Screening for Preeclampsia: A Systematic Evidence Review for the U.S. Preventive Services Task Force

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

Contract No. HHSA-290-2012-00151-I, Task Order No. 4

Prepared by:
Kaiser Permanente Research Affiliates Evidence-based Practice Center
Kaiser Permanente Center for Health Research
Portland, OR

Investigators:
Jillian T. Henderson, PhD, MPH
Jamie H. Thompson, MPH
Brittany U. Burda, MPH
Amy Cantor, MD, MPH
Tracy Beil, MPH
Evelyn P. Whitlock, MD, MPH

AHRQ Publication No. 14-05211-EF-1
September 2016
This report is based on research conducted by the Kaiser Permanente Research Affiliates Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. HHSA-290-2012-00151-I, Task Order No. 4). The findings and conclusions in this document are those of the authors, who are responsible for its contents, and do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information (i.e., in the context of available resources and circumstances presented by individual patients).

The final report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report.

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this report: Iris Mabry-Hernandez, MD, MPH, at AHRQ; current and former members of the U.S. Preventive Services Task Force who contributed to topic deliberations; Aaron B. Caughey, MD, PhD, Liona Poon, MBBS, MRCOG, MD(Res), and James M. Roberts, MD, who provided expert review; the National Institute of Child Health and Human Development, the U.S. Department of Veterans Affairs, and the Indian Health Service, who provided federal partner review; and Smyth Lai, MLS, Elizabeth Hess, ELS(D), Elizabeth O’Connor, PhD, Ning Smith, PhD, and Maya G. Rowland, MPH, at the Kaiser Permanente Center for Health Research.
Structured Abstract

**Background:** The U.S. Preventive Services Task Force (USPSTF) does not currently have an active recommendation for preeclampsia screening. Preeclampsia is a complex disease occurring in the second half of pregnancy and is estimated to affect nearly 4 percent of pregnancies in the United States. Nearly 9 percent of maternal deaths in the United States are directly attributed to preeclampsia and eclampsia, and it is a leading cause of induced preterm birth and low birth weight. Early detection through general or high-risk screening approaches may help reduce the health-related consequences, particularly for infants.

**Purpose:** We conducted a systematic review to assess the direct evidence of benefits and harms of preeclampsia screening on health outcomes; to evaluate the effectiveness of routine blood pressure and urine protein screening tests to identify women with preeclampsia; to estimate the accuracy of screening tests for proteinuria; and to evaluate the performance of multivariable risk assessment tools used during the first trimester to identify women at increased risk of preeclampsia as well as the potential harms of risk assessment.

**Data Source:** MEDLINE, PubMed, and Cochrane Central Register of Controlled Trials from 1990 through September 1, 2015. We included all references from the 1996 USPSTF recommendation and examined reference lists of relevant systematic reviews.

**Study Selection:** English-language trials and observational studies of screening effectiveness, test accuracy, and harms. Two investigators independently reviewed identified abstracts and full-text articles against a set of *a priori* inclusion and quality criteria.

**Data Analysis:** One investigator abstracted details about study design, patient population, setting, screening method, followup, and results. Two investigators independently applied prespecified criteria to rate study quality. Discrepancies were resolved through consensus, and poor quality studies were excluded. Due to small numbers of studies and methodological shortcomings, meta-analysis was not attempted for any outcome measure other than urine protein:creatinine tests performed as point-of-care screening.

**Results:** A fair-quality randomized controlled trial of 2,764 “low risk” pregnant U.S. women found no statistically significant differences in health outcomes among women assigned to fewer prenatal screening visits compared to usual care at a large managed care organization in 1996 (mean number of visits 12.0 versus 14.7, *p*<.001). A fair-quality before-after study of 1,952 low-income, Hispanic pregnant women did not identify harms related to preeclampsia diagnosis and birth outcomes when protein urine screening was used for specific indications instead of on a routine basis in prenatal care. We found no evidence to evaluate the effectiveness of routine screening tests in identifying women with preeclampsia and limited evidence on various screening approaches for establishing the presence of proteinuria (a diagnostic criterion for preeclampsia). Fourteen diagnostic test accuracy studies (4 good quality, 10 fair quality) compared point-of-care tests used to screen for proteinuria versus the gold standard (24-hour protein excretion). Included studies of test accuracy were conducted in women with suspected preeclampsia, while studies with healthy, asymptomatic patients seeking routine care were lacking. Twelve studies evaluated the performance of protein:creatinine tests. High heterogeneity
precluded pooling of test performance (k=11). Sensitivity for the protein:creatinine test ranged from 0.65 to 0.96 ($I^2$=80.5% of 11 studies) and specificity ranged from 0.49 to 1.00 ($I^2$=91.8% of 11 studies). Statistical heterogeneity of test sensitivity was partly explained by differences in the study populations; studies with a positive protein dipstick result as an inclusion criterion had higher sensitivity ($p<0.05$). Two studies of the albumin:creatinine ratio spot test had high sensitivity ($\geq 0.94$, [95% confidence intervals [CI] range, 0.75 to 1.00]). Four studies of quantitatively read protein dipstick tests had widely variable sensitivity (0.22 to 1.00) and specificity (0.36 to 1.00). Four studies validated five first-trimester risk assessment models with good-to-excellent discrimination, primarily for predicting early preeclampsia requiring delivery. No externally validated multivariable risk prediction models were based only on patient history measures that could be collected in a routine prenatal care visit; all included serum markers and uterine artery Doppler ultrasound measure of the pulsatility index, or both. Five models had good discrimination of preeclampsia cases (c index > 0.80), but had very low positive predictive values and did not provide necessary information on model calibration.

**Conclusions:** Changes in diagnostic criteria, patient demographics, and treatment recommendations affect the applicability of previous trials, precluding conclusions about the optimal screening approach. Most studies for detecting proteinuria, one of the diagnostic criteria for preeclampsia, tested the protein:creatinine ratio in urine samples; however, all studies were among patients with prescreened suspicion of preeclampsia and none evaluated the performance of repeat testing of urine protein for screening. Due to limited and variable evidence, different urine protein screening tests cannot be compared. There was no clear evidence of the performance, clinical benefits, or harms for any externally validated models for risk prediction, and the clinical performance and impact of risk prediction models could not be extrapolated to relevant patient settings. Current screening practices are considered routine and represent relatively minor burdens to patients, clinicians, and healthcare systems, but evidence is limited for determining the benefits and harms of preeclampsia screening.
Table of Contents

Chapter 1. Introduction ............................................................................................................... 1
  Condition Definition .............................................................................................................. 1
  Prevalence and Burden of Preeclampsia ............................................................................. 2
  Etiology and Natural History .............................................................................................. 3
  Risk Factors ....................................................................................................................... 4
  Rationale and Strategies for Screening .............................................................................. 5
  Disease Management and Treatment .................................................................................. 5
  Rationale and Strategies for Risk Assessment ..................................................................... 6
  Current Risk Assessment and Screening Practice in the United States ................................ 6
  Previous USPSTF Recommendation .................................................................................. 7

Chapter 2. Methods ...................................................................................................................... 8
  Scope and Purpose ............................................................................................................... 8
  Key Questions and Analytic Framework ............................................................................. 8
  Data Sources and Searches ............................................................................................... 9
  Study Selection .................................................................................................................. 9
  Quality Assessment and Data Abstraction ......................................................................... 11
  Approach to Review of Clinical Risk Prediction Studies .................................................. 11
  Measures of Risk Prediction Model Performance .............................................................. 12
  Data Synthesis and Analysis ............................................................................................ 12
  Expert Review and Public Comment .................................................................................. 13
  USPSTF Involvement ......................................................................................................... 13

Chapter 3. Results ....................................................................................................................... 14
  Literature Search Results .................................................................................................. 14
  Results of Included Studies ............................................................................................... 14
  Key Question 1. How Effectively Does Screening for Preeclampsia Reduce Maternal and
                  Perinatal Morbidity and Mortality? ........................................................................ 14
  Key Question 1a. Does Effectiveness Differ by Screening Protocol (e.g., Tests Used,
                  Timing of Tests, Rescreen Intervals) or Preeclampsia Risk Status? ................. 14
  Key Question 2. What Is the Effectiveness of Risk Assessment in Early Pregnancy for
                  Identifying Women at High Risk for Preeclampsia? .......................................... 15
  Key Question 3. What Are the Harms of Preeclampsia Risk Assessment? ...................... 17
  Key Question 4. How Effectively Do Routine Screening Tests (i.e., Blood Pressure,
                  Proteinuria) Identify Women With Preeclampsia? ........................................... 18
  Key Question 4a. How Accurate Are Different Point-of-Care Screening Tests for
                  Proteinuria (a Diagnostic Criterion for Preeclampsia)? .................................... 18
  Key Question 4b. How Effective for Identifying Women With Preeclampsia Are Different
                  Screening Protocols (e.g., Instruments, Test Procedures, Timing of Tests, Rescreen
                  Intervals)? ........................................................................................................ 21
  Key Question 4c. How Should Women at High Risk for Preeclampsia Be Screened
                  Differently From Women at Low or Average Risk? .......................................... 22
  Key Question 5. What Are the Harms of Preeclampsia Screening and Do They Differ by
                  Risk Status or Screening Protocol? .................................................................... 22

Chapter 4. Discussion ................................................................................................................. 24
  Summary of Evidence ........................................................................................................ 24
Chapter 1. Introduction

Condition Definition

Preeclampsia is a multi-system syndrome that is primarily defined by the development of new-onset hypertension, persistent systolic blood pressure [SBP] of 140 millimeters of mercury [mm Hg] or higher, or diastolic blood pressure [DBP] of 90 mm Hg or higher after 20 weeks of gestation in a woman with previously normal blood pressure. Although preeclampsia is usually accompanied by new-onset proteinuria, the American Congress of Obstetrics and Gynecology (ACOG) recently revised the diagnostic criteria for preeclampsia so that the presence of proteinuria for diagnosis was no longer required, noting that elevated blood pressure accompanied by other signs and symptoms is sufficient for diagnosis. These other signs are also included in new terminology proposed by ACOG to identify cases with severe features. Those severe features are: very high blood pressure, thrombocytopenia, renal insufficiency, impaired liver function, pulmonary edema, and cerebral or visual symptoms. The proportion of women who develop preeclampsia without proteinuria or who have proteinuria without hypertension preceding preeclampsia is unclear— with inconsistent definitions and approaches to measurement, and few studies examining these atypical presentations. Proteinuria levels among women diagnosed with preeclampsia, however, are not found to be consistently associated with adverse outcomes.

Systems for diagnosing and classifying the severity of disease vary across professional societies and organizations, including ACOG, the American Society of Hypertension (ASH), and obstetrics and gynecology professional organizations in the United Kingdom, Canada, New Zealand, and Australia. Fetal complications of preeclampsia include intrauterine growth restriction (IUGR) and can occur due to placental perfusion problems. Preeclampsia often remains stable until delivery, but sometimes it can rapidly and unpredictably take a more serious turn. Severe hypertension, eclampsia, or hemolysis, elevated liver enzymes, and low platelet counts (HELLP syndrome) and organ and systemic complications can lead to maternal or fetal injury and death. For this reason, the term ‘mild preeclampsia’ has been recommended against in the new ACOG diagnostic criteria.

Other hypertensive conditions overlap and can coexist with preeclampsia. Chronic hypertension is defined as predating the pregnancy and/or continuing beyond 12 weeks postpartum. Women with chronic hypertension are diagnosed with superimposed preeclampsia if proteinuria develops after 20 weeks’ gestation. Pregnant women who develop hypertension during pregnancy (without proteinuria) that subsides within 12 weeks postpartum are defined as having gestational hypertension.

The concepts of early- and late-onset preeclampsia have been used to define different manifestations of the syndrome and may reflect differences in pathophysiology as well as long-term outcomes. These timing categories usually distinguish between cases developing prior to 34 weeks’ gestation versus later in pregnancy. Early-onset preeclampsia is associated with more severe maternal and fetal outcomes and may be especially influenced by aberrations in the placentation process, particularly in the remodeling of the maternal uterine spiral arteries.
Later-onset disease may also involve placental dysfunction, but it often occurs in women with proinflammatory maternal constitutional and environmental factors, such as multiple pregnancy, high body mass index (BMI), comorbidities, and chronic hypertension. Although pathophysiologic markers and processes underlying the development of preeclampsia are becoming more clearly understood, there remain considerable gaps in science that confer challenges for diagnosis and treatment.10

**Prevalence and Burden of Preeclampsia**

Approximately 2 to 8 percent of pregnancies are affected by preeclampsia, which is the second leading cause of maternal mortality worldwide.11,12 In the United States, the rate of preeclampsia increased from 3.4 percent in 1980 to 3.8 percent in 201013, and was accompanied by a rise in severe cases, which increased from 0.3 percent in 1980 to 1.4 percent in 2010.

Based on the most recent analysis of the Centers for Disease Control and Prevention (CDC) Pregnancy Mortality Surveillance System data, 9 percent of maternal deaths are directly attributed to preeclampsia and eclampsia.14 Complications of preeclampsia also indirectly contribute to approximately one in 10 pregnancy-related deaths attributed to anesthesia, cardiomyopathy, or placental abruption.15 Serious morbidity is far more common than mortality; it has been estimated that over one third of severe obstetric morbidities are related to preeclampsia.16 Significant maternal morbidities include cerebrovascular bleeding, retinal detachment, and complications from HELLP syndrome, such as major organ damage and failure.8 Eclampsia occurs in approximately 1 to 2 percent of preeclampsia cases, with complications such as brain damage, aspiration pneumonia, pulmonary edema, placental abruption, disseminated coagulopathy, acute renal failure, cardiopulmonary arrest, and coma.12 Cohort data from obstetric patients attending 25 U.S. medical centers comprising the Maternal-Fetal Medicine Units Network in 2008 to 2011 indicated that at least 21 percent of severe maternal morbidity was related to hypertension-related complications.17 The prevalence of hospitalizations from severe preeclampsia or eclampsia rose from 9.4 to 12.4 per 1,000 deliveries in the United States between 1998 and 2006,18 and there some recent indication that hospitalizations due to eclampsia may be falling.19

Preeclampsia also dramatically increases risks to the fetus or neonate. These risks include IUGR, small for gestational age (SGA), low birth weight, premature birth, oligohydramnios, placental abruption, low Apgar score, neonatal intensive care unit admission, stillbirth, and neonatal death.6,20 Because the treatment for preeclampsia is delivery, it is a leading cause of induced preterm birth and low birth weight. It has been estimated that preeclampsia is an indication in 6 percent of preterm births and 19 percent of medically indicated preterm births.21 Infants born before term (<37 weeks gestation) are at increased risk of morbidity and mortality, with risks rising dramatically with earlier delivery. When preeclampsia occurs before 34 weeks, maternal and perinatal risks are greater and management decisions have to balance maternal and perinatal health risks. The majority of preeclampsia cases occur after 34 weeks, but morbidity and mortality is greater for early onset disease.22

Obstetric interventions are more common in pregnancies complicated by preeclampsia, and can
include induction of labor (preterm or term), intravenous magnesium sulfate treatment, and emergency or planned caesarean section. Early delivery interventions can improve maternal health, and reduce some risks to the neonate, such as stillbirth, while increasing others, depending on the severity of disease and gestational timing. There is evidence that preeclampsia itself is associated with poor psychosocial outcomes, post-traumatic stress syndrome, and postpartum depression, with fetal or neonatal morbidity or mortality contributing to, but not entirely accounting for, the relationship.

In the United States, the prevalence of preeclampsia reveals marked disparities by race ethnicity. The rate of pregnancy-related death is four times greater among non-Hispanic black women, and preeclampsia is a major contributor to this disparity. National data on chronic and gestational hypertension show the conditions are more common and increasing over time among non-Hispanic black women, least common among Asian and Pacific Islanders and Hispanic women, and intermediate for non-Hispanic white women. Case fatality rates from preeclampsia are three times higher among black non-Hispanic women than among whites, contributing to the large mortality disparity. Approximately one third of the disparity in mortality from preeclampsia among black women stems from higher prevalence, and the remainder is due to a higher case fatality rate. Disparities in risk factors for preeclampsia, such as chronic hypertension, diabetes, and high BMI, contribute to a higher prevalence of preeclampsia among black women, and disparities in access to adequate prenatal care decrease opportunities to intervene before preeclampsia becomes more severe. Inadequate prenatal care is associated with higher case fatality from preeclampsia for all women, which is likely due to the reduced opportunity for monitoring, detection, and early intervention. Nevertheless, even in a large population (n=35,529) provided with early access to prenatal care, race/ethnic disparities have been observed, with minority women experiencing higher rates of preeclampsia than non-Hispanic white women do. Finally, recurrent preeclampsia in subsequent pregnancies is often more severe for black women than for white or Hispanic women.

**Etiology and Natural History**

Preeclampsia is a complex disease with multiple causes and interactions leading to its clinical manifestation. The intractability to effective treatment (apart from delivery of the placenta) makes it an area of considerable scientific inquiry with important implications for women’s health worldwide. The heightened risk of preeclampsia in first pregnancies and in women who undergo *in vitro* fertilization with donor eggs have led to numerous investigations regarding a potential role of the immune system and paternal genetic influences. Preeclampsia is generally understood to be an immunological and inflammatory condition that involves the process of placentation, but the underlying causes, precipitating factors, and conditions are not fully understood.

Preeclampsia may develop through different processes that can occur either alone or in combination. “Placental” disease may lead to earlier onset and more severe disease, while “maternal” disease may result in later-onset disease. Placental preeclampsia is thought to arise from problems with the process of placentation whereby trophoblast cells fail to fully activate transformation of uterine spiral arteries (about 12 to 16 weeks of pregnancy), resulting in
placental ischemia. This relative ischemia and lowered placental perfusion could cause the release of damaging factors (e.g., cellular debris, oxidized lipids, anti-angiogenic factors) into the maternal bloodstream, resulting in inflammation and oxidative stress. In contrast, maternal preeclampsia is thought to involve overactive inflammatory responses to normal placentation. Preexisting hypertension, diabetes, and other inflammatory conditions (e.g., lupus) as well as twin or higher order pregnancies are thought to precipitate a systemic inflammatory response and oxidative stress process. Consistent with this theory, women with early-onset placental preeclampsia exhibit abnormal uterine artery ultrasound Doppler readings and placental morphology compared to women without preeclampsia or later-onset disease.7,8,38 Adding to the complexity, maternal and environmental factors also may contribute to the risk of developing placental preeclampsia.

Preeclampsia can occur without immediate apparent adverse health consequences for the mother or infant. Challenges in preventing and treating the disease are heightened by the difficulty in determining who will develop preeclampsia and go on to experience severe or life-threatening complications. Preeclampsia also may pose a longer term risk factor for poor cardiovascular health for mothers and their offspring.39 The association may be explained by common risk factors, and it is unclear whether or not preventing preeclampsia would benefit the long-term cardiovascular health for women or children of mothers with preeclampsia is the subject of ongoing inquiry.

Despite intensive research, understanding of this complex disease is not complete. Most recently, the possibility that preeclampsia is a syndrome comprised of multiply subtypes has been proposed to explain its diverse etiology and unpredictable course.40,41

**Risk Factors**

There are a number of well-established clinical and historical risk factors for preclampsia.42 Chronic health conditions with increasing prevalence in the United States, such as essential hypertension and diabetes, affect the risk for preeclampsia. Women with preexisting hypertension and or new-onset hypertension in pregnancy are at elevated risk of developing preeclampsia.43 A recent systematic review of over 50 studies, including nearly 800,000 pregnancies, estimated the incidence of superimposed preeclampsia among women with chronic hypertension to be 26 percent (95% CI, 21% to 32%).44 In pregnant women with preexisting diabetes, the incidence of preeclampsia increased from 18 percent in women without preexisting proteinuria or chronic hypertension to 28 percent when one or both of these conditions were present (odds ratio [OR], 1.75 [95% CI, 1.02 to 3.01]).45

Other risk factors based on medical history are also used for risk-stratified clinical preventive services. For example, the U.S. Preventive Services Task Force (USPSTF) recommends a pragmatic approach to identify patients at low, moderate, or high risk for preeclampsia in their guideline for low-dose aspirin prophylaxis in pregnancy.46 The National Institute for Health and Care Excellence (NICE) risk assessment approach is similar (Table 2).47 High-risk pregnant women include those with a history of preeclampsia, multifetal gestation, chronic hypertension, type 1 or 2 diabetes, renal disease, or autoimmune disease.46 Moderate-level risk factors include
nulliparity, obesity, family history of preeclampsia, sociodemographic characteristics such as African American race or low socioeconomic status, age, or personal history factors (including low birthweight, previous adverse pregnancy outcome, and a pregnancy interval of more than 10 years). A pregnant woman with a previous uncomplicated full-term delivery is at lower risk for preeclampsia.

Rationale and Strategies for Screening

Screening for preeclampsia occurs periodically throughout pregnancy for all women receiving prenatal care. The aim of screening is to identify and diagnose the condition early in its course, to allow closer monitoring and effective disease management. Blood pressure measurement and testing for proteinuria have long been routine primary care screening tools for preeclampsia, and are core components of the diagnostic criteria. The timing of prenatal care visits, and inclusion of both of these tests at every visit on a routine or indicated basis is variable and not well described in the United States.

The gestational age at the time of diagnosis has a strong relationship with maternal and neonatal outcomes. Once diagnosed, care can be managed according to protocols that have been found to reduce the likelihood of maternal and neonatal harm. Depending on the timing of disease occurrence and spacing of visits, preeclampsia occurring at or very near term may not be detected during prenatal screening visits, but are likely to be detected when women present for delivery. The detection of earlier term preeclampsia, particularly cases that develop before 34 weeks’ gestation, is particularly important given the risks to the mother and neonate if severe disease features emerge.

Disease Management and Treatment

Effective management and treatment of diagnosed preeclampsia can prevent complications and poor health outcomes. Identification of women with preeclampsia allows health care providers to reduce the risk of eclampsia and cerebral, vascular, hepatic, and renal maternal complications. These most commonly occur among women who develop severe features of preeclampsia, but can unexpectedly develop even in cases without severe features. The clinically proven approaches for management of preeclampsia to reduce the likelihood of poor maternal and perinatal health outcomes include delivery, intravenous administration of magnesium sulfate, and treatment of high blood pressure. Importantly, delivery is the only curative treatment for preeclampsia once the condition develops; depending on the gestational timing of diagnosis and the seriousness of the maternal and fetal condition, induction of labor can reduce the risk of major morbidity and mortality. For women who develop severe preeclampsia, intravenous administration of magnesium sulfate is effective for reducing the risk of eclamptic seizures. Pharmacological treatment of very high blood pressure is recommended to reduce the risk of stroke and cerebral vascular events. These treatments are supported by a broad range of medical and public health organizations, including the World Health Organization, ACOG, and NICE. While there is variability in the strength of scientific evidence underlying different aspects of treatment, there is broad consensus that diagnosis and treatment improve perinatal
and maternal health outcomes.\textsuperscript{1,53,54} A detailed discussion of the trial evidence for these interventions is provided in Appendix A. Recent developments in prognostic evaluation using the fullPIERS model and other markers also hold promise for improving disease management.\textsuperscript{55,56}

Once a diagnosis of preeclampsia is made, increased maternal and fetal surveillance and often referral to specialty care begin, with recommended treatments undertaken as needed during the course of monitoring. The clinical evidence for some management practices is less established than for others, but treatment is clearly associated with better outcomes in the case of magnesium sulfate as well as early delivery in defined circumstances. Preeclampsia is among the most preventable causes of maternal mortality. Analyses of causes of maternal mortality from preeclampsia suggest that substandard clinical care often contributes to poor outcomes.\textsuperscript{57} Delayed responses to clinical warning signs and ineffective management have been found to be contributing factors in the majority of cases.\textsuperscript{31,58} Data based on medical charts from the United Kingdom suggests that fewer than half of women who developed eclampsia were diagnosed with both hypertension and proteinuria in the week preceding the event (38\%), and that preeclampsia with atypical presentation comprises a greater proportion of eclampsia cases, owing in part to delays in diagnosis.\textsuperscript{50,59} Inaccuracies in proteinuria tests or record-keeping, however, could also account for a portion of these findings.

**Rationale and Strategies for Risk Assessment**

Early identification of women most likely to develop preeclampsia is potentially important for at least two reasons. First, women at higher risk may benefit from heightened surveillance and timely interventions if severe features of the disease appear, to mitigate the risk of negative health consequences for the mother and fetus.\textsuperscript{8} Second, low-dose aspirin for women at high risk when commenced after the first trimester of pregnancy, ideally before 16 weeks gestation, reduces the incidence of preeclampsia and the likelihood of experiencing serious complications such as preterm birth and intrauterine growth restriction.\textsuperscript{60,61} Risk assessment generally relies on factors that are known to be associated with preeclampsia.\textsuperscript{42,62} However, no recommendations currently specify the use of clinical risk prediction tools for estimating a patient’s individual risk for developing preeclampsia. Reviews of serum markers and uterine artery Doppler testing as singular prediction tools do not support their use in routine clinical care to identify women at great risk for preeclampsia.\textsuperscript{1,63,64} Efforts to develop multivariable predictive models for identifying women who will develop preeclampsia and its adverse consequences are ongoing.\textsuperscript{65,66}

**Current Risk Assessment and Screening Practice in the United States**

In the United States, the number of prenatal care visits and the specific tests used for screening may vary across clinical settings and populations. Risk factor assessment via medical history taking is a routine part of prenatal care and is included in Medicaid standards of practice for prenatal care.\textsuperscript{67} Data from the National Maternal and Infant Health survey found that 80 percent of pregnant women had a medical history taken and 98 percent had their weight and height
measured in the first or second prenatal visit. Routine blood pressure measurement at prenatal visits is recommended by most prenatal care guidelines, including ACOG, and is considered best practice. In one survey, 96 percent of women reported receiving a blood pressure measurement at their first or second prenatal visit. According to national data on health care utilization from 2004 through 2006, pregnant women received an estimated 4.26 (95% CI, 3.35 to 5.16) urinalysis tests per pregnancy, but the number varied by patient race, insurance status, geographic location, and risk status. A survey of residency programs from 2007 found that urine dipstick tests at every prenatal care visit was taught at 94 percent. None of the prevailing clinical guidelines provide specific recommendations for the optimal number or interval of screening tests for preeclampsia in routine care (Table 1). Descriptions of the timing of routine prenatal care visits for healthy patients vary, but generally suggest 8 to 14 visits, with greater frequency closer to term.

General recommendations to screen for proteinuria screening throughout pregnancy are offered, although the value of proteinuria as a predictor of disease has been questioned. A systematic review including 11 studies (n=4,388) found low predictive accuracy for adverse outcomes from total proteinuria: sensitivity was 35 percent and specificity 89 percent. It also has been suggested that due to the lack of sensitivity and specificity, routine protein dipstick testing should be reconsidered. The gold-standard 24 hour urine protein test is not a practical screening tool, and there is increasing interest in the potential for using spot urine tests to determine the protein:creatinine ratio for diagnosis and for point-of-care screening.

Several organizations have guidelines to identify women at greatest risk of developing preeclampsia that are based on the clinical judgment of providers guided by an evaluation of risk factors that can be identified during a routine medical history (Table 2). No guidelines recommend the use of a specific risk assessment instrument or tool.

**Previous USPSTF Recommendation**

In 1996, the USPSTF recommended screening for preeclampsia with office-based blood pressure measurement using sphygmomanometry for all pregnant women at the first prenatal visit and periodically throughout the remainder of the pregnancy (B recommendation). The 1996 report did not contain an analytic framework or Key Questions (KQs), and it was not conducted as a systematic review according to current USPSTF procedures.
Chapter 2. Methods

Scope and Purpose

This report will be used by the USPSTF to update their 1996 recommendation on screening for preeclampsia and was developed using standard USPSTF procedures. Investigators in collaboration with the USPSTF and the Agency for Healthcare Research and Quality (AHRQ) created an analytic framework incorporating the KQs and outlining the patient populations, interventions, outcomes, and potential adverse effects for this review (Figure 1). The target population includes pregnant women, including adolescents. All KQs include studies of high- and low-risk populations unless otherwise specified.

We sought high-level evidence on the value of screening for preventing health outcomes compared to no screening (KQ1) and comparing different approaches to screening (KQ1a). We also reviewed multivariable tools for assessing the risk of preeclampsia that could be used to identify women for whom screening and clinical care could differ (KQ3). Evidence on the performance of routine screening tests for detecting preeclampsia (KQ4) and on harms of risk assessment (KQ3) and screening (KQ5) were also reviewed to balance any negative, inadvertent effects that clinical screening might incur.

Key Questions and Analytic Framework

An analytic framework (Figure 1) and five KQs (listed below) were developed in consultation with USPSTF members.

1. How effectively does screening for preeclampsia reduce maternal and perinatal morbidity and mortality?
   a. Does effectiveness differ by screening protocol (e.g., 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 (e.g., blood pressure, proteinuria) identify women with preeclampsia?
   a. How accurate are different screening tests for proteinuria?
   b. How effective are different screening protocols (e.g., 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?
Data Sources and Searches

We conducted an initial search for existing systematic reviews published from 1995 to February 17, 2014, in MEDLINE, PubMed, the Database of Abstracts of Reviews of Effects, the Cochrane Database of Systematic Reviews, and the Centre for Reviews and Dissemination Health Technology Assessment. We also reviewed systematic reviews and reports published by AHRQ, British Medical Journal Clinical Evidence, the Institute of Medicine, Clinical Key, DynaMed, and NICE. The literature search strategies can be found in Appendix B.

We performed comprehensive literature searches for primary literature in the MEDLINE, PubMed, and Cochrane Central Register of Controlled Trials databases from 1990 through September 1, 2015 (Appendix B). We excluded studies published before 1990, as changes in diagnostic criteria and treatments made data from the preceding era less applicable to practices in the past 25 years. Before 1990, relative increases in blood pressure or mean arterial pressure (MAP) and the presence of edema were included in the diagnostic criteria, so those screening approaches would not be applicable to current practices or the disease definition. Moreover, important developments in the prevention of eclampsia with magnesium sulfate and in the treatment of early preterm babies with lung surfactants occurred in the late 1980s and early 1990s, changing the balance of harms and benefits with regard to health outcomes for different preeclampsia screening and management strategies. Nevertheless, we remained attentive to the possibility that a landmark study predating 1990 could be relevant to our review and therefore included references from the 1996 USPSTF report. We reviewed reference lists of relevant studies and reviews to identify additional potentially relevant studies that were not identified by our literature searches. Additional references were obtained from expert reviewers.

Study Selection

Two investigators independently reviewed titles and abstracts using an online screening platform, Abstrackr.77 The same investigators reviewed full-text articles against prespecified inclusion and exclusion criteria (Appendix B Table 1).

The diagnostic criteria for preeclampsia have changed over time and there is international variability (Appendix C).78 Any standard diagnostic criterion for preeclampsia was allowed, as defined by the study. For evidence on KQs pertaining to the benefits and harms (KQ1 and KQ5) of preeclampsia screening for maternal, fetal, and infant health outcomes and the accuracy of screening for detecting women with preeclampsia (KQ4), we considered studies of screening occurring from 20 weeks of gestation until delivery since widely used diagnostic criteria for preeclampsia specify that it occurs after 20 weeks’ gestation.1,79 For KQs relating to risk assessment (KQ2 and KQ3), we included studies evaluating risk assessment tools applied in the first 20 weeks of pregnancy, prior to the onset of disease as defined by the diagnostic criteria for preeclampsia. One purpose of risk assessment would be to identify high-risk patients for low-dose aspirin prophylaxis, which may be more beneficial when begun early in the second trimester.80,81 Eligible study populations were pregnant women without a diagnosis of preeclampsia and asymptomatic for the condition. We did not exclude studies that included pregnant women with common chronic conditions often seen in primary care settings (i.e.,
chronic hypertension and diabetes mellitus) or those at elevated risk for preeclampsia. We did, however, exclude studies that solely focused on women seeking high-risk obstetric care, infertility treatment, inpatients, and other non-generalizable populations with select preexisting health conditions.

The health outcomes we considered for KQ1 were any benefits or harms related to maternal, fetal, or infant health. We placed priority on health outcomes known to be directly related to preeclampsia, such as eclampsia, HELLP syndrome, organ damage or failure, fetal growth restriction, preterm delivery, low birth weight, stillbirth, and placental abruption, some of which are associated with both short- or long-term health and developmental consequences.

The screening and risk assessment interventions we considered were point-of-care tests, measures, and evaluations conducted in routine primary prenatal care. For screening, point-of-care blood pressure measurements using manual or automated devices and point-of-care urine tests for proteinuria with qualitative, quantitative, visual, or automated readings were included. When assessing the diagnostic accuracy of point-of-care urine tests used to detect proteinuria (KQ4a), we excluded studies that did not use a 24-hour urine test as the reference standard. Secondary evaluations and tests used to assess preeclampsia severity or to confirm diagnosis were not included.

For risk assessment (KQ2), our a priori inclusion/exclusion criteria aimed to include externally validated multiple variable risk assessment tools that used patient history and routinely collected clinical measures (e.g., BMI, weight, blood pressure). We subsequently broadened our inclusion/exclusion criteria for risk assessment tools based on the low yield of potentially eligible studies, since most externally validated models to date also include serum markers, Doppler measures, or both. The serum marker and Doppler reading could be available for women who opt for aneuploidy testing in the first or early second trimester of pregnancy. Approximately 67 to 72 percent of pregnant women in the United States participate in aneuploidy screening, so these markers could be available for a large proportion of prenatal care patients.

Our inclusion criteria specified that we would exclude models that were not externally validated (i.e., models not tested in another population than the derivation study, assessing either performance or impact); internal validation studies are suited best to the population where they were derived, and most biases of internal validation studies are in the direction of overly optimistic results. Testing the predictive algorithm in a new population, ideally with a new study team, gives an approximation of the performance to be expected with broader application of a tool.

In the hierarchy of risk prediction tool validation, randomized impact studies are most valuable for purposes of clinical implementation since they give evidence on the expected effect of clinical application of the tool on health outcomes. The ideal impact study design would compare standard care to care with the risk assessment tool in a randomized controlled trial (RCT). The second-best level of evidence for a risk assessment tool comes from well-conducted external validation studies of a model in a different cohort (i.e., time, place, or both) by different investigators. External validation is important given that many of the threats to validity in model
development studies lead to overestimation of model performance (e.g., overfitting, selection of parameters).

We did not include studies not published in English or where the majority of participants were from countries that are not designated as having a very high Human Development Index, as defined by the United Nations Development Programme (2014). Studies conducted in other settings are less likely to offer evidence that would translate to the U.S. primary care setting in terms of the laboratory testing operations, screening modes, access and use of health care, treatments, and health outcomes. A list of excluded studies and reason for exclusion are provided in Appendix D.

**Quality Assessment and Data Abstraction**

Two investigators independently assessed the quality of all included studies using criteria predefined by the USPSTF and supplemented them with other criteria from the Quality Assessment of Diagnostic Accuracy (QUADAS) II for diagnostic accuracy studies (KQ4a), the Newcastle-Ottawa Scale (NOS), and the Before After Quality Assessment tool for observational studies (KQ3 and KQ5) (Appendix B Table 2). The critical appraisal of risk prediction models (KQ2) was developed for this review based on recent guidance on reporting on and quality appraisal for multivariable risk prediction models. Each included study received a final quality rating of good, fair or poor, and disagreements in quality were resolved through discussion. We excluded poor-quality studies (i.e., attrition >40%, differential attrition >20%, or other fatal flaws or cumulative effects of multiple minor flaws or missing information significant enough to limit our confidence in the validity of results). Good-quality studies met all or most of the assessment criteria. We rated studies as fair if they did not meet most of the good quality criteria.

One investigator abstracted data from all included studies into a Microsoft Access® database (Microsoft Corporation, Redmond, WA). A second investigator checked the data for accuracy. We abstracted study design characteristics, baseline population demographics, screening and risk assessment characteristics, health outcomes, adverse events, and diagnostic accuracy where applicable.

**Approach to Review of Clinical Risk Prediction Studies**

Our approach to the appraisal of the external prediction models identified in our literature search was informed by recent guidance articulated in the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) Statement and the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies (CHARMS). We used domains defined in both the TRIPOD statement and CHARMS tool, as well as related methods reports, to evaluate the externally validated risk prediction tools we identified. No published tool for quality appraisal of external model validation studies is currently available; existing guidance is focused on risk of bias in model development and internal validation studies. Since we included only models that have been subjected to external validation, the bias risks outlined in CHARMS regarding problems with
overfitting and overoptimistic performance were not as relevant, since these are generally addressed through the process of external validation. Therefore, we did not exclude any studies for quality concerns and instead described the performance of all of the externally validated models we identified, providing detailed information on the validation studies and reported measures of model performance. More details on methodological considerations that informed our evaluation of risk prediction models are available in Appendix E.

**Measures of Risk Prediction Model Performance**

Discrimination was consistently reported using the concordance index (c statistic), or area under a receiver operator curve plot, which represents the probability that a case will have a higher risk score than a non-case will. The degree to which a model correctly orders true positive and true negative results is represented by this statistic. Discrimination also is described with sensitivity, specificity, positive predictive values (PPVs), and negative predictive values (NPVs). *A priori* risk-level cutpoints are optimal, but in this literature “detection rates,” analogous to sensitivity, were commonly reported with risk cutpoints corresponding to a 10 percent false-positive rate (90% specificity).83

Calibration is another measure of model performance that reflects the extent to which the model predictions match the observed outcomes for individuals across risk levels. Goodness-of-fit tests (e.g., Hosmer-Lemeshow test) are sometimes used to report calibration, but calibration plots that graphically depict the observed outcome frequencies against predicted probabilities are more informative.85 Principles that are the basis for the TRIPOD statement indicate that discrimination and calibration are both “fundamental” for evaluating model performance in validation studies.85

**Data Synthesis and Analysis**

We created summary evidence tables for each of the KQs that include important population characteristics and study design features.

Our analysis was primarily qualitative owing to the low number of included studies identified for each KQ. For KQ2, we focused our reporting on models having at least good discrimination but extracted data from all models. In evaluating c index values we defined values below 0.70 as inadequate, from 0.70 to 0.79 as adequate, and 0.80 or higher as good to excellent.92 Our description of externally validated models focused on those with good to excellent discrimination.

Quantitative meta-analysis was conducted only for KQ4a. We calculated the diagnostic accuracy of point-of-care urine tests using the 24-hour urine collection as the reference standard (significant proteinuria defined as 300 mg with exertion over the 24-hour collection). We converted all urine excretion ratios to milligrams per millimole (mg/mmol) by converting any values to milligrams per gram (mg/g) and multiplying this value by 0.113.75 We stratified results by the type of point-of-care index test: protein:creatinine, albumin:creatinine, and protein dipstick. For studies reporting multiple thresholds, we selected the most clinically acceptable cutoff for each urine test to be used when pooling (30 mg/mmol,93 2.0 mg/mmol,94 and 1+,
respectively). Due to the small number of studies evaluating the diagnostic accuracy of the albumin:creatinine tests and protein dipstick, we qualitatively described the results and visually displayed the data. For studies evaluating the diagnostic accuracy of protein:creatinine tests, we ran bivariate analyses to simultaneously examine sensitivity and specificity using the *midas* meta-analysis for diagnostic accuracy command in Stata version 13.1 (Stata Corp LP, College Station, TX). The pooled sensitivity and specificity, heterogeneity, and pooled receiver operator curve (ROC) were calculated when enough studies using the same index test and reference standard were identified (≥8 studies). Our strategy for exploring heterogeneity was to examine patterns in the results based on population, intervention, comparator, outcomes, timing, and setting factors that might be associated with different findings. Factors we expected to be important *a priori* included the index test threshold, clinical setting, and inclusion criteria. For example, we stratified women by the entry criteria of enrollment due to high blood pressure, enrollment due to high blood pressure or other indications for proteinuria screening, and previous positive (+1) urine protein dipstick result. We conducted random-effects meta-regression using the *metareg* procedure in Stata to statistically test the contribution of factors likely to explain heterogeneity based on the patterns observed in the stratified test accuracy results. We used the restricted maximum likelihood method with the Knapp-Hartung adjustment.

We conducted sensitivity analyses, excluding studies that did not provide complete 2x2 results (e.g., studies reporting sensitivity and specificity only), those with cutoffs beyond ±5 mg/mmol from 30 mg/mmol, and those that were not at a threshold of 30 mg/mmol.

Statistical meta-analysis was not performed for all outcomes because of methodological limitations of the studies and heterogeneity in study designs, interventions, populations, and other factors, but we did conduct them when appropriate. Studies included in prior reviews were reviewed for consistency with current results; however, lack of studies and differences in scope, KQs, and inclusion criteria limited aggregate synthesis with the updated evidence.

**Expert Review and Public Comment**

A draft Research Plan for this review was available for public comment from May 22 to June 18, 2014. The draft version of this report will be reviewed by content experts, USPSTF members, AHRQ Medical Officers, and Federal Partners.

**USPSTF Involvement**

This research was funded by AHRQ under a contract to support the USPSTF. We consulted with four USPSTF liaisons during the development of the research plan. An AHRQ Medical Officer provided project oversight and reviewed the draft review. The USPSTF and AHRQ had no role in the study selection, quality assessment, or the writing of the systematic review.
Chapter 3. Results

Literature Search Results

We screened 10,082 abstracts and 378 full-text articles to identify 21 included studies reported in 35 publications (Appendix B Figure 1).94,97-118 We did not identify any studies that directly compared the effectiveness of preeclampsia screening in a screened population versus an unscreened population (KQ1). We included one RCT97 on the benefits and harms of differing visit schedules for preeclampsia screening for both KQ1a and KQ5 and one observational before-after study that assessed potential harms of different screening strategies (KQ5).115 We identified four studies reporting on the external validation of preeclampsia risk prediction models (KQ2)98-101 and a single observational study evaluating the harms of risk assessment (KQ3).102 We did not identify any studies evaluating the effectiveness of screening tests for identifying women with preeclampsia (KQ4). We included 14 studies examining the diagnostic accuracy of urine tests for proteinuria (KQ4a), which included comparisons of the test accuracy of different approaches to urine protein screening (KQ4b).94,103-114,119

Results of Included Studies

Key Question 1. How Effectively Does Screening for Preeclampsia Reduce Maternal and Perinatal Morbidity and Mortality?

No studies directly compared the effectiveness of preeclampsia screening in a screened population versus an unscreened population.

Key Question 1a. Does Effectiveness Differ by Screening Protocol (e.g., Tests Used, Timing of Tests, Rescreen Intervals) or Preeclampsia Risk Status?

Summary

One RCT conducted among 2,764 “low-risk” insured women seeking prenatal care in the first trimester in the early 1990s found that five fewer scheduled prenatal care visits (and thus fewer preeclampsia screenings) did not result in worse birth outcomes. However, the difference in the number of visits between groups was smaller than intended and the power to detect differences was insufficient for some important health outcomes.

Evidence

One fair-quality RCT randomized 2,764 pregnant women, enrolled from 1992 to 1994, aged 18 to 39 presenting for their intake visit in the first trimester to a routine number of prenatal care visits (14 visits) or a schedule of fewer visits (9 visits) (Appendix F Tables 1–3).97 A total of
2,328 women completed the study: 1,163 in the control group and 1,165 in the intervention group. The study sought to enroll “low-risk” women; the most common reasons for exclusion were: presenting too late for the first prenatal care visit (17% of women making their first visit arrived after 13 weeks gestation), being outside included age range, and having a previous high risk obstetric condition or a medical condition. Past high-risk conditions included preterm delivery, preterm labor, abruptio, severe preeclampsia, caesarean section (vertical incision), gestational diabetes, incompetent cervix, uterine anomaly, diethylstilbestrol exposure, isoimmunization, more than one second trimester abortion, fetal anomaly, or small for gestational age neonate. Current high-risk conditions included diabetes, chronic hypertension, drug or alcohol abuse, multiple gestation, conception through assisted reproductive technology, and large (≥4 centimeters[cm]) leiomyomata.

Routine prenatal care consisted of visits 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 number of visits was reduced to nine; they occurred at 8, 12, 16, 24, 28, 32, 36, 38, and 40 weeks of gestation. In both groups, the initial visit consisted of routine blood analysis, Pap smear, and gonorrhea and chlamydia testing; later tests included serum alpha-fetoprotein screening (15 to 18 weeks’ gestation), diabetic screening with the 1-hour glucose tolerance test and hematocrit (24 to 28 weeks’ gestation), and antibody screening (28 weeks’ gestation). Ongoing risk assessment occurred at each visit, which included blood pressure screening and urine testing for glucose and protein.

At baseline, there were no statistically significant differences between groups on maternal characteristics. During pregnancy, women in the control group had more healthcare visits in total (p<0.001), with a provider (p<0.001), and with a nurse (p=0.04) than did women in the intervention group, although the mean difference in the number of visits between the two study groups was smaller than intended (12.0 ± 4.2 versus 14.7 ± 4.2; p<0.001). At the time of delivery, there were no statistically significant differences between groups on maternal outcomes (e.g., gestational diabetes, preeclampsia), delivery complications (e.g., preterm delivery, cesarean section, postpartum hemorrhage), or neonatal outcomes (e.g., birthweight, gestational age, stillbirth) (Table 3). At 6 weeks postpartum, there were also no statistically significant differences between groups on satisfaction with prenatal care. More women in the intervention group felt the number of prenatal care visits was “just right” (p=0.002). Overall, reducing the number of prenatal care visits did not clearly affect health outcomes.

**Key Question 2. What Is the Effectiveness of Risk Assessment in Early Pregnancy for Identifying Women at High Risk for Preeclampsia?**

**Summary**

We identified five multivariable risk prediction models whose external validation studies (k=4) indicated, based on the $c$ statistic, good or better discrimination. Three of the models (Models A, B, C) aimed to predict early-onset preeclampsia (requiring delivery) and two to predict preeclampsia occurring or requiring delivery after 34 weeks (Models D and E). The models predicting early preeclampsia included clinical indicators, serum markers, and the uterine artery...
pulsatility index as parameters upon which risk prediction was based, whereas models for predicting later pre eclampsia did not include serum markers. Detection (i.e., sensitivities of the risk-based prediction models at a 90% specificity cut-point) was generally low (52 to 92%) with wide confidence intervals, and PPV was low (4 to 39%) for all of the models. The relationship between the predicted probabilities and observed outcomes (calibration) was not reported, so we could not evaluate model performance with this important metric. There was no clear evidence of high performance or clinical benefits for any of the externally validated models.

Evidence

Seven articles reported on results for four external validation studies (for 16 models) (Table 4; Appendix B Figure 1; Appendix G).98-101,116-118 Six models were developed for prediction of pre eclampsia requiring delivery before 34 weeks of gestation,120-125 one before 37 weeks’ gestation,126 seven after 34 weeks of gestation,121,122,125,127-130 and two at any time.126,127 External validation of the five models in the Italian cohort98 was for prediction of pre eclampsia diagnosed after 34 weeks’ gestation, regardless of delivery status. An additional eleven articles reported on the model development studies related to these external validations.120-130 Each external validation study98-101 had at least one model with c statistics indicating discrimination of 0.80 or higher, with a total of five models meeting this standard (Table 5).121,129 124,126 Models were labelled A through E for clarity of communication.

The external validation of models A through E was conducted using prospective cohort data collected in the United States by Oliveira and colleagues (n=2,962),99 Australia by Park and colleagues (n=3,014),100 Italy by Farina and colleagues (n=554),98 and Norway by Skrastad and colleagues (n=541).101 The Farina and Park studies enrolled women with singleton pregnancies presenting for aneuploidy screening, the Oliveira study enrolled women with singleton pregnancies presenting for prenatal care in the first trimester, and the Skrastad study enrolled nulliparous women. All of the cohorts were enrolled sometime between 2007 and 2012.

The external validation study evaluating the performance of Models A and B within a U.S. cohort of women presenting in the first trimester for prenatal care with singleton pregnancies at one of four centers in Baltimore, Maryland, was identified as the most likely to provide results applicable to U.S. prenatal care patients.99 Of these two models, Model B had better discrimination and detection and had also been developed in a U.S. population by Odibo and colleagues.124 Model B used clinical history, placental protein 13 (PP-13), pregnancy associated plasma protein A (PAPP-A), and the mean artery pulsatility index (c=0.86) to predict pre eclampsia- required delivery before 34 weeks’ gestation. It was validated with a smaller subset of the available cohort (n=871; 29% of the 2,969 women in the external validation cohort) because not all women had data on one of the serum markers needed for the model. Model B was initially developed and internally validated in a cohort of women presenting for aneuploidy screening, where discrimination and detection were similar to the external validation results.124

Model A was the only risk tool externally validated in more than one setting.99,100 In the U.S. cohort validation study, discrimination was moderate (c statistic, 0.80 [95% CI, 0.71 to 0.89]) and the detection (52%) and PPV (4.2) were low based on 29 cases (1% incidence). The same model was externally validated in the Australian cohort of women with singleton pregnancies
attending aneuploidy screening (n=3,014), where only 12 cases of early preeclampsia occurred. In that cohort, discrimination was high (c statistic, 0.93 [95% CI, 0.92 to 0.94]) as was the detection (91.7% [95% CI, 61.5 to 98.6]), but the PPV was low (3.6). Model A also was evaluated in an observational study occurring alongside the Australian external validation cohort study to assess the impact of using the model to assign women to low-dose aspirin prophylaxis.

High discrimination was seen for Model C when validated in a small Norwegian cohort of nulliparous women. The model was used to predict any preeclampsia requiring delivery before 37 weeks gestation (c statistic, 0.94 [95% CI, 0.86 to 1.00]). There were five cases of early preeclampsia requiring delivery (incidence 0.9%). Detection was 80 percent and the PPV was 6.8 percent. Model D and Model E used clinical history and uterine Doppler measures to detect later-onset preeclampsia; when validated with a small Italian cohort they had good to excellent discrimination, with detection of 85 percent and 74 percent, and PPVs of 39.3 and 36.3, respectively.

Information on model calibration was not provided in any of the model external validation studies, precluding a complete assessment of model performance. There were no randomized impact studies evaluating the effects of these models when used as risk assessment tools in clinical application relative to usual care.

Key Question 3. What Are the Harms of Preeclampsia Risk Assessment?

Summary

One fair-quality prospective cohort study (n=255) found no differences in anxiety before and after counseling on preeclampsia risk and categorization as high or low risk based on results of a multivariable risk prediction model. High-risk women were subject to changes in their clinical care, with usual care for the low-risk group. Measures of anxiety over time did not change but were collected from less than half of the study participants.

Evidence

We identified one fair-quality, prospective cohort study conducted in Spain that examined whether first-trimester risk assessment and clinical care protocols based on risk status increased anxiety in pregnant women (Appendix F Tables 1–3). Risk for early preeclampsia requiring delivery before 34 weeks was assessed using a model developed in Spain and externally validated in a U.S. cohort (Appendix G). Pregnant women screened as high risk were recruited and matched with the next visited low-risk screened woman in the first trimester screening unit (n=255: 135 low risk, 120 high risk).

After risk assessment, all participating women were provided counseling on the potential risks of preeclampsia. Women at high risk underwent a followup management protocol that included recommended daily intake of acetylsalicylic acid (150 mg) from the day of screening until 36 weeks’ gestation, and second-trimester ultrasonography at 20 to 22 weeks that included uterine...
artery Doppler velocimetry. There were no statistically significant differences in the demographic characteristics of women screened high risk and those screened low risk at the enrollment visit.

Anxiety levels were assessed within one hour post-counseling using a self-reported anxiety questionnaire (Spielberg State-Trait Anxiety Inventory [STAI]) to measure trait (STAI-T) and state (STAI-S) anxiety. Study participants completed the STAI-T on the day of preeclampsia screening (prior to risk counseling) and were asked to answer the STAI-S immediately after the counseling visit. In a subgroup of women (51 low risk and 50 high risk), anxiety levels were also measured during the second and third trimesters (data not shown).

At baseline, low- and high-risk women did not differ in STAI-T scores: 41.2 (standard deviation [SD] 6.7) versus 40.4 (SD 8.1); p=0.35. After risk assessment and counseling, the STAI-S scores for low- and high-risk women were 35.0 (SD 9.9) and 34.6 (SD 10.1), respectively (p=0.77). The proportion of women with high anxiety was also not significantly different between the two groups (28/134 [20.7%] vs. 24/120 [20%]; p=0.88). Measurements of anxiety levels taken during the second and third trimesters indicated no differences in the low- and high-risk subgroups, but less than half of the baseline participants provided data for these comparisons (data not shown). Overall, first-trimester preeclampsia risk assessment and counseling did not increase maternal anxiety, but risk assessment was coupled with counseling for all women and changes in clinical care for high-risk patients.

**Key Question 4. How Effectively Do Routine Screening Tests (i.e., Blood Pressure, Proteinuria) Identify Women With Preeclampsia?**

We found no evidence to evaluate the effectiveness of routine screening tests in identifying women with preeclampsia. Such studies would have evaluated how accurately a clinical blood pressure measurement or urinalysis identified women with the diagnosis of preeclampsia at that time. The only available evidence was for various screening approaches for establishing the presence of proteinuria (one diagnostic criterion for preeclampsia).

**Key Question 4a. How Accurate Are Different Point-of-Care Screening Tests for Proteinuria (a Diagnostic Criterion for Preeclampsia)?**

**Summary**

Fourteen studies evaluated the diagnostic accuracy of urine tests in detecting proteinuria compared to 24-hour urine collection (reference standard). All of the studies examined the test performance of protein urine testing in women with suspected preeclampsia. None considered accuracy in the context of repeated testing. The limited and variable evidence on test accuracy did not support conclusions about which test would perform best for routine screening. Twelve studies evaluated protein:creatinine test sensitivity (range, 0.65 to 0.96; I²=80.5% of 11 studies) and specificity (range, 0.49 to 1.00; I²=91.8% of 11 studies). Studies conducted among women with ≥1+ dipstick protein at enrollment had higher test sensitivity those with women seen for other indications of suspected preeclampsia (e.g., de novo hypertension). Two studies evaluating
albumin:creatinine urine tests had similarly high sensitivity (0.94 and 1.00) and disparate specificity (0.94 and 0.68). Four studies tested the accuracy of automated protein dipstick test (≥1+), with variable results—only one protein dipstick test had evidence of both specificity and sensitivity above 80 percent. Spectrum bias likely influenced all of the study results, so these are probably overestimates of the likely test performance in routine prenatal care.

Evidence

We identified 14 studies (four good-quality and 10 fair-quality) evaluating the accuracy of different screening tests for proteinuria (Appendix F Tables 1 and 2). Twelve studies evaluated protein:creatinine urine tests, two studies evaluated albumin:creatinine urine tests, and four studies evaluated urine protein dipstick tests (Figure 2; Table 6). Most studies evaluated the test performance characteristics for significant proteinuria across a wide range of thresholds (Appendix F Table 4). The reference standard for inclusion was the 24-hour urine collection; all studies used a urinary protein excretion threshold of 300 mg over 24 hours to diagnose significant proteinuria. The prevalence of significant proteinuria ranged from 8.7 percent to 93.8 percent (Appendix F Table 3). Four studies evaluated test performance characteristics for identifying severe proteinuria (e.g., urinary protein excretion threshold of ≥3,000 mg over 24 hours); all urine tests performed well as expected and are not further discussed (Appendix F Table 5).

All studies were conducted in pregnant women with suspected preeclampsia, which included those referred to 24-hour urine collection for suspected preeclampsia, chronic or de novo hypertension, and/or proteinuria (i.e., positive on at least one previous protein dipstick). Study participants were generally in their late twenties and early thirties and were evaluated near the end of their third trimester of pregnancy. The majority of studies were conducted in predominantly white populations; other racial and ethnic groups were underrepresented. Personal or family history of preeclampsia was not reported. Six studies were conducted in the United States, four in the United Kingdom, one in New Zealand, one in Canada, one in Chile, and one in the Netherlands.

Twelve studies evaluated the accuracy of protein:creatinine urine tests in 1,516 pregnant women (Figure 2). One study did not provide the raw data necessary for pooled analyses and we were unable to retrieve the data from the study authors, but the reported sensitivity and specificity are reported in our results narrative. Only one study reported the make or manufacturer of the protein:creatinine urine test being evaluated, which was the Albustix (Siemens Healthcare Diagnostics, Malvern, PA). The sensitivities of the protein:creatinine urine tests ranged from 0.65 (95% CI not calculable) to 0.96 (0.88 to 0.99), with most falling above 0.81; the combined sensitivity (k=11) was 0.85 (95% CI, 0.78 to 0.90) with high heterogeneity (I²=80.5%) (Appendix F Figure 1). The specificities ranged from 0.49 (95% CI, 0.36 to 0.63) to 1.00 (0.16 to 1.00); the combined specificity (k=11) was 0.85 (95% CI, 0.72 to 0.92) with very high heterogeneity (I²=91.8%). In pooled ROC analysis (Appendix F Figure 2), the area under the curve was 0.91 (95% CI, 0.88 to 0.93), with similar numbers of studies in the space above the curve and below. The dispersion of study data points reflects the considerable heterogeneity seen for both sensitivity and specificity and limits drawing summary conclusions about overall performance from the pooled ROC analysis.
To evaluate statistical and clinical heterogeneity, we examined the data using forest plots sorted by theoretically plausible factors. A correlation between test accuracy and the degree of likely spectrum bias was observed. Studies enrolling patients with a ≥1+ dipstick result\textsuperscript{103,108,111} or including any patients already diagnosed with preeclampsia\textsuperscript{110} would likely be subject to greater spectrum bias than those recruiting all patients with 24-hour tests for suspected preeclampsia, regardless of indication.\textsuperscript{105,107,109,119} Three studies fell somewhere in the middle, where patients could be enrolled with a ≥1+ protein dipstick reading, but also for other specific indications (e.g., worsening hypertension) that did not require a positive protein dipstick result.\textsuperscript{104,106,113} Sorting by the population enrolled, we observed that most of the highest sensitivities were in studies that enrolled only patients with a positive dipstick test; differences between groups were statistically significant in meta-regression (Appendix F Figure 3). Heterogeneity was low after adjustment for enrollment criteria ($I^2=15.9\%$). For specificity, however, these differences did not explain the high statistical heterogeneity, nor were alternative explanations found. Poor reporting across the included studies on the specific type test assay limited our ability to evaluate the possible contribution of the index test used against the observed heterogeneity.

Two studies evaluated the accuracy of albumin:creatinine urine tests in 321 pregnant women.\textsuperscript{94,106} Both studies evaluated the DCA 2000 point-of-care system (Bayer Healthcare, Whippany, NJ) that estimates albumin:creatinine in 7 minutes from a 40 microliter urine sample utilizing immunoturbidimetric (albumin) and colorimetric (creatinine) assays. The samples were collected before the 24-hour urine collection was initiated\textsuperscript{94} or in the early morning before the final 24-hour specimen.\textsuperscript{106} The sensitivities were similar at the common threshold of 2.0 mg/mmol, but the specificities differed (Figure 2). The sensitivity was 0.94 (95% CI, 0.85 to 0.98) and the specificity was 0.94 (95% CI, 0.87 to 0.98) in the good-quality study of 171 pregnant women (45% significant proteinuria).\textsuperscript{94} In the fair-quality study of 150 pregnant women (8.7% significant proteinuria), the sensitivity was 1.00 (95% CI, 0.75 to 1.00) and the specificity was 0.68 (95% CI, 0.59 to 0.76).\textsuperscript{106} This study also evaluated the albumin:creatinine urine test at two other thresholds (3.5 mg/mmol and 8.0 mg/mmol) and found similar sensitivities as (1.00) and higher specificities than those of the 2.0 mg/mmol threshold (0.88 and 0.96, respectively) (Appendix F Table 4).

The good-quality study in 171 pregnant women also evaluated the Microalbustix and Clinitek albumin:creatinine urine dipsticks (Bayer Healthcare, Whippany, NJ).\textsuperscript{94} The Microalbustix was read visually by two observers while the Clinitek dipstick only could be read on the Clinitek 50 urine chemistry analyzer. The visually read Microalbustix dipstick had lower sensitivity (0.49 [95% CI, 0.38 to 0.61]) than the automated Clinitek dipstick did (0.58 [95% CI, 0.47 to 0.70]), but identical specificities (0.83 [95% CI, 0.74 to 0.90]) (Appendix F Table 4).

Four studies evaluated the accuracy of protein dipsticks in 634 pregnant women with mixed test performance characteristics.\textsuperscript{94,105,106,112} None of the protein dipsticks was the same make or model; only the reference standard assay (benzethonium chloride [BEC] assay) was similar between studies. Two studies obtained the urinalysis sample before the initiation of the 24-hour collection.\textsuperscript{105,106} The other two studies used aliquots from the thoroughly mixed 24-hour collection.\textsuperscript{94,112} The sensitivities ranged from 0.22 to 1.00 and the specificities from 0.36 to 1.00 (Figure 2). Of these, only one study had both sensitivity and specificity above 0.80.\textsuperscript{94} This good-quality study with 171 pregnant women evaluated the visual reading by two observers and the
automated reading using the Clinitek 50 of the Multistix 8SG dipstick (Bayer Healthcare, Whippany, NJ). The visually read protein dipstick had a lower sensitivity (0.51 [95% CI, 0.39 to 0.62]) than the automated reading did (0.82 [95% CI, 0.71 to 0.90]), but the specificities were similar (0.78 and 0.81, respectively). Apart from that study, the others had very high sensitivity and low specificity or vice versa (Appendix F Table 4).

Key Question 4b. How Effective for Identifying Women With Preeclampsia Are Different Screening Protocols (e.g., Instruments, Test Procedures, Timing of Tests, Rescreen Intervals)?

Summary

There was no evidence evaluating the effectiveness of blood pressure or urine screening for identifying women with preeclampsia and no evidence to inform comparisons among various screening protocols using these tests. However, a few of the studies for KQ4a on various screening approaches for detecting proteinuria, a diagnostic criterion for preeclampsia, reported on the effect of variations in urine sample collection, assay methods, and reading approaches for urine screening tests. On the basis of the included studies for KQ4a, within-study comparisons suggest that automated rather than visually read tests have higher test performance, the time of day of testing is not predictive of performance for the protein:creatinine test, and the sensitivity of tests depends on the 24-hour test assay used.

Evidence

Visual Versus Automated Readings

One good-quality study compared visual and automated readings of protein dipsticks and albumin:creatinine urine tests in 171 pregnant women. Two trained observers, who were blinded to each other’s readings, visually read the results of the tests and the same samples were then re-tested using the automated analyzers. For both types of tests, the visual reading performed lower than the automated reading (Appendix F Table 4).

Time of Urine Sample Collection

One good-quality study evaluated the test performance characteristics of the Albustix (Siemens Healthcare Diagnostics, Malvern, PA) protein:creatinine urine test when collected from 105 pregnant women in the morning (8 AM), afternoon (12 PM), and evening (5 PM). Using a protein:creatinine excretion threshold of 30 mg/mmol, there were no statistically significant differences in the sensitivities (p=0.12) and specificities (p=0.89) between timepoints (Appendix F Table 4).

24-Hour Assay

One fair-quality study evaluated the test performance characteristics of the BM-Test-5L urine protein dipstick (Boehringer Mannhein, East Sussex, UK) and compared it with two standard qualitative protein assays using the pooled 24-hour urine collection (BEC assay and the Bradford
The BEC assay is an immunoturbidometric assay that is the most frequently used method in clinical practice to assess proteinuria in a 24-hour collection aliquot. The Bradford assay is based on the ability of proteins in the urine to bind to Coomassie blue dye and is more frequently used in laboratories. The prevalence of significant proteinuria among 197 women varied greatly between assays (70.1% for BEC assay and 24.9% for Bradford assay); therefore, the sensitivities of the dipstick were markedly different between assays (0.22 vs. 0.57, respectively) while the specificities were similar (0.98 and 0.97, respectively) (Appendix F Table 4).

### Key Question 4c. How Should Women at High Risk for Preeclampsia Be Screened Differently From Women at Low or Average Risk?

We found no evidence that compared different screening strategies among women at high risk for preeclampsia versus women at low or average risk.

### Key Question 5. What Are the Harms of Preeclampsia Screening and Do They Differ by Risk Status or Screening Protocol?

#### Summary

Two fair-quality studies were identified that reported on potential harms of different approaches to preeclampsia screening. Neither found evidence of harms, but both were underpowered to provide evidence on rare, but important, clinical outcomes.

#### Evidence

The fair-quality trial included for KQ1a found no difference in birth outcomes (e.g., low birth weight, preterm birth, number of cesarean sections) with an intended reduction in the number of prenatal care visits from 14 to nine visits (Table 3). The difference in the mean number of visits was not as great as expected (12.0 visits in the intervention group and 14.7 in the control group; p<0.001). As previously noted, power was not sufficient to detect differences for rare outcomes related to preeclampsia, particularly serious adverse maternal events such as progression to eclampsia, organ failure, stroke, and death.

We also identified one fair-quality retrospective before-and-after comparison cohort study (N=1,952) that evaluated the differences in health outcomes after a change in the standard of care at a hospital-based nurse midwifery practice that primarily served low-income Hispanic women (74% of eligible study) from routine prenatal dipstick urine testing to “clinically indicated” urine testing (Appendix F Tables 1–3). All women in the study received urine tests at their first prenatal visit, but 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=1,019) had subsequent urine screening only when certain conditions were indicated: symptoms of a urinary tract infection, severe vomiting, weight loss ≥0.9 kilograms since previous visit, SBP ≥140 mm Hg, DBP ≥90 mm Hg, a condition requiring periodic urine testing (e.g., chronic hypertension, renal disease). Women who
were enrolled before but delivered after August 15, 2002, were excluded (n=570). The two cohorts delivering before and after the practice change were similar in terms of baseline characteristics except for insurance payment source (p<0.0001).

Women in the routine urine testing group had used an average of 7.8 (range, 0 to 19) chemical reagent strips—equivalent to the number of tests—while women in the indicated testing group had used an average of 1.4 (range, 0 to 16). Among the indicated testing group, the reasons for urine testing were urinary tract infection (UTI) or vaginitis symptoms (31.5%) and elevated blood pressure or significant preeclampsia-related symptomatology (35.6%).

Since the purpose of the study was to evaluate whether changes in the urine screening approach were safe (i.e., did not change preeclampsia or other adverse condition diagnosis rates or other health outcomes), statistical tests were designed to evaluate noninferiority; thus, statistically significant p-values indicated equivalent diagnosis rates between the two groups (Table 3). These results suggest there were no differences in diagnosis rates of preeclampsia/eclampsia, high blood pressure, or gestational hypertension or in cesarean deliveries. Preterm delivery rates were not equivalent (p=0.14) but were lower with indicated testing (7.7% with routine testing, 4.9% with indicated testing). Overall, there was no evidence of reduced diagnosis of preeclampsia or of adverse health outcomes when changing from routine to clinically indicated urine testing. There also was no evidence suggesting underdiagnosis of adverse outcomes related to urinalysis for bacteriuria.
Chapter 4. Discussion

Summary of Evidence

We reviewed externally validated multivariable models for predicting preeclampsia, most focusing on the prediction of early-onset disease, but did not identify a model with supporting evidence indicating readiness for clinical use. A few models had good discrimination (c index ≥ 0.80), but CIs on estimates were wide, PPVs were low, and calibration statistics or plots were not provided so it is not possible to determine likely performance in clinical use. As others have emphasized, at minimum predictive models must be externally validated and shown to have acceptable discrimination and calibration before they might be ready for clinical practice. Beyond that, it is desirable to determine the likely performance or clinical impact of these models. These evidence standards were not achieved by the currently available studies of predictive models in preeclampsia. Further, the serum markers and Doppler ultrasound tests used in the models require resource-intensive collection and evaluation using complex algorithms. Efforts are under way to externally validate models using more easily collected clinical history information.

Although we found considerable evidence on the accuracy of protein urine screening at a single time point in pregnancy, the studies were conducted in only women with suspected preeclampsia and thus are not representative of women presenting for routine prenatal visits. Of the different protein urine tests that can be conducted with point-of-care urine samples, protein dipstick tests are easy and low cost, but we identified only four eligible studies, which had highly variable results. The accuracy of albumin:creatinine tests was high, but only two studies contributed evidence. The majority of the evidence on accuracy was for protein:creatinine ratio tests, which can be conducted using a variety of tools (e.g., automated dipstick or aliquot readers, laboratory assays), but high heterogeneity precluded summary generalizations about performance. While these tests are often evaluated as a potential alternative to 24-hour urine collection for diagnostic confirmation, they have also been proposed for use in routine screening, particularly to rule out proteinuria. The available evidence, however, does not facilitate conclusions about which point-of-care proteinuria test would be optimal for this purpose. There was considerable variation among tests, and the variation is difficult to explain given limitations in the evidence. Based on our findings and likely spectrum bias, we would expect quite variable and generally lower test accuracy performance in routine care of general prenatal care populations. Finally, no studies evaluated the performance of urine protein screening in the context of repeated testing, where false-negative results might be corrected over time.

Preeclampsia Screening for Reducing Morbidity and Mortality

We found no evidence that directly compared health outcomes in a screened population compared to an unscreened population (Table 7). One large trial, conducted in 1996 in a large health maintenance organization, suggested a schedule of somewhat fewer prenatal risk assessment and screening visits (and thus blood pressure and urine screening tests), which resulted in similar maternal and infant health outcomes and patient satisfaction among selected...
women determined to be low risk for adverse birth outcomes. The extensive inclusion and exclusion criteria, including required initiation of prenatal care in the first trimester of pregnancy, limited the potential relevance of this large trial. Generally, women presenting for prenatal care in the first trimester of pregnancy differ from those presenting at later gestations; tending to have higher socioeconomic status, more planned pregnancy, and higher rates of health insurance. Even in this study conducted among insured women, 17 percent of women presenting for care were excluded because they made their first prenatal care visit after 13 weeks. The applicability of the study, conducted over 20 years ago, in the context of updated diagnostic criteria and treatment algorithms may be limited for current clinical practice settings and populations. The study does, however, represent an important attempt to provide evidence for a specific screening approach relative to standard care.

Two fair-quality studies reporting on rates of adverse events with different screening protocols did not identify any harms of preeclampsia screening (Table 7). The 1996 trial of reduced prenatal visits found that preeclampsia diagnoses and related adverse outcomes did not differ with fewer screening visits. The difference in the number of visits obtained between the two groups ultimately was not as large as the study had aimed to generate, however, and the study’s age limits its value for current populations and practices. The study was also underpowered to detect rare but serious health outcomes related to preeclampsia. The other study provided some reassurance that there is no harm associated with a change in protocol where point-of-care urine tests are conducted for specific indications rather than on a routine basis, resulting in fewer tests on average. The before-after study design was subject to more potential threats to bias, including secular trends across the study period, than a randomized study would be. Reported differences in the study groups would tend to bias results in a conservative direction—toward worse preeclampsia-related outcomes in the indicated testing group—but this was not observed.

Effectiveness of Routine Screening for Preeclampsia Detection

No studies directly evaluated the test accuracy of blood pressure screening or urine protein screening for detecting the presence of preeclampsia. In a screening test performance paradigm, evidence to answer KQ4 would compare results from screening blood pressure measurements and protein urine tests to confirmatory gold standard diagnostic tests. These data would populate a test accuracy table for assessing sensitivity and specificity of the individual or combined use of these screening tests for detection of preeclampsia. Since screening for preeclampsia is ongoing throughout pregnancy and provided at multiple time points, screening test accuracy must consider the timing of the test, cumulative clinical results, or both. A single positive proteinuria screening test is followed up with other diagnostic tests and, if the result is not confirmed, routine screening continues. Estimating how often false-positive and false-negative readings occur for elevated blood pressure and proteinuria has not been a research priority, likely owing to the low-resource nature of the screening tests and the low risk to patients of the confirmatory tests (e.g., additional blood pressure measurements, repeat point of care urine test, diagnostic urine test).

Although multiple studies have tested the associations of continuous levels of blood pressure and proteinuria with the likelihood of developing preeclampsia at future time points in pregnancy, these do not directly address questions about screening test effectiveness in detecting
preeclampsia. High blood pressure is an important sign of preeclampsia, hence the importance of the current clinical practice of repeat measurement at clinical visits. The accuracy of individual blood pressure readings is optimized if conducted in accordance with guidance on clinical blood pressure measurement in generally and during pregnancy.

There is less evidence to support the relationship between proteinuria levels and adverse preeclampsia outcomes. Recent changes to the ACOG guidelines regarding the role of protein urine screening in the definition and management of preeclampsia highlight the importance of other signs and symptoms that might be used to diagnose preeclampsia in the absence of proteinuria. Because the disease is not well understood and unexpected cases of preeclampsia that are not preceded by high blood pressure do occur, in the absence of an alternative evidence-based screening strategy this historically important and relatively inexpensive aspect of prenatal health care will likely continue. We found no evidence of harms and no evidence to evaluate whether the use of one type of urine protein screening test versus another contributes to improved pregnancy outcomes.

**Accuracy of Urine Screening Tests for Proteinuria**

Most of the available evidence assessed how well point-of-care urine protein screening tests detect proteinuria, a diagnostic criterion for preeclampsia diagnosis, compared to the 24-hour collection gold standard. The most commonly studied test was protein:creatinine ratio, which has been cited in recent diagnostic criteria as a reasonable alternative to 24-hour urine collection. The range of sensitivities and specificities was wide, and heterogeneity was high: sensitivity and specificity near 80 percent for studies using thresholds at or near 30 mg/mmol likely represents a best case scenario, but the high and largely unexplained heterogeneity limits conclusions. In routine prenatal care for women presenting outside of monitored study conditions, performance could be considerably lower and is likely to be more variable across clinical settings given diversity in the tests available and 24-hour test assays. Our findings on the protein:creatinine ratio test are consistent with those conducted by others in recent years. A review by Côté and colleagues concluded that although this ratio was not sufficiently accurate to replace the 24-hour test for diagnostic confirmation, it does perform well enough to serve as a screening test to rule out significant proteinuria (0.3 g/day).

Based on relatively limited evidence available on other point-of-care tests, quantitative rather than qualitative readings of various urine dipstick tests appear more accurate. In two studies, quantitative albumin:creatinine ratio urine testing could obtain sensitivity and specificity in ranges similar to those of the protein:creatinine tests. These two studies’ populations exhibited a very different prevalence of proteinuria, one with less than 10 percent and the other with nearly 50 percent having significant proteinuria according to the 24-hour test. In both studies, however, albumin:creatinine ratio testing had higher sensitivity and specificity values than were reported in studies of the protein:creatinine ratio test. As in our own review, another recent review noted the limited evidence but also the potential promise of the albumin:creatinine ratio test, deeming the test deserving of further investigation.

Evidence on test performance of point-of-care urine screening tests is both statistically and clinically heterogeneous. Different assays for the 24-hour reference test, manufacturers of tests
kits and readers, laboratory procedures, and protocols for collection and reading of the index and gold standard tests could all contribute to the range of sensitivities and specificities observed. We sought to identify potential patterns in test performance based on these methodological characteristics but found few clear signals. The degree of spectrum bias, which is based on the extent to which preeclampsia was suspected, may have contributed to the heterogeneity in sensitivity, but the heterogeneity in specificity could not be explained. Limited information on the types of index tests and 24-hour assays used did not permit us to examine the role of different tests and procedures on performance.

None of the proteinuria test accuracy studies were conducted in a general primary care screening population of pregnant women. Instead, all of the available evidence was from pregnant women in the later part of the third trimester of pregnancy already identified with suspected preeclampsia and undergoing evaluation or diagnostic confirmation (Table 7). Tested populations were also predominantly white, further limiting the applicability of this review to women who bear a disproportionate burden of preeclampsia-related morbidity and mortality. Thus, the test accuracy estimates are subject to spectrum bias, wherein the higher pretest probability for the condition results in higher test performance.96 Furthermore, nine of the studies included or were conducted among inpatients, where the fidelity to collection protocols is higher than that in routine outpatient care, where problems with 24-hour gold standard tests have been well articulated.144,145 Incomplete collection and erratic patterns of protein excretion can influence test accuracy results, making the clinical feasibility and reliability the standard not entirely “golden”. It is likely that the test performance results we obtained are an upper boundary for performance and that studies in general populations or high-risk populations that do not already have a positive screen result, would be more informative.

Considered broadly, protein urine screening testing occurs throughout pregnancy and false positive results may lead to greater surveillance or further confirmatory testing. As noted in the review by Morris and colleagues,142 the implications of the sensitivity and specificity of tests depend on the clinical actions taken with positive and negative results. Additionally, test performance for both would likely improve when cumulative rather than single test performance is considered, as false negative readings can later be identified as positive and false positive findings become subject to disconfirmation with followup testing. Thus, maximizing single-test performance for a relatively inexpensive and noninvasive test has limited value.

**Effectiveness of Risk-Based Screening**

We did not identify any studies that assessed the performance of different preeclampsia screening strategies with high- or low-risk women or compared screening effectiveness between women considered to be at high or low risk. We reviewed multivariable risk prediction models that could be useful for risk-based screening approaches or for targeting other clinical preventive services, such as aspirin chemoprophylaxis for preeclampsia prevention. We identified 16 models externally validated in four studies that assessed the performance of multivariable risk assessment tools for use in pregnancy before 20 weeks’ gestation (Table 7). Of the five models found to have good or better discrimination based on the c statistic, two had high detection (10% false-positive rate) but all models had low PPVs. Importantly, information on model calibration was not provided for any of the risk prediction models and we could not determine the likely
performance or impact of the available validated risk prediction models that would be expected in routine clinical use. Two recent systematic reviews,\textsuperscript{99,141} several methodological critiques,\textsuperscript{90,146-148} and recent guidance from ACOG\textsuperscript{134} supported our assessment of the current state of the evidence aimed at developing a model for preeclampsia risk prediction.

Our review identified one small, fair-quality prospective cohort study that evaluated potential psychological harms of preeclampsia risk assessment using a multivariable risk assessment model (Table 7) and found no difference in anxiety before and after clinical risk assessment.\textsuperscript{102} A recent qualitative study, which analyzed in-depth interviews with women who had been identified as low or high-risk using a published risk assessment tool,\textsuperscript{149} found that some women had strong negative perceptions of being labeled at high risk.\textsuperscript{150} We did not find any evidence on harms related to risk assessment for health or pregnancy outcomes.

Owing partly to the rarity and complexity of preeclampsia and partly to limitations in methodological rigor, the current evidence for preeclampsia risk prediction does not support conclusions regarding likely performance, benefits, or harms. Recent efforts to establish reporting guidelines for prediction models and the growing maturity of risk assessment in other clinical areas will likely improve efforts to develop and validate risk prediction tools for preeclampsia.\textsuperscript{85} All of the models we identified included clinical tests that are not routinely collected for all pregnant women in early pregnancy. While the collection of serum marker data might be feasible in the context of aneuploidy screening, first-trimester ultrasound scans to calculate the uterine artery pulsatility index for those who seek it are not currently recommended in routine prenatal care. The additional resources needed to conduct clinical tests and calculate risk might be worthwhile if there were a clear net benefit of a risk assessment tool relative to usual practice, where current risk assessment recommendations are based on established risk factors and the clinician’s judgment.

**Evaluating Clinical Performance of Preeclampsia Risk Prediction**

To inform clinical practice, determining the net effect of risk assessment and the clinical actions that follow is necessary.\textsuperscript{133} High sensitivity may be more important for preeclampsia risk assessment because false-negative results could be more detrimental than false-positive results;\textsuperscript{151} a lower risk threshold, lower PPV, and possibly low-dose aspirin prophylaxis may be reasonable to consider for heightened surveillance.\textsuperscript{152} Harms could occur among women misclassified as low risk, not offered aspirin prophylaxis, and assigned to lower-intensity prenatal care schedules. Heightened surveillance and allocation of aspirin to women at low risk for the disease could also lead to harms.

In the absence of studies comparing model-based risk assessment to usual care risk assessment, it is not clear whether a formal risk assessment algorithm would improve performance or health outcomes beyond currently recommended risk assessment practice.\textsuperscript{133,135,153} The USPSTF has issued an evidence-based recommendation (Grade B) for the use of low-dose aspirin (81 mg) to prevent preeclampsia and associated morbidity and mortality.\textsuperscript{46,154} The recommendation specifies that benefits relative to potential harms be observed for women at increased risk of preeclampsia. The USPSTF, like NICE, provides a list of risk factors to assist clinicians in determining which patients are at elevated risk for preeclampsia (Table 2). These risk factors are based on the
strongest epidemiologic associations and the most common risk factors used to select high-risk participants in trials of aspirin for preeclampsia. These approaches to risk stratification might be improved upon with further development of validated tools for combining individual risk factors into an algorithm to provide more personalized risk assessment but, as we noted above, no externally validated models we evaluated provided evidence of performance that supports their clinical use.

Comparing risk prediction model effectiveness to NICE and USPSTF clinical risk assessment criteria would help provide a means of evaluating relative performance and whether incremental gains are worth implementation costs.\textsuperscript{135,153} We identified a few risk prediction models that compared detection in the proposed model with detection that would be achieved for the same population using the NICE criteria\textsuperscript{65,155-157} based on maternal characteristics and maternal history (Table 2). A recent study by Wright and colleagues reported better detection using a new (but not yet externally validated model) compared with the NICE guidelines for risk stratification.\textsuperscript{155} Although detection was higher with the model than with the NICE criteria (67% vs. 58%), models derived and tested in the same dataset tend to overestimate discrimination and the observed difference may not be reproduced in external validation. We did not identify any studies comparing model performance to the risk factors in the USPSTF clinical considerations for assessing risk for purposes of low-dose aspirin prophylaxis.\textsuperscript{46}

Another risk prediction model recently reported by Macdonald-Wallis et al. employed two different general population cohorts to establish a model for predicting preeclampsia, preterm birth, and small for gestational age babies.\textsuperscript{158} The study more closely adhered to TRIPOD guidance for development and reporting of prediction modeling results, including provision of calibration plots for their best performing model. Because our review was scoped to identify models for use in the first half of pregnancy, this study was not included because the optimal model in external validation, for which calibration, recalibration, and classification were reported, included measures added to the model at 28 weeks gestation and beyond. While not included in this review, our conclusions would not differ substantially with its findings. It does, however, represent an informative effort. First, the model proposed does not rely on serum markers or Doppler ultrasound measures, but instead uses common clinical history measures and an estimate of MAP based on blood pressure measurements (i.e., SBP times 2 divided by DBP), which would be easily collected in primary care. Secondly, it is conceptually worthwhile to consider whether preeclampsia risk prediction might be beneficial on an ongoing basis across pregnancy, alongside screening. Clinicians may informally assess risk at each prenatal care visit based on the signs and symptoms they observe – with more or less concern that the patient is likely to develop preeclampsia, and decisions regarding additional visits or screening tests informed by these evaluations. The results of this model validation study do not provide evidence on how the proposed model would improve upon current approaches.

The levels of discrimination and classification achieved with the proposed models are modest until adding the measurement of MAP at 28 weeks gestation, at which point the AUC is >0.80 ($c = 0.84, 95\% CI 0.79$ to $0.88$).\textsuperscript{158} The AUC is highest when MAP at 36 weeks is added to the model (0.88, 95\% CI 0.84 to 0.93). Model calibration and recalibration plots and classification tables are reported for the model that includes the 28 week MAP measurement. Similar to the included models in our review, the PPV was very low. Although models using the same
approach were developed to predict preeclampsia related health outcomes (i.e., preterm birth and small for gestational age), but had lower performance than for prediction of preeclampsia.

Despite better adherence to rigorous conduct and reporting of model development and validation, including a recalibrated model based on calibration plot observations, the clinical implications for the model remain unclear. The study authors conclude that the model could be used to risk stratify the clinical care of women that need more intensive monitoring from those likely to have a normal pregnancy. Whether the model would improve upon current practices for adjusting the intensity of prenatal monitoring is unknown, however, since blood pressure readings at prenatal visits already serve as a key indicator of the need for closer monitoring. An accompanying editorial questioned what would be gained or lost with reduced monitoring that might be proposed for those with a low risk score using the model, particularly given other activities that occur in prenatal visits. Conversely, there is an absence of information on harms for risk prediction that could occur for the many women unnecessarily assigned to heightened surveillance given the high false positive rates for the prediction model. Without comparisons of proposed models to current clinical practices, the potential benefits and harms of risk prediction cannot be determined, even with clear reporting of validated models. Testing different prenatal care algorithms against usual care, possibly incorporating use of the best performing and most feasible models would be valuable to the field.

An example is a nonrandomized impact study conducted by Park and colleagues that tested whether risk prediction Model A would improve the clinical process for assigning women at high risk for preeclampsia to low-dose aspirin prophylaxis in order to improve health outcomes. The simulation resulted in a statistically significant reduction in early preeclampsia cases (12 cases vs. 1 case, p=0.01). The findings are encouraging but are limited by substantial, unexplained differences in the demographic characteristics of the two cohorts that could introduce confounding. Secular changes over the study time period (April 2010 to June 2013) might also have influenced the results, as would the approaches to risk assessment and aspirin prophylaxis in the original comparison cohort. Similar work to assess the clinical impact of risk prediction tools using randomized study designs should be pursued, ideally, with comparisons to current usual care.

**Applicability**

The only study we identified that compared different screening strategies in a randomized study was conducted over 20 years ago in a large health maintenance organization. In the context of updated diagnostic criteria and management protocols for preeclampsia, application of this study to current clinical practice settings and populations is limited.

We did not identify any studies assessing the accuracy of urine protein tests in general prenatal care populations nor for the common practice of repeated testing over the course of pregnancy. The role of urine protein in the detection and diagnosis of preeclampsia is undergoing reevaluation as new guidelines and understanding of the disease process emerge. This may warrant consideration of different approaches to screening for women with hypertension in the absence of proteinuria in future reviews, as evidence accrues on the use of newer diagnostic
Limitations of the Review

The absence of rigorous studies on different approaches to preeclampsia screening, including risk-based approaches to care, is the most notable limitation. Newer studies are needed, in the current population of prenatal care patients with higher prevalence of obesity and other preeclampsia risk factors, to improve the evidence base for conducting screening. Moreover, with recent changes to diagnostic criteria additional screening tests for women with hypertension but not proteinuria require study to estimate potential clinical benefits, harms and performance and to develop evidence-based approaches to screening.160

One reason for the absence of data on the test performance of screening with blood pressure and urine protein tests is the fact that these have until very recently also been the primary diagnostic criteria for the condition, thus repeat tests to confirm initial readings are commonly conducted. The need to evaluate the proportion of women misclassified as having or not having preeclampsia at a single point in time, or even over the course of pregnancy may be less important when the result of a false negative is to continue screening, and the result of a false positive is enhanced surveillance, also resulting in continued screening. The complexity of the condition and limited tools currently available for screening may, have necessarily focused scientific resources on efforts to better understand the pathogenesis and potential new disease markers. Advances in proteomics and genomics may yield better precursors and markers, as well as new treatments for preeclampsia.161-164

Multivariable risk prediction tools that have been developed have aimed to combine serum tests and ultrasonography measures with known clinical history risk factors. Only recently have efforts been undertaken to combine a more robust set of known maternal history risk factors into a prediction model,136 and this work has not yet been externally validated. Very large cohorts are needed to develop and test models for early preeclampsia, when the risk of poor outcomes is greatest. Many of the studies we identified had very few cases to classify, so the confidence intervals for performance estimates were wide. Transparency and completeness of reporting has shortcomings in the literature on preeclampsia prediction modeling, as noted by others.90,146 In particular, the absence of calibration statistics limited our ability to comprehensively evaluate and compare model performance.148 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.165

Eleven included studies were pooled for meta-analysis of the accuracy of tests for protein:creatinine ratio, with results indicating high, mostly unexplainable heterogeneity, which limited our ability to generalize the accuracy of test performance. Heterogeneity was due in part to the degree to which preeclampsia was suspected. A recent methodological review for diagnostic studies noted limitations in the methods for determining the overall degree of heterogeneity in test accuracy studies, in part because of the interrelatedness of sensitivity and specificity.96 The bivariate random-effects model can take these correlations into account, but interpretation is challenging and better approaches are needed.
**Future Research Needs**

The complexity of preeclampsia and the variety of ways it can present with regard to timing, signs, and symptoms make it difficult to identify a highly effective and broadly applicable risk assessment and screening strategy. Studies aimed at testing risk assessment and screening algorithms are needed to provide a more robust evidence base with which to inform prenatal care screening practices. 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. Basic descriptive studies characterizing variations in current preeclampsia screening practices in different types of health care settings would be helpful for identifying alternative screening approaches to evaluate in clinical studies. Basic research into the pathophysiology of preeclampsia will also help identify better tools for risk assessment and screening. If preeclampsia is comprised of several distinct syndromes, new screening approaches will need to be developed to distinguish subtypes, especially those most likely to result in serious morbidity or mortality without intervention. If, as others have proposed, preeclampsia is a single condition with a spectrum of severity and speed of progression, it will remain important to identify the best tools and clinical protocols for capturing all cases as soon as they arise for enhanced monitoring and evidence-based treatment protocols. Additional risk assessment measures in the second half of pregnancy may further predict the likelihood of developing disease. For example, uterine artery Doppler measurement in later pregnancy and serum uric acid levels are potentially useful for identifying women at risk for adverse outcomes resulting from preeclampsia. Future validation of these techniques may help guide more individualized surveillance of women at highest risk.

Protein dipstick tests are an initial screen for preeclampsia in many health care settings and are used for diagnosis when other tests are not available. Urine dipstick tests for proteinuria have poor test performance, particularly with visual rather than automated readings. Further research on the extent to which these tests are used to guide clinical decision-making and whether variations in practice explain differences in health outcomes could inform investigations into best practices. Assessing the protein:creatinine ratio in point-of-care urine samples appears to have more evidence suggestive of better performance, but further evaluation of accuracy in general populations, and with repeat testing could better estimate its optimal role for proteinuria detection for routine preeclampsia screening.

Recently published models based on maternal characteristics and clinical history may hold promise if external validation supports their reproducibility. These studies have focused on a larger set of health behaviors, clinical measurements, health history, and maternal characteristics. A model developed by North and colleagues in the international cohort study Screening for Pregnancy Endpoints (SCOPE) sought a clinical tool for risk prediction, for use at around 15 weeks’ gestation in nulliparous women, based on clinical risk factors that could be easily collected in routine care. Based on an international cohort of healthy nulliparous women from New Zealand, Australia, the United Kingdom, and Ireland (n=3,529) the study methods more closely aligned with TRIPOD guidance. Overall, 5 percent of the cohort developed preeclampsia. The performance of the model was modest ($c$ statistic=0.71), but was adjusted statistically with a tenfold cross-validation technique to adjust for optimism/overfitting bias. Unlike the existing models we identified, the North model considered a larger set of potential
predictors and found some predictive factors that were not included in previous algorithms, such as family history of coronary heart disease, vaginal bleeding during pregnancy for at least 5 days, and protective factors that decreased risk (e.g., high fruit intake, prior miscarriage with the same partner).

Given the heterogeneity of the disease, screening tools aimed at different subtypes of preeclampsia, once they are more clearly defined, and for different study populations may be necessary. The SCOPE model would not be applicable to parous women entering prenatal care, but the majority of cases of preeclampsia occur among nulliparous women. Efforts to test and recalibrate a clinically feasible tool for other targeted or broader populations could be undertaken. Well-designed impact studies are needed to determine what level of performance improves upon current clinical risk assessment practices using to compare usual care to more complex risk prediction model based tools. Without these impact studies, the value of new instruments for improving processes of care and health outcomes cannot be quantified.

**Conclusion**

There is limited evidence available to determine the health benefits and harms of preeclampsia screening or the test performance of different screening and risk assessment strategies over the course of pregnancy. Despite the lack of empiric evidence, routine preeclampsia screening as currently conducted in prenatal care (i.e., blood pressure measurement and urine protein testing as part of routine pregnancy monitoring) is an established and feasible practice that is unlikely to be harmful or expensive. This is particularly true since the result of a positive screening measurement is repeat or similar testing for diagnostic confirmation and determination of severity to inform management. For most cases that will not develop into severe preeclampsia, enhanced monitoring is the most common initial clinical management. Given the rarity of preeclampsia and the potentially devastating consequences, especially in early-onset disease requiring preterm delivery, the focus of scientific inquiry has emphasized understanding the complex condition in order to more accurately identify those who will develop severe disease.

The complex pathophysiology of preeclampsia and its diverse outcomes present challenges for research aimed at improving health outcomes through evidence-based risk assessment and screening strategies. Research on the effectiveness of longstanding screening practices may be a lower research priority relative to efforts to better define the condition, to understand its physiological and causal underpinnings, and to develop new markers, tools, or tests for early identification and disease treatment. Broadly considered, screening recommendations for preeclampsia, including prior USPSTF guidance, highlight the low resource requirements of screening for high blood pressure and proteinuria. Efforts to identify the patients most likely to have severe or early-onset preeclampsia hold promise for better targeting of enhanced screening and preventive interventions. None of the existing validated models to estimate preeclampsia risk are sufficiently supported by evidence of performance that would warrant clinical application to general populations of pregnant women. Additional development, validation, and implementation research is needed to derive a tool ready for preeclampsia risk assessment in routine prenatal care and define its uses for improving health outcomes.
Periodic blood pressure and proteinuria measurements are routinely collected in primary obstetric care. Because of the long history of use of blood pressure and urine protein screening for preeclampsia screening, few studies have assessed their benefits and harms. Changes to diagnostic criteria in conjunction with evolving evidence on preeclampsia pathophysiology may foster new opportunities for improving clinical practice.
References


147. Sibai B. Subfertility/infertility and assisted reproductive conception are independent risk factors for pre-eclampsia. BJOG 2014 Oct 7 PMID: 25288131.


166. Roberts JM, Bell MJ. If we know so much about preeclampsia, why haven't we cured the disease? J Reprod Immunol 2013 Sep;99(1-2):1-9. PMID: 23890710.


Figure 1. Analytic Framework

Pregnancy duration

< 20 weeks gestation

≥ 20 weeks gestation

Risk assessment

Screening

All asymptomatic pregnant women

2

High risk

3

Harms

Low/normal risk

4

Preeclampsia

5

Harms

1

Screening

Health Outcomes

Eclampsia
Maternal morbidity and mortality
Perinatal and neonatal morbidity and mortality

Established protocols for disease management and treatment
Figure 2. Diagnostic Accuracy of Point of Care Tests for Proteinuria (Key Question 4a)

Note: One study\(^\text{114}\) is not plotted as it did not provide enough information to determine a 2x2 table.

**Abbreviations:** A = albumin; CI = confidence interval; Cr = creatinine; mg = milligram(s); mmol = millimole(s); P = protein.
<table>
<thead>
<tr>
<th>Organization</th>
<th>Year</th>
<th>Recommendation</th>
</tr>
</thead>
</table>
| Society of Obstetricians & Gynecologists of Canada (SOGC) | 2014  | The diagnosis of hypertension should be based on office or in-hospital BP measurements. (II-B; Low/Strong) [Additional detailed recommendations for blood pressure measurement and diagnosis are provided.]
All pregnant women should be assessed for proteinuria. (II-2B; Low/Weak) [Comments: We suggest screening with urinary dipstick at each antenatal visit. Proteinuria should be quantified by PrCr or 24h collection if preeclampsia is suspected.]
Significant proteinuria should be suspected when urinary dipstick proteinuria is ≥1+. (II-2A; Moderate/Strong)
Screening using biomarkers or Doppler ultrasound velocimetry of the uteroplacental circulation cannot be recommended routinely at present for women at low or increased risk of preeclampsia until such screening has been shown to improve pregnancy outcome. (II-2C; Very low/Weak) |
| American Congress of Obstetricians & Gynecologists (ACOG) | 2013  | Specific preeclampsia screening recommendations are not provided (e.g., type or frequency of screening tests to use) guidelines are primarily focused on diagnostic criteria and disease management. |
| National Institute of Health and Care Excellence (NICE) | 2008  | Blood pressure measurement and urinalysis for protein should be carried out at each antenatal visit to screen for preeclampsia. |
| Royal College of Obstetricians and Gynaecologists (RCOG) | 2003  | All women with blood pressure >140/90 mm Hg with or without proteinuria should be referred to a day assessment or obstetric unit. (Grade A)
All women with persistent proteinuria, even in the absence of hypertension, should be referred for further investigation. (Grade A)
Although pregnancies associated with an abnormal uterine artery Doppler waveform are at significant risk of adverse outcome (particularly severe preeclampsia requiring early delivery), its introduction as a screening test for all women cannot currently be recommended other than in clinical trials. (Grade B)
Automated instruments for BP measurement are generally not validated for use during pregnancy and preeclampsia. Therefore, the use of mercury sphygmomanometers remains preferable. (Grade B)
Due to the variation in urine concentration, largely determined by hydration, all urine screening in obstetric day units should be by protein:creatinine ratio; this can be by laboratory test or at point of care. (Grade C)
The definition of gestational proteinuria is derived from studies calculating the 95th centile for an uncomplicated population. A protein loss of >300 mg in 24 hours is associated with an increased morbidity to the mother and her baby. (Grade B) |
| United States Preventive Services Task Force (USPSTF) | 1996, reaffirmed 2002, no longer posted – out of date | Screening for preeclampsia with blood pressure measurement is recommended for all pregnant women at the first prenatal visit and periodically throughout the remainder of the pregnancy. Further diagnostic evaluation and clinical monitoring, including frequent BP monitoring and urine testing for protein, are indicated if BP does not decrease normally during the middle trimester, if the SBP increases 30 mm Hg above baseline, if the DBP increases 15 mm Hg above baseline, or if the blood pressure exceeds 140/90 mm Hg above baseline. (B Recommendation) |

**Abbreviations:** BMI = body mass index; BP = blood pressure; DBP = diastolic blood pressure; mg = milligram(s); mm Hg = millimeters of mercury; SBP = systolic blood pressure.
Table 2. Factors for Clinical Assessment of Preeclampsia Risk*

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>USPSTF Risk Assessment for Low-dose Aspirin Prophylaxis*</th>
<th>National Institute of Health and Care Excellence (NICE)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primiparity</td>
<td>High risk†</td>
<td>High risk†</td>
</tr>
<tr>
<td>Previous preeclamptic pregnancy</td>
<td>History of preeclampsia</td>
<td>Hypertensive disease during previous pregnancy</td>
</tr>
<tr>
<td>Chronic hypertension or chronic renal disease or both</td>
<td>Multifetal gestation</td>
<td>Chronic hypertension</td>
</tr>
<tr>
<td>History of thrombophilia</td>
<td>Chronic hypertension</td>
<td>Type 1 or 2 diabetes</td>
</tr>
<tr>
<td>Multifetal pregnancy</td>
<td>Type 1 or 2 diabetes</td>
<td>Chronic kidney disease</td>
</tr>
<tr>
<td>In vitro fertilization</td>
<td>Renal disease</td>
<td>Autoimmune disease (i.e., systemic lupus erythematosus,</td>
</tr>
<tr>
<td>Family history of preeclampsia</td>
<td>Autoimmune disease (i.e., systemic lupus erythematosus,</td>
<td>antiphospholipid syndrome)</td>
</tr>
<tr>
<td>Type I diabetes mellitus or type II diabetes mellitus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obesity</td>
<td>Moderate risk‡</td>
<td>Moderate risk‡</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
<td>Nulliparity</td>
<td>First pregnancy</td>
</tr>
<tr>
<td>Advanced maternal age (older than 40 years)</td>
<td>Obesity (BMI &gt;30 kg/m2)</td>
<td>Obesity (BMI ≥35 kg/m2)</td>
</tr>
<tr>
<td></td>
<td>Family history of preeclampsia</td>
<td>Family history of preeclampsia</td>
</tr>
<tr>
<td></td>
<td>Sociodemographic characteristics</td>
<td>Age ≥40 years</td>
</tr>
<tr>
<td></td>
<td>(African American race, low socioeconomic status)</td>
<td>Pregnancy interval &gt;10 years</td>
</tr>
<tr>
<td></td>
<td>Age ≥35 years</td>
<td>Multiple pregnancy</td>
</tr>
<tr>
<td></td>
<td>Personal history factors (e.g., low birthweight or small for gestational age, previous adverse pregnancy outcome, &gt;10 year pregnancy interval)</td>
<td></td>
</tr>
<tr>
<td>Low risk</td>
<td>Previous uncomplicated full-term delivery</td>
<td></td>
</tr>
</tbody>
</table>

* Includes only risk factors that can be obtained from the patient medical history. Clinical measures, such as uterine artery Doppler ultrasound, may additionally be used by some clinicians to evaluate risk.
† The USPSTF and NICE recommend low-dose aspirin if the patient has ≥1 high-risk factor.
‡ The USPSTF recommends considering low-dose aspirin if the patient has several of the listed moderate-risk factors. NICE recommends low-dose aspirin if the patient has at least two moderate-risk factors.

**Abbreviations:** BMI = body mass index; kg = kilogram(s); m = meter(s)
Table 3. Differences in Health Outcomes During Pregnancy, at Time of Delivery or 6 Weeks Postpartum (Key Questions 1a and 5)

<table>
<thead>
<tr>
<th>Study &amp; Quality</th>
<th>Category</th>
<th>Outcomes</th>
<th>Group</th>
<th>Results</th>
<th>Between Group Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>McDuffie, 1996</td>
<td>Preeclampsia</td>
<td>Mild PE, n (%)</td>
<td>IG</td>
<td>59 (5.1)</td>
<td>RR 0.94 (95% CI, 0.78 to 1.14), p=0.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>66 (5.7)</td>
<td></td>
</tr>
<tr>
<td>Fair</td>
<td></td>
<td>Severe PE, n (%)</td>
<td>IG</td>
<td>10 (0.9)</td>
<td>RR 1.05 (95% CI, 0.68 to 1.62), p=0.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>9 (0.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Preterm Birth</td>
<td>Preterm delivery &lt; 32 weeks, n (%)</td>
<td>IG</td>
<td>10 (0.9)</td>
<td>RR 1.11 (95% CI, 0.73 to 1.68), p=0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>8 (0.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Preterm delivery &lt; 37 weeks, n (%)</td>
<td>IG</td>
<td>73 (6.3)</td>
<td>RR 1.08 (95% CI, 0.92 to 1.27), p=0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>63 (5.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Delivery Complications</td>
<td>Abruptio placentae, n (%)</td>
<td>IG</td>
<td>17 (1.5)</td>
<td>RR 1.21 (95% CI, 0.90 to 1.64), p=0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>11 (0.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Apgar score at 5 minutes &lt; 7, n (%)</td>
<td>IG</td>
<td>18 (1.6)</td>
<td>RR 0.77 (95% CI, 0.53 to 1.10), p=0.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>29 (2.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chorioamnionitis, n (%)</td>
<td>IG</td>
<td>9 (0.8)</td>
<td>RR 0.90 (95% CI, 0.55 to 1.46), p=0.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>11 (0.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Placenta previa, n (%)</td>
<td>IG</td>
<td>7 (0.6)</td>
<td>RR 0.87 (95% CI, 0.50 to 1.52), p=0.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>9 (0.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Postpartum hemorrhage with cesarean delivery, n (%)</td>
<td>IG</td>
<td>2 (1.3)</td>
<td>RR 0.77 (95% CI, 0.26 to 2.27), p=0.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>3 (2.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Postpartum hemorrhage with vaginal delivery, n (%)</td>
<td>IG</td>
<td>32 (3.2)</td>
<td>RR 0.98 (95% CI, 0.77 to 1.27), p=0.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>33 (3.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Preterm labor, n (%)</td>
<td>IG</td>
<td>79 (6.8)</td>
<td>RR 1.01 (95% CI, 0.86 to 1.18), p=0.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>77 (6.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Preterm premature rupture of membranes, n (%)</td>
<td>IG</td>
<td>38 (3.3)</td>
<td>RR 1.00 (95% CI, 0.80 to 1.25), p=0.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>38 (3.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cesarean Section</td>
<td>Cesarean delivery, overall, n (%)</td>
<td>IG</td>
<td>151 (13.0)</td>
<td>RR 1.04 (95% CI, 0.93 to 1.17), p=0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>140 (12.0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Perinatal/Neonatal Mortality</td>
<td>Stillbirth, n (%)</td>
<td>IG</td>
<td>5 (0.4)</td>
<td>RR 1.00 (95% CI, 0.54 to 1.86), p=0.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>5 (0.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Birthweight</td>
<td>Birthweight (g), mean (SD)</td>
<td>IG</td>
<td>3286 (520)</td>
<td>NR, p=0.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>3295 (536)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Very low birthweight (&lt; 1,500 g), n (%)</td>
<td>IG</td>
<td>7 (0.3)</td>
<td>RR 1.08 (95% CI, 0.65 to 1.79), p=0.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>6 (0.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low birthweight (&lt; 2,500 g), n (%)</td>
<td>IG</td>
<td>64 (5.4)</td>
<td>RR 0.94 (95% CI, 0.78 to 1.12), p=0.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>72 (6.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SGA, n (%)</td>
<td>IG</td>
<td>36 (3.1)</td>
<td>RR 1.13 (95% CI, 0.91 to 1.41), p=0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>28 (2.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Healthcare Use during Pregnancy</td>
<td>Total number of visits, mean (SD)</td>
<td>IG</td>
<td>12.0 (4.2)</td>
<td>NR, p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>14.7 (4.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Satisfaction with Prenatal Care at 6 weeks Postpartum</td>
<td>Number of prenatal visits, just right</td>
<td>IG</td>
<td>494 (89.2)</td>
<td>NR, p=0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>473 (82.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Number of prenatal visits, too few</td>
<td>IG</td>
<td>49 (8.8)</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>6 (1.1)</td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Differences in Health Outcomes During Pregnancy, at Time of Delivery or 6 Weeks Postpartum (Key Questions 1a and 5)

<table>
<thead>
<tr>
<th>Study &amp; Quality</th>
<th>Category</th>
<th>Outcomes</th>
<th>Group</th>
<th>Results</th>
<th>Between Group Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhode, 2007</td>
<td>Number of prenatal visits, too many</td>
<td>IG 11 (2.0)</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CG 92 (16.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fair</td>
<td>Quality of prenatal care, excellent or good, n (%)</td>
<td>IG 574 (97.5)</td>
<td>NR, p=0.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CG 587 (97.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Preeclampsia</td>
<td>IG 23 (2.3)</td>
<td>NR, p=0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CG 36 (3.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Preterm Birth</td>
<td>IG 50 (4.9)</td>
<td>NR, p=0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CG 72 (7.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cesarean Section</td>
<td>IG 181 (17.8)</td>
<td>NR, p=0.029</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CG 173 (18.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other Maternal Morbidity</td>
<td>IG 33 (3.3)</td>
<td>NR, p&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CG 15 (1.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gestational diabetes, n (%)</td>
<td>IG 42 (4.2)</td>
<td>NR, p=0.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CG 81 (9.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gestational hypertension, n (%)</td>
<td>IG 58 (5.7)</td>
<td>NR, p&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CG 38 (4.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High blood pressure, n (%)</td>
<td>IG 81 (6.0)</td>
<td>NR, p=0.0005</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CG 74 (7.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pyelonephritis, n (%)</td>
<td>IG 4 (0.40)</td>
<td>NR, p&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CG 4 (0.40)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asymptomatic bacteriuria, n (%)</td>
<td>IG 67 (6.8)</td>
<td>NR, p=0.051</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CG 79 (8.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Urinary tract infection, n (%)</td>
<td>IG 141 (14.2)</td>
<td>NR, p=0.043</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CG 140 (15.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Rhode, 2007 used statistical tests for non-inferiority. A p-value less than 0.05 indicates rates are statistically equivalence (no greater than 0.04 in one direction)

**Abbreviations:** CG = control group; CI = confidence interval; g = gram(s); IG = intervention group; NR = not reported; PE = preeclampsia; RR = relative risk; SD = standard deviation; SGA = small for gestational age.
| External Validation Studies | Oliveira 2014<sup>99</sup> Baltimore, Maryland, United States  
PE requiring delivery:  
<34 weeks gestation (early)  
>34 weeks gestation (late) | Park 2013<sup>100</sup> Sydney, Australia  
PE requiring delivery:  
<34 weeks gestation (early)  
<42 weeks gestation (any)  
>34 weeks gestation (late) | Skrastad 2014<sup>101</sup> Trondheim, Norway  
PE requiring delivery:  
<37 weeks gestation (early)  
<42 weeks gestation (any)  
>34 weeks gestation (late) | Farina 2011<sup>88</sup> Bologna, Italy  
PE diagnosis:  
>34 weeks gestation |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Study design</td>
<td>Prospective cohort</td>
<td>Prospective cohort</td>
<td>Prospective cohort</td>
<td>Prospective cohort</td>
</tr>
<tr>
<td>Study population</td>
<td>Women with singleton pregnancies</td>
<td>Women with singleton pregnancies presenting for aneuploidy screening</td>
<td>Nulliparous women</td>
<td>Women with singleton pregnancies enrolled at screening visit for early diagnosis of chromosomal and other fetal abnormalities, and delivery in tertiary care center</td>
</tr>
</tbody>
</table>
| Sample size (n) | 871–2,962  
(model n depended on availability of variables needed for each predictive model) | 3,066 | 541 | 554 |
| Outcome prevalence (%) | Early PE: 1.0–1.2  
(10-30 cases)  
Late PE: 4.1-5.0  
(78-116 cases) | Early PE: 0.4 (12 cases) | Any PE: 3.9 (21 cases)  
Preterm PE requiring delivery (<37 wks): 0.9 (5 cases) | Late PE: 7.0 (39 cases) |
| Funding | Diagnostic Technologies Limited and PerkinElmer | NR | Norwegian University of Science and Technology; National Center for Fetal Medicine | Ricerca Fondamentale Orientata |

**Abbreviations:** PE = preeclampsia; wks = weeks.
### Table 5. External Validation Performance of Five Preeclampsia Risk Prediction Models With Good or Better Discrimination (c Index ≥0.80) (Key Question 2)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cohort study for external validation of model</strong></td>
<td>Oliveira 2014**</td>
<td>Park 2013**</td>
<td>Oliveira 2014**</td>
<td>Skrastad 2014**</td>
<td>Farina 2011**</td>
</tr>
<tr>
<td></td>
<td>Baltimore, Maryland, United States</td>
<td>Sydney, Australia</td>
<td>Baltimore, Maryland, United States</td>
<td>Trondheim, Norway</td>
<td>Bologna, Italy</td>
</tr>
<tr>
<td></td>
<td>N (validation cohort)</td>
<td>2,833</td>
<td>3,014</td>
<td>871</td>
<td>541</td>
</tr>
<tr>
<td><strong>PE timing</strong></td>
<td>PE requiring early delivery (&lt;34 weeks)</td>
<td>PE requiring early delivery (&lt;34 weeks)</td>
<td>PE requiring early delivery (&lt;34 weeks)</td>
<td>PE requiring early delivery (&lt;37 weeks)</td>
<td>Late PE diagnosis (&gt;34 weeks)</td>
</tr>
<tr>
<td></td>
<td>1.0 (29 cases)</td>
<td>0.4 (12 cases)</td>
<td>1.2 (10 cases)</td>
<td>0.9 (5 cases)</td>
<td>7.0 (39 cases)</td>
</tr>
<tr>
<td><strong>% PE outcome (n cases)</strong></td>
<td>0.80 (0.71-0.89)</td>
<td>0.93 (0.92-0.94)</td>
<td>0.86 (0.73-0.99)</td>
<td>0.94 (0.86-1.00)</td>
<td>0.93 (0.88-0.98)</td>
</tr>
<tr>
<td><strong>Calibration</strong></td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td><strong>Detection % (95% CI)</strong></td>
<td>52 (CI not reported)</td>
<td>91.7 (61.5-98.6)</td>
<td>80 (CI not reported)</td>
<td>80.0 (28.4 – 99.5)</td>
<td>84.6 (73.3-95.9)</td>
</tr>
<tr>
<td><strong>PPV</strong></td>
<td>4.2 (2.6-6.5)</td>
<td>3.6 (2.0-7.0)</td>
<td>11.3 (5.3-21.5)</td>
<td>6.8 (1.9-16.5)</td>
<td>39.3</td>
</tr>
<tr>
<td><strong>NPV</strong></td>
<td>99.6 (99.0-100.0)</td>
<td>99.9 (99.7-99.9)</td>
<td>99.8 (99.0-100.0)</td>
<td>99.8 (98.8-100.0)</td>
<td>98.7</td>
</tr>
<tr>
<td><strong>Model parameters</strong></td>
<td>history, MAP, PAPP-A, uterine artery pulsatility index</td>
<td>history, MAP, PAPP-A, uterine artery pulsatility index</td>
<td>history, PAPP-A, PP-13, uterine artery pulsatility index</td>
<td>history, MAP, serum tests, uterine artery pulsatility index Model algorithm not clearly specified</td>
<td>history, MAP, uterine artery pulsatility index Model variables listed but algorithm not provided</td>
</tr>
<tr>
<td></td>
<td>history, MAP, PAPP-A, uterine artery pulsatility index</td>
<td>history, MAP, PAPP-A, uterine artery pulsatility index</td>
<td></td>
<td></td>
<td>history, MAP, uterine artery pulsatility index Model variables listed but algorithm not provided</td>
</tr>
</tbody>
</table>

---

1. Preeclampsia defined as requiring delivery, with the exception of the Farina external validation which defined the outcome as the diagnosis of PE.
2. c index = A test performance statistic (equivalent to the AUC) used to assess discrimination, a model performance measure that refers to how well a model differentiates between those with and without the outcome.
3. Calibration = A model performance measure that refers to how well predicted risks compare to observed outcomes preferably evaluated graphically by calibration plots and supplemented by a formal statistical test, the Hosmer-Lemeshow test for logistic regression and its equivalent for Cox regression.
4. Detection %: Analogous to sensitivity. The percent of cases correctly classified based on a predefined false positive risk threshold. 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.
5. Positive predictive value = A test performance statistic used to measure what proportion of patients who test positive have the disease. PPV not reported in the Farina external validation study and was calculated by hand.
6. Negative predictive value = A test performance statistic used to measure what proportion of patients who test negative do not have the disease. NPV not reported in the Farina external validation study and was calculated by hand.
7. Maternal characteristics, medical history, and routine clinical measures can include family history of PE, personal history of PE, parity, race/ethnicity, prior preterm labor, CHTN, diabetes, thrombophilia, renal disease, mode of conception, smoking status, DBP, SBP, weight, MAP, BMI. Serum markers include PAPP-A, PIGF, PP13. Doppler ultrasound includes UtA-PI. Some variables are expressed as adjusted Multiples of the Median.
8. Derived from the Fetal Medicine Foundation Algorithm

**Abbreviations:** AUC = area under the curve; BMI = body mass index; CHTN = chronic hypertension; CI = confidence intervals; DBP = diastolic blood pressure; MAP = mean arterial pressure; NPV = negative predictive value; NR = not reported; PAPP-A = pregnancy-associated plasma protein-A; PE = preeclampsia; PIGF = placental growth factor; PPV = positive predictive value; SBP = systolic blood pressure; UtA PI = uterine artery pulsatility index; vs = versus.
### Table 6. Index Tests and Reference Standard Characteristics of Included Diagnostic Accuracy Studies (Key Question 4a)

<table>
<thead>
<tr>
<th>Type of Index Test</th>
<th>Study &amp; Quality</th>
<th>Index Test (Make, Manufacturer)</th>
<th>Index Test Sampling Methods</th>
<th>Index Test Machine &amp; Assay</th>
<th>Index Test Operator &amp; Reader</th>
<th>24-hour Ref Stand Collection Method</th>
<th>Ref Stand Machine &amp; Assay</th>
<th>Ref Stand Operator &amp; Reader</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein:Creatinine</td>
<td>Tun, 2012 Fair</td>
<td>P/Cr spot test (NR)</td>
<td>Number of samples: 1</td>
<td>NR</td>
<td>NR</td>
<td>Started at time of admission; collected in 2 consecutive 12 hour collections; total protein a combination of both 12-hour urine specimens</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Initial urine specimen (which was otherwise discarded). Some pts had it collected from the timed specimen (24-hour) or immediately after the timed collection.</td>
<td></td>
<td></td>
<td>NR, sent to Health Network Lab</td>
<td>ADVIA total protein urine assay (modified Fujita method)</td>
<td>NR, sent to Health Network Lab</td>
</tr>
<tr>
<td></td>
<td>Stout, 2013 Fair</td>
<td>P/Cr spot ratio (NR)</td>
<td>Number of samples: 1</td>
<td>NR</td>
<td>NR</td>
<td>The first 24-hour urine collection was used for each patient.</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ratio assessment prior to initiation of 24-hour collection</td>
<td></td>
<td></td>
<td>Enzymatic creatinase</td>
<td>Benzethonium chloride</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wheeler, 2007 Fair</td>
<td>P/Cr spot ratio (NR)</td>
<td>Number of samples: 1</td>
<td>NR</td>
<td>NR</td>
<td>No specimen collected as the first void of the morning.</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Samples obtained at the beginning of the 24 hour urine collection</td>
<td></td>
<td></td>
<td>Johnson &amp; Johnson Vitros 250 Biuret method (protein); 2-point rate method (creatinine)</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>Young, 1996 Fair</td>
<td>P/Cr spot test (NR)</td>
<td>Number of samples: 1</td>
<td>Beckman analyzer Standard spectrophotometric technique</td>
<td>NR</td>
<td>No specimen collected as the first void of the morning.</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Single voided specimen collected when patient emptied bladder to begin the timing of the 24 hour urine collection (in 66% of samples); otherwise, single voided sample collected after the completion of the 24-hour collection (34% of samples). No specimen collected as the first void of the morning.</td>
<td></td>
<td></td>
<td>Standard spectrophotometric technique</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Type of Index Test</td>
<td>Study &amp; Quality</td>
<td>Index Test (Make, Manufacturer)</td>
<td>Index Test Sampling Methods</td>
<td>Index Test Machine &amp; Assay</td>
<td>Index Test Operator &amp; Reader</td>
<td>24-hour Ref Stand Collection Method</td>
<td>Ref Stand Machine &amp; Assay</td>
<td>Ref Stand Operator &amp; Reader</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------</td>
<td>--------------------------------</td>
<td>----------------------------</td>
<td>---------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------------</td>
<td>-------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Verdonk, 2014</td>
<td>Good</td>
<td>P/Cr spot test (Albustix, Siemens Healthcare Diagnostics)</td>
<td>Number of samples: 3 Began at midnight w/ 5 mL aliquots saved for P/Cr testing from requested spontaneous voids at approximately 8 AM, 12 PM (noon), and 5 PM; visually analyzed.</td>
<td>CREA plus, Roche Diagnostics Enzymatic assay (Cr), colorimetric assay (P)</td>
<td>NR</td>
<td>Began at midnight, nurses monitored for completeness and when errors occurred, the procedure was stopped and restarted at midnight the next day.</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Sethuram, 2011</td>
<td>Fair</td>
<td>P/Cr spot test (NR)</td>
<td>Number of samples: 1 10 mL sample of urine collected before 24-hour collection; avoided the first void sample.</td>
<td>Abbott Diagnostics analyzer Benzethonium chloride turbidometric method (protein); Jaffe method (Cr)</td>
<td>NR</td>
<td>Avoided the first void sample.</td>
<td>Abbott Diagnostics analyzer Benzethonium chloride turbidometric method (protein)</td>
<td>NR</td>
</tr>
<tr>
<td>Lamontagne, 2014</td>
<td>Good</td>
<td>P/Cr spot ratio (NR)</td>
<td>Number of samples: 1 When women entered study, urinalysis, urine culture and P/Cr calculated on same urine sample provided at any moment during the day before 24-hour urine collection; not collected w/ catheter.</td>
<td>Beckman Coulter multianalyzer w/ the Synchron LX system Colorimetric method using pyrogallol red-molybdate (P); Jaffe method (Cr)</td>
<td>NR</td>
<td>Inpatients instructed on how to proceed by a nurse, while ambulatory women given oral and written instructions; not collected w/ catheter.</td>
<td>Beckman Coulter multianalyzer w/ the Synchron LX system Colorimetric method using pyrogallol red-molybdate (P); Jaffe method (Cr)</td>
<td>NR</td>
</tr>
<tr>
<td>Kyle, 2008</td>
<td>Fair</td>
<td>P/Cr spot test (NR)</td>
<td>Number of samples: 1 Aliquot from midstream urine specimen before 24-hour, performed at health lab</td>
<td>Abbott Ci8200 Analyzer</td>
<td>Research midwife</td>
<td>24-hour urine collection as an outpatient. Discard first void in the toilet and write date/time of the sample on the request form, all subsequent voids were collected. Final void collected 24 hours later and placed in the specimen container.</td>
<td>NR</td>
<td>Benzethonium chloride assay</td>
</tr>
<tr>
<td>Type of Index Test</td>
<td>Study &amp; Quality</td>
<td>Index Test (Make, Manufacturer)</td>
<td>Index Test Sampling Methods</td>
<td>Index Test Machine &amp; Assay</td>
<td>Index Test Operator &amp; Reader</td>
<td>24-hour Ref Stand Collection Method</td>
<td>Ref Stand Machine &amp; Assay</td>
<td>Ref Stand Operator &amp; Reader</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------</td>
<td>---------------------------------</td>
<td>----------------------------</td>
<td>--------------------------</td>
<td>-----------------------------</td>
<td>---------------------------------</td>
<td>------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Bhide, 2015&lt;sup&gt;103&lt;/sup&gt;</td>
<td>Fair</td>
<td>P/Cr spot test (NR)</td>
<td>Number of samples: 1 Usually collected and sent to the lab on the day and time of presentation to day assessment unit (but if not w/in 48 hours from attendance at no specific time of day), before 24-hour. Only data from the first attendance included in study.</td>
<td>NR</td>
<td>NR</td>
<td>24-hour urine collection</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Dwyer, 2008&lt;sup&gt;105&lt;/sup&gt;</td>
<td>Good</td>
<td>P/Cr spot test (NR)</td>
<td>Number of samples: 1 Obtained before 24-hour collection, taken immediately after 24-hour. All samples collected via clean catch unless the membranes had been ruptured, in which case specimens were captured by catheter.</td>
<td>Synchron LX Systems Pyrogallol red/molybdate (P) and Jaffe rate (Cr)</td>
<td>Lab technician</td>
<td>Most collected as outpatients, all samples collected via clean catch unless the membranes had been ruptured, in which case specimens were captured by catheter.</td>
<td>NR</td>
<td>Lab technician</td>
</tr>
<tr>
<td>Durnwald, 2003&lt;sup&gt;104&lt;/sup&gt;</td>
<td>Fair</td>
<td>P/Cr spot ratio (NR)</td>
<td>Number of samples: 1 Random urine collection before initiation of 24-hour collection</td>
<td>NR</td>
<td>NR</td>
<td>Outpatients collected all urine in a container for 24 hours and returned it to the outpatient laboratory; inpatients who had vaginal bleeding and/or active labor, were receiving Mg sulfate seizure prophylaxis, who had delivered collected by Foley catheter.</td>
<td>NR</td>
<td>NR</td>
</tr>
</tbody>
</table>
Table 6. Index Tests and Reference Standard Characteristics of Included Diagnostic Accuracy Studies (Key Question 4a)

<table>
<thead>
<tr>
<th>Type of Index Test</th>
<th>Study &amp; Quality</th>
<th>Index Test (Make, Manufacturer)</th>
<th>Index Test Sampling Methods</th>
<th>Index Test Machine &amp; Assay</th>
<th>Index Test Operator &amp; Reader</th>
<th>24-hour Ref Stand Collection Method</th>
<th>Ref Stand Machine &amp; Assay</th>
<th>Ref Stand Operator &amp; Reader</th>
</tr>
</thead>
<tbody>
<tr>
<td>P/Cr spot test (NR)</td>
<td>Valdes, 2015 Fair</td>
<td>P/Cr spot test (NR)</td>
<td>Number of samples: 1 An additional urine sample (15-20 mL) collected for storage at -20 degrees Celsius for quantification of P:Cr concentrations upon completion of the study period</td>
<td>NR</td>
<td>NR</td>
<td>Upon admission, pts underwent a 24-hour proteinuria test.</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Dipstick - Clinitek Microalbumin (automated) (Clinitek Microalbumin, Bayer)</td>
<td>Waugh, 2005 Good</td>
<td>Number of samples: 1 Early morning sample before the final 24-hour specimen was added to the 24-hour collection, a mixed 10 mL aliquot was removed for urinalysis.</td>
<td>Clinitek 50 Two semi-quantitative immunoassays for albumin and Cr</td>
<td>NR</td>
<td>Two semi-quantitative immunoassays for albumin and Cr</td>
<td>On waking, the first void was discarded and the sample started with the second urine specimen; the final specimen was the first void the following day.</td>
<td>NR</td>
<td>Benzethonium chloride assay</td>
</tr>
<tr>
<td>Dipstick - Microalbumstix (visual) (Microalbumstix, Bayer)</td>
<td>Albumin/Creatinine</td>
<td>Number of samples: 1 Early morning sample before the final 24-hour specimen was added to the 24-hour collection, a mixed 10 mL aliquot was removed for urinalysis.</td>
<td>NA</td>
<td>Two semi-quantitative immunoassays for albumin and Cr</td>
<td>Two observers</td>
<td>On waking, the first void was discarded and the sample started with the second urine specimen; the final specimen was the first void the following day.</td>
<td>NR</td>
<td>Benzethonium chloride assay</td>
</tr>
<tr>
<td>DCA 2000 - POC test (DCA 2000, Bayer)</td>
<td></td>
<td>Number of samples: 1 Early morning sample before the final 24-hour specimen was added to the 24-hour collection, a mixed 10 mL aliquot was removed for urinalysis. Utilizes a cartridge system and 40 uL of sample.</td>
<td>AliveNano</td>
<td>Immunoturbidometric assay (albumin), colorimetric assay (Cr)</td>
<td>NR</td>
<td>On waking, the first void was discarded and the sample started with the second urine specimen; the final specimen was the first void the following day.</td>
<td>NR</td>
<td>Benzethonium chloride assay</td>
</tr>
<tr>
<td>Type of Index Test</td>
<td>Study &amp; Quality</td>
<td>Index Test (Make, Manufacturer)</td>
<td>Index Test Sampling Methods</td>
<td>Index Test Machine &amp; Assay</td>
<td>Index Test Operator &amp; Reader</td>
<td>24-hour Ref Stand Collection Method</td>
<td>Ref Stand Machine &amp; Assay</td>
<td>Ref Stand Operator &amp; Reader</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------</td>
<td>---------------------------------</td>
<td>-----------------------------</td>
<td>----------------------------</td>
<td>-----------------------------</td>
<td>--------------------------------</td>
<td>--------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Kyle, 2008&lt;sup&gt;106&lt;/sup&gt;</td>
<td>Fair</td>
<td>Albumin:Cr spot test (DCA 2000, Bayer Healthcare LLC)</td>
<td>Number of samples: 1 Aliquot from midstream urine specimen before 24-hour; performed at antenatal clinic</td>
<td>NR</td>
<td>Research midwife</td>
<td>24-hour urine collection as an outpatient. Discard first void into the toilet and write date/time of the sample on the request form, all subsequent voids were collected. Final void collected 24 hours later and placed in the specimen container.</td>
<td>NR</td>
<td>Benzethonium chloride assay</td>
</tr>
<tr>
<td>Waugh, 2001&lt;sup&gt;101&lt;/sup&gt;</td>
<td>Fair</td>
<td>Dipstick (BM-Test-5L, Boehringer Mannheim UK)</td>
<td>Number of samples: 2 Two 10 mL aliquots of thoroughly mixed urine from the 24-hour urine; removed for dipstick analysis and protein assays.</td>
<td>NR</td>
<td>Observer</td>
<td>Collections performed between 8 AM and 8 AM on consecutive days; women instructed regarding collection procedures.</td>
<td>ExcelGel w/ silver staining kit</td>
<td>Benzethonium chloride or Bradford assay</td>
</tr>
<tr>
<td>Kyle, 2008&lt;sup&gt;106&lt;/sup&gt;</td>
<td>Fair</td>
<td>Dipstick (NR)</td>
<td>Number of samples: 1 Aliquot from midstream urine specimen before 24-hour</td>
<td>NR</td>
<td>Research midwife</td>
<td>24-hour urine collection as an outpatient. Discard first void into the toilet and write date/time of the sample on the request form, all subsequent voids were collected. Final void collected 24 hours later and placed in the specimen container.</td>
<td>NR</td>
<td>Benzethonium chloride assay</td>
</tr>
<tr>
<td>Waugh, 2005&lt;sup&gt;84&lt;/sup&gt;</td>
<td>Good</td>
<td>Dipstick - Multistix 8SG (visual) (Multistix 8SG, Bayer)</td>
<td>Number of samples: 1 Early morning sample before the final 24-hour specimen was added to the 24-hour collection, a mixed 10 mL aliquot was removed for urinalysis.</td>
<td>NA</td>
<td>Two observers</td>
<td>On waking, the first void was discarded and the sample started w/ the second urine specimen; the final specimen was the first void the following day.</td>
<td>NR</td>
<td>Benzethonium chloride assay</td>
</tr>
<tr>
<td>Type of Index Test</td>
<td>Study &amp; Quality</td>
<td>Index Test (Make, Manufacturer)</td>
<td>Index Test Sampling Methods</td>
<td>Index Test Machine &amp; Assay</td>
<td>Index Test Operator &amp; Reader</td>
<td>24-hour Ref Stand Collection Method</td>
<td>Ref Stand Machine &amp; Assay</td>
<td>Ref Stand Operator &amp; Reader</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------</td>
<td>--------------------------------</td>
<td>-----------------------------</td>
<td>---------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------------</td>
<td>--------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Dipstick - Multistix 8SG (automated) (Multistix 8SG, Bayer)</td>
<td>Good</td>
<td>Number of samples: 1 Early morning sample before the final 24-hour specimen was added to the 24-hour collection, a mixed 10 mL aliquot was removed for urinalysis.</td>
<td>Clinitek 50 NR</td>
<td>NR</td>
<td>On waking, the first void was discarded and the sample started w/ the second urine specimen; the final specimen was the first void the following day.</td>
<td>Benzethonium chloride assay</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>P/Cr automated dipstick (Iris test strips, IRIS Inc or Arcray Inc)</td>
<td>Dwyer, 2008</td>
<td>Number of samples: 1 Urinalysis; obtained before 24-hour collection, if a sample unavailable, taken immediately after 24-hour. All samples collected via clean catch unless the membranes had been ruptured, in which case specimens were captured by catheter.</td>
<td>Autoanalyzers 3’3”5’5” tetrachlorophehnol-3,4,5,6-tetabromosufon-phalein (protein error of pH indicator)</td>
<td>Lab technician</td>
<td>Most collected as outpatients, All samples collected via clean catch unless the membranes had been ruptured, in which case specimens were captured by catheter.</td>
<td>Lab technician</td>
<td>NR</td>
<td>Lab technician</td>
</tr>
</tbody>
</table>

**Abbreviations:** Cr = creatinine; mL = milliliter(s); NR = not reported; P = protein; pt = participant; ref = reference; stand = standard; w/ = with.
Table 7. Overall Summary of Evidence by Key Question

<table>
<thead>
<tr>
<th>Key Question</th>
<th>No. of Studies (k), No. of Observations (n), Design</th>
<th>Quality</th>
<th>Limitations*</th>
<th>Consistency</th>
<th>US Primary Care Applicability</th>
<th>Summary of Findings†</th>
</tr>
</thead>
<tbody>
<tr>
<td>KQ1</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>KQ1a</td>
<td>k=1 n=2,764 RCT</td>
<td>Fair</td>
<td>Differences in the number of visits between the control and intervention groups where not as pronounced as intended by the design (12.0 vs. 14.7 visits). Insufficient power to detect differences in rare adverse outcomes such as very low birthweight, eclampsia and stillbirth.</td>
<td>NA</td>
<td>Low</td>
<td>Fewer prenatal care visits did not have a beneficial or harmful effect on rates of health outcomes or diagnoses of preeclampsia, but evidence has limited relevance to current clinical practice or population.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KQ2</td>
<td>k=4 external validation studies 16 externally validated risk assessment models 5 externally validated models with good to excellent discrimination (c index ≥0.80) n= 541-3,066 Prospective cohort</td>
<td>External validation in prospective cohort studies addresses common sources of bias in model development, we therefore included all externally validated risk prediction models in our review“ Calibration was not reported for any of the models, and risk cutpoints were not established \textit{a priori} Many prediction models were developed by the same team of investigators, and the Australian cohort external validation studies were not independent from that group</td>
<td>Moderate/low</td>
<td>Moderate</td>
<td>No externally validated model was supported by evidence of good performance or clinical benefits Important model performance statistics were not available. Positive predictive values were low.</td>
<td></td>
</tr>
<tr>
<td>Key Question</td>
<td>No. of Studies (k), No. of Observations (n), Design</td>
<td>Quality</td>
<td>Limitations*</td>
<td>Consistency</td>
<td>US Primary Care Applicability</td>
<td>Summary of Findings†</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------------</td>
<td>---------</td>
<td>--------------</td>
<td>-------------</td>
<td>-----------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td><strong>KQ3</strong> Harms of preeclampsia multivariable risk assessment</td>
<td>k=1 n=255 Prospective cohort study</td>
<td>Fair</td>
<td>The risk assessment tool used was not clearly described, risk assessment occurred alongside intensive counseling and changes to clinical care and was not clearly described—cannot disentangle effects of risk assessment and clinical care Insufficient power to assess differences in effects of risk assessment for false negative results compared to others.</td>
<td>NA</td>
<td>Low</td>
<td>Study was conducted in Spain and Italy among women undergoing aneuploidy screening The risk assessment tool is unlikely to be used in practice based on external validation in U.S. cohort.</td>
</tr>
<tr>
<td><strong>KQ4</strong> Effectiveness of screening tests in identifying women with preeclampsia</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Anxiety was not different between women screened low and high risk for preeclampsia, but study groups were not equivalent. Comparisons in anxiety levels for women screened false negative could not be ascertained due to the timing of outcome measurement and insufficient power.</td>
</tr>
</tbody>
</table>
Table 7. Overall Summary of Evidence by Key Question

<table>
<thead>
<tr>
<th>Key Question</th>
<th>No. of Studies (k), No. of Observations (n), Design</th>
<th>Quality</th>
<th>Limitations*</th>
<th>Consistency</th>
<th>US Primary Care Applicability</th>
<th>Summary of Findings†</th>
</tr>
</thead>
<tbody>
<tr>
<td>KQ4a</td>
<td>Diagnostic accuracy of urine tests for proteinuria</td>
<td>k=14</td>
<td>Fair</td>
<td>Moderate/Low</td>
<td>Moderate</td>
<td>Protein:creatinine tests (k=12) had the most evidence on test accuracy; sensitivity ranged from 0.65 to 0.96 and specificity from 0.49 to 1.00. Evidence of spectrum bias and high heterogeneity limit conclusions that can be drawn about test performance in routine clinical care.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n=1,888</td>
<td></td>
<td></td>
<td></td>
<td>Dipstick urine tests least accurate (k=4). Sensitivity ranged from 0.22 to 1.00 and specificities from 0.36 to 1.00.</td>
</tr>
<tr>
<td></td>
<td>Diagnostic accuracy studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Albumin:creatinine spot urine tests (k=2); sensitivity (0.94 and 1.00) and specificity (0.94 and 0.68).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No studies comparing the performance of different approaches to routine preeclampsia screening were identified.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Within study comparisons from individual studies included for KQ4a provided limited evidence that: automated readings may be more accurate than visual; urine samples taken at different times of day have similar performance for the Albustix protein:creatinine ratio test; and different assays used for evaluating 24 hour protein gold-standard give different results for test sensitivity.</td>
</tr>
<tr>
<td>KQ4b</td>
<td>Effectiveness of different screening tests in identifying women with preeclampsia</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No studies comparing the performance of different approaches to routine preeclampsia screening were identified.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NA</td>
<td></td>
<td></td>
<td>Within study comparisons from individual studies included for KQ4a provided limited evidence that: automated readings may be more accurate than visual; urine samples taken at different times of day have similar performance for the Albustix protein:creatinine ratio test; and different assays used for evaluating 24 hour protein gold-standard give different results for test sensitivity.</td>
</tr>
</tbody>
</table>
### Table 7. Overall Summary of Evidence by Key Question

<table>
<thead>
<tr>
<th>Key Question</th>
<th>No. of Studies (k), No. of Observations (n), Design</th>
<th>Quality</th>
<th>Limitations*</th>
<th>Consistency</th>
<th>US Primary Care Applicability</th>
<th>Summary of Findings†</th>
</tr>
</thead>
<tbody>
<tr>
<td>KQ4c Effectiveness of different screening tests in identifying women at high- or low-risk for preeclampsia</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>KQ5 Harms of preeclampsia screening</td>
<td>k=2 1 RCT n=2,764 1 before-after study n=1,952</td>
<td>Fair</td>
<td>Insufficient power to detect differences in rare adverse outcomes such as very low birthweight and stillbirth RCT powered to detect differences of 2% or more between groups, but 1% differences for some outcomes could be clinically important Before-after study found a statistical difference in the source of payment for care over the study period, suggesting secular changes over time</td>
<td>NA</td>
<td>Only 2 studies, similar finding that reductions in routine preeclampsia screening did not increase adverse maternal and infant health outcomes.</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

*Includes reporting bias  †Includes precision  ‡ See methods for full explanation of the prediction model appraisal approach

**Abbreviations:** KQ = key question; NA = not available; RCT = randomized controlled trial; U.S. = United States.
Appendix A. Evidence on Preeclampsia Interventions to Reduce Morbidity and Mortality

**Induction of labor/early delivery.** Upon delivery, blood pressure and laboratory readings generally return to normal range values within a few days, although some women experience persistent high blood pressure that usually resolves within six weeks. While delivery of the placenta is the only definitive treatment for preeclampsia, there are potential disadvantages including increased risk of neonatal complications and increased risk of caesarean section. Clinical decisions are based on the balance between risks of expectant management versus the risks of immediate induction of labor. In cases of severe preeclampsia before 34 weeks of gestation, the effect of early delivery on neonatal and maternal outcomes is uncertain, with the exception of a decrease in the proportion of small for gestational age infants. Between 34 and 37 weeks of gestation, little is known about the risks of continuing pregnancy versus immediate delivery in women with preeclampsia.

To date, the only published prospective trials of gestational hypertension management are the HYPITAT and HYPITAT-II trials, both large multicenter open-label randomized controlled trials (RCTs) from the Netherlands evaluating the induction of labor versus expectant monitoring of hypertensive disorders at specific times during pregnancy. The first HYPITAT study found that immediate delivery reduced the risk of composite adverse maternal outcomes for women with preeclampsia after 37 weeks (RR, 0.71 [95% CI, 0.59 to 0.86]; p<0.0001), with no differences in rates of cesarean section or neonatal outcomes. Although trial evidence for assessing differences in outcomes with expectant management versus induced delivery was limited, recommendations generally favor induction of labor rather than continuing observation in women with preeclampsia at term (≥ 37 weeks) as it reduces the time the mother and fetus are at risk of injury from preeclampsia complications, such as eclampsia and placental abruption.

For pregnant women diagnosed with preeclampsia without severe features at less than 37 weeks, expectant management, monitoring for disease progression rather than immediate delivery, is recommended. The HYPITAT-II trial found that for pregnant women with nonsevere hypertensive disorders (systolic blood pressure less than 170 mm Hg or diastolic blood pressure less than 110 mm Hg) between 34 to 36 weeks gestation, immediate delivery may slightly reduce the small risk of adverse maternal outcomes, but significantly increase the risk of neonatal respiratory distress syndrome. Recommendations for preeclampsia with severe features between 34 and 37 weeks gestation are varied. Both the Society of Obstetricians and Gynecologists of Canada (SOGC) and National Institute for Health and Care Excellence (NICE) recommend delivery for all women who have preeclampsia with severe hypertension after 34 weeks, while the World Health Organization (WHO) advises a policy of expectant management provided that uncontrolled maternal hypertension, maternal organ dysfunction or fetal distress are absent.

Trial evidence on the health outcomes associated with delivery versus expectant management of severe preeclampsia occurring before 34 weeks is limited and inconclusive.

**Magnesium sulfate.** Magnesium sulfate (MgSO₄) has been routinely used for the prevention of eclampsia seizures since the middle of the 20th century. The Magpie Trial (n=10,141), an important international randomized placebo-controlled trial of magnesium sulfate to prevent eclampsia, established clear evidence of a benefit for preventing eclampsia with magnesium.
sulfate among women for whom there was clinical uncertainty as to whether it should be administered.49 Pregnant women given magnesium sulfate (n=40) had a 58 percent lower risk of eclampsia (95% CI, 40 to 71) than those allocated placebo (n=96).49 Maternal mortality was also lower in the treatment group, although there were few cases (n=11; 0.2%) and the difference was not statistically significant (RR, 0.55 [95% CI, 0.26 to 1.14]).49 Placental abruption was significantly lower in the treatment group, and there was no evidence of short term or longer term (up to 2 years) harms to the mother or offspring from the treatment. Following the trial, clinical management protocols have unequivocally recommended treatment of women with worsening or severe manifestations of preeclampsia to receive magnesium sulfate during delivery. A followup study of the Magpie Trial found that the use of magnesium sulfate in preeclamptic women demonstrated a 16 percent reduction in mortality and morbidity risk related to preeclampsia two to three years after delivery,187 and no association with any difference in mortality and morbidity risk in children (18 months) whose mothers were recruited to the trial.188 A recent Cochrane review of anticonvulsant management of preeclampsia found that magnesium sulfate more than halved the risk of eclampsia and likely reduced maternal death.181

Guideline groups strongly agree on the importance of magnesium sulfate as first-line treatment of eclampsia as well as prophylaxis against eclampsia in women with severe preeclampsia.1,47,51,143

**Antihypertensive medications.** Severe hypertension in pregnancy, regardless of the diagnosis of preeclampsia, poses a serious health risk to a pregnant woman and her fetus, and the use of antihypertensive medications is sometimes necessary to lower blood pressure to a safe range.189,190 Antihypertensive medications are used to prevent potential cardiovascular, renal, or cerebrovascular complications related to uncontrolled severe hypertension. The American Heart Association/American Stroke Association (AHA/ASA)43 along with other professional guideline groups1,47,51,143 strongly recommend that women with severe hypertension during pregnancy be treated with safe and effective antihypertensive medications. The benefit of antihypertensive therapy is highlighted by the results of a 2014 retrospective chart review of over 1.2 million women delivering in a U.S. hospital system.57 The analysis found a significant reduction in deaths from preeclampsia (15 to 3; p=0.02) following implementation of an automatic protocol for antihypertensive treatment during pregnancy.57 There is not consensus, however, regarding the management of nonsevere hypertension.190 A recent Cochrane review of 49 trials assessed the effects of antihypertensive drug treatments for pregnant women with mild to moderate hypertension and found no statistically significant difference in preeclampsia risk, and no evidence of benefit or harm to the fetus.191
### Literature Search Strategies – Systematic Reviews

Database: Ovid MEDLINE(R) without Revisions <1996 to November Week 2 2013>, Ovid MEDLINE(R) Daily Update <November 20, 2013>, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <November 20, 2013>

Search Strategy:

| 1  | Pre-Eclampsia/ ()                                      |
| 2  | Hypertension, Pregnancy-Induced/ ()                   |
| 3  | Eclampsia/ ()                                         |
| 4  | Pregnancy/ ()                                         |
| 5  | Hypertension/ ()                                      |
| 6  | 4 and 5 ()                                            |
| 7  | 1 or 2 or 3 or 6 ()                                   |
| 8  | Mass screening/ ()                                    |
| 9  | Biological markers/ ()                                |
| 10 | Ultrasonography, Doppler/ ()                          |
| 11 | "Predictive Value of Tests"/ ()                      |
| 12 | "Sensitivity and Specificity"/ ()                    |
| 13 | Diagnostic errors/ ()                                 |
| 14 | Risk factors/ ()                                      |
| 15 | Risk assessment/ ()                                   |
| 16 | (screen$ or diagnos$ or predict$ or detect$ or risk$).ti. () |
| 17 | 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 ()    |
| 18 | 7 and 17 ()                                           |
| 19 | Pre-Eclampsia/di ()                                   |
| 20 | Hypertension, Pregnancy-Induced/di ()                 |
| 21 | Eclampsia/di ()                                       |
| 22 | 18 or 19 or 20 or 21 ()                               |
| 23 | preeclamps$.ti,ab. ()                                 |
| 24 | eclamps$.ti,ab. ()                                    |
| 25 | gestosis.ti,ab. ()                                    |
| 26 | ((gestational or pregnan$) adj5 (tox?emi$ or hypertens$)).ti,ab. () |
| 27 | 23 or 24 or 25 or 26 ()                               |
| 28 | (screen$ or diagnos$ or predict$ or detect$ or risk$).ti,ab. () |
| 29 | 27 and 28 ()                                          |
| 30 | limit 29 to ("in data review" or in process or "pubmed not medline") () |
| 31 | 22 or 30 ()                                           |
| 32 | limit 31 to systematic reviews ()                     |
| 33 | limit 32 to (english language and yr="2009 -Current") () |
| 34 | remove duplicates from 33 ()                          |
Appendix B. Detailed Methods

Literature Search Strategies – Primary Literature

**MEDLINE**
Database: Ovid MEDLINE(R) <1946 to March Week 4 2015>, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations < April 2, 2015>, Ovid MEDLINE(R) Daily Update <April 2, 2015>
Search Strategy:

1. Pre-Eclampsia/ ()
2. Hypertension, Pregnancy-Induced/ ()
3. Eclampsia/ ()
4. Pregnancy/ ()
5. Pregnancy Trimester, First/ ()
6. Pregnancy Trimester, Second/ ()
7. Pregnancy Trimester, Third/ ()
8. Hypertension/ ()
9. (4 or 5 or 6 or 7) and 8 ()
10. (preeclamp$ or pre eclamp$).ti. ()
11. eclamp$.ti. ()
12. gestosis.ti. ()
13. ((gestational or pregnan$) and (tox?emi$ or hypertens$ or blood pressure)).ti. ()
14. 1 or 2 or 3 or 9 or 10 or 11 or 12 or 13 ()
15. Blood pressure/ ()
16. Blood pressure determination/ ()
17. Blood pressure monitoring, Ambulatory/ ()
18. Blood pressure monitors/ ()
19. Urinalysis/ ()
20. Uric acid/ ()
21. Proteinuria/ ()
22. Pregnancy Proteins/ ()
23. Uterine Artery/us ()
24. Ultrasonography, Doppler/ ()
25. Creatinine/ur ()
26. Biological Markers/ ()
27. Pregnancy-Associated Plasma Protein-A/ ()
28. ((blood or systolic or diastolic) adj pressure).ti,ab. ()
29. urinalys$.ti,ab. ()
30. (urine adj (measur$ or analy$ or test$ or collect$)).ti,ab. ()
31. uric acid.ti,ab. ()
32. (proteinuria or albuminuria or urine albumin).ti,ab. ()
33. (ultrasound or ultrasonography).ti,ab. ()
34. uterine artery doppler.ti,ab. ()
35. ((biological or serum) adj3 (marker$ or biomarker$)).ti,ab. ()
36. plasma protein a.ti,ab. ()
37. or/15-36 ()
38. Mass screening/ ()
Appendix B. Detailed Methods

39 screen$.ti,ab. ()
40 (detect$ or predict$ or identif$).ti. ()
41 38 or 39 or 40 ()
42 14 and (37 or 41) ()
43 clinical trials as topic/ or controlled clinical trials as topic/ or randomized controlled trials as topic/ or meta-analysis as topic/ ()
44 (clinical trial or controlled clinical trial or meta analysis or randomized controlled trial).pt. ()
45 Random$.ti,ab. ()
46 control groups/ or double-blind method/ or single-blind method/ ()
47 clinical trial$.ti,ab. ()
48 controlled trial$.ti,ab. ()
49 meta analy$.ti,ab. ()
50 epidemiologic studies/ or cohort studies/ or longitudinal studies/ or follow-up studies/ or prospective studies/ or retrospective studies/ ()
51 cohort$.ti,ab. ()
52 longitudinal.ti,ab. ()
53 incidence stud$.ti,ab. ()
54 retrospective.ti,ab. ()
55 (follow-up or followup).ti,ab. ()
56 prospective.ti,ab. ()
57 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 ()
58 42 and 57 ()
59 limit 58 to (english language and yr="1990 -Current") ()
60 remove duplicates from 59 ()
61 Risk/ ()
62 Risk factors/ ()
63 Risk assessment/ ()
64 risk$.ti,ab. ()
65 multivariable prediction.ti,ab. ()
66 61 or 62 or 63 or 64 or 65 ()
67 14 and 66 ()
68 limit 67 to (english language and yr="1990 -Current") ()
69 remove duplicates from 68 ()
70 "Sensitivity and Specificity"/ ()
71 "Predictive Value of Tests"/ ()
72 ROC Curve/ ()
73 False Negative Reactions/ ()
74 False Positive Reactions/ ()
75 Diagnostic Errors/ ()
76 "Reproducibility of Results"/ ()
77 Reference Values/ ()
78 Reference Standards/ ()
79 Observer Variation/ ()
80 Receiver operat$.ti,ab. ()
81 ROC curve$.ti,ab. ()
Appendix B. Detailed Methods

82 sensitivit$.ti,ab. ()
83 specificit$.ti,ab. ()
84 predictive value.ti,ab. ()
85 accuracy.ti,ab. ()
86 false positive$.ti,ab. ()
87 false negative$.ti,ab. ()
88 miss rate$.ti,ab. ()
89 error rate$.ti,ab. ()
90 70 or 71 or 72 or 74 or 75 or 76 or 77 or 78 or 79 or 80 or 81 or 82 or 83 or 84 or 85
or 86 or 87 or 88 or 89 ()
91 42 and 90 ()
92 limit 91 to (english language and yr="1990 -Current") ()
93 remove duplicates from 92 ()
94 Mortality/ ()
95 Morbidity/ ()
96 Death/ ()
97 safety.ti,ab. ()
98 harm$.ti,ab. ()
99 mortality.ti,ab. ()
100 complication$.ti,ab. ()
101 (death or deaths).ti,ab. ()
102 ((adverse or unintended or negative) adj (effect$ or event$ or reaction$ or
outcome$)).ti,ab. ()
103 (adverse effects or mortality).fs. ()
104 Cesarean Section/ ()
105 Magnesium Sulfate/to ()
106 Anxiety/ ()
107 Stress, Psychological/ ()
108 Premature Birth/ ()
109 (cesarean$ or c-section$).ti,ab. ()
110 hypermagnesemi$.ti,ab. ()
111 (anxiety or anxious).ti,ab. ()
112 ((psychological or psychosocial or mental) adj (stress or distress or outcome$)).ti,ab. ()
113 ((preterm or premature$) adj (birth$ or deliver$)).ti,ab. ()
114 misdiagnos$.ti,ab. ()
115 overdiagno$.ti,ab. ()
116 misclassification$.ti,ab. ()
117 ((unnecessary or unneeded) adj3 (treat$ or induc$ or monitor$)).ti,ab. ()
118 (increase$ adj3 monitor$).ti,ab. ()
119 or/94-118 ()
120 42 and 119 ()
121 limit 120 to (english language and yr="1990 -Current") ()
122 remove duplicates from 121 ()
123 60 or 69 or 93 or 122 ()
124 Animal/ not (Animal/ and Human/) ()
125 123 not 124 ()
Appendix B. Detailed Methods

**PUBMED**
Query
Search (((#24) AND publisher[sb]) AND English[Language]) AND ("1990"[Date - Publication] : "3000"[Date - Publication])
Search (#10 AND #23)
Search #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22
Search "multivariable prediction"[tiab]
Search risk*[tiab]
Search "plasma protein a"[tiab]
Search biological marker*[tiab] OR biological biomarker*[tiab] OR serum marker*[tiab] OR serum biomarker*[tiab]
Search "uterine artery doppler"[tiab]
Search (ultrasound[tiab] or ultrasonography[tiab])
Search (proteinuria[tiab] or albuminuria[tiab] or "urine albumin"[tiab])
Search urinalys*[tiab]
Search blood pressure[tiab] OR systolic pressure[tiab] OR diastolic pressure[tiab]
Search (detect*[title] OR predict*[title] OR identif*[title])
Search screen*[tiab]
Search (#8 OR #9)
Search (hypertens*[title] OR blood pressure[title] OR toxemi*[title] OR toxaemi*[title]) AND (gestational[title] OR pregnan*[title])
Search pre eclampsia[title] OR preeclampsia[title] OR pre eclamptic[title] OR preeclamptic[title]
or eclampsia[title] or eclamptic[title] or gestosis[title])

**Cochrane Central Register of Controlled Clinical Trials (CENTRAL)**

#1 preeclam*:ti,ab,kw
#2 (pre-eclampsia or pre-eclamptic):ti,ab,kw
#3 eclamp*:ti,ab,kw
#4 gestosis:ti,ab,kw
#5 #1 or #2 or #3 or #4
#6 hypertension:ti,ab,kw
#7 hypertensive:ti,ab,kw
#8 (toxemi*:ti,ab,kw or toxaemi*:ti,ab,kw)
#9 "blood pressure":ti,ab,kw near/5 (high or elevated or abnormal):ti,ab,kw
#10 #6 or #7 or #8 or #9
#11 "pregnancy":ti,ab,kw
#12 "pregnant":ti,ab,kw
#13 gestational:ti,ab,kw
#14 #11 or #12 or #13
#15 #10 and #14
#16 #5 or #15
#17 screen*:ti,ab,kw
#18 (detect* or predict* or identif*):ti
Appendix B. Detailed Methods

#19 (blood or systolic or diastolic):ti,ab,kw next pressure:ti,ab,kw
#20 urinalys*:ti,ab,kw
#21 urine:ti,ab,kw next (measur* or analy* or test* or collect*):ti,ab,kw
#22 (proteinuria or albuminuria or "urine albumin"):ti,ab,kw
#23 (ultrasound or ultrasonography):ti,ab,kw
#24 "uterine artery doppler":ti,ab,kw
#25 (biological or serum):ti,ab,kw near/3 (marker* or biomarker*):ti,ab,kw
#26 "plasma protein a":ti,ab,kw
#27 risk*:ti,ab,kw
#28 "multivariable prediction":ti,ab,kw
#29 #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28
#30 #16 and #29 Publication Year from 1990 to 2014, in Trials
Appendix B Figure 1. Literature Flow Diagram

Abbreviations: KQ = Key Question

We included four studies (seven articles) on externally validated risk prediction models. We also identified 11 articles that represent the model development studies related to the external validation studies.

Screening for Preeclampsia

Kaiser Permanente Research Affiliates EPC
### Appendix B Table 1. Inclusion and Exclusion Criteria

<table>
<thead>
<tr>
<th>Category</th>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Populations</strong></td>
<td>All pregnant women without a diagnosis of preeclampsia and asymptomatic for preeclampsia including pregnant women with common chronic conditions seen in primary care (i.e., hypertension, diabetes mellitus) and at elevated risk for preeclampsia</td>
<td>Studies that exclusively include individuals seeking high-risk obstetric care (e.g., in-vitro fertilization); inpatients or hospitalized; other selected non-generalizable populations or populations with other preexisting health conditions (e.g., HIV, HPV, hepatitis, autoimmune disorders, polycystic ovarian syndrome, renal disease, organ transplant recipients, sickle cell trait)</td>
</tr>
<tr>
<td><strong>Disease/Condition</strong></td>
<td>KQ1, 1a: Eclampsia, maternal morbidity and mortality, perinatal/neonatal morbidity and mortality KQ2, KQ4: Preeclampsia KQ3, KQ5: Eclampsia, maternal morbidity and mortality (including psychological effects of risk assessment and screening), perinatal/neonatal morbidity and mortality</td>
<td></td>
</tr>
<tr>
<td><strong>Interventions:</strong></td>
<td>KQs 1, 4, 5: Screening occurs from 20 weeks gestation to delivery. Screening tests for preeclampsia are blood pressure measurement and urine protein tests KQ4a: Point of care urine tests (e.g., dipstick or random urine spot test) KQs 2, 3: Risk assessment occurring before 20 weeks gestation using multivariable prediction tools for the identification of women at high risk for preeclampsia</td>
<td>Experimental tests that are not routinely used for preeclampsia screening in clinical practice Secondary evaluations and tests used to assess preeclampsia severity or confirm diagnosis in symptomatic women Urine screening tests requiring ongoing collection of urine 24 hour ambulatory blood pressure measurements Risk assessment occurring after 20 weeks gestation Non-routine screening tests: Serum markers (e.g., angiogenic factors, activated protein C, calcium, HCG, HCy, hormones, lipids, thyroid hormone levels) Genetic susceptibility markers (e.g., fetal DNA) Ultrasound measurements (e.g., Doppler ultrasound pulsatility index or resistance index; pulse wave velocity or notching)</td>
</tr>
<tr>
<td><strong>Comparisons</strong></td>
<td>KQ 1: No screening, different screening protocols (e.g., modality, timing, rescreen interval) KQ3: Usual care; low risk KQs 4, 5: Different blood pressure and proteinuria screening protocols (e.g., instrument, procedure, timing, frequency) and screening protocols KQ4a: Reference standard is 24 hour urine collection for protein measurement</td>
<td>Reference standard other than 24 hour urine collection for protein measurement</td>
</tr>
</tbody>
</table>
## Appendix B Table 1. Inclusion and Exclusion Criteria

<table>
<thead>
<tr>
<th>Category</th>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcomes</strong></td>
<td>KQ1 (maternal and perinatal health outcomes): Maternal mortality and serious morbidity (e.g., organ or system failure or injury, eclampsia) and perinatal or neonatal mortality and serious morbidity (e.g., intrauterine growth restriction, low birth weight, brain injury)</td>
<td>Nonclinical health outcomes, such as length of hospital stay (without indication), intensive care unit admission, or neonatal intensive care unit admission.</td>
</tr>
<tr>
<td></td>
<td>KQs 2 (intermediate outcome): Prediction, discrimination, calibration outcomes for preeclampsia risk prediction model (e.g., AUC, Brier score)</td>
<td>KQ 4: Bivariable or multivariable regression (e.g., correlations)</td>
</tr>
<tr>
<td></td>
<td>KQ 4 (intermediate outcomes): Test performance characteristics, sensitivity, specificity, for accuracy and effectiveness of screening.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>KQs 3, 5 (harms): Misclassification, increased monitoring, false positives, overdiagnosis, overtreatment (e.g., failed induction, Cesarean section, induced preterm birth, hypermagnesemia), and patient stress and anxiety</td>
<td></td>
</tr>
<tr>
<td><strong>Setting</strong></td>
<td>Primary care outpatient settings for obstetric care (e.g., obstetrician gynecologists, family physicians, certified nurse midwives)</td>
<td>Clinics and study sites treating only high risk maternity patients</td>
</tr>
<tr>
<td></td>
<td>Countries categorized as “Very High” or equivalent on the Human Development Index (as defined by the World Health Organization, 2014)</td>
<td>Countries not categorized as “Very High” on the Human Development Index or not applicable to U.S. clinical settings or populations</td>
</tr>
<tr>
<td><strong>Study Designs</strong></td>
<td>KQ1: RCTs</td>
<td>KQ1: Case-control study, editorial, narrative review, commentary, postmarketing surveillance, and case report.</td>
</tr>
<tr>
<td></td>
<td>KQ2: Nested case-control or cohort study aiming to externally validate a multivariable clinical risk prediction tool; randomized impact studies comparing clinical risk prediction based care to usual care</td>
<td>KQ4: Case-control study, editorial, narrative review, commentary, postmarketing surveillance, and case reports</td>
</tr>
<tr>
<td></td>
<td>KQ4: RCTs, cohort studies, instrument validation studies, and test accuracy studies</td>
<td>KQs 3, 5: Editorial, narrative review, commentary, postmarketing surveillance, and case reports</td>
</tr>
<tr>
<td></td>
<td>KQs 3, 5: RCTs or observational studies (e.g., nested case control, case series, cohort, registry, survey data)</td>
<td></td>
</tr>
<tr>
<td><strong>Publication Dates</strong></td>
<td>Studies published after January 1990, all references from the previous USPSTF review, and eligible studies identified through a bridge search</td>
<td>Studies published before 1990</td>
</tr>
<tr>
<td><strong>Study Quality</strong></td>
<td>Good and fair quality according to USPSTF design-specific criteria</td>
<td>Poor quality according to USPSTF design-specific criteria</td>
</tr>
<tr>
<td><strong>Language</strong></td>
<td>English</td>
<td>Non-English studies</td>
</tr>
</tbody>
</table>

*Settings: Included Countries: All countries listed as “very high” or equivalent on human development on the Human Development Index, 2014: Andorra, Argentina, Australia, Austria, Barbados, Belgium, Brunei Darussalam, Canada, Chile, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hong Kong, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Netherlands, New Zealand, Norway, Poland, Portugal, Qatar, Seychelles, Singapore, Slovakia, Slovenia, Spain, Sweden, Switzerland, Taiwan, United Arab Emirates, United Kingdom, United States

**Abbreviations:** RCT = randomized, controlled trial.
### Appendix B Table 2. Quality Assessment Criteria

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Adapted Quality Criteria</th>
</tr>
</thead>
</table>
| Randomized controlled trials, adapted from the U.S. Preventive Services Task Force methods<sup>87</sup> | Was there valid random assignment?  
Was allocation concealed?  
Was eligibility criteria specified?  
Were groups similar at baseline?  
Were the outcome assessors blinded?  
Was there intervention fidelity?  
Was there adequate adherence to the intervention?  
Were measurements equal, valid and reliable?  
Was there acceptable followup?  
Was there a difference between those who completed the study and those who withdrew?  
Was the handling of missing data appropriate?  
Were the statistical methods acceptable?  
Was there evidence of selective reporting of outcomes? |
| Observational studies (e.g., prospective cohort studies), adapted from the Newcastle-Ottawa Scale (NOS)<sup>89</sup> | Was there representativeness of the exposed cohort?  
Was the non-exposed systematically selected?  
Was the ascertainment of exposure reported?  
Was eligibility criteria specified?  
Comparability of cohorts on the basis of design or analysis?  
Was the outcome of interest not present at baseline?  
Were measurements equal, valid and reliable?  
Were outcome assessors blinded?  
Was followup long enough for the outcome to occur?  
Was there adequate followup of cohorts?  
Was there adjustment for confounders?  
Were the statistical methods acceptable?  
Was the handling of missing data appropriate? |
| Diagnostic accuracy studies, adapted from the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) II instrument<sup>88</sup> | Could the selection of patients have introduced bias?  
Was a consecutive or random sample of patients enrolled?  
Was a case-control design avoided?  
Did the study avoid inappropriate exclusions?  
Could the conduct or interpretation of the index test have introduced bias?  
Was the index test interpreted without knowledge of the reference standard results?  
If a threshold was use, was it pre-specified?  
Was the fidelity of the index test monitored and/or reported?  
Could the conduct or interpretation of the reference standard have introduced bias?  
Was the reference standard interpreted without knowledge of the index test results?  
Was the fidelity of the reference test monitored and/or reported?  
Could the patient flow have introduced bias?  
Was there an appropriate interval between the index test and reference standard?  
Did all patients receive the same reference standard?  
Did all patient complete all tests?  
Were all patients completing both tests included in the analysis? |
| Before-After<sup>89</sup> | Is the post-intervention group representative?  
Is the pre-intervention group representative?  
Are the pre- and post-intervention groups comparable on the basis of design or analysis?  
Was the assessment of outcomes valid?  
Was the assessment of outcomes reliable?  
Was the method of outcome assessment the same for the pre- and post-intervention groups?  
Did the study report the point of time when the intervention occurred?  
Was the intervention clearly described?  
Were the data collected during a similar timeframe? |
### Appendix C. Preeclampsia Diagnostic Criteria Included in Major Guidelines and Recommendations, 1972–2013

<table>
<thead>
<tr>
<th>Organization</th>
<th>Hypertension</th>
<th>Proteinuria</th>
<th>Other Diagnostic Indicators (Symptoms, Blood Test Results, or Health Outcomes)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>United States</strong></td>
<td><strong>Preeclampsia must include one of the following:</strong> SBP ≥140 mmHg or DBP ≥90 mmHg on two occasions &gt;4 hours apart if pregnancy after 20 weeks of gestation in a previously normotensive woman. If SBP ≥160 mmHg or DBP ≥110 mmHg, hypertension can be confirmed within a short interval to facilitate timely delivery of antihypertensive therapy. <strong>Severe preeclampsia may include:</strong> SBP ≥160 mmHg or DBP ≥110 mmHg on two occasions &gt;4 hours apart while the patient is on bed rest (unless antihypertensive therapy is initiated before this time).</td>
<td><strong>Preeclampsia may include one of the following:</strong> ≥300 mg protein per 24 hour urine collection or Protein/creatinine ratio ≥0.3 mg/dL Dipstick reading of 1+ (used only if other quantitative methods not available). <strong>Severe preeclampsia may include:</strong> Serum creatinine concentration &gt;1.1 mg/dL or a doubling of the serum creatinine concentration in the absence of other renal disease.</td>
<td><strong>In the absence of proteinuria, preeclampsia can be confirmed by new onset hypertension and one of the following:</strong> Thrombocytopenia Renal insufficiency Impaired liver function Pulmonary edema Cerebral or visual symptoms. <strong>Severe preeclampsia may include:</strong> Thrombocytopenia Progressive renal insufficiency Impaired liver function Pulmonary edema Cerebral or visual disturbances.</td>
</tr>
<tr>
<td>The American College of Obstetricians and Gynecologists (ACOG) 2013, U.S.</td>
<td><strong>Preeclampsia must include:</strong> SBP ≥140 mmHg or DBP ≥90 mmHg on two occasions &gt;4 hours apart after 20 weeks of gestation in a previously normotensive woman.</td>
<td><strong>Preeclampsia must include:</strong> ≥0.3 g protein per 24 hour urine collection (correlates with ≥1+ reading on dipstick but should be confirmed using a random urine evaluation). <strong>Severe preeclampsia may include:</strong> ≥5 g protein per 24 hour urine collection or dipstick ≥3+ on two random urine samples collected ≥4 hours apart.</td>
<td><strong>Preeclampsia may include:</strong> Edema Visual disturbances Headache Epigastric pain Hemolysis Elevated liver enzymes Low platelet counts (HELLP syndrome). <strong>Severe preeclampsia may include one of the following:</strong> Oliguria of &lt; 500 mL in 24 hours Cerebral or visual disturbances Pulmonary edema or cyanosis Epigastric or right upper-quadrant pain Impaired liver function Thrombocytopenia Fetal growth restriction.</td>
</tr>
<tr>
<td>The American College of Obstetricians and Gynecologists (ACOG) 2002, U.S.</td>
<td><strong>Preeclampsia must include:</strong> SBP ≥140 mmHg or DBP ≥90 mmHg presenting after 20 weeks of gestation in a previously normotensive woman. <strong>Severe preeclampsia may include:</strong> SBP ≥160 mmHg or DBP ≥110 mmHg on two occasions ≥6 hours apart while on bed rest.</td>
<td><strong>Preeclampsia must include:</strong> ≥0.3 g protein per 24 hour urine collection (correlates with ≥1+ reading on dipstick but should be confirmed using a random urine evaluation). <strong>Severe preeclampsia may include:</strong> ≥5 g protein per 24 hour urine collection or dipstick ≥3+ on two random urine samples collected ≥4 hours apart.</td>
<td><strong>Preeclampsia may include:</strong> Edema Visual disturbances Headache Epigastric pain Hemolysis Elevated liver enzymes Low platelet counts (HELLP syndrome). <strong>Severe preeclampsia may include one of the following:</strong> Oliguria of &lt; 500 mL in 24 hours Cerebral or visual disturbances Pulmonary edema or cyanosis Epigastric or right upper-quadrant pain Impaired liver function Thrombocytopenia Fetal growth restriction.</td>
</tr>
<tr>
<td>Based on NHLBI Working Group 2000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Appendix C. Preeclampsia Diagnostic Criteria Included in Major Guidelines and Recommendations, 1972–2013

<table>
<thead>
<tr>
<th>Organization</th>
<th>Hypertension</th>
<th>Proteinuria</th>
<th>Other Diagnostic Indicators (Symptoms, Blood Test Results, or Health Outcomes)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>National Heart, Lung, and Blood Institute (NHLBI) Working Group 2000, U.S.</strong>&lt;sup&gt;192&lt;/sup&gt;</td>
<td>Preeclampsia must include: SBP &gt;140 mmHg or DBP &gt;90 mmHg presenting after 20 weeks of gestation in a previously normotensive woman</td>
<td>Preeclampsia must include: ≥0.3 g protein per 24 hour urine collection (correlates with ≥30 mg/dL in a random urine determination or ≥1+ reading on dipstick)</td>
<td>In the absence of proteinuria, preeclampsia is highly suspected when hypertension appears with the following: Headache Blurred vision Abdominal pain Low platelet counts Abnormal liver enzyme values Edema occurs in too many women with normal pregnancies and has been removed as a marker in the classification of preeclampsia.</td>
</tr>
<tr>
<td><strong>National Heart, Lung, and Blood Institute (NHLBI) Working Group 1990, U.S.</strong>&lt;sup&gt;193&lt;/sup&gt;</td>
<td>Preeclampsia must include one of the following: SBP ≥140 mmHg or DBP ≥90 mmHg presenting after 20 weeks of gestation in a previously normotensive woman SBP increases of ≥30 mmHg or DBP increases of ≥15 mmHg from early values before 20 weeks of gestation</td>
<td>Preeclampsia may include: ≥0.3 g protein per 24 hour urine collection (correlates with ≥30 mg/dL in a random urine determination or ≥1+ reading on dipstick)</td>
<td>Preeclampsia may include: Edema</td>
</tr>
<tr>
<td><strong>United Kingdom</strong></td>
<td>Preeclampsia must include: SBP ≥140 mmHg or DBP ≥90 mmHg presenting after 20 weeks of gestation in a previously normotensive woman</td>
<td>Preeclampsia must include one of the following: &gt;300 mg protein per 24 hour urine collection Protein/creatinine ratio &gt;30 mg/mmol</td>
<td>In the absence of severe hypertension, features of severe preeclampsia include mild/moderate hypertension and proteinuria with at least one of the following: Severe headache Problems with vision such as blurring or flashing Severe pain just below ribs or vomiting Papilloedema Signs of clonus (≥ 3 beats) Liver tenderness HELLP syndrome Platelet count falls to &lt; 100 x 10&lt;sup&gt;9&lt;/sup&gt;/liter abnormal liver enzymes</td>
</tr>
<tr>
<td><strong>National Institute for Health and Care Excellence (NICE) 2010, U.K.</strong>&lt;sup&gt;47&lt;/sup&gt;</td>
<td>Severe preeclampsia must include one of the following: SBP ≥160 mmHg or DBP ≥110 mmHg SBP ≥140 mmHg or DBP ≥90 mmHg (mild hypertension) or SBP ≥150 mmHg or DBP ≥100 mmHg (moderate hypertension) with other diagnostic indicators</td>
<td>Severe preeclampsia must include one of the following: &gt;300 mg protein per 24 hour urine collection Protein/creatinine ratio &gt;30 mg/mmol</td>
<td>Screening for Preeclampsia</td>
</tr>
</tbody>
</table>
### Appendix C. Preeclampsia Diagnostic Criteria Included in Major Guidelines and Recommendations, 1972–2013

<table>
<thead>
<tr>
<th>Organization</th>
<th>Hypertension</th>
<th>Proteinuria</th>
<th>Other Diagnostic Indicators (Symptoms, Blood Test Results, or Health Outcomes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Society of Obstetricians and Gynecologists of Canada (SOGC) 2014, Canada(^{145})</td>
<td><em>Preeclampsia must include:</em> SBP ≥140 mmHg or DBP ≥90 mmHg (based on the average of ≥2 measurements taken ≥15 minutes apart) after 20 weeks of gestation in a previously normotensive woman</td>
<td><em>Preeclampsia may include one of the following:</em> ≥0.3 g protein per 24 hour urine collection ≥30 mg/mmol urinary creatinine in a spot (random) urine sample</td>
<td>In the absence of proteinuria, preeclampsia can be confirmed by new onset hypertension and one of the following: Adverse condition (e.g., headache, visual symptoms, chest pain, low platelet count, nausea or vomiting, epigastric pain) Severe complication (e.g., eclampsia, stroke, uncontrolled severe hypertension, platelet count &lt; 50 x 10^9/liter, acute kidney injury, hepatic dysfunction, abruption with evidence of maternal fetal compromise)</td>
</tr>
</tbody>
</table>

**Abbreviations:** DBP = diastolic blood pressure; HELLP = hemolysis, elevated liver enzymes, low platelet count; SBP = systolic blood pressure.
Appendix D. Excluded Studies

<table>
<thead>
<tr>
<th>Code</th>
<th>Reason for Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>Setting&lt;br&gt;Clinics and study sites treating only high risk maternity patients&lt;br&gt;Countries not categorized as “very high HDI” equivalent or not applicable to U.S. clinical settings</td>
</tr>
<tr>
<td>E2</td>
<td>Population&lt;br&gt;Patients seeking high risk obstetric care or those with known chronic conditions (other than hypertension or diabetes)&lt;br&gt;Hospitalized patients</td>
</tr>
<tr>
<td>E3</td>
<td>Study Design&lt;br&gt;Editorial, narrative review, commentary, post-marketing surveillance, case reports&lt;br&gt;Not approved study design for the KQ&lt;br&gt;Risk factor screening occurred after 20 weeks gestation (KQ2)&lt;br&gt;Case-control study (KQ4)&lt;br&gt;Case-control study, but not nested (KQ1, KQ2)&lt;br&gt;N too small (&lt;100) (KQ2)</td>
</tr>
<tr>
<td>E4</td>
<td>Outcomes&lt;br&gt;Non-clinical health outcomes, such as length of hospital stay (without indication), ICU admission, or NICU admission</td>
</tr>
<tr>
<td>E5</td>
<td>Disease/Condition&lt;br&gt;Not preeclampsia or eclampsia (KQ1, KQ2)&lt;br&gt;Not proteinuria (KQ4a)&lt;br&gt;Not hypertension (KQ4b)</td>
</tr>
<tr>
<td>E6</td>
<td>Interventions&lt;br&gt;Serum markers or ultrasound measurements not routinely collected&lt;br&gt;Secondary evaluations or diagnostic tests&lt;br&gt;Experimental tests&lt;br&gt;Not a screening tool (e.g., prognostic assessment)</td>
</tr>
<tr>
<td>E7</td>
<td>Comparisons&lt;br&gt;Comparators (e.g., not the appropriate reference standard)</td>
</tr>
<tr>
<td>E8</td>
<td>Language&lt;br&gt;Non-English publication</td>
</tr>
<tr>
<td>E9</td>
<td>Publication Date&lt;br&gt;Published before 1990</td>
</tr>
<tr>
<td>E10</td>
<td>Study Quality&lt;br&gt;Poor</td>
</tr>
<tr>
<td>E11</td>
<td>Unable to locate article</td>
</tr>
<tr>
<td>E12</td>
<td>Study Aim&lt;br&gt;Not applicable/relevant to key question</td>
</tr>
<tr>
<td>E13</td>
<td>Non-externally validated risk prediction models (KQ2)</td>
</tr>
</tbody>
</table>


Appendix D. Excluded Studies


Appendix D. Excluded Studies


Appendix D. Excluded Studies


Appendix D. Excluded Studies


Appendix D. Excluded Studies


Appendix D. Excluded Studies


Appendix D. Excluded Studies


Appendix D. Excluded Studies


133. Hermida RC, Ayala DE. Diagnosing gestational hypertension and preeclampsia with the 24-hour mean of blood pressure. Hypertension 1997 Dec;30(6):1531-7. PMID: 9403578. KQ1E6, KQ1aE6, KQ2E12, KQ3E12, KQ4E6, KQ4aE6, KQ4bE6, KQ4cE6, KQ5E6.


Appendix D. Excluded Studies


Appendix D. Excluded Studies


Appendix D. Excluded Studies


179. LA BM, Okido MM, Barbieri MA, et al. [240-POS]: Risk factors influencing the development of hypertension in pregnancy in a convenience cohort. Pregnancy Hypertens 2015 Jan;5(1):121. PMID: 25787590. KQ1E1, KQ1aE1, KQ2E12, KQ3E12, KQ4E1, KQ4aE1, KQ4bE1, KQ4cE1, KQ5E1.


Appendix D. Excluded Studies


188. Lee LC, Sheu BC, Shau WY, et al. Mid-trimester beta-hCG levels incorporated in a multifactorial model for the prediction of severe pre-eclampsia. Prenat Diagn 2000 Sep;20(9):738-43. PMID: 11015703. KQ1E12, KQ1aE12, KQ2E13, KQ3E4, KQ4E12, KQ4aE12, KQ4bE12, KQ4cE12, KQ5E12.


191. Levine RJ, Ewell MG, Hauth JC, et al. Should the definition of preeclampsia include a rise in diastolic blood pressure of >/=15 mm Hg to a level <90 mm Hg in association with proteinuria? Am J Obstet Gynecol 2000 Oct;183(4):787-92. PMID: 11035314. KQ1E6, KQ1aE6, KQ2E12, KQ3E12, KQ4E6, KQ4aE6, KQ4bE6, KQ4cE6, KQ5E6.


Appendix D. Excluded Studies


Appendix D. Excluded Studies


233. Ohkuchi A, Iwasaki R, Ojima T, et al. Increase in systolic blood pressure of > or = 30 mm Hg and/or diastolic blood pressure of > or = 15 mm Hg during pregnancy: is it pathologic? Hypertens Pregnancy 2003;22(3):275-85. PMID: 14572364. KQ1E6, KQ1aE6, KQ2E12, KQ3E12, KQ4E6, KQ4aE6, KQ4bE6, KQ4cE6, KQ5E6.


Appendix D. Excluded Studies


Appendix D. Excluded Studies


263. Poon LC, Kametas N, Bonino S, et al. Urine albumin concentration and albumin-to-creatinine ratio at 11(+0) to 13(+6) weeks in the prediction of pre-eclampsia. BJOG 2008 Jun;115(7):866-73. PMID: 18485165. KQ1E12, KQ1aE12, KQ2E13, KQ3E4, KQ4E12, KQ4aE12, KQ4bE12, KQ4cE12, KQ5E12.


Appendix D. Excluded Studies


310. Shaarawy M, Salem ME. The clinical value of microtransferrinuria and microalbuminuria in the prediction of preeclampsia. Clin Chem Lab Med 2001 Jan;39(1):29-34. PMID: 11256797. KQ1E1, KQ1aE1, KQ2E12, KQ3E12, KQ4aE1, KQ4bE1, KQ4cE1, KQ5E1.


Appendix D. Excluded Studies


328. Sukur YE, Yalcin I, Kahraman K, et al. Predictive value of 3+ spot urinary protein value measured by dipstick in hypertensive pregnant patients. Hypertens Pregnancy 2013 May;32(2):139-45. PMID: 23725079. KQ1E1, KQ1aE1, KQ2E12, KQ3E12, KQ4E1, KQ4aE1, KQ4bE1, KQ4cE1, KQ5E1.

Appendix D. Excluded Studies


Appendix D. Excluded Studies


KQ1E3, KQ1aE3, KQ2E12, KQ3E12, KQ4E4, KQ4bE4, KQ4cE4, KQ5E4.


KQ1E2, KQ1aE2, KQ2E12, KQ3E12, KQ4E2, KQ4aE3, KQ4bE2, KQ4cE2, KQ5E2.


KQ1E3, KQ1aE3, KQ2E12, KQ3E12, KQ4E4, KQ4aE4, KQ4bE4, KQ4cE4, KQ5E4.


KQ1E4, KQ1aE4, KQ2E12, KQ3E12, KQ4E4, KQ4aE4, KQ4bE4, KQ4cE4, KQ5E4.


KQ1E3, KQ1aE3, KQ2E12, KQ3E12, KQ4E4, KQ4aE4, KQ4bE4, KQ4cE4, KQ5E4.


KQ1E3, KQ1aE3, KQ2E12, KQ3E12, KQ4E3, KQ4aE3, KQ4bE3, KQ4cE3, KQ5E3.


KQ1E3, KQ1aE3, KQ2E12, KQ3E12, KQ4E4, KQ4bE4, KQ4cE4, KQ5E4.


KQ1E4, KQ1aE4, KQ2E12, KQ3E12, KQ4E4, KQ4aE4, KQ4bE4, KQ4cE4, KQ5E4.


KQ1E2, KQ1aE2, KQ2E12, KQ3E12, KQ4E2, KQ4bE2, KQ4cE2, KQ5E2.


KQ1E12, KQ1aE12, KQ2E6, KQ3E6, KQ4E12, KQ4aE12, KQ4bE12, KQ4cE12, KQ5E12.


KQ1E3, KQ1aE3, KQ2E12, KQ3E12, KQ4E12, KQ4aE12, KQ4bE12, KQ4cE12, KQ5E12.


KQ1E1, KQ1aE1, KQ2E12, KQ3E12, KQ4E1, KQ4aE1, KQ4bE1, KQ4cE1, KQ5E1.


KQ1E3, KQ1aE3, KQ2E12, KQ3E12, KQ4E3, KQ4aE3, KQ4bE3, KQ4cE3, KQ5E3.
Appendix D. Excluded Studies


377. Zanello M, Sekizawa A, Purwosunu Y, et al. Circulating mRNA for the PLAC1 gene as a second trimester marker (14-18 weeks' gestation) in the screening for late preeclampsia. Fetal Diagn Ther 2014;36(3):196-201. PMID: 25138310. KQ1E1, KQ1aE1, KQ2E1, KQ3E1, KQ4E1, KQ4aE1, KQ4bE1, KQ4cE1, KQ5E1.

1. The source of data for risk prediction models is ideally from cohort, nested case-control or case-cohort studies. Prospective cohort data are less prone to missing data and predictors than are retrospective cohort data. RCT and registry databases also are sometimes used for risk prediction model validation, but they are subject to greater risk of bias relative to cohort and nested designs.

2. The size of the study and incidence of preeclampsia determines the number of outcome events available for the model development process. For model internal validation, a rule of thumb is that there should be ten events for every predictor in the model. The low prevalence of preterm preeclampsia poses model development challenges. Models developed with this risk of bias are less likely to perform well in external validation.

3. Discrimination and calibration characterize the model performance, and without information on both it is difficult to determine the degree to which a model correctly classifies those who ultimately develop the condition of interest, and to compare models as new predictors are added and removed during model development and validation. As stated in the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies (CHARMS) checklist, the absence of either calibration or discrimination hinders the full appraisal of models.

4. The discriminatory capacity of predictive tools is often represented with the Area under the Receiver Operator Curve (AUC) which in general terms depicts the degree to which a test correctly classifies those with and without the condition of interest. Values of the AUC range from 0.5 to 1.0, with 0.5 representing discrimination no better than a coin toss and 1.0 representing perfect classification. Guidelines for interpretation of discriminatory power are as follows: AUC values below 0.70 are generally considered as inadequate, from 0.70 to 0.79 as adequate, and 0.80 or higher as good to excellent. The AUC, or c statistic, is the most commonly reported measure of performance, specifically model discrimination, and provides a general basis for evaluating the models we identified. An AUC of at least 0.70 is often cited as the lowest threshold for a clinically useful test.

5. Based on the CHARMS checklist, for external validity studies the risk of bias is greater when the research team and setting engaged for the external validation study is independent from the team that developed the model and conducted internal validation.

6. Models developed following established methods for reducing the risk of bias are more likely to be reproducible in similar populations. Information on differences in the derivation study population and the validation study population is important for interpreting differences in the performance of a model in external validation. If calibration and discrimination are not upheld in a population with similar characteristics to the model development study, overfitting and risk of bias in the model development process likely account for the lack of reproducibility. When the performance is not upheld in a population with different prevalence of important predictors and the outcome, then a model is not considered transportable – and further adjustments for application in different populations are likely necessary.

7. Clarity in reporting of the risk prediction algorithm and how it would be applied in the clinical setting are important to appraisal of models for potential clinical use recommendations. Clear reporting on the risk algorithm and its calculation, including the necessary variables, coding rules and risk cut-offs are practical requirements for model application.
### Appendix F Table 1. Study Design Characteristics of Included Studies (Key Questions 1a, 3, 4a, and 5)

<table>
<thead>
<tr>
<th>Study &amp; Quality</th>
<th>Study Design</th>
<th>Study Period</th>
<th>Country</th>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
<th>n</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Key Questions 1a and 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McDuffie, 1996</td>
<td>RCT</td>
<td>1992-1994 (enrollment)</td>
<td>United States</td>
<td>Healthy women in first trimester of pregnancy who presented for an intake visit</td>
<td>Age &lt;18 or &gt;39 years of age; had completed 13 weeks of gestation; had a past (e.g., preterm delivery, preterm labor, abruptio placenta, severe PE, classical c-section, gestational diabetes, incompetent cervix, uterine anomaly, diethylstilbestrol exposure, isoimmunization, &gt;1 2nd trimester abortion, fetal anomaly, or SGA neonate) or current (e.g., multiple gestation, assisted pregnancy, large leiomyoma) high-risk obstetrical condition; had a current medical condition (e.g., diabetes, chronic HTN, drug or alcohol abuse, any ongoing medical or psychiatric illness requiring treatment or monitoring); were non-English speaking; or were planning to change insurance carriers during pregnancy</td>
<td>2764</td>
<td>IG: Scheduled of less perinatal visits (nine visits)</td>
</tr>
<tr>
<td></td>
<td>Fair</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CG: Routine number of prenatal care visits (14 visits)</td>
</tr>
<tr>
<td></td>
<td>Fair</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CG: Routine urine screening</td>
</tr>
<tr>
<td><strong>Key Question 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simeone, 2015</td>
<td>Prospective cohort study</td>
<td>July 2013-February 2014 (recruitment)</td>
<td>Spain and Italy</td>
<td>Consecutive women if they had a singleton pregnancy, absence of psychiatric disorder, and low-risk Down syndrome screening (&lt;1/240); high-risk woman matched with the next visited low-risk woman in the 1st trimester screening unit.</td>
<td>NR</td>
<td>255</td>
<td>IG: High-risk women</td>
</tr>
<tr>
<td></td>
<td>Fair</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CG: Low-risk women</td>
</tr>
</tbody>
</table>
Table 1. Study Design Characteristics of Included Studies (Key Questions 1a, 3, 4a, and 5)

<table>
<thead>
<tr>
<th>Study &amp; Quality</th>
<th>Study Design</th>
<th>Study Period</th>
<th>Country</th>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
<th>n</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tun, 2012 110  Fair</td>
<td>Diagnostic Accuracy</td>
<td>July 1, 2010 to December 31, 2011</td>
<td>United States</td>
<td>Pregnant women age 18-55 years, &gt;20 weeks gestation who were admitted to the Lehigh Valley Health Network antepartum unit, undergoing 24-hour urine collection for diagnosis and/or management of PE</td>
<td>Prepregnancy renal disease (24-hour urine protein ≥300 mg), clinical indication for delivery at time of admission, outside maternal or gestational age, did not speak English, did not give informed consent, or had been enrolled previously in the study</td>
<td>90</td>
<td>P/Cr spot test</td>
</tr>
<tr>
<td>Stout, 2013 110  Fair</td>
<td>Diagnostic Accuracy</td>
<td>2005-2007</td>
<td>United States</td>
<td>Pregnant women after 20 weeks gestation who underwent evaluation for suspected PE</td>
<td>Proteinuria (i.e., ≥300 mg in 24 hours) before 20 weeks of gestation</td>
<td>356</td>
<td>P/Cr spot ratio</td>
</tr>
<tr>
<td>Wheeler, 2007 113  Fair</td>
<td>Diagnostic Accuracy</td>
<td>December 2000 to July 2002</td>
<td>United States</td>
<td>Pregnant women admitted to the Greenville Hospital System University Medical Center for evaluation of preeclampsia, which was in general new-onset persistent hypertension, worsening hypertension, or proteinuria; new-onset hypertension was systolic blood pressure &gt;140 mm Hg or diastolic blood pressure &gt;90 mm Hg after 20 weeks' gestation in a previously normotensive patient, whereas worsening hypertension was an increase in blood pressure from baseline taken before 20 weeks' gestation in a patient with baseline blood pressure of &lt;140/90 mm Hg</td>
<td>Women who had bacteriuria on microscopy or who were on more than 24 hours’ bed rest, because of a potential poor correlation between spot P:C and 24 hour urine collections for protein after prolonged recumbency</td>
<td>126</td>
<td>P/Cr spot ratio</td>
</tr>
<tr>
<td>Young, 1996 114  Fair</td>
<td>Diagnostic Accuracy</td>
<td>June 1992-June 1993; December 1993-August 1994</td>
<td>United States</td>
<td>Ambulatory women suspected of having PIH (i.e., BP &gt;140/90 mm Hg, SBP &gt;30 mm Hg above baseline or DBP 15 mm Hg above baseline)</td>
<td>Previously diagnosed as having PIH and had been placed on long-term bed rest at home or strict bed rest in the hospital for more than 36 hours</td>
<td>45</td>
<td>P/Cr spot test</td>
</tr>
<tr>
<td>Kyle, 2008 116  Fair</td>
<td>Diagnostic Accuracy</td>
<td>NR</td>
<td>New Zealand</td>
<td>Pregnant women attending a high-risk obstetric medical antenatal clinic if they had automated dipstick analysis of ≥1+ of new-onset proteinuria on a midstream urine specimen; group of negative or trace proteinuria women on automated</td>
<td>Positive urine culture for UTI, underlying proteinuric renal disease, those w/ diabetes w/ an abnormal albumin/Cr in the first trimester.</td>
<td>150</td>
<td>P/Cr spot test</td>
</tr>
<tr>
<td>Study &amp; Quality</td>
<td>Study Design</td>
<td>Study Period</td>
<td>Country</td>
<td>Inclusion Criteria</td>
<td>Exclusion Criteria</td>
<td>n</td>
<td>Group</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------</td>
<td>--------------</td>
<td>---------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----</td>
<td>--------------------</td>
</tr>
<tr>
<td>Sethuram, 2011&lt;sup&gt;108&lt;/sup&gt; Fair</td>
<td>Diagnostic Accuracy</td>
<td>January-September 2007</td>
<td>United Kingdom</td>
<td>Women w/pregnancies of &gt;24 weeks gestation undergoing evaluation for PE (i.e., BP &gt;140/90 mm Hg and urine protein &gt;1+ on dipstick); women w/secondary PE to HTN or GDM also included</td>
<td>UTI, renal pathologies, delivered before they could complete their 24-hour urine collection</td>
<td>32</td>
<td>P/Cr spot test</td>
</tr>
<tr>
<td>Bhide, 2015&lt;sup&gt;103&lt;/sup&gt; Fair</td>
<td>Diagnostic Accuracy</td>
<td>NR</td>
<td>United Kingdom</td>
<td>Pregnant women w/suspected PE (SBP ≥140 mm Hg and/or DBP ≥90 mm Hg in the antenatal clinic or in the community when checked by midwives or doctors and a spot urine dipstick proteinuria of 1+ or more)</td>
<td>Duration of &gt;72 hours between commencing the 24-hour urine collection and spot P/Cr ratio sample being taken; 24-hour urinary Cr excretion &lt;97 uM/kg (to avoid undercollection) or &gt;220 uM/kg (to avoid overcollection over 24 hours); known or suspected UTI; documented proteinuria at booking; delivered elsewhere</td>
<td>117</td>
<td>P/Cr spot test</td>
</tr>
<tr>
<td>Lamontagne, 2014&lt;sup&gt;107&lt;/sup&gt; Good</td>
<td>Diagnostic Accuracy</td>
<td>November 2005-November 2006</td>
<td>Canada</td>
<td>Women ≥18 years, were in their second or third trimester of pregnancy, were ambulatory, and had an indication for a 24-hour urine collection as part of an investigation for PE</td>
<td>Serum Cr &gt;150 umol/L, history of renal transplant, pre-existing microalbuminuria or proteinuria, macroscopic hematuria, or known UTI. Specimens discarded if UTI, hematuria, vaginal bleeding, rupture of membranes, labor, or induction of labor occurred during 24-hour collection; incomplete urine defined as Cr &lt;10 mmol/kg of pre-pregnancy weight</td>
<td>91</td>
<td>P/Cr spot ratio</td>
</tr>
<tr>
<td>Verdonk, 2014&lt;sup&gt;111&lt;/sup&gt; Good</td>
<td>Diagnostic Accuracy</td>
<td>NR, but 2 year study</td>
<td>Netherlands</td>
<td>Women w/suspected PE (i.e., de novo HTN w/BP ≥140/90 mm Hg after 20 weeks of gestation and a urine protein dipstick reading ≥1+) admitted as inpatients; pregnant women w/chronic HTN who developed new onset proteinuria after mid-gestation also invited to participate</td>
<td>UTI, pre-existing proteinuria, having a delivery before the 24-hour urinary collection was completed</td>
<td>105</td>
<td>P/Cr spot test</td>
</tr>
</tbody>
</table>
## Appendix F Table 1. Study Design Characteristics of Included Studies (Key Questions 1a, 3, 4a, and 5)

<table>
<thead>
<tr>
<th>Study &amp; Quality</th>
<th>Study Design</th>
<th>Study Period</th>
<th>Country</th>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
<th>n</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dwyer, 2008&lt;sup&gt;110&lt;/sup&gt; Good</td>
<td>Diagnostic Accuracy</td>
<td>September 2002-March 2004</td>
<td>United States</td>
<td>All pregnant women being evaluated for PE, regardless of the alerting signs or symptoms, suspected severity or comorbid conditions</td>
<td>If urinalysis contained &gt;10 white blood cells per high-power field, if a catheter was not used after membrane rupture or if an outpatient 24-hour collection was incomplete (complete collection defined as total Cr &gt;1000 mg [850 mg for obese women] or total Cr 13 mg per kg body weight); in general, if a 24-hour urine protein not done, urinalysis not done, P/Cr ratio not done.</td>
<td>116</td>
<td>P/Cr spot test</td>
</tr>
<tr>
<td>Durnwald, 2003&lt;sup&gt;104&lt;/sup&gt; Fair</td>
<td>Diagnostic Accuracy</td>
<td>January 2001 to July 2002</td>
<td>United States</td>
<td>Women w/pregnancies ≥24 weeks gestation who were undergoing evaluation for suspected PE (≥1 of the following findings: HTN, edema, new-onset proteinuria on urinary dipstick)</td>
<td>Concurrent diagnosis of chronic HTN, DM, or preexisting renal disease; documented preexisting proteinuria (1+ urine dipstick on initial office visit)</td>
<td>220</td>
<td>P/Cr spot ratio</td>
</tr>
<tr>
<td>Valdes, 2015&lt;sup&gt;119&lt;/sup&gt; Fair</td>
<td>Diagnostic Accuracy</td>
<td>January-December 2012</td>
<td>Chile</td>
<td>Women admitted to the Maternity Unit of the Hospital Clinico Universidad de Chile w/a diagnosis of pregnancy hypertensive disorder</td>
<td>Twin pregnancies, fetal birth defects (w/antenatal diagnosis or diagnosed during the neonatal period), chronic nephropathies, maternal age &lt;18 years, gestational age &lt;20 weeks, incomplete demographic and perinatal data</td>
<td>72</td>
<td>P/Cr spot test</td>
</tr>
<tr>
<td>Waugh, 2005&lt;sup&gt;101&lt;/sup&gt; Good</td>
<td>Diagnostic Accuracy</td>
<td>October 2000-June 2001</td>
<td>United Kingdom</td>
<td>Pregnant women &gt;20 weeks gestation referred for assessment of de novo HTN occurring for the first time to the day assessment unit if they had an estimated and sustained SBP &gt;140 mm Hg or a DBP of &gt;90 mm Hg using mercury sphygmomanometry</td>
<td>Pre-existing HTN</td>
<td>171</td>
<td>Dipstick - Microalbumustix (visual) Dipstick - Clinitek Microalbumin (automated) DCA 2000 - POC test</td>
</tr>
<tr>
<td>Kyle, 2008&lt;sup&gt;106&lt;/sup&gt; Fair</td>
<td>Diagnostic Accuracy</td>
<td>NR</td>
<td>New Zealand</td>
<td>Pregnant women attending a high-risk obstetric medical antenatal clinic if they had automated dipstick analysis of ≥1+ of new-onset proteinuria on a midstream urine specimen; group of negative or trace proteinuria women on automated</td>
<td>Positive urine culture for UTI, underlying proteinuric renal disease, those w/diabetes w/an abnormal albumin/Cr in the first trimester.</td>
<td>150</td>
<td>Albumin/Cr spot test</td>
</tr>
</tbody>
</table>

*Albumin:creatinine spot urine tests*
## Appendix F Table 1. Study Design Characteristics of Included Studies (Key Questions 1a, 3, 4a, and 5)

<table>
<thead>
<tr>
<th>Study &amp; Quality</th>
<th>Study Design</th>
<th>Study Period</th>
<th>Country</th>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
<th>n</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>dipstick also recruited; women attending clinic included those w/ pre-existing HTN, pre-existing DM, gestational DM, renal disease, connective tissue disorders and other high-risk obstetric and fetal conditions.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein dipsticks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waugh, 200112</td>
<td>Diagnostic Accuracy</td>
<td>NR</td>
<td>United Kingdom</td>
<td>Pregnant women who presented either for assessment of HTN in pregnancy or as referrals to the antenatal HTN clinic, over 20 weeks gestation</td>
<td>NR</td>
<td>197</td>
<td>Dipstick</td>
</tr>
<tr>
<td>Kyle, 2008106</td>
<td>Diagnostic Accuracy</td>
<td>NR</td>
<td>New Zealand</td>
<td>Pregnant women attending a high-risk obstetric medical antenatal clinic if they had automated dipstick analysis of ≥1+ of new-onset proteinuria on a midstream urine specimen; group of negative or trace proteinuria women on automated dipstick also recruited; women attending clinic included those w/ pre-existing HTN, pre-existing DM, gestational DM, renal disease, connective tissue disorders and other high-risk obstetric and fetal conditions.</td>
<td>Positive urine culture for UTI, underlying proteinuric renal disease, those w/ diabetes w/ an abnormal albumin/Cr in the first trimester.</td>
<td>150</td>
<td>Dipstick</td>
</tr>
<tr>
<td>Waugh, 2005114</td>
<td>Diagnostic Accuracy</td>
<td>October 2000-June 2001</td>
<td>United Kingdom</td>
<td>Pregnant women &gt;20 weeks gestation referred for assessment of de novo HTN occurring for the first time to the day assessment unit if they had an estimated and sustained SBP &gt;140 mm Hg or a DBP of &gt;90 mm Hg using mercury sphygmomanometry</td>
<td>Pre-existing HTN</td>
<td>171</td>
<td>Dipstick - Multistix 8SG (automated) Dipstick - Multistix 8SG (visual) DCA 2000 - POC test</td>
</tr>
</tbody>
</table>
### Appendix F Table 1. Study Design Characteristics of Included Studies (Key Questions 1a, 3, 4a, and 5)

<table>
<thead>
<tr>
<th>Study &amp; Quality</th>
<th>Study Design</th>
<th>Study Period</th>
<th>Country</th>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
<th>n</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dwyer, 2008</td>
<td>Diagnostic Accuracy</td>
<td>September 2002-March 2004</td>
<td>United States</td>
<td>All pregnant women being evaluated for PE, regardless of the alerting signs or symptoms, suspected severity or comorbid conditions</td>
<td>If urinalysis contained &gt;10 white blood cells per high-power field, if a catheter was not used after membrane rupture or if an outpatient 24-hour collection was incomplete (complete collection defined as total Cr &gt;1000 mg [850 mg for obese women] or total Cr 13 mg per kg body weight); in general, if a 24-hour urine protein not done, urinalysis not done, P/Cr ratio not done.</td>
<td>116</td>
<td>P/Cr automated dipstick</td>
</tr>
</tbody>
</table>

**Abbreviations:** BP = blood pressure; CG = control group; Cr = creatinine; DBP = diastolic blood pressure; DM = diabetes mellitus; GDM = gestational diabetes mellitus; HTN = hypertension; IG = intervention group; kg = kilogram(s); L = liter(s); mg = milligram(s); mm Hg = milligrams of mercury; mmol = millimole(s); NR = not reported; P = protein; PE = preeclampsia; PIH = pregnancy-induced hypertension; POC = point of care; RCT = randomized controlled trial; SBP = systolic blood pressure; SGA = small for gestational age; UTI = urinary tract infection; umol = micromole(s); w/ = with.
### Appendix F Table 2. Baseline Characteristics of Included Studies (Key Questions 1a, 3, 4a, and 5)

<table>
<thead>
<tr>
<th>Study &amp; Quality</th>
<th>Maternal age, years (range)</th>
<th>Race/Ethnicity (%)</th>
<th>Gestational age, weeks (range)</th>
<th>HTN (%)</th>
<th>Diabetes (%)</th>
<th>BMI (kg/m^2)</th>
<th>Nulliparity (%)</th>
<th>Singleton pregnancy (%)</th>
<th>Suspected PE (%) or Significant Proteinuria (%)</th>
<th>Inpatient (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Key Questions 1a and 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Key Question 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Key Question 4a (sorted by type of test)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein:creatinine spot urine tests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tun, 2012 Fair</td>
<td>29 (19-42)</td>
<td>White: 78.9 Black: 5.6 Hispanic: 2.2 Asian: 3.3</td>
<td>33.8 (24.0-39.0)</td>
<td>Chronic: 22.2 Gestational: 24.4</td>
<td>Pre-existing: 5.6 Gestational: 15.6</td>
<td>34.1</td>
<td>45.6</td>
<td>87.8</td>
<td>31.4</td>
<td>100</td>
</tr>
<tr>
<td>Kyle, 2008 Fair</td>
<td>NR (NR)</td>
<td>White: 90.7 Black: NR Hispanic: NR Asian: NR</td>
<td>34.0 (20.1-39.7)</td>
<td>Chronic: 12.7 Gestational: NR</td>
<td>Pre-existing: 4.7 Gestational: 9.3</td>
<td>32.5</td>
<td>36.7</td>
<td>92.0</td>
<td>8.7</td>
<td>0</td>
</tr>
<tr>
<td>Study &amp; Quality</td>
<td>Maternal age, years (range)</td>
<td>Race/Ethnicity (%)</td>
<td>Gestational age, weeks (range)</td>
<td>HTN (%)</td>
<td>Diabetes (%)</td>
<td>BMI (kg/m^2)</td>
<td>Nulliparity (%)</td>
<td>Singleton pregnancy (%)</td>
<td>Suspected PE (%) or Significant Proteinuria (%)</td>
<td>Inpatient (%)</td>
</tr>
<tr>
<td>----------------</td>
<td>-----------------------------</td>
<td>-------------------</td>
<td>-------------------------------</td>
<td>---------</td>
<td>--------------</td>
<td>--------------</td>
<td>-----------------</td>
<td>--------------------------</td>
<td>---------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Good</td>
<td></td>
<td>Black: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hispanic: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Asian: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fair</td>
<td></td>
<td>Black: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hispanic: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Asian: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lamontagne, 2014</td>
<td>31.8 (≥ 18)</td>
<td>White: 62.6</td>
<td>32.3 (NR)</td>
<td>Chronic: 37.4 Gestational: NR</td>
<td>Pre-existing: 8.8 Gestational: 16.5</td>
<td>29.0</td>
<td>46.2</td>
<td>95.6</td>
<td>47.3</td>
<td>62.6</td>
</tr>
<tr>
<td>Good</td>
<td></td>
<td>Black: 28.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hispanic: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Asian: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verdonk, 2014</td>
<td>31 (IQR 28-34)</td>
<td>White: NR</td>
<td>31 (IQR 29-35)</td>
<td>Chronic: NR Gestational: NR</td>
<td>Pre-existing: NR Gestational: NR</td>
<td>NR</td>
<td>56.2</td>
<td>NR</td>
<td>69.5</td>
<td>100</td>
</tr>
<tr>
<td>Good</td>
<td></td>
<td>Black: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hispanic: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Asian: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dwyer, 2008</td>
<td>30.8 (NR)</td>
<td>White: 40.5</td>
<td>NR (NR)</td>
<td>Chronic: 22.4 Gestational: NR</td>
<td>Pre-existing: 6.9 Gestational: NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>48.3</td>
<td>100</td>
</tr>
<tr>
<td>Good</td>
<td></td>
<td>Black: 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hispanic: 31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Asian: 16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fair</td>
<td></td>
<td>Black: 43.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hispanic: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Asian: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valdes, 2015</td>
<td>30.5 (NR)</td>
<td>White: NR</td>
<td>NR (NR)</td>
<td>Chronic: 100 Gestational: NR</td>
<td>Pre-existing: NR Gestational: NR</td>
<td>32.2</td>
<td>44.4</td>
<td>100</td>
<td>58.3</td>
<td>100</td>
</tr>
<tr>
<td>Fair</td>
<td></td>
<td>Black: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hispanic: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Asian: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Albumin:creatinine spot urine tests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waugh, 2005</td>
<td>29 (19-40)</td>
<td>White: 97.7</td>
<td>NR (NR)</td>
<td>Chronic: NR Gestational: 100</td>
<td>Pre-existing: NR Gestational: NR</td>
<td>NR</td>
<td>58</td>
<td>NR</td>
<td>45.0</td>
<td>NR</td>
</tr>
<tr>
<td>Good</td>
<td></td>
<td>Black: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hispanic: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Asian: 2.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kyle, 2008</td>
<td>NR (NR)</td>
<td>White: 90.7</td>
<td>34.0 (20.1-39.7)</td>
<td>Chronic: 12.7 Gestational: NR</td>
<td>Pre-existing: 4.7 Gestational: 9.3</td>
<td>32.5</td>
<td>36.7</td>
<td>92.0</td>
<td>8.7</td>
<td>0</td>
</tr>
<tr>
<td>Fair</td>
<td></td>
<td>Black: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hispanic: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Asian: NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Screening for Preeclampsia 118 Kaiser Permanente Research Affiliates EPC
## Appendix F Table 2. Baseline Characteristics of Included Studies (Key Questions 1a, 3, 4a, and 5)

<table>
<thead>
<tr>
<th>Study &amp; Quality</th>
<th>Maternal age, years (range)</th>
<th>Race/Ethnicity (%)</th>
<th>Gestational age, weeks (range)</th>
<th>HTN (%)</th>
<th>Diabetes (%)</th>
<th>BMI (kg/m²)</th>
<th>Nulliparity (%)</th>
<th>Singleton pregnancy (%)</th>
<th>Suspected PE (%) or Significant Proteinuria (%)</th>
<th>Inpatient (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Protein dipsticks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kyle, 2008&lt;sup&gt;106&lt;/sup&gt; Fair</td>
<td>NR (NR)</td>
<td>White: 90.7 Black: NR Hispanic: NR Asian: NR</td>
<td>34.0 (20.1-39.7)</td>
<td>Chronic: 12.7 Gestational: NR</td>
<td>Pre-existing: 4.7 Gestational: 9.3</td>
<td>32.5</td>
<td>36.7</td>
<td>92.0</td>
<td>8.7</td>
<td>0</td>
</tr>
<tr>
<td>Waugh, 2005&lt;sup&gt;104&lt;/sup&gt; Good</td>
<td>29 (19-40)</td>
<td>White: 97.7 Black: NR Hispanic: NR Asian: 2.3</td>
<td>NR (NR)</td>
<td>Chronic: NR Gestational: 100</td>
<td>Pre-existing: NR Gestational: NR</td>
<td>NR</td>
<td>58</td>
<td>NR</td>
<td>45.0</td>
<td>NR</td>
</tr>
<tr>
<td>Dwyer, 2008&lt;sup&gt;105&lt;/sup&gt; Good</td>
<td>30.8 (NR)</td>
<td>White: 40.5 Black: 12 Hispanic: 31 Asian: 16</td>
<td>NR (NR)</td>
<td>Chronic: 22.4 Gestational: NR</td>
<td>Pre-existing: 6.9 Gestational: NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>48.3</td>
<td>100</td>
</tr>
</tbody>
</table>

*For Key Question 4a only, all pregnant women had suspected pre-eclampsia; the data in this column reflects those with significant proteinuria according to the 24-hour urine collection (reference standard)

**Abbreviations**: BMI = body mass index; HTN = hypertension; kg = kilogram(s); m = meter(s); NR = not reported; PE = preeclampsia.
### Table 3. Intervention Characteristics of Included Trials (Key Questions 1a, 3, and 5)

<table>
<thead>
<tr>
<th>Study &amp; Quality</th>
<th>Group</th>
<th>n</th>
<th>Group Name</th>
<th>Description</th>
<th>Provider</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Key Questions 1a and 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McDuffie, 1996</td>
<td>IG</td>
<td>1165</td>
<td>Nine perinatal visits</td>
<td>Experimental schedule consisted of visits at 8, 12, 16, 24, 28, 32, 36, 38 and 40 weeks (total of nine visits) with ongoing risk assessment. For parous women, a telephone call was scheduled at 12 weeks instead of a visit. Since not all women presented at 8 weeks of gestation: 7-8 weeks seen according to schedule; 9-10 weeks asked to return at 14 weeks and have blood drawn at 16 weeks; 11-12 weeks asked to return at 16 weeks. Visits ranged from 45 minutes (intake) to 10-15 minutes with practitioners or physicians.</td>
<td>OBGYN, NPs, PA, or nurse midwives</td>
</tr>
<tr>
<td></td>
<td>CG</td>
<td>1163</td>
<td>Usual care</td>
<td>Routine clinical schedule consisted of visits every 4 weeks from 8 to 28 weeks, then every 2 weeks until 36 weeks and weekly thereafter (total of 14 visits) with ongoing risk assessment. Since not all women presented at 8 weeks of gestation: 7-8 weeks seen according to schedule; 9-10 weeks asked to return at 14 weeks and have blood drawn at 16 weeks; 11-12 weeks asked to return at 16 weeks. Visits ranged from 45 minutes (intake) to 10-15 minutes with practitioners or physicians.</td>
<td>OBGYN, NPs, PA, or nurse midwives</td>
</tr>
<tr>
<td>Rhode, 2007</td>
<td>IG</td>
<td>1251</td>
<td>Indicated urine testing</td>
<td>Women who were enrolled and delivered on or after August 15, 2002. Indicated urine testing was substituted for routine urine screening; a urine specimen was obtained prior to the patient’s visits with a care provider whenever any of the criteria were present (first prenatal visit, patient complaint of symptoms of UTI, patient complaint of severe vomiting, weight loss ≥ 0.9 kg since previous visit, SBP ≥ 140 mm Hg, DBP ≥ 90 mm Hg, or any pregnancy requiring periodic urine testing such as chronic HTN and renal disease). Chemical reagent strips use for all urine tests, mean (SD) number of test strips: 1.4 (1.3), range, 0-16. Indications for urine test also reported.</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>CG</td>
<td>1160</td>
<td>Routine urine screening</td>
<td>Women who were enrolled and delivered prior to August 15, 2002. First prenatal visit included routine urine screening, a urine culture and blood pressure determination; urine screening and BP determination were included in all subsequent visits. Chemical reagent strips use for all urine tests, mean (SD) number of test strips: 7.8 (3.4), range, 0-19.</td>
<td>NR</td>
</tr>
<tr>
<td><strong>Key Question 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simeone, 2015</td>
<td>IG</td>
<td>140</td>
<td>High-risk women</td>
<td>At screening, women were informed about PE and its consequences by trained midwives; counseling concerned the concept of risk, the parental expectations of the screening, and the consequences of a positive test. High-risk women underwent a followup protocol consisting of daily aspirin (150 mg) from the day of screening until 36 weeks gestation and second trimester ultrasound at 20-22 weeks including UIA Doppler velocimetry. Dietary calcium intake was evaluated in each case and when &lt; 3 daily products/day, a supplementation w/ 1 g/day was recommended. Pts w/ normal second trimester UtA mean pulsatility index (&lt;95th percentile) underwent a subsequent ultrasound and blood/urine test at 28 and 32 weeks, whereas those w/ abnormal results underwent same evaluation at 24, 28, 32 and 36 weeks.</td>
<td>Trained midwives</td>
</tr>
<tr>
<td></td>
<td>CG</td>
<td>140</td>
<td>Low-risk women</td>
<td>At screening, women were informed about PE and its consequences by trained midwives; counseling concerned the concept of risk, the parental expectations of the screening, and the consequences of a positive test.</td>
<td>Trained midwives</td>
</tr>
</tbody>
</table>

**Abbreviations:** BP = blood pressure; CG = control group; DBP = diastolic blood pressure; g = gram(s); HTN = hypertension; IG = intervention group; mg = milligram(s); mm Hg = millimeters of mercury; kg = kilogram(s); NP = nurse practitioner; NR = not reported; OBGYN = obstetrician/gynecologist; PA = physician’s assistant; PE = preeclampsia; pts = participants; SD = standard deviation; SBP = systolic blood pressure; UIA = uterine artery; UTI = urinary tract infection; w/ = with.
## Appendix F Table 4. Results of Diagnostic Accuracy Studies (Key Question 4a): Significant Proteinuria

<table>
<thead>
<tr>
<th>Study &amp; Quality</th>
<th>Threshold*</th>
<th>n</th>
<th>tp</th>
<th>fp</th>
<th>fn</th>
<th>tn</th>
<th>Sensitivity†</th>
<th>Specificity†</th>
<th>PPV†</th>
<th>NPV†</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P:Cr Spot Urine Tests</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tun, 2012 Fair</td>
<td>17.0 mg/mmol</td>
<td>86</td>
<td>24</td>
<td>30</td>
<td>3</td>
<td>29</td>
<td>89</td>
<td>49</td>
<td>32</td>
<td>91</td>
</tr>
<tr>
<td>Stout, 2013 Fair</td>
<td>&gt; 9.0 mg/mmol</td>
<td>356</td>
<td>140</td>
<td>180</td>
<td>4</td>
<td>32</td>
<td>97</td>
<td>15</td>
<td>44</td>
<td>86</td>
</tr>
<tr>
<td>Stout, 2013 Fair</td>
<td>&gt; 13.6 mg/mmol</td>
<td>356</td>
<td>130</td>
<td>129</td>
<td>14</td>
<td>83</td>
<td>90</td>
<td>39</td>
<td>50</td>
<td>86</td>
</tr>
<tr>
<td>Stout, 2013 Fair</td>
<td>&gt; 21.5 mg/mmol</td>
<td>356</td>
<td>112</td>
<td>64</td>
<td>32</td>
<td>148</td>
<td>78</td>
<td>70</td>
<td>64</td>
<td>82</td>
</tr>
<tr>
<td>Stout, 2013 Fair</td>
<td>&gt; 45.2 mg/mmol</td>
<td>356</td>
<td>72</td>
<td>17</td>
<td>72</td>
<td>195</td>
<td>50</td>
<td>64</td>
<td>81</td>
<td>74</td>
</tr>
<tr>
<td>Stout, 2013 Fair</td>
<td>&gt; 50.9 mg/mmol</td>
<td>356</td>
<td>68</td>
<td>8</td>
<td>76</td>
<td>204</td>
<td>47</td>
<td>96</td>
<td>88</td>
<td>73</td>
</tr>
<tr>
<td>Stout, 2013 Fair</td>
<td>&gt; 134.5 mg/mmol</td>
<td>356</td>
<td>44</td>
<td>2</td>
<td>98</td>
<td>212</td>
<td>31</td>
<td>&gt;99</td>
<td>96</td>
<td>67</td>
</tr>
<tr>
<td>Wheeler, 2007 Fair</td>
<td>23.7 mg/mmol</td>
<td>126</td>
<td>59</td>
<td>13</td>
<td>9</td>
<td>45</td>
<td>86.8</td>
<td>77.6</td>
<td>81.9</td>
<td>83.3</td>
</tr>
<tr>
<td>Young, 1996 Fair</td>
<td>≥ 5.7 mg/mmol</td>
<td>45</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>100</td>
<td>0</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Young, 1996 Fair</td>
<td>≥ 11.3 mg/mmol</td>
<td>45</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>96</td>
<td>18</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Young, 1996 Fair</td>
<td>≥ 17.0 mg/mmol</td>
<td>45</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>91</td>
<td>41</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Young, 1996 Fair</td>
<td>≥ 22.6 mg/mmol</td>
<td>45</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>78</td>
<td>59</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Young, 1996 Fair</td>
<td>≥ 28.3 mg/mmol</td>
<td>45</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>65</td>
<td>82</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Young, 1996 Fair</td>
<td>≥ 33.9 mg/mmol</td>
<td>45</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>57</td>
<td>100</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Young, 1996 Fair</td>
<td>≥ 39.6 mg/mmol</td>
<td>45</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>48</td>
<td>100</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Kyle, 2008 Fair</td>
<td>≥ 30 mg/mmol</td>
<td>150</td>
<td>12</td>
<td>4</td>
<td>1</td>
<td>133</td>
<td>92.3</td>
<td>97.1</td>
<td>75.0</td>
<td>99.3</td>
</tr>
<tr>
<td>Sethuram, 2011 Fair</td>
<td>&gt; 30 mg/mmol</td>
<td>32</td>
<td>25</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>83</td>
<td>92</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Bhide, 2015 Fair</td>
<td>≥ 18 mg/mmol</td>
<td>117</td>
<td>76</td>
<td>17</td>
<td>0</td>
<td>24</td>
<td>100</td>
<td>59.6</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Bhide, 2015 Fair</td>
<td>≥ 30 mg/mmol</td>
<td>117</td>
<td>72</td>
<td>9</td>
<td>4</td>
<td>32</td>
<td>95</td>
<td>78</td>
<td>78</td>
<td>NR</td>
</tr>
<tr>
<td>Bhide, 2015 Fair</td>
<td>≥ 60 mg/mmol</td>
<td>117</td>
<td>57</td>
<td>0</td>
<td>19</td>
<td>41</td>
<td>74.5</td>
<td>100</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Lamontagne, 2014 Good</td>
<td>≥ 30 mg/mmol§</td>
<td>26</td>
<td>7</td>
<td>1</td>
<td>5</td>
<td>13</td>
<td>58</td>
<td>93</td>
<td>88</td>
<td>72</td>
</tr>
<tr>
<td>Lamontagne, 2014 Good</td>
<td>≥ 30 mg/mmol¶</td>
<td>91</td>
<td>36</td>
<td>1</td>
<td>5</td>
<td>47</td>
<td>81</td>
<td>98</td>
<td>97</td>
<td>86</td>
</tr>
<tr>
<td>Verdonk, 2014 Good</td>
<td>≥ 30 mg/mmol¶</td>
<td>104</td>
<td>68</td>
<td>7</td>
<td>4</td>
<td>25</td>
<td>94</td>
<td>78</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Verdonk, 2014 Good</td>
<td>≥ 35.4 mg/mmol¶</td>
<td>104</td>
<td>67</td>
<td>3</td>
<td>5</td>
<td>29</td>
<td>93</td>
<td>91</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Verdonk, 2014 Good</td>
<td>≥ 30 mg/mmol**</td>
<td>105</td>
<td>65</td>
<td>8</td>
<td>8</td>
<td>24</td>
<td>89</td>
<td>75</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Verdonk, 2014 Good</td>
<td>≥ 50.4 mg/mmol**</td>
<td>105</td>
<td>59</td>
<td>1</td>
<td>14</td>
<td>31</td>
<td>81</td>
<td>97</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Verdonk, 2014 Good</td>
<td>≥ 30 mg/mmol††</td>
<td>105</td>
<td>70</td>
<td>7</td>
<td>3</td>
<td>25</td>
<td>96</td>
<td>78</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Verdonk, 2014 Good</td>
<td>≥ 42.4 mg/mmol††</td>
<td>105</td>
<td>69</td>
<td>4</td>
<td>4</td>
<td>28</td>
<td>95</td>
<td>88</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Dwyer, 2008 Good</td>
<td>≥ 17.0 mg/mmol</td>
<td>116</td>
<td>54</td>
<td>28</td>
<td>2</td>
<td>32</td>
<td>96</td>
<td>53</td>
<td>66</td>
<td>94</td>
</tr>
<tr>
<td>Dwyer, 2008 Good</td>
<td>≥ 19.2 mg/mmol</td>
<td>116</td>
<td>51</td>
<td>25</td>
<td>5</td>
<td>35</td>
<td>91</td>
<td>58</td>
<td>67</td>
<td>88</td>
</tr>
<tr>
<td>Dwyer, 2008 Good</td>
<td>≥ 21.5 mg/mmol</td>
<td>116</td>
<td>50</td>
<td>18</td>
<td>6</td>
<td>42</td>
<td>89</td>
<td>70</td>
<td>74</td>
<td>88</td>
</tr>
<tr>
<td>Dwyer, 2008 Good</td>
<td>≥ 27.1 mg/mmol</td>
<td>116</td>
<td>41</td>
<td>8</td>
<td>15</td>
<td>52</td>
<td>73</td>
<td>87</td>
<td>84</td>
<td>78</td>
</tr>
</tbody>
</table>
### Appendix F Table 4. Results of Diagnostic Accuracy Studies (Key Question 4a): Significant Proteinuria

<table>
<thead>
<tr>
<th>Study &amp; Quality</th>
<th>Threshold*</th>
<th>n</th>
<th>tp</th>
<th>fp</th>
<th>fn</th>
<th>tn</th>
<th>Sensitivity†</th>
<th>Specificity†</th>
<th>PPV†</th>
<th>NPV†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Durnwald, 2003 ³</td>
<td>≥ 31.6 mg/mmol</td>
<td>116</td>
<td>37</td>
<td>3</td>
<td>19</td>
<td>57</td>
<td>66</td>
<td>95</td>
<td>93</td>
<td>75</td>
</tr>
<tr>
<td>Fair</td>
<td>≥ 44.1 mg/mmol</td>
<td>116</td>
<td>31</td>
<td>0</td>
<td>25</td>
<td>60</td>
<td>55</td>
<td>100</td>
<td>100</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>≥ 17.0 mg/mmol</td>
<td>220</td>
<td>156</td>
<td>35</td>
<td>12</td>
<td>17</td>
<td>92.9</td>
<td>32.7</td>
<td>81.7</td>
<td>58.6</td>
</tr>
<tr>
<td></td>
<td>≥ 22.6 mg/mmol</td>
<td>220</td>
<td>152</td>
<td>27</td>
<td>16</td>
<td>25</td>
<td>90.5</td>
<td>48.1</td>
<td>84.9</td>
<td>61.0</td>
</tr>
<tr>
<td></td>
<td>≥ 33.9 mg/mmol</td>
<td>220</td>
<td>136</td>
<td>23</td>
<td>32</td>
<td>29</td>
<td>61.0</td>
<td>55.8</td>
<td>85.5</td>
<td>47.5</td>
</tr>
<tr>
<td></td>
<td>≥ 44.1 mg/mmol</td>
<td>220</td>
<td>122</td>
<td>14</td>
<td>46</td>
<td>38</td>
<td>72.6</td>
<td>73.1</td>
<td>89.7</td>
<td>45.2</td>
</tr>
<tr>
<td></td>
<td>≥ 45.2 mg/mmol</td>
<td>220</td>
<td>120</td>
<td>12</td>
<td>48</td>
<td>40</td>
<td>71.4</td>
<td>76.9</td>
<td>90.9</td>
<td>45.5</td>
</tr>
<tr>
<td></td>
<td>≥ 56.5 mg/mmol</td>
<td>220</td>
<td>106</td>
<td>9</td>
<td>62</td>
<td>43</td>
<td>63.1</td>
<td>82.7</td>
<td>92.2</td>
<td>41.0</td>
</tr>
<tr>
<td>Valdes, 2015 ³ ⁴</td>
<td>≥ 40.7 mg/mmol</td>
<td>72</td>
<td>31</td>
<td>3</td>
<td>11</td>
<td>27</td>
<td>73</td>
<td>91</td>
<td>95</td>
<td>62</td>
</tr>
<tr>
<td>A:Cr Spot Urine Tests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waugh, 2005 ³ ⁵</td>
<td>2.0 mg/mmol</td>
<td></td>
<td>171</td>
<td>72</td>
<td>6</td>
<td>5</td>
<td>88</td>
<td>94</td>
<td>94</td>
<td>92</td>
</tr>
<tr>
<td>Good</td>
<td>≥ 3.4 mg/mmol† †</td>
<td>171</td>
<td>45</td>
<td>16</td>
<td>32</td>
<td>78</td>
<td>58</td>
<td>83</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>≥ 3.4 mg/mmol† †</td>
<td>171</td>
<td>38</td>
<td>16</td>
<td>39</td>
<td>78</td>
<td>49</td>
<td>83</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Kyle, 2008 ³ ⁶ ³</td>
<td>≥ 2.0 mg/mmol</td>
<td>150</td>
<td>13</td>
<td>44</td>
<td>0</td>
<td>93</td>
<td>100</td>
<td>67.9</td>
<td>22.8</td>
<td>100</td>
</tr>
<tr>
<td>Fair</td>
<td>≥ 3.5 mg/mmol</td>
<td>150</td>
<td>13</td>
<td>17</td>
<td>0</td>
<td>120</td>
<td>100</td>
<td>87.6</td>
<td>43.3</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>≥ 8.0 mg/mmol</td>
<td>150</td>
<td>13</td>
<td>5</td>
<td>0</td>
<td>132</td>
<td>100</td>
<td>96.4</td>
<td>72.2</td>
<td>100</td>
</tr>
<tr>
<td>Dipstick</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waugh, 2001 ³ ⁶ ³</td>
<td>≥ 1+† †</td>
<td>197</td>
<td>31</td>
<td>1</td>
<td>76</td>
<td>89</td>
<td>29.0</td>
<td>98.9</td>
<td>96.9</td>
<td>53.9</td>
</tr>
<tr>
<td>Fair</td>
<td>≥ 1+† †</td>
<td>197</td>
<td>30</td>
<td>2</td>
<td>13</td>
<td>152</td>
<td>69.8</td>
<td>98.7</td>
<td>93.8</td>
<td>92.1</td>
</tr>
<tr>
<td></td>
<td>≥ 1+</td>
<td>197</td>
<td>31</td>
<td>1</td>
<td>107</td>
<td>58</td>
<td>22.5</td>
<td>98.3</td>
<td>96.9</td>
<td>35.2</td>
</tr>
<tr>
<td></td>
<td>≥ 1+</td>
<td>197</td>
<td>28</td>
<td>4</td>
<td>21</td>
<td>144</td>
<td>57.1</td>
<td>97.3</td>
<td>87.5</td>
<td>87.3</td>
</tr>
<tr>
<td>Kyle, 2008 ³ ⁶ ³</td>
<td>1+</td>
<td>150</td>
<td>13</td>
<td>87</td>
<td>0</td>
<td>50</td>
<td>100</td>
<td>36.5</td>
<td>13.0</td>
<td>100</td>
</tr>
<tr>
<td>Fair</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waugh, 2005 ³ ⁶ ³</td>
<td>1+† †</td>
<td>171</td>
<td>63</td>
<td>18</td>
<td>14</td>
<td>76</td>
<td>82</td>
<td>81</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Good</td>
<td>1+***</td>
<td>171</td>
<td>39</td>
<td>21</td>
<td>38</td>
<td>73</td>
<td>51</td>
<td>78</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Dwyer, 2008 ³ ⁶ ³</td>
<td>≥ 1+</td>
<td>116</td>
<td>23</td>
<td>0</td>
<td>33</td>
<td>60</td>
<td>41</td>
<td>100</td>
<td>100</td>
<td>65</td>
</tr>
<tr>
<td>Good</td>
<td>≥ 2+</td>
<td>116</td>
<td>13</td>
<td>0</td>
<td>43</td>
<td>60</td>
<td>23</td>
<td>100</td>
<td>100</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>≥ 3+</td>
<td>116</td>
<td>7</td>
<td>0</td>
<td>49</td>
<td>60</td>
<td>11</td>
<td>100</td>
<td>100</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>≥ negative</td>
<td>116</td>
<td>56</td>
<td>60</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>48</td>
<td>NC</td>
</tr>
</tbody>
</table>

*We converted all urine excretion ratios to g/mmol by converting any values to mg/g and multiplying this value by 0.113
†Study-reported
‡Reference standard used concentration, not excretion
§First morning void sample
∥Other void samples
¶Evening sample (5 PM)
**Morning sample (8 AM)
††Noon sample (12 PM)
‡‡Clinitek Microalbumin (automated) dipstick (ACR)
Table 4. Results of Diagnostic Accuracy Studies (Key Question 4a): Significant Proteinuria

<table>
<thead>
<tr>
<th>Test Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>§§ Microalbumin (visual) dipstick (ACR)</td>
</tr>
<tr>
<td>DCA 2000 - POC test</td>
</tr>
<tr>
<td>¶¶ Multistix 8SG (automated) (ACR) dipstick</td>
</tr>
<tr>
<td>*** Multistix 8SG (visual) (ACR) dipstick</td>
</tr>
</tbody>
</table>

Abbreviations: A = albumin; Cr = creatinine; fn = false negatives; fp = false positives; mg = milligram(s); mmol = millimole(s); P = protein; tn = true negatives; tp = true positives.
### Appendix F Table 5. Results of Diagnostic Accuracy Studies (Key Question 4a): Severe Proteinuria

<table>
<thead>
<tr>
<th>Study &amp; Quality</th>
<th>Index Test</th>
<th>Threshold*</th>
<th>n</th>
<th>tp</th>
<th>fp</th>
<th>fn</th>
<th>tn</th>
<th>Sensitivity†</th>
<th>Specificity†</th>
<th>PPV†</th>
<th>NPV†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Durnwald, 2003</td>
<td>P/Cr spot</td>
<td>≥ 214.7 mg/mmol</td>
<td>220</td>
<td>15</td>
<td>34</td>
<td>3</td>
<td>168</td>
<td>83.7</td>
<td>83.3</td>
<td>31.3</td>
<td>98.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 565.0 mg/mmol</td>
<td>220</td>
<td>13</td>
<td>8</td>
<td>5</td>
<td>194</td>
<td>72.2</td>
<td>96.0</td>
<td>61.9</td>
<td>97.5</td>
</tr>
<tr>
<td>Dwyer, 2008</td>
<td>Dipstick</td>
<td>≥ 1+</td>
<td>116</td>
<td>3</td>
<td>18</td>
<td>0</td>
<td>95</td>
<td>100</td>
<td>83</td>
<td>14</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 2+</td>
<td>116</td>
<td>3</td>
<td>9</td>
<td>0</td>
<td>104</td>
<td>100</td>
<td>92</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 3+</td>
<td>116</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>111</td>
<td>100</td>
<td>98</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ negative</td>
<td>116</td>
<td>3</td>
<td>113</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>3</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td>P/Cr spot</td>
<td>≥ 226.0 mg/mmol</td>
<td>116</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>108</td>
<td>100</td>
<td>96</td>
<td>38</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 339.0 mg/mmol</td>
<td>116</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>110</td>
<td>100</td>
<td>97</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 452.0 mg/mmol</td>
<td>116</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>111</td>
<td>100</td>
<td>98</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 565.0 mg/mmol</td>
<td>116</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>113</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 1528.9 mg/mmol</td>
<td>116</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>113</td>
<td>67</td>
<td>100</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>Valdes, 2015</td>
<td>P/Cr spot test</td>
<td>≥ 517.5 mg/mmol</td>
<td>72</td>
<td>17</td>
<td>3</td>
<td>0</td>
<td>52</td>
<td>100</td>
<td>95</td>
<td>73</td>
<td>100</td>
</tr>
<tr>
<td>Wheeler, 2007</td>
<td>P/Cr spot</td>
<td>≥ 52.0 mg/mmol</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>87.5</td>
<td>82.4</td>
<td>53.8</td>
<td>96.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 92.7 mg/mmol</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>100</td>
<td>94.8</td>
<td>62.5</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 339.0 mg/mmol</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

*We converted all urine exertion ratios to g/mmol by converting any values to mg/g and multiplying this value by 0.113
†Study-reported

**Abbreviations:** A = albumun; Cr = creatinine; fn = false negatives; fp = false positives; mg = milligram(s); mmol = millimole(s); P = protein; tn = true negatives; tp = true positives.
Appendix F Figure 1. Diagnostic Accuracy of Protein:Creatinine Spot Urine Tests (Key Question 4a), Sorted by Threshold

Abbreviations: CI = confidence interval; df = degrees of freedom; mg = milligram(s); mmol = millimole(s); N = number of participants.
Appendix F Figure 2. Summary Receiver Operative Characteristics of Protein:Creatinine Spot Urine Tests (Key Question 4a)

Abbreviations: AUC = area under curve; sens = sensitivity; spec = specificity; SROC = summary receiver operating characteristics

Appendix F Figure 3. Diagnostic Accuracy of Protein:Creatinine Spot Urine Tests (Key Question 4a), Sorted by Study Population

Abbreviations: CI = confidence interval; df = degrees of freedom; mg = milligram(s); mmol = millimole(s); N = number of participants.
### Appendix G Table 1. Test Performance Characteristics of Externally Validated Preeclampsia Risk Prediction Models

| External Validation Studies Models | Oliveira 2014[^9]  
Baltimore, Maryland, USA  
**PE requiring delivery:**  
<34 weeks gestation (early)  
>34 weeks gestation (late) | Park 2013[^10]  
Sydney, Australia  
**PE requiring delivery:**  
<34 weeks gestation (early) | Skrastad 2014[^11]  
Trondheim, Norway  
**PE requiring delivery:**  
<37 weeks gestation (early)  
<42 weeks gestation (any)  
>34 weeks gestation (late) | Farina 2011[^12]  
Bologna, Italy  
**PE diagnosis:**  
>34 weeks gestation |
|---|---|---|---|---|
| Caradeux 2013[^13]  
Santiago and Valdivia, Chile  
[clinical history and Doppler][^5]  
Women presenting for 11 to 14 weeks ultrasound in pregnancy  
N = 627  
% early PE = 1.4 (9 cases) | N = 2,962  
% early PE = 1.0 (30 cases)  
C* = NR  
AUC = 0.69 (0.59-0.80)  
early  
DR = 30 (CI not reported) [history, Doppler] | | |
| Odibo 2011[^14]  
St. Louis, Missouri, USA  
[clinical history, serum, and Doppler][^5]  
Women presenting for first trimester aneuploidy screening in pregnancy  
N = 452  
% early PE = 2.7 (12 cases) | N = 871  
% early PE = 1.2 (10 cases)  
C = NR  
AUC = 0.86 (0.73-0.99)  
early  
DR = 80 (CI not reported) [history, serum, Doppler] | | |
| Parra-Cordero 2013[^15]  
Santiago, Chile  
[clinical history, serum, and Doppler][^5]  
Asymptomatic women undergoing routine Doppler scan at 11+0 to 13+6 weeks in pregnancy  
N = 5,367  
% early PE = 0.3 (17 cases) | N = 1,558  
% early PE = 1.1 (17 cases)  
C = NR  
AUC = 0.70 (0.58-0.83)  
early  
DR = 29 (CI not reported) [history, serum, Doppler] | | |

[^9]: Screening for Preeclampsia  
[^10]: Kaiser Permanente Research Affiliates EPC
Table 1. Test Performance Characteristics of Externally Validated Preeclampsia Risk Prediction Models

| External Validation Studies Models | Oliveira 2014<sup>a</sup>  
Baltimore, Maryland, USA  
PE requiring delivery:  
<34 weeks gestation (early)  
>34 weeks gestation (late)  
| Park 2013<sup>b</sup>  
Sydney, Australia  
PE requiring delivery:  
<34 weeks gestation (early)  
>34 weeks gestation (late)  
| Skrastad 2014<sup>c</sup>  
Trondheim, Norway  
PE requiring delivery:  
<37 weeks gestation (early)  
<42 weeks gestation (any)  
>34 weeks gestation (late)  
| Farina 2011<sup>d</sup>  
Bologna, Italy  
PE diagnosis:  
>34 weeks gestation  
|
|———|———|———|———|
| Parra-Cordero 2013<sup>e</sup>  
Santiago, Chile  
[clinical history, serum, and Doppler] §  
Asymptomatic women undergoing routine Doppler scan at 11+0 to 13+6 weeks in pregnancy  
N = 5,367  
% late PE = 1.0 (53 cases)  
| N = 1,558  
% late PE = 5.0 (78 cases)  
C = NR  
AUC = 0.61 (0.55-0.68) late  
DR = 18 (CI not reported) late  
[history, serum, Doppler]  
| N = 2,962  
% early PE = 1.0 (30 cases)  
C = NR  
AUC = 0.78 (0.69-0.88) early  
DR = 53 (CI not reported) early  
[history, Doppler]  
| N = 2,962  
% early PE = 1.0 (30 cases)  
C = NR  
AUC = 0.77 (0.67-0.86) early  
DR = 43 (CI not reported) early  
[history, Doppler]  
| N = 2,833  
% late PE = 4.1 (116 cases)  
C = NR  
AUC = 0.69 (0.64-0.75) late  
DR = 31 (CI not reported) late  
[history, Doppler]  
<p>|</p>
<table>
<thead>
<tr>
<th>External Validation Studies Models</th>
<th>Oliveira 2014(^{22})</th>
<th>Park 2013(^{20})</th>
<th>Skrastad 2014(^{10})</th>
<th>Farina 2011(^{28})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baltimoe, Maryland, USA</td>
<td>Sydney, Australia</td>
<td>Trondheim, Norway</td>
<td>Bologna, Italy</td>
</tr>
<tr>
<td></td>
<td>PE requiring delivery: &lt;34 weeks gestation (early)</td>
<td>PE requiring delivery: &lt;34 weeks gestation (early)</td>
<td>PE requiring delivery: &lt;37 weeks gestation (early)</td>
<td>PE diagnosis: &gt;34 weeks gestation</td>
</tr>
<tr>
<td>Poon 2010(^{12})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>London, United Kingdom</td>
<td>[clinical history, serum, and Doppler] §</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women presenting for first routine hospital visit in pregnancy</td>
<td>N = 2,833</td>
<td>N = 3,014</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% early PE = 1.0 (29 cases)</td>
<td>C = NR</td>
<td>C = NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUC = 0.80 (0.71-0.89) early</td>
<td>AUC = 0.93 (0.92-0.94) early</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DR = 52 (CI not reported)</td>
<td>DR = 91.7 (61.5-98.6) early</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[history, serum, Doppler]</td>
<td>[history, serum, Doppler]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poon 2010(^{12})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>London, United Kingdom</td>
<td>[clinical history, serum, and Doppler] §</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women presenting for first routine hospital visit in pregnancy</td>
<td>N = 8,061</td>
<td>N = 541</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% early PE = 0.5 (37 cases)</td>
<td>% preterm PE (&lt;37 wks) = 0.9 (5 cases)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akolekar 2013(^{12})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>London and Gillingham, United Kingdom</td>
<td>[clinical history, serum, and Doppler] §</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women with singleton pregnancies presenting for first trimester aneuploidy screening</td>
<td>N = 58,884</td>
<td>N = 541</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% preterm PE (&lt;37 wks) = 1.0 (568 cases)</td>
<td>% any PE = 3.9 (21 cases)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akolekar 2013(^{12})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>London and Gillingham, United Kingdom</td>
<td>[clinical history, serum, and Doppler] §</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women with singleton pregnancies presenting for first trimester aneuploidy screening</td>
<td>N = 58,884</td>
<td>N = 541</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% any PE = 2.4 (1,426 cases)</td>
<td>% any PE = 3.9 (21 cases)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix G Table 1. Test Performance Characteristics of Externally Validated Preeclampsia Risk Prediction Models

<table>
<thead>
<tr>
<th>External Validation Studies Models</th>
<th>Oliveira 2014&lt;sup&gt;92&lt;/sup&gt; &lt;br&gt;Baltimore, Maryland, USA &lt;br&gt;PE requiring delivery: &lt;br&gt;&lt;34 weeks gestation (early) &lt;br&gt;&gt; 34 weeks gestation (late)</th>
<th>Park 2013&lt;sup&gt;100&lt;/sup&gt; &lt;br&gt;Sydney, Australia &lt;br&gt;PE requiring delivery: &lt;br&gt;&lt;34 weeks gestation (early) &lt;br&gt;&lt;42 weeks gestation (early) &lt;br&gt;&gt;34 weeks gestation (late)</th>
<th>Skrastad 2014&lt;sup&gt;101&lt;/sup&gt; &lt;br&gt;Trondheim, Norway &lt;br&gt;PE requiring delivery: &lt;br&gt;&lt;37 weeks gestation (early) &lt;br&gt;&lt;42 weeks gestation (any) &lt;br&gt;&gt;34 weeks gestation (late)</th>
<th>Farina 2011&lt;sup&gt;98&lt;/sup&gt; &lt;br&gt;Bologna, Italy &lt;br&gt;PE requiring delivery: &gt;34 weeks gestation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PREDICTOR algorithm &lt;br&gt;[clinical history, serum, and Doppler]&lt;sup&gt;§&lt;/sup&gt; Proprietary model, derived from multiple studies, not reported in detail</td>
<td>N = 541&lt;br&gt;% any PE = 3.9 (21 cases)</td>
<td>C = NR&lt;br&gt;AUC = 0.74 (0.63-0.84) any&lt;br&gt;DR = 30.0 (11.9-54.3) any</td>
<td>[history, serum, Doppler]</td>
<td></td>
</tr>
<tr>
<td>Onwudiwe 2008&lt;sup&gt;129&lt;/sup&gt; &lt;br&gt;London, United Kingdom &lt;br&gt;[clinical history and Doppler]&lt;sup&gt;§&lt;/sup&gt; Women with singleton pregnancies presenting for routine antenatal care</td>
<td>N = 3,347&lt;br&gt;% late PE = 2.3 (78 cases)</td>
<td>N = 554&lt;br&gt;% late PE = 7.0 (39 cases)</td>
<td>[history, Doppler]</td>
<td></td>
</tr>
<tr>
<td>Plasencia 2008&lt;sup&gt;130&lt;/sup&gt; &lt;br&gt;London, United Kingdom &lt;br&gt;[clinical history and Doppler]&lt;sup&gt;§&lt;/sup&gt; Women with singleton pregnancies presenting for routine antenatal care</td>
<td>N = 3,107&lt;br&gt;% late PE = 2.3 (71 cases)</td>
<td>N = 554&lt;br&gt;% late PE = 7.0 (39 cases)</td>
<td>[history, Doppler]</td>
<td></td>
</tr>
<tr>
<td>Plasencia 2007&lt;sup&gt;127&lt;/sup&gt; &lt;br&gt;London, United Kingdom &lt;br&gt;[clinical history only]&lt;sup&gt;§&lt;/sup&gt; Women with singleton pregnancies presenting for routine assessment of risk for chromosomal abnormalities</td>
<td>N = 6,015&lt;br&gt;% any PE = 1.8 (107 cases)</td>
<td>N = 554&lt;br&gt;% late PE = 7.0 (39 cases)</td>
<td>[history]</td>
<td></td>
</tr>
</tbody>
</table>

<sup><br>§</sup> Women with singleton pregnancies presenting for routine antenatal care.
### External Validation Studies

<table>
<thead>
<tr>
<th>Models</th>
<th>Studies</th>
<th>Location</th>
<th>Early PE</th>
<th>Late PE</th>
<th>N</th>
<th>% Late PE</th>
<th>C</th>
<th>AUC (CI)</th>
<th>DR (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oliveira 2014&lt;sup&gt;99&lt;/sup&gt;</td>
<td>Baltimore, Maryland, USA</td>
<td>PE requiring delivery: &lt;34 weeks gestation (early)</td>
<td>&gt;34 weeks gestation (late)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Park 2013&lt;sup&gt;100&lt;/sup&gt;</td>
<td>Sydney, Australia</td>
<td>PE requiring delivery: &lt;34 weeks gestation (early)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skrastad 2014&lt;sup&gt;101&lt;/sup&gt;</td>
<td>Trondheim, Norway</td>
<td>PE requiring delivery: &lt;37 weeks gestation (early)</td>
<td>&lt;42 weeks gestation (any)</td>
<td>&gt;34 weeks gestation (late)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farina 2011&lt;sup&gt;102&lt;/sup&gt;</td>
<td>Bologna, Italy</td>
<td>PE diagnosis: &gt;34 weeks gestation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poon 2010&lt;sup&gt;124&lt;/sup&gt;&lt;sup&gt;¶&lt;/sup&gt;</td>
<td>London, United Kingdom</td>
<td>[clinical history and Doppler]&lt;sup&gt;§&lt;/sup&gt;</td>
<td>Women presenting for first routine hospital visit in pregnancy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poon 2009&lt;sup&gt;120&lt;/sup&gt;</td>
<td>London, United Kingdom</td>
<td>[clinical history and serum]&lt;sup&gt;§&lt;/sup&gt;</td>
<td>Women with singleton pregnancies presenting for routine assessment of risk for chromosomal abnormalities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Notes:

- Calibration = A model performance measure that refers to how well predicted risks compare to observed outcomes preferably evaluated graphically by calibration plots and supplemented by a formal statistical test, the Hosmer-Lemeshow test for logistic regression and its equivalent for Cox regression.<sup>83</sup>
- Area under the curve = A test performance statistic (equivalent to the c-statistic) used to assess discrimination, a model performance measure that refers to how well a model differentiates between those with and without the outcome.<sup>83</sup>
- Detection rate = The percent of cases correctly classified based on a predefined false positive probability threshold.<sup>83</sup> Detection rates for preeclampsia in this table are based on a fixed 10% false positive rate, which was the most commonly reported.
- Clinical history includes maternal characteristics, medical history, and routine clinical measures (e.g., family history of PE, personal history of PE, parity, race/ethnicity, prior preterm labor, CHTN, diabetes, thrombophilia, renal disease, mode of conception, smoking status, DBP, SBP, weight, MAP, BMI). Serum markers include PAPP-A, PIGF, PP13. Doppler ultrasound includes UtA-PI. Some variables are expressed as adjusted Multiples of the Median.
- Derived from the Fetal Medicine Foundation Algorithm
- Leona Poon and Kypros Nicolaides are authors on many of the model development papers for preeclampsia risk assessment. They also hold patents related to the use of biological markers for prenatal screening. The Fetal Medicine Foundation (founded by Nicolaides) and PerkinElmer are assignees on several patents on prenatal screening held by Poon and Nicolaides.
- Clinical history algorithm described in Poon 2010<sup>65</sup> and Poon 2009<sup>120</sup>
- ** This model was used in the study included for KQ3 to evaluate potential harms of risk assessment.

#### Abbreviations:

- AUC = area under the curve; BMI = body mass index; CHTN = chronic hypertension; CI = confidence intervals; DBP = diastolic blood pressure; DR = detection rate; MAP = mean arterial pressure; NR = not reported; PE = preeclampsia; PIGF = placental growth factor; SBP = systolic blood pressure; wks = weeks.