several prediction models developed by the same team of investigators; Australian cohort external validation studies were not independent from model development group	Moderate or low: Only 1 model validated >1 cohort, both with positive predictive value 4%, and inconsistent discrimination and detection	Moderate: Some models validated in a diverse cohort of US women with singleton pregnancies seeking routine prenatal care potentially applicable, but effect of the model in other settings and in routine clinical use unknown	No applicable model supported by evidence of good performance or clinical benefits Important model calibration performance statistics not reported
KQ3: Harms of preeclampsia Multivariable risk prediction	1 (prospective cohort) n = 255 0	Fair	Risk prediction model not clearly described; intensive counseling and changes to clinical care based on risk not clearly described; unable to isolate effects of risk prediction from clinical interventions Insufficient power to assess potential health outcome harms of false-negative risk-prediction results	NA	Low: Study conducted in Spain and Italy among women undergoing aneuploidy screening Risk-prediction model used in study unlikely to be used in US clinical practice, based on external validation performance in US cohort[33] Specially trained midwives conducted risk-prediction counseling	Anxiety no different between women screened as at low and high risk for preeclampsia; study groups not equivalent. Comparisons of anxiety levels to women with false-negative risk prediction could not be assessed owing to timing of outcome measurement and insufficient power
KQ4: Effectiveness of screening tests in identifying women with preeclampsia	NA	NA	NA	NA	NA	NA
KQ4a: Diagnostic accuracy of urine tests for proteinuria	14 (diagnostic accuracy) n = 1888	Fair	Spectrum bias high; all studies among women with suspected preeclampsia (eg, de novo hypertension, ≥1+ dipstick) rather than general screening populations of pregnant women High heterogeneity across studies; limited information on tests, assays, and collection methods Few studies evaluating diagnostic accuracy of dipsticks and albumin to creatinine ratio spot urine tests; not possible to pool performance estimates	Moderate or low: Similar test accuracy for protein to creatinine and albumin to creatinine ratio spot urine test studies; dipstick test performance highly variable	Moderate: Six studies conducted in the United States; remainder conducted in European or South American countries with representative samples of pregnant women All studies conducted in women with suspected preeclampsia	Protein to creatinine ratio tests (n = 12) had the most evidence on test accuracy; sensitivity range, 65%-96% and specificity, 49%-100%; spectrum bias and high heterogeneity limit conclusions on test performance in routine clinical care Dipstick urine tests least accurate (n = 4); sensitivity range, 22%-100% and specificity, 36%-100% Albumin:creatinine spot urine tests (n = 2); sensitivity range, 94%-100%; specificity, 94% and 68%

(continued)

Table 5. Overall Summary of Evidence by Key Question (continued)

Key Question Topic	No. of Studies (Study Design), Sample Size	Quality	Limitations[a]	Consistency	US Primary Care Applicability	Summary of Findings[b]
KQ4b: Effectiveness of different screening tests in identifying women with preeclampsia	0	NA	NA	NA	NA	No studies found comparing performance of different approaches to routine preeclampsia screening Within-study comparisons from KQ4a-included studies suggest that automated readings more accurate than visual; urine samples taken at different times of day have similar performance (Albustix protein to creatinine ratio test); different assays for evaluating 24-h protein gold standard give different test sensitivity results
KQ4c: Effectiveness of different screening tests in identifying women at high or low risk for preeclampsia	0	NA	NA	NA	NA	NA
KQ5: Harms of preeclampsia screening	2 (1 RCT, 1 before-after) n = 4712	Fair	Insufficient power to detect rare adverse outcomes (eg, very low birthweight, stillbirth) RCT powered to detect differences of ≥2% between groups; 1% differences for some outcomes could be clinically important Statistical difference in the source of payment for care over study period, likely risk of bias from secular changes over time	NA: Two studies found evidence of neither increased maternal nor infant adverse health outcomes with reductions in routine preeclampsia screening	Moderate: RCT published nearly 20 y ago; changes to clinical practices in the United States since original study conducted Before-after study among primarily publicly insured, racial/ethnic minority patients obtaining care in a US hospital-based midwifery practice	No differences in maternal, delivery, or perinatal/neonatal health outcomes with fewer preeclampsia screening visits during pregnancy[31] or when urine testing conducted only when indicated (vs at every visit)[50] Limitations in relevance to current clinical practice[31] and weakness in study design[50] preclude strong conclusions about potential harms of different screening approaches

Abbreviations: KQ, key question; NA, not available; RCT, randomized clinical trial.

[a] Includes reporting bias.

[b] Includes precision.

[c] See Methods for full explanation of the prediction model appraisal approach.

Evidence on the net effect of risk prediction and the clinical actions that follow identification of a woman at risk for preeclampsia is needed to fully evaluate the effect of clinical risk prediction.[85] High sensitivity may be more important for prediction of preeclampsia risk, because false-negative results arguably are more detrimental than false-positive results[82]; a lower risk threshold and lower PPV may be reasonable[86] to consider for low-dose aspirin prophylaxis and heightened surveillance.[87] Evidence is limited for determining whether model-based risk prediction would be beneficial for preeclampsia health outcomes, beyond risk-assessment approaches currently practiced by clinicians.[80,85,88,89] Rigorous validation and well-designed clinical impact studies are needed to determine the likely performance and effect on health outcomes for multivariable risk assessment models.

Limitations

The review was limited to externally validated models, a minimum level of evidence necessary for estimating performance of a model before considering routine use. Quality appraisal of prediction models was not conducted as part of the synthesis. The risk of bias inherent to model development studies is addressed to some extent with external validation and impact studies, so this review focused on this higher level of evidence.

The relatively short time frame of pregnancy, rarity and unpredictability of severe preeclampsia, and interwoven maternal and fetal risks pose challenges to straightforward estimation of screening performance, benefits, and harms. No studies were identified on the effectiveness of screening for preeclampsia, including risk-based approaches to care. No studies assessing the accuracy of urine protein testing in general prenatal care populations or of the common practice of repeated testing over the course of pregnancy were identified. Consideration of different approaches to screening among women with hypertension in the absence of proteinuria may be important for future reviews as more evidence becomes available on the clinical use of newer diagnostic criteria. Large study populations are required to compare different approaches to screening and effects on maternal and perinatal health outcomes, as well as longer-term sequelae.

The need to evaluate the proportion of women misclassified as having or not having preeclampsia at a single point or over the course of pregnancy may not be clinically important when the result of a false-negative finding is to continue screening and when the result of a false-positive finding is enhanced surveillance and additional noninvasive diagnostic tests. Trials of different approaches to screening would provide more informative data for improving clinical screening practices and preeclampsia outcomes. The absence of information on potential harms of risk prediction, considering the high false-positive rates, is a notable shortcoming of the risk prediction literature. Without comparisons of proposed models to current clinical practices, the potential benefits and harms of risk prediction cannot be determined. Testing different prenatal care algorithms against usual care, possibly incorporating use of the best-performing and most feasible models, would be valuable.

Many of the studies identified had very few cases to classify, so the confidence intervals for performance estimates were wide. Shortcomings in the literature on preeclampsia prediction related to transparency and completeness of reporting modeling have been noted by others.[26,79] The absence of calibration statistics limited the ability to comprehensively evaluate and compare model performance.[29,81] Values of area under the curve do not provide a solid basis for determining how well, and at what level of risk, a risk prediction model would perform.[90]

Conclusions

Evidence to estimate benefits and harms of preeclampsia screening and the test performance of different screening approaches over the course of pregnancy was limited. Externally validated risk prediction models had limited applicability and lacked calibration and clinical implementation data needed to support routine use. Further research is needed to better inform risk-based screening approaches and improve screening strategies, given the complex pathophysiology and clinical unpredictability of preeclampsia.

ARTICLE INFORMATION

Author Contributions: Dr Henderson had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: All authors.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Henderson, Burda.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Burda.
Administrative, technical, or material support: Thompson, Burda, Cantor.
Supervision: Henderson.

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

Funding/Support: This research was funded under contract HHSA-290-2012-00151-I, Task Order 4, from the Agency for Healthcare Research and Quality (AHRQ), US Department of Health and Human Services, under a contract to support the USPSTF.

Role of the Funder/Sponsor: Investigators worked with USPSTF members and AHRQ staff to develop the scope, analytic framework, and key questions for this review. AHRQ had no role in study selection, quality assessment, or synthesis. AHRQ staff provided project oversight; reviewed the report to ensure that the analysis met methodological standards; and distributed the draft for peer review. Otherwise, AHRQ had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript findings. The opinions expressed in this document are those of the authors and do not reflect the official position of AHRQ or the US Department of Health and Human Services.

Additional Contributions: We gratefully acknowledge the following individuals for their contributions: AHRQ staff; the US Preventive Services Task Force; and EPC staff Evelyn P. Whitlock, MD, MPH, Jennifer Lin, MD, Tracy Beil, MPH, Smyth Lai, MLS, and Elizabeth Hess, ELS(D). USPSTF members, peer reviewers, and federal partner reviewers did not receive financial compensation for their contributions.

Additional Information: A draft version of this evidence report underwent external peer review from 4 content experts (Gregory E. Simon, MD, Group Health Research Institute; Barbara Yawn, MD, Department of Research, Olmsted Medical Center; Marian McDonagh, PharmD, Oregon Health and Science University; Ramin Mojtabai, MD, John Hopkins Bloomberg School of Public Health) and 4 federal partners (Centers for Disease Control and Prevention, National Institute of Mental Health, Substance Abuse and Mental Health Services Administration, and the US Air Force). Comments were presented to the USPSTF during its deliberation of the evidence and were considered in preparing the final evidence review.

Editorial Disclaimer: This evidence report is presented as a document in support of the accompanying USPSTF Recommendation Statement. It did not undergo additional peer review after submission to *JAMA*.

REFERENCES

1. *Hypertension in Pregnancy*. Washington, DC: American College of Obstetricians and Gynecologists; 2013.

2. Huppertz B. Placental origins of preeclampsia: challenging the current hypothesis. *Hypertension*. 2008;51(4):970-975.

3. Ghulmiyyah L, Sibai B. Maternal mortality from preeclampsia/eclampsia. *Semin Perinatol*. 2012;36 (1):56-59.

4. Ananth CV, Keyes KM, Wapner RJ. Pre-eclampsia rates in the United States, 1980-2010: age-period-cohort analysis. *BMJ*. 2013;347:f6564.

5. Duley L, Meher S, Jones L. Drugs for treatment of very high blood pressure during pregnancy. *Cochrane Database Syst Rev*. 2013;7(7):CD001449.

6. Too GT, Hill JB. Hypertensive crisis during pregnancy and postpartum period. *Semin Perinatol*. 2013;37(4):280-287.

7. Altman D, Carroli G, Duley L, et al; Magpie Trial Collaboration Group. Do women with pre-eclampsia, and their babies, benefit from magnesium sulphate? the Magpie Trial: a randomised placebo-controlled trial. *Lancet*. 2002;359(9321):1877-1890.

8. Magpie Trial Follow-Up Study Collaborative Group. The Magpie Trial: a randomised trial comparing magnesium sulphate with placebo for pre-eclampsia: outcome for children at 18 months. *BJOG*. 2007;114(3):289-299.

9. Magpie Trial Follow-Up Study Collaborative Group. The Magpie Trial: a randomised trial comparing magnesium sulphate with placebo for pre-eclampsia: outcome for women at 2 years. *BJOG*. 2007;114(3):300-309.

10. Duley L, Gülmezoglu AM, Henderson-Smart DJ, Chou D. Magnesium sulphate and other anticonvulsants for women with pre-eclampsia. *Cochrane Database Syst Rev*. 2010;(11):CD000025.

11. Koopmans CM, Bijlenga D, Groen H, et al; HYPITAT Study Group. Induction of labour versus expectant monitoring for gestational hypertension or mild pre-eclampsia after 36 weeks' gestation (HYPITAT): a multicentre, open-label randomised controlled trial. *Lancet*. 2009;374(9694):979-988.

12. Broekhuijsen K, van Baaren GJ, van Pampus MG, et al; HYPITAT-II Study Group. Immediate delivery versus expectant monitoring for hypertensive disorders of pregnancy between 34 and 37 weeks of gestation (HYPITAT-II): an open-label, randomised controlled trial. *Lancet*. 2015;385 (9986):2492-2501.

13. US Preventive Services Task Force. Screening for preeclampsia. In: DiGiuseppi C, Atkins D, Woolf S, Kamerow D, eds. *Guide to Clinical Preventive Services*. 2nd ed. Baltimore, MD: Lippincott Williams & Wilkins; 1996:419-424.

14. Harris RP, Helfand M, Woolf SH, et al; Methods Work Group, Third US Preventive Services Task Force. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med*. 2001;20(3)(suppl):21-35.

15. US Preventive Services Task Force. *Procedure Manual*. Rockville, MD: Agency of Healthcare Research and Quality; 2014.

16. United Nations Development Programme. *Human Development Report 2014*. Washington, DC: United Nations Development Programme; 2014.

17. Gillon TE, Pels A, von Dadelszen P, MacDonell K, Magee LA. Hypertensive disorders of pregnancy: a systematic review of international clinical practice guidelines. *PLoS One*. 2014;9(12):e113715.

18. ACOG Committee on Practice Bulletins—Obstetrics. ACOG practice bulletin: diagnosis and management of preeclampsia and eclampsia: number 33, January 2002. *Obstet Gynecol*. 2002;99(1):159-167.

19. Bujold E, Roberge S, Lacasse Y, et al. Prevention of preeclampsia and intrauterine growth restriction with aspirin started in early pregnancy: a meta-analysis. *Obstet Gynecol*. 2010;116(2, pt 1): 402-414.

20. Moore GS, Allshouse AA, Post AL, Galan HL, Heyborne KD. Early initiation of low-dose aspirin for reduction in preeclampsia risk in high-risk women: a secondary analysis of the MFMU High-Risk Aspirin Study. *J Perinatol*. 2015;35(5):328-331.

21. Askie LM, Duley L, Henderson-Smart DJ, Stewart LA; PARIS Collaborative Group. Antiplatelet agents for prevention of pre-eclampsia: a meta-analysis of individual patient data. *Lancet*. 2007;369(9575):1791-1798.

22. LeFevre ML; U.S. Preventive Services Task Force. Low-dose aspirin use for the prevention of morbidity and mortality from preeclampsia: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;161(11):819-826.

23. Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536.

24. Wells GA, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analysis. http://www.ohri.ca/programs /clinical_epidemiology/oxford.asp. 2014. Accessed October 26, 2015.

25. Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KG. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol*. 2015;68(3):279-289.

26. Kleinrouweler CE, Cheong-See FM, Collins GS, et al. Prognostic models in obstetrics: available, but far from applicable. *Am J Obstet Gynecol*. 2016;214 (1):79-90.

27. Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med*. 2014;11(10):e1001744.

28. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1-W3.

29. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128-138.

30. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ*. 2004;329(7458):168-169.

31. McDuffie RS Jr, Beck A, Bischoff K, Cross J, Orleans M. Effect of frequency of prenatal care visits on perinatal outcome among low-risk women: a randomized controlled trial. *JAMA*. 1996;275(11): 847-851.

32. Farina A, Rapacchia G, Freni Sterrantino A, Pula G, Morano D, Rizzo N. Prospective evaluation of ultrasound and biochemical-based multivariable models for the prediction of late pre-eclampsia. *Prenat Diagn*. 2011;31(12):1147-1152.

33. Oliveira N, Magder LS, Blitzer MG, Baschat AA. First-trimester prediction of pre-eclampsia: external validity of algorithms in a prospectively enrolled cohort. *Ultrasound Obstet Gynecol*. 2014;44(3): 279-285.

34. Park FJ, Leung CH, Poon LC, Williams PF, Rothwell SJ, Hyett JA. Clinical evaluation of a first trimester algorithm predicting the risk of hypertensive disease of pregnancy. *Aust N Z J Obstet Gynaecol*. 2013;53(6):532-539.

35. Skråstad RB, Hov GG, Blaas HG, Romundstad PR, Salvesen KÅ. Risk assessment for preeclampsia in nulliparous women at 11-13 weeks gestational age: prospective evaluation of two algorithms [published online December 4, 2014]. *BJOG*. doi:10.1111/1471-0528

36. Simeone S, Lojo C, Garcia-Esteve L, et al. Psychological impact of first-trimester prevention for preeclampsia on anxiety. *Prenat Diagn*. 2015;35 (1):60-64.

37. Bhide A, Rana R, Dhavilkar M, Amodio-Hernandez M, Deshpande D, Caric V. The value of the urinary protein:creatinine ratio for the detection of significant proteinuria in women with suspected preeclampsia. *Acta Obstet Gynecol Scand*. 2015;94(5):542-546.

38. Durnwald C, Mercer B. A prospective comparison of total protein/creatinine ratio versus 24-hour urine protein in women with suspected preeclampsia. *Am J Obstet Gynecol*. 2003;189(3): 848-852.

39. Dwyer BK, Gorman M, Carroll IR, Druzin M. Urinalysis vs urine protein-creatinine ratio to predict significant proteinuria in pregnancy. *J Perinatol*. 2008;28(7):461-467.

40. Kyle PM, Fielder JN, Pullar B, Horwood LJ, Moore MP. Comparison of methods to identify significant proteinuria in pregnancy in the outpatient setting. *BJOG*. 2008;115(4):523-527.

41. Lamontagne A, Côté AM, Rey E. The urinary protein-to-creatinine ratio in Canadian women at risk of preeclampsia: does the time of day of testing matter? *J Obstet Gynaecol Can*. 2014;36(4):303-308.

42. Sethuram R, Kiran TS, Weerakkody AN. Is the urine spot protein/creatinine ratio a valid diagnostic test for pre-eclampsia? *J Obstet Gynaecol*. 2011;31 (2):128-130.

43. Stout MJ, Scifres CM, Stamilio DM. Diagnostic utility of urine protein-to-creatinine ratio for identifying proteinuria in pregnancy. *J Matern Fetal Neonatal Med*. 2013;26(1):66-70.

44. Tun C, Quiñones JN, Kurt A, Smulian JC, Rochon M. Comparison of 12-hour urine protein and protein:creatinine ratio with 24-hour urine protein for the diagnosis of preeclampsia. *Am J Obstet Gynecol*. 2012;207(3):233.e1-233.e8.

45. Verdonk K, Niemeijer IC, Hop WC, et al. Variation of urinary protein to creatinine ratio during the day in women with suspected pre-eclampsia. *BJOG*. 2014;121(13):1660-1665.

46. Waugh J, Bell SC, Kilby M, Lambert P, Shennan A, Halligan A. Effect of concentration and biochemical assay on the accuracy of urine dipsticks

in hypertensive pregnancies. *Hypertens Pregnancy*. 2001;20(2):205-217.

47. Waugh JJ, Bell SC, Kilby MD, et al. Optimal bedside urinalysis for the detection of proteinuria in hypertensive pregnancy: a study of diagnostic accuracy. *BJOG*. 2005;112(4):412-417.

48. Wheeler TL II, Blackhurst DW, Dellinger EH, Ramsey PS. Usage of spot urine protein to creatinine ratios in the evaluation of preeclampsia. *Am J Obstet Gynecol*. 2007;196(5):465.e1-465.e4.

49. Young RA, Buchanan RJ, Kinch RA. Use of the protein/creatinine ratio of a single voided urine specimen in the evaluation of suspected pregnancy-induced hypertension. *J Fam Pract*. 1996;42(4):385-389.

50. Rhode MA, Shapiro H, Jones OW III. Indicated vs. routine prenatal urine chemical reagent strip testing. *J Reprod Med*. 2007;52(3):214-219.

51. E Holanda Moura SB, Park F, Murthi P, et al. TNF-R1 as a first trimester marker for prediction of pre-eclampsia. *J Matern Fetal Neonatal Med*. 2016; 29(6):897-903.

52. Oliveira N, Doyle LE, Atlas RO, Jenkins CB, Blitzer MG, Baschat AA. External validity of first-trimester algorithms in the prediction of pre-eclampsia disease severity. *Ultrasound Obstet Gynecol*. 2014;44(3):286-292.

53. Park F, Russo K, Williams P, et al. Prediction and prevention of early-onset pre-eclampsia: impact of aspirin after first-trimester screening. *Ultrasound Obstet Gynecol*. 2015;46(4):419-423.

54. Valdés E, Sepúlveda-Martínez Á, Tong A, Castro M, Castro D. Assessment of protein:creatinine ratio versus 24-hour urine protein in the diagnosis of preeclampsia [published online June 3, 2015]. *Gynecol Obstet Invest*. doi:10.1159/000381773

55. Poon LC, Stratieva V, Piras S, Piri S, Nicolaides KH. Hypertensive disorders in pregnancy: combined screening by uterine artery Doppler, blood pressure and serum PAPP-A at 11-13 weeks. *Prenat Diagn*. 2010; 30(3):216-223.

56. Onwudiwe N, Yu CK, Poon LC, Spiliopoulos I, Nicolaides KH. Prediction of pre-eclampsia by a combination of maternal history, uterine artery Doppler and mean arterial pressure. *Ultrasound Obstet Gynecol*. 2008;32(7):877-883.

57. Odibo AO, Zhong Y, Goetzinger KR, et al. First-trimester placental protein 13, PAPP-A, uterine artery Doppler and maternal characteristics in the prediction of pre-eclampsia. *Placenta*. 2011;32(8): 598-602.

58. Akolekar R, Syngelaki A, Poon L, Wright D, Nicolaides KH. Competing risks model in early screening for preeclampsia by biophysical and biochemical markers. *Fetal Diagn Ther*. 2013;33(1): 8-15.

59. Poon LC, Karagiannis G, Leal A, Romero XC, Nicolaides KH. Hypertensive disorders in pregnancy: screening by uterine artery Doppler imaging and blood pressure at 11-13 weeks. *Ultrasound Obstet Gynecol*. 2009;34(5):497-502.

60. Scazzocchio E, Figueras F, Crispi F, et al. Performance of a first-trimester screening of preeclampsia in a routine care low-risk setting. *Am J Obstet Gynecol*. 2013;208(3):203.e1-203.e10.

61. Caradeux J, Serra R, Nien JK, et al. First trimester prediction of early onset preeclampsia using demographic, clinical, and sonographic data: a cohort study. *Prenat Diagn*. 2013;33(8):732-736.

62. Parra-Cordero M, Rodrigo R, Barja P, et al. Prediction of early and late pre-eclampsia from maternal characteristics, uterine artery Doppler and markers of vasculogenesis during first trimester of pregnancy. *Ultrasound Obstet Gynecol*. 2013;41(5): 538-544.

63. Plasencia W, Maiz N, Bonino S, Kaihura C, Nicolaides KH. Uterine artery Doppler at 11 + 0 to 13 + 6 weeks in the prediction of pre-eclampsia. *Ultrasound Obstet Gynecol*. 2007;30(5):742-749.

64. Poon LCY, Maiz N, Valencia C, Plasencia W, Nicolaides KH. First-trimester maternal serum pregnancy-associated plasma protein-A and pre-eclampsia. *Ultrasound Obstet Gynecol*. 2009; 33(1):23-33.

65. Plasencia W, Maiz N, Poon L, Yu C, Nicolaides KH. Uterine artery Doppler at 11 + 0 to 13 + 6 weeks and 21 + 0 to 24 + 6 weeks in the prediction of pre-eclampsia. *Ultrasound Obstet Gynecol*. 2008;32 (2):138-146.

66. Wright D, Akolekar R, Syngelaki A, Poon LC, Nicolaides KH. A competing risks model in early screening for preeclampsia. *Fetal Diagn Ther*. 2012; 32(3):171-178.

67. Poon LC, Kametas NA, Chelemen T, Leal A, Nicolaides KH. Maternal risk factors for hypertensive disorders in pregnancy: a multivariate approach. *J Hum Hypertens*. 2010;24(2):104-110.

68. Woelkers D, Barton J, Dadelszen P, Sibai B. [71-OR]: the revised 2013 ACOG definitions of hypertensive disorders of pregnancy significantly increase the diagnostic prevalence of preeclampsia [abstract]. *Pregnancy Hypertens*. 2015;5:38.

69. Bramham K, Parnell B, Nelson-Piercy C, Seed PT, Poston L, Chappell LC. Chronic hypertension and pregnancy outcomes: systematic review and meta-analysis. *BMJ*. 2014;348:g2301.

70. Siu AL; U.S. Preventive Services Task Force. Screening for high blood pressure in adults: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2015;163(10):778-786.

71. Higgins JR, de Swiet M. Blood-pressure measurement and classification in pregnancy. *Lancet*. 2001;357(9250):131-135.

72. Pickering TG, Hall JE, Appel LJ, et al. Recommendations for blood pressure measurement in humans and experimental animals, part 1: blood pressure measurement in humans: a statement for professionals from the Subcommittee of Professional and Public Education of the American Heart Association Council on High Blood Pressure Research. *Circulation*. 2005;111(5): 697-716.

73. Thangaratinam S, Coomarasamy A, O'Mahony F, et al. Estimation of proteinuria as a predictor of complications of pre-eclampsia: a systematic review. *BMC Med*. 2009;7:10.

74. Morris RK, Riley RD, Doug M, Deeks JJ, Kilby MD. Diagnostic accuracy of spot urinary protein and albumin to creatinine ratios for detection of significant proteinuria or adverse pregnancy outcome in patients with suspected pre-eclampsia: systematic review and meta-analysis. *BMJ*. 2012; 345:e4342.

75. Meyer NL, Mercer BM, Friedman SA, Sibai BM. Urinary dipstick protein: a poor predictor of absent or severe proteinuria. *Am J Obstet Gynecol*. 1994; 170(1, pt 1):137-141.

76. Waugh J, Maybury H, Shennan A. A prospective comparison of total protein/creatinine ratio versus 24-hour urine protein in women with suspected preeclampsia. *Am J Obstet Gynecol*. 2004;191(3): 1049-1050.

77. Côté AM, Brown MA, Lam E, et al. Diagnostic accuracy of urinary spot protein:creatinine ratio for proteinuria in hypertensive pregnant women: systematic review. *BMJ*. 2008;336(7651):1003-1006.

78. Papanna R, Mann LK, Kouides RW, Glantz JC. Protein/creatinine ratio in preeclampsia: a systematic review. *Obstet Gynecol*. 2008;112(1): 135-144.

79. Brunelli VB, Prefumo F. Quality of first trimester risk prediction models for pre-eclampsia: a systematic review. *BJOG*. 2015;122(7):904-914.

80. Al-Rubaie Z, Askie LM, Ray JG, Hudson HM, Lord SJ. The performance of risk prediction models for pre-eclampsia using routinely collected maternal characteristics and comparison with models that include specialised tests and with clinical guideline decision rules: a systematic review. *BJOG*. 2016;123(9):1441-1452.

81. Bouwmeester W, Zuithoff NP, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med*. 2012;9(5): 1-12.

82. Cnossen JS, ter Riet G, Mol BW, et al. Are tests for predicting pre-eclampsia good enough to make screening viable? a review of reviews and critical appraisal. *Acta Obstet Gynecol Scand*. 2009;88(7): 758-765.

83. Chappell LC, Sandall J, Barnard AM, McManus RJ. Predicting pre-eclampsia. *BMJ*. 2015;351:h6349.

84. American College of Obstetricians and Gynecologists. Committee Opinion No. 638: first-trimester risk assessment for early-onset preeclampsia. *Obstet Gynecol*. 2015;126(3):e25-e27.

85. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med*. 2006;144(3):201-209.

86. Bartsch E, Medcalf KE, Park AL, Ray JG; High Risk of Pre-eclampsia Identification Group. Clinical risk factors for pre-eclampsia determined in early pregnancy: systematic review and meta-analysis of large cohort studies. *BMJ*. 2016;353:i1753.

87. Bartsch E, Park AL, Kingdom JC, Ray JG. Risk threshold for starting low dose aspirin in pregnancy to prevent preeclampsia: an opportunity at a low cost. *PLoS One*. 2015;10(3):e0116296.

88. Ioannidis JP, Tzoulaki I. What makes a good predictor? the evidence applied to coronary artery calcium score. *JAMA*. 2010;303(16):1646-1647.

89. Grady D, Berkowitz SA. Why is a good clinical prediction rule so hard to find? *Arch Intern Med*. 2011;171(19):1701-1702.

90. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007;115(7):928-935.