

# Protocol: Screening for asymptomatic bacteriuria in pregnancy

December 1, 2016

ERSC Project Lead Investigator: Lisa Hartling

ERSC Project Staff: Aileen Wingert, Jennifer Pillay, Robin Featherstone (MLIS), and Ben Vandermeer (Statistician)

Suggested citation: Wingert A, Pillay J, Featherstone R, Vandermeer B, and Hartling L. Screening for asymptomatic bacteriuria in pregnancy: protocol for evidence review. Edmonton, AB; Evidence Review Synthesis Centre. Available at: <http://canadiantaskforce.ca/ctfphc-guidelines/overview/>

## Author Contributions

AW and JP drafted the protocol and RF developed the search strategy and provided text for the protocol. AW, JP, and LH contributed to discussions with the CTFPHC and PHAC on the scope for this work. LH and BV critically reviewed the protocol. All of the authors approved the final version of this protocol.

## Acknowledgements

We would like to acknowledge the contribution of the following individuals for their clinical and methodological input and/or their review of the protocol:

CTFPHC Asymptomatic Bacteriuria Working Group: Ainsley Moore (Chair), Roland Grad, Stéphane Groulx, Kevin Pottie, Brett Thombs, and Marcello Tonelli

PHAC Global Health & Guidance Division Staff: Marion Doull, Alejandra Jaramillo Garcia, Susan Courage, Nicki Sims-Jones

## Declaration of funding

Funding for this protocol and evidence review is provided by the Public Health Agency of Canada. This funding will support the collection of the data, data management, analyses, and writing of the protocol and the technical report.

# **Section I. Background and Purpose**

## **Asymptomatic Bacteriuria in Pregnancy**

Asymptomatic bacteriuria (ASB) - synonymous with asymptomatic urinary tract infection (UTI) - signifies a significant quantitative count of bacteria in the urine without symptoms of a lower (acute cystitis) or upper urinary tract (acute pyelonephritis) infection (1, 2). There is a 2-10% prevalence of ASB in premenopausal, ambulatory women (1), but due to anatomical and physiological changes (e.g., urinary stasis - difficulty emptying the bladder due to extended accumulation of urine) to the urinary tract in pregnancy there are theoretical reasons to suspect a greater chance of progression to symptomatic UTI and other pregnancy complications (e.g., maternal kidney infection, preterm delivery) (1, 3). Numerous risk factors for ASB in pregnancy have been identified, with low socioeconomic status, parity, a history of recurrent UTI, diabetes, and anatomical abnormalities of the urinary tract most cited (1, 2, 4).

### **Consequences of Untreated Bacteriuria in Pregnancy and Rationale for Review of Screening**

There is a potentially greater risk in pregnant women compared to other populations for ASB developing into pyelonephritis (upper urinary tract infection) (3) with its associated inflammation of the renal parenchyma, calices and pelvis (5), although controversy exists. There is significant heterogeneity in reports of the incidence of pyelonephritis in untreated ASB during pregnancy. Some reports suggest low incidences of 1% or less after the introduction of screening and treatment for ASB and 4% or higher before the era of screening and treatment of ASB in pregnancy. Historical reports prior to 1966 indicated up to 40% of pregnant women with ASB developed pyelonephritis. These higher rates were before modern obstetrical care; however, these numbers continue to be cited in current systematic reviews (4) and guidelines (6) of ASB in pregnancy (1, 7). Furthermore, whether there is evidence to support a causal link between ASB and pyelonephritis in contemporary practice is uncertain.

There is an association between clinical signs of pyelonephritis and maternal respiratory insufficiency, septicemia, renal dysfunction and anemia, as well as evidence of a 20 to 50% higher incidence of preterm birth and low birth weight (4, 8). However, controversy exists over the direct link between ASB, pyelonephritis, and adverse perinatal outcomes (e.g., whether ASB affects pregnancy and neonatal outcomes solely through pyelonephritis or also other mechanisms) (2, 4), and also about whether treatment of ASB will reduce the risk of such adverse outcomes. A 2015 Cochrane review (4) found that antibiotic treatment for ASB in pregnancy may greatly reduce the incidence of pyelonephritis, preterm birth, and low birth weight babies. However, the authors' confidence in the findings were low due to poor quality evidence. A preliminary search identified a recent cohort study (9) with an embedded RCT, which found no statistically significant difference between ASB-positive women who were untreated or placebo-treated compared to ASB-negative women in terms of both pyelonephritis and preterm birth (6/208 [2.9%] vs 77/4035 [1.9%]; adjusted odds ratio [OR] 1.5, 95% CI 0.6–3.5).

Although the direct link between pyelonephritis and adverse perinatal outcomes may not be easily resolved (4), some main issues to examine include: 1) which, if any, screening tests and

methods (e.g., collection methods, timing) are most accurate, and; 2) whether screening of all pregnant women and treatment for positive cases is effective (9). The effectiveness of screening for reducing risk of pyelonephritis and neonatal and maternal complications need to be examined in an era of modern obstetrical care.

### **Issues to Consider for Screening Tests**

Significant bacteriuria is usually defined by the presence of at least  $10^5$  colony-forming units (CFU) per mL of urine of a single uropathogen, in two consecutive clean-catch specimens (4, 7). Acceptable thresholds and repetitions considered positive for bacteriuria in pregnancy may vary in practice. The quantitative urine culture is considered to be the gold standard for accurate detection of ASB. However, it is costlier, more labor intensive and more time-consuming compared with other rapid urine screening tests (urinalysis, dipstick nitrite tests) which reportedly have lower sensitivity<sup>1</sup> (1, 2). A preliminary search for recent literature identified a systematic review of onsite tests (point-of-care tests that are widely available in resource-limited settings) compared with urine culture that concluded specificity<sup>2</sup> was high overall but sensitivity was low and therefore onsite tests were not reliable in detecting pregnant women with ASB (10). There is no consistent recommendation for urine specimen collection in pregnancy (clean-catch with or without perineal cleansing) or optimal timing and frequency of screening tests or follow-up cultures (2). It is unclear whether universal screening (with subsequent treatment) for ASB confers benefits, and whether available screening tests for ASB are comparable to the current gold standard (urine culture) for identifying bacteriuric patients. The standard urine culture protocol is evolving with the testing of emerging techniques that may improve the detection of uropathogens (11, 12). However, at this time, urine culture is considered the reference standard. Resource needs for screening may be an important factor to consider. For example, an economic analysis indicated that screening with a dipstick and providing screen positive women treatment with antibiotics remained cost-beneficial for reducing pyelonephritis when prevalence of ASB is <2% or when the proportion of patients with ASB who develop pyelonephritis dropped to 10%, but the cost-benefit was not seen for culture diagnostics where the absolute clinical benefit was shown to be reduced (13).

<sup>1</sup>Sensitivity is a diagnostic test accuracy outcome that refers to how well a test correctly identifies individuals with a disease/condition; <sup>2</sup>Specificity is a diagnostic test accuracy outcome that refers to how well a test correctly identifies individuals without a disease/condition.

### **Issues to Consider for Harms of Screening**

Patients may have preferences for avoiding harms due to screening and treatment in asymptomatic conditions (e.g., test anxiety/distress). Although the harms from screening tests may be considered minimal, harms from antibiotic treatment need to be considered when making decisions about screening practices for ASB in pregnancy. Some sources have outlined concerns with incidence and reporting on adverse effects of antibiotic treatment for ASB, UTIs, or antibiotic use in general during pregnancy (2, 4, 14). Some trials evaluating treatment versus no treatment/placebo of ASB in pregnancy have been critiqued for poorly reporting harms (4), such that making judgments on the net balance of benefits and harms may be difficult. The significance of the expected side effects from a short course of antibiotics may be small although increasingly there are concerns about the effect of antibiotics on the human microbiome and the immune system. Antimicrobial resistance has certainly made the selection of an antibiotic for an individual woman more difficult (4). Additionally, patients may have preferences for avoiding treatment harms in asymptomatic conditions that need to be considered.

The goal of this review is to determine the effectiveness of screening for ASB among pregnant women. This evidence synthesis will inform recommendations on screening for ASB made by the Canadian Task Force for Preventive Healthcare (CTFPHC). As part of the guideline development process, the CTFPHC will also engage organizational stakeholders and peer-reviewers to gather information on key implementation considerations, such as strategies to help address potential health inequities and any concerns about the acceptability and feasibility of the guideline.

## **Section II. Recommendations in Other Guidelines and Current Practice**

### **Canadian Organizations**

The Society of Obstetricians and Gynecologists of Canada (SOGC), concerned over maternal and perinatal risks associated with ASB, recommends to treat single-strain colony counts of  $10^5$  CFU/mL (or  $10^8$  CFU/L) or greater with appropriate antibiotics during pregnancy to prevent adverse outcomes such as pyelonephritis and preterm birth (15). They support a single quantitative culture in any trimester as sufficient and recommend re-treatment with sensitivities for women with recurrent bacteriuria although they do not make recommendations for timing or frequency of re-testing. Similar recommendations apply when group B streptococcal (GBS) bacteria is detected in the urine during screening in pregnancy; separate recommendations (not relevant for this review) are made for screening and treating GBS (at any colony counts) at time of labour or rupture of membranes for prevention of early-onset neonatal GBS disease.

### **Guidelines from International Organizations**

The U.S. Preventive Services Task Force 2008 guideline (16) on screening of ASB in adults recommends all pregnant women be screened at 12 to 16 weeks' gestation (or first prenatal visit) for ASB using a urine culture, and that treatment with antibiotics significantly reduces the incidence of symptomatic maternal urinary tract infections. The evidence informing this reaffirmation of the original recommendation from 2004 is mainly drawn from a Cochrane review of treatment effectiveness (17). The American Academy of Family Physicians (AAFP) (18) endorses the recommendations of the USPSTF. The Infectious Diseases Society of America (6) recommends screening for bacteriuria by urine culture for pregnant women in early pregnancy, and treatment if results are positive, with periodic re-testing for recurrent bacteriuria after therapy. The American Academy of Pediatrics (AAP), jointly with the American College of Obstetricians and Gynecologists (ACOG) recommend to treat ASB and then to test for cure (19).

The UK's National Institute for Health and Care Excellence (NICE) states that women should be offered routine screening for ASB by midstream urine culture early in pregnancy to reduce the risk of developing pyelonephritis (20).

The Scottish Intercollegiate Guidelines Network (SIGN) recommends that pregnant women be tested for ASB by urine culture at the first antenatal visit and culture-positive patients be treated with an antibiotic (21).

## **Current Practice**

Several major healthcare organizations in North America (USPSTF, IDSA, ACOG, AAP, AAFP) advocate screening of pregnant women, and nearly all recommend treating patients who have been confirmed with ASB using antibiotics. In Canada, the current usual practice is to obtain a urine sample at each prenatal visit, where testing may typically be done by culture early in pregnancy and then followed with subsequent testing if indicated. It is clear there is diversity in which of these samples are collected for the presence of significant bacteriuria, how the sample is collected, how presence of bacteriuria is determined, and when sample(s) for ASB is/are collected in pregnancy. It is unclear whether and to what degree practices use screening methods incorporating tests other than urine culture.

## **Section III. Review Approach and Scope**

This review will be completed by the Evidence Review and Synthesis Centre (ERSC) at the University of Alberta. The review will be developed, conducted, and prepared according to the CTFPHC methods (<http://canadiantaskforce.ca/methods/methods-manual/>). A working group of CTFPHC members was formed for development of the topic, refinement of the key questions and scope, and rating of patient-important outcomes considered most important for creating a recommendation. The CTFPHC will not be involved in the conduct of the review including selection of studies and data analysis, but will comment on the draft report and provide input on the interpretations of findings. The Global Health and Guidelines Division science team at the Public Health Agency of Canada provided assistance and input on CTFPHC methodological considerations during the topic refinement and development of the protocol. Perspectives of patients, and members of the public have been incorporated regarding prioritization of outcomes (benefits and harms), as well as other aspects of guideline development. A draft version of this protocol was reviewed by nine external topic experts and stakeholders and all comments were considered when finalizing this protocol. This final version of the protocol has been approved by the entire CTFPHC and will be posted on the CTFPHC website and registered with the International Prospective Registry of Systematic Reviews (PROSPERO) database.

## **Analytical Framework and Staged Approach**

Figure 1 is an analytical framework that depicts the structure used to address the Key Questions (KQs) for evaluating the benefits and harms of screening asymptomatic women during pregnancy for bacteriuria.

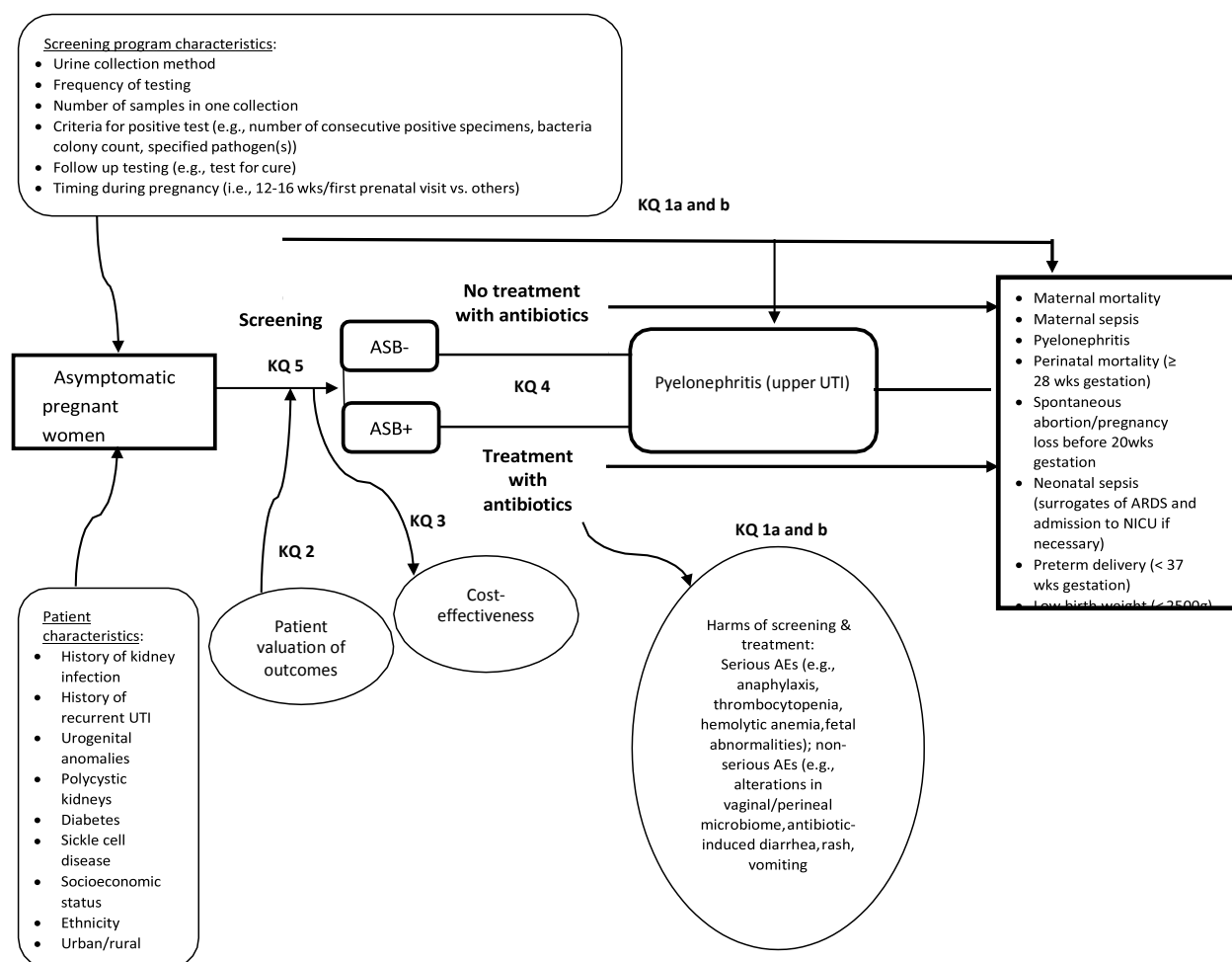
A staged approach will be followed based on the availability and quality of the body of evidence. Quality of evidence (classified as high, moderate, low, very low) will be assessed using methods developed by the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group (<http://www.gradeworkinggroup.org/>) whereby high quality evidence relies on precise and consistent effect estimates from studies having few limitations on internal validity (i.e., low bias) and examining directly relevant populations, interventions, comparators, and outcomes (i.e., PICO) (see Section IV for more details). The staging approach of the CTFPHC relies on choices made when considering, primarily, the GRADE domains of study limitations and indirectness. Moreover, decisions made during the evidence review are based on the information needs of the CTFPHC for making a screening recommendation based on the balance of critical patient-important benefits and harms.

The most direct and least biased evidence for the effectiveness of screening for ASB will be prioritized. This review will start by examining evidence from randomized-controlled trials (RCTs) on the clinical effectiveness of screening on patient-important outcomes. Staging beyond this point will require careful deliberation with documentation of rationale. If data from the initial stage is scarce for critical benefits or harms the CTFPHC will consider searching for data from (potentially) more biased study designs or indirect evidence (e.g., evidence from observational studies treatment RCTs, test accuracy studies. In cases where evidence on test accuracy and treatment effects will be used to provide indirect evidence on screening effectiveness, the limitations of such an indirect approach will be described. Examining both accuracy and treatment data may not be useful in all cases; for example, if the CTFPHC becomes confident that treatment is ineffective there would be no need to further examine test accuracy. In general, subsequent stages will only be conducted when the evidence from the previous stage(s) is non-existent or of too poor quality (e.g., very low quality based on GRADE tables) for the Task Force to make a screening recommendation based on the balance of patient-important benefits and harms.

For this review, the first stage will focus on identifying and using data from studies directly linking screening for ASB to patient-important benefits and harms (KQ1). Study designs providing the highest internal validity (e.g., RCTs) for this KQ will be preferred with a hierarchy of evidence used after this point if necessary. After RCTs we will consider controlled clinical trials (CCTs; defined for this review as *experimental trials without random allocation but where intervention(s) are introduced, standardized, and allocated objectively [e.g., by date of birth, but not using subjective means such as patient or clinician preferences] by investigators and blinding of participants is typically possible*) and then prospective and retrospective controlled observational studies. This stage will also include examination of KQ2 on women's valuation of benefit and harm outcomes of screening for ASB (and more broadly/indirectly treatment with antibiotics) in pregnancy. The cost-effectiveness of screening for ASB (KQ3) will also be considered only if there is evidence from KQ1 indicating a favorable benefit-harm ratio such that screening may be recommended.

If this first stage does not provide high enough quality of evidence for making a recommendation, the CTFPHC will carefully consider pursuing stage two with documentation of rationale before proceeding. Stage two will commence with examination of effectiveness of treatment of ASB in pregnancy (KQ4). If there is sufficient quality evidence indicating favorable treatment effectiveness from KQ4, an examination of KQ5 on diagnostic test accuracy will be considered in stage 3. Due to the indirectness of evidence provided by KQs 4 and 5 for making recommendations for the clinical effectiveness of screening, we will only seek data from study designs offering the greatest potential for high internal validity. That is, for KQ4 (treatment) we will focus on RCTs, and for KQ5 (test accuracy) we will exclude case-control designs. Where high quality systematic reviews exist examining these indirect evidence links, we will utilize these when possible.

## **Figure 1. Analytical Framework**



AEs: adverse events; ARDS: acute respiratory distress syndrome; ASB: asymptomatic bacteriuria; d: day; g: grams; KQ: key question; NICU: neonatal intensive care unit; UTI: urinary tract infection; wks: weeks

## Key Questions (KQs)\*

### Stage 1:

#### Benefits and harms of screening

**KQ1a:** What are the benefits and harms of screening compared with no screening for asymptomatic bacteriuria in pregnancy? Are there subgroup differences with SES or other patient characteristics?

**KQ1b:** What are the comparative benefits and harms of screening with different screening tests/algorithms for asymptomatic bacteriuria in pregnancy?

#### Outcome valuation

**KQ2a:** How do women weigh the benefits and harms of screening and treatment of asymptomatic bacteriuria in pregnancy?

**KQ2b:** How do women's valuation of benefits and harms of screening and treatment inform their decisions to undergo screening?

**Resource use\*\***

**KQ3:** What is the cost-effectiveness of screening for asymptomatic bacteriuria in pregnancy?

**Stage 2:****Treatment**

**KQ4:** What are the benefits and harms of antibiotic treatment compared with no treatment for asymptomatic bacteriuria in pregnancy?

**Stage 3:****Diagnostic accuracy of screening tests**

**KQ5:** What is the diagnostic accuracy of screening tests for asymptomatic bacteriuria in pregnancy?

*\*Decision process for staging outlined in section on Analytical Framework and Staged Approach*

*\*\*Conducted if benefit-harm ratio deemed beneficial based on KQ1*

## **Section IV. Review Methods**

### **Literature Search**

The literature search strategy will be developed and implemented by a research librarian. The search strategy will consist of both controlled vocabulary, such as the National Library of Medicine's MeSH (Medical Subject Headings), and keywords, and will be peer-reviewed. Methodological filters will not be applied to limit retrieval by study design; study designs included for each KQ are identified in the section on inclusion and exclusion criteria. Searches will be restricted by language to include full texts published in English and French, without a publication date restriction.

We will conduct comprehensive searches in bibliographic databases most relevant for each KQ. For evidence informing stage 1 of our review we will perform comprehensive searches for studies meeting our inclusion criteria as described below. For KQ1, we will search MEDLINE (1946-) via Ovid; Embase (1974-) via Ovid; Cochrane Library; CINAHL (1937-present) via EBSCOhost; and PubMed via NCBI Entrez. The detailed search strategy for MEDLINE is reported in Appendix 1 and will be adapted to accommodate the controlled vocabularies of each database. For KQs 2 (women's outcome valuation) and 3 (cost-effectiveness of screening), we will modify the search to include relevant terms and will add suitable databases (e.g. PsycINFO for patient preferences, NHS Economic Evaluation Database [EED] for cost effectiveness). Full search strategies for all databases will be included in the final report.

For evidence used in stages 2 and 3, we are aware of at least one high-quality systematic review for KQs 4 (4) and 5 (10) which we may rely on. For KQ4 on effectiveness of antibiotic treatment compared with no treatment, we anticipate updating a recent Cochrane review of treatment for asymptomatic bacteriuria in pregnancy (4); if an update is not possible, we will follow methods adopted by the CTFPHC for integrating systematic reviews (see Appendix 2). If we update this



review, the original search will be updated. For KQ5 (test accuracy), we anticipate using a recent review of screening tests for asymptomatic bacteriuria in pregnancy (10) and any additional reviews that may be identified as similar in scope. While multiple reviews may be considered for KQ5 (test accuracy) if found, we will not attempt to update the search(es) to identify more recent studies. If the scope of any review is narrower (e.g., does not include all interventions applicable to our topic), we may screen the excluded studies list(s) to identify potentially relevant studies for inclusion. To ensure we have identified all potentially relevant systematic reviews relevant to KQs 4 and 5, we will conduct a database search for systematic reviews. We will search PubMed (1946-) via NCBI Entrez, the Cochrane Database of Systematic Reviews (inception-) and the Database of Abstracts of Reviews of Effects (DARE) (inception-2013) via Wiley Cochrane Library to identify systematic reviews, meta-analyses and health technology assessments. Our PubMed search will utilize a search filter from CADTH (<https://www.cadth.ca/resources/finding-evidence>)

Grey literature will be searched and documented according to CTFPHC methods and will include internet-based searches (via adapted Canadian Agency for Drugs and Therapeutics in Health [CADTH] checklists; <https://www.cadth.ca/resources/finding-evidence/grey-matters>) electronic libraries (e.g., Health Canada Library, Canadian Electronic Library), and trial registries (ClinicalTrials.gov, World Health Organization International Clinical Trials Registry Platform). Based on consultation with clinical experts, the following highly relevant conference proceedings will be hand-searched for recent studies not yet published (2014-present): Society of Obstetricians and Gynaecologists of Canada, Association of Medical Microbiology and Infectious Disease Canada, ID Week, and American Society for Microbiology meeting (ICAAC). Clinical and content experts identified by the CTFPHC will be contacted and invited to identify relevant research reports for consideration; websites of relevant Canadian stakeholder organizations will be searched.

## Eligibility Criteria

Table 1 outlines the inclusion and exclusion criteria for all KQs, and details are provided below.

### Population

Studies will be considered for inclusion in all KQs if they examine pregnant women at any stage of pregnancy where the population represents a “routine screening” scenario (e.g., the majority of patients do not have a degree of signs or symptoms prompting diagnostic testing and/or treatment for upper or lower UTI). It is recognized that many women experience nocturnal and increased frequency of urination, or other symptoms, which do not necessarily indicate bacteriuria or infections. We will include studies where a proportion of, but not all, women have risk factors for UTIs or other outcomes of the review. KQ2 on women’s outcome valuation, we will include studies of women of child-bearing age if no evidence is found from studies with pregnant women; studies will still be required to examine screening or antibiotic treatment during pregnancy.

We will exclude studies *exclusively* including women with conditions that place them at substantially higher than average risk for bacteriuria (i.e., kidney infection, urogenital anomalies, polycystic kidneys, recurrent urinary tract infections [UTI], diabetes, sickle-cell disease), or with symptoms of UTI.

**Population subgroups of interest** history of kidney infection, urogenital anomalies, polycystic kidneys, recurrent urinary tract infection (UTI), diabetes, sickle cell disease, socioeconomic status (i.e., education, income), ethnicity (i.e., percent South Asian versus others), and urban/rural setting.

### **Interventions & Comparators**

For clinical effectiveness of screening (KQ1), any screening test/algorithm for ASB will be eligible for inclusion and the comparator is absence of screening (1a) or a different urine test or screening algorithm (1b). Studies that compare urine cultures of differing criteria (e.g., threshold  $10^3$  CFU/mL versus  $10^4$  CFU/mL) will also be eligible for inclusion. For women's outcome valuation (KQ2), any screening test for ASB during pregnancy will be eligible for inclusion; indirect evidence about antibiotic treatment during pregnancy broadly will be used if needed. For cost-effectiveness (KQ3), any screening test compared with no screening or another screening test (i.e., urine culture) will be eligible for inclusion; costs must be compared with outcomes/effects such that studies examining costs only will be excluded. For treatment effectiveness (KQ4), any antibiotic treatment for ASB compared to no treatment or placebo will be eligible for inclusion. For diagnostic accuracy (KQ5), any index test compared with a urine culture for detecting ASB will be eligible for inclusion. For all KQs, studies that include screening or treatment for group B streptococcus (GBS) at any time of pregnancy for any of the outcomes of interest will be included.

We will exclude studies exclusively examining urine tests used for screening for other conditions (e.g., proteinuria, glycosuria), and non-urine screening tests (e.g., vaginal/rectal swab culture for GBS testing).

**Screening subgroups of interest** urine collection methods (e.g., clean-catch and/or midstream; excluding catheter methods/samples), frequency of testing, number of samples in one collection, criteria for a positive test (including number of consecutive positive specimens, bacterial colony count, and specified pathogen(s)), follow-up testing during pregnancy, and timing during pregnancy.

### **Outcomes**

As with the KQs, the outcomes for inclusion for KQ1 (screening effectiveness) and KQ4 (treatment) will be staged to some extent, if necessary. Each outcome has been rated independently by members of the CTFPHC and by women, as per the patient engagement activities of an independent group with expertise in knowledge translation from St. Michael's Hospital in Toronto, Ontario. All patient-important outcomes rated as critical (7 to 9 out of 9) or important (4 to 6 out of 9) for decision making were considered for inclusion. From these ratings, the eight outcomes were rated as critical will be included in stage 1; of three outcomes rated as important, low birth weight (but not hypertension or acute kidney injury) will be included because in the past (i.e. older studies) this was conceptually considered the same as "pre-term birth", which both the CTFPHC members and patients rated as critical. Considering harms separately, if no evidence is found for any of the outcomes (serious adverse events [AEs]) in stage 1, there will be inclusion of the outcomes (non-serious AEs) from stage 2. This grouped and staged approach to harms will address infrequent reporting, reporting of different harms across studies, and also uncertainty regarding all the potential harms that may be reported. Non-serious AEs, particularly if frequent or severe, are considered important but not critical for decision making by the

CTFPHC. This approach acknowledges guidance to limit the number of total outcomes (maximum 7) to those which can be successfully managed cognitively by guideline panels when balancing multiple benefits and harms.

Outcomes for KQs 1 and 4 with ratings:

*Benefits (reduced incidence for all):*

1. maternal mortality (9)
2. maternal sepsis (8)
3. pyelonephritis (7)
4. perinatal mortality ( $\geq 28$  weeks of gestation (e.g., intrauterine demise, stillbirth, early neonatal death)) (9)
5. spontaneous abortion/pregnancy loss before 20 weeks of gestation (8)
6. neonatal sepsis (if not reported will include surrogate outcomes of acute respiratory distress syndrome [ARDS] or admission to neonatal intensive care unit [NICU]) (8)
7. preterm delivery (live fetus passed  $< 37$  weeks of gestation) (7)
8. low birth weight ( $< 2500\text{g}$ ) (6)

*Harms:*

1. serious adverse event(s)<sup>a</sup> associated with antibiotic treatment, *including but not limited to:* (7)
  - a. anaphylaxis,
  - b. thrombocytopenia,
  - c. hemolytic anemia,
  - d. fetal abnormalities; and,
2. non-serious adverse event(s) associated with treatment, *including but not limited to:* (4)
  - a. alterations in vaginal/perineal microbiome (e.g., candidiasis, vaginitis),
  - b. antibiotic-induced diarrhea,
  - c. rash,
  - d. vomiting

<sup>a</sup>Serious adverse event (experience) or reaction is any untoward medical occurrence that: a) results in death, b) is life-threatening, c) requires in-patient hospitalisation or prolongation of existing hospitalisation, d) results in persistent or significant disability/incapacity, or e) is a congenital anomaly/birth defect (Health Canada, 2011);

We will exclude studies that screen pregnant women for group B streptococcus near delivery or at time of rupture of membranes for the prevention or treatment of chorioamnionitis or neonatal GBS (without other outcomes of interest listed above).

Women's outcome valuation (KQ2) include several possible outcomes related to the weighing of benefits and harms of screening and treatment (KQs 1 and 4) and how this may affect their decisions to undergo screening (e.g., relative weight/utilities of benefit and harms; willingness to be screened based on relative value placed on benefits and harms of screening programs or treatment); these outcomes will be based on considerations of the possibility or perceived/expected magnitude of effects for the outcomes identified for KQs 1 and 4.

During focus groups, women identified an additional outcome - psychological distress/anxiety -  
ASB Proposal - Appendix Page

and rated this as critical (7 out of 9), although it was interpreted differently by some women as either a benefit (e.g., reduction in psychological distress/anxiety by knowing the health status of themselves and their baby) or a harm (e.g., another of many tests and potential worries during pregnancy). Anxiety as a critical outcome will be sought and synthesized within findings from KQ2 on women's valuation of benefits and harms of screening and treatment, as well as within interpretation of test accuracy outcomes from KQ5 (TP, TN, FP, FN) which will be *interpreted based on the CTFPHC judgments* on the magnitude of potential consequences of each (e.g., unnecessary anxiety from high FP, loss of potential benefit in FN) as identified in the section below "Assessment of the Overall Quality of the Evidence using GRADE".

Cost-effectiveness (KQ3) outcomes include cost per quality-adjusted life year (QALYs), incremental cost-effectiveness ratios (ICERs), and net benefit (in dollars from cost-benefit studies).

Diagnostic test accuracy (KQ5) outcomes include: sensitivity, specificity, false positives, false negatives, positive predictive value, negative predictive value, positive likelihood ratio, and negative likelihood ratio.

### **Setting, Study Design & Timing**

Studies conducted in primary care, or relevant clinical settings (e.g., prisons, remote stations, community health centers, midwifery practice) will be included. For KQ3 on cost-effectiveness we will limit studies to those conducted using data relevant to Canada, thus within countries having a very high Human Development Index (22).

For KQ1 (screening effectiveness), we will include RCTs initially and then, if needed based on the GRADE assessment of overall quality of the evidence, we will search for CCTs (defined in Section III) and then controlled observational studies (i.e., prospective and retrospective cohort, case-control, controlled before-after). For KQ2 (outcome valuation), we will include any study where women are asked to balance the benefits and harms of screening and treatment for ASB and state/choose their willingness to be screened and treated; surveys, experimental designs (e.g., contingent valuation), and qualitative research are examples. Cost-effectiveness (KQ3) will look at any study comparing effects and costs (e.g., cost-effectiveness, cost-utility, cost-benefit) and may include modelling of effects and/or costs. For KQ4 (treatment), we will rely on RCTs. For KQ5 (test accuracy), we will rely on prospective and retrospective studies where a consecutive or random sample of participants receive both the index test(s) and reference standard, or where participants are randomized to different index tests but all receive the reference standard, and assessment in a cross-sectional manner. We will exclude case-control studies and studies with longitudinal assessment of the reference standard.

For all KQs, case reports and case series (i.e., group of patients selected based on particular outcome) will be excluded as will papers not reporting primary research (e.g. editorials, commentaries, opinion pieces). Conference abstracts will not be eligible for inclusion, but will be captured and serve to help identify full study reports and assess the quality of evidence in relation to potential publication and reporting biases. No limits will be applied to publication year.

## Additional considerations

We do not have a minimum sample size for inclusion, nor do we have a minimum threshold for extent of incomplete follow-up or participant attrition; these factors will be considered during assessment of the quality of evidence (e.g., precision domain accounts for sample size across studies), and during sensitivity analyses in cases of substantial heterogeneity in findings at the data synthesis stage (see relevant sections).

## Tables 1 to 5. Inclusion and Exclusion Criteria for Key Questions

**Table 1. KQ1a, b: Benefits and harms of screening**

|                      |   |
|----------------------|---|
| <b>Population</b>    | Asymptomatic pregnant women at any stage of pregnancy who are not at high risk for bacteriuria.   |
|                      | <p><b>Excluded:</b> thrombocytopenia, hemolytic anemia, b. urogenital anomalies, polycystic kidneys, recurrent urinary tract infection [UTI], diabetes, sickle cell disease, socioeconomic status, ethnicity, urban/rural fetal abnormalities; and, d.</p> <p><b>Exclude:</b> non-serious adverse event(s) associated with treatment, including but not limited to: (4) a. alterations in vaginal/perineal microbiome (e.g., candidiasis, vaginitis), b. antibiotic-induced diarrhea, c. urinary tract infection [UTI], diabetes, and sickle cell disease), or with symptoms of UTI</p>   |
| <b>Interventions</b> | Any screening program or test/vomiting  |
|                      | Screening subgroups/algorithms, including: urine collection method, frequency of testing, number of samples in one collection, criteria for a positive test (including number of consecutive positive specimens, bacterial colony count, and specified pathogen(s)), follow-up testing during pregnancy, timing during pregnancy  |
| <b>Study Designs</b> | Based on RCTs, CCTs, controlled observational (i.e., prospective and retrospective cohorts, case-control, controlled before-after)  |
| <b>Language</b>      | English and French  |
| <b>Setting</b>       | Primary care and clinical settings (e.g., birthing rooms, remote locations, community (GBS) testing, midwifery practices)   |
| <b>Comparator</b>    | No public screening (but may include indicated/targeted testing and/or treatment upon development of  |
| <b>Outcomes</b>      | <p>CCT: controlled clinical trial</p> <p>Non-serious adverse event (expedited) or difference in screening test or maternal or neonatal (see intervention results table)</p> <p>KQ1b: A difference in screening test or maternal or neonatal (see intervention results table)</p> <p>c) Outcomes: in-patient hospitalization or prolongation of existing hospitalisation, d) results in persistent or significant disability/incapacity, or e) is a congenital anomaly/birth defect (Health Canada, 2011)</p> <p>1. maternal mortality (9)</p> <p>2. maternal sepsis (8)</p> <p>3. pyelonephritis (7)</p> <p>4. perinatal mortality (<math>\geq 28</math> week's gestation (e.g., intrauterine demise, stillbirth, early neonatal death)) (9)</p> <p>5. spontaneous abortion/pregnancy loss before 20 week's gestation (8)</p> <p>6. neonatal sepsis (if not reported will include surrogate outcomes of acute respiratory distress syndrome [ARDS] or admission to neonatal intensive care unit [NICU]) (8)</p> <p>7. preterm delivery (five fetus passed &lt; 37 week's gestation) (7)</p> <p>8. low birth weight (<math>&lt; 2500</math> g) (6)</p> |

**Table 2. KQ2: Outcome valuation**

|                                      |   |
|--------------------------------------|---|
| <b>Population</b>                    | <p>Asymptomatic pregnant women at any stage of pregnancy who are not at high risk for bacteriuria; will also accept asymptomatic women who are not pregnant if necessary</p> <p><b>Excluded:</b> thrombocytopenia, hemolytic anemia, b. urogenital anomalies, polycystic kidneys, recurrent urinary tract infection [UTI], diabetes, sickle cell disease, socioeconomic status, ethnicity, urban/rural fetal abnormalities; and, d.</p> <p><b>Exclude:</b> non-serious adverse event(s) associated with antibiotic treatment, including but not limited to: (7) a. alterations in vaginal/perineal microbiome (e.g., candidiasis, vaginitis), b. antibiotic-induced diarrhea, c. urinary tract infection [UTI], diabetes, and sickle cell disease), or with symptoms of UTI</p> |
| <b>Interventions/Index Test</b>      | <p>Any screening program or test, and any antibiotic; will accept studies on treatment for any bacterial condition in pregnancy</p> <p><b>Screening subgroups/algorithms, including:</b> urine collection method, frequency of testing, criteria for a positive test (including number of consecutive positive specimens, bacteria colony count, and specified pathogen(s)), follow-up testing during pregnancy, timing during pregnancy</p> <p><b>Exclude:</b> urine screening is done for other conditions (e.g., proteinuria, glycosuria), non-urine screening test (e.g., vaginal/rectal swab culture for GBS testing)</p>  |
| <b>Comparator/Reference Standard</b> | Not applicable  |
| <b>Outcomes<sup>s</sup></b>          | Several possible outcomes (e.g., relative weight/utilities of benefit and harms; willingness to be screened based on relative value placed on benefits and harms of screening programs or treatment)  |

|                      |   |
|----------------------|---|
| <b>Study Designs</b> | Qualitative, mixed methods, surveys/cross-sectional   |
| <b>Language</b>      | English and French  |
| <b>Setting</b>       | Primary care and clinical settings (e.g., prisons, remote stations, community centers, midwifery practices) |
| <b>Time frame</b>    | No publication date limits  |

KQ: key question

§If there is a very limited quality of evidence base for KQ2 (i.e., in terms of quantity/sample size, methodological quality, inconsistency between studies, or applicability to our population or setting) we will consider including studies examining women's valuation of harms *or* benefits rather than the trade-off between the two. For example, studies examining women's acceptance of screening and/or treatment for ASB when only considering their perspectives on the potential risks of antibiotic treatment to their baby, or the importance placed on reassurance about the potential to prevent preterm delivery et cetera, could offer some indirect evidence to help the CTFPHC in their deliberations. Likewise, the relative value placed on different benefit or harm outcomes (e.g., serious versus non-serious AEs) could be informative.

**Table 3. KQ3: Cost-effectiveness of screening**

|                                      |   |
|--------------------------------------|---|
| <b>Population</b>                    | Asymptomatic pregnant women at any stage of pregnancy who are not at high risk for bacteriuria.<br><br><u>Patient subgroups</u> : women with kidney infection, urogenital anomalies, polycystic kidneys, recurrent urinary tract infection [UTI], diabetes, sickle cell disease, socioeconomic status, ethnicity, urban/rural<br><br><u>Exclude</u> : studies <i>exclusively</i> including women with conditions that place them at substantially higher than average risk of bacteriuria (kidney infection, urogenital anomalies, polycystic kidneys, recurrent urinary tract infection [UTI], diabetes, and sickle cell disease), or with symptoms of UTI |
| <b>Interventions/Index Test</b>      | Any screening program or test<br><br><u>Screening subgroups/algorithms, including</u> : urine collection method, frequency of testing, number of samples in one collection, criteria for a positive test (including number of consecutive positive specimens, bacterial colony count, specified pathogen(s)), follow-up testing during pregnancy, timing during pregnancy<br><br><u>Exclude</u> : urine screening is done for other conditions (e.g., proteinuria, glycosuria, Chlamydia), non-urine screening test (e.g., vaginal/rectal swab culture for GBS testing)   |
| <b>Comparator/Reference Standard</b> | No screening (but may include indicated/targeted testing and/or treatment upon development of symptoms or for high-risk groups), or a different screening test or algorithm (see intervention subgroups)  |
| <b>Outcomes</b>                      | Cost per quality-adjusted life-years (cost per QALY), incremental cost-effectiveness ratio (ICER), net benefit/cost   |
| <b>Study Designs</b>                 | Economic evaluations  |
| <b>Language</b>                      | English and French  |
| <b>Setting</b>                       | Primary care and clinical settings (e.g., prisons, remote stations, community centers, midwifery practices); limited to countries rated as having very high Human Development Index (22)  |
| <b>Time frame</b>                    | No publication date limits  |

KQ: key question

**Table 4. KQ4: Treatment**

|                   |   |
|-------------------|---|
| <b>Population</b> | Asymptomatic pregnant women at any stage of pregnancy who are not at high risk for bacteriuria.<br><br><u>Patient subgroups</u> : women with kidney infection, urogenital anomalies, polycystic kidneys, recurrent urinary tract infection [UTI], diabetes, sickle cell disease, socioeconomic status, ethnicity, urban/rural<br><br><u>Exclude</u> : studies <i>exclusively</i> including women with conditions that place them at substantially higher than average risk of bacteriuria (kidney infection, urogenital anomalies, polycystic kidneys, recurrent urinary tract infection [UTI], diabetes, and sickle cell disease), or with symptoms of UTI |
|-------------------|---|

|                                      |   |
|--------------------------------------|---|
| <b>Interventions/Index Test</b>      | Any antibiotic<br><br><u>Screening subgroups/algorithms, including:</u> urine collection method, frequency of testing, number of samples in one collection, criteria for a positive test (including number of consecutive positive specimens, bacterial colony count, and specified pathogen(s)), follow-up testing during pregnancy, timing during pregnancy<br><br><u>Exclude:</u> urine screening is done for other conditions (e.g., proteinuria, glycosuria, Chlamydia.), non-urine screening test (e.g., vaginal/rectal swab culture for GBS testing) |
| <b>Comparator/Reference Standard</b> | No treatment or placebo   |
| <b>Outcomes*</b>                     | <i>Benefits (reduced incidence for all):</i><br>1. maternal mortality (9)   |

|                      |   |
|----------------------|---|
|                      | 2. maternal sepsis (8)\<br>3. pyelonephritis (7)<br>4. perinatal mortality ( $\geq 28$ week's gestation (e.g., intrauterine demise, stillbirth, early neonatal death)) (9)<br>5. spontaneous abortion/pregnancy loss before 20 week's gestation (8)<br>6. neonatal sepsis (if not reported will include surrogate outcomes of acute respiratory distress syndrome [ARDS] or admission to neonatal intensive care unit [NICU]) (8)<br>7. preterm delivery (live fetus passed $< 37$ week's gestation) (7)<br>8. low birth weight ( $< 2500$ g) (6)<br><i>Harms:</i><br>1. serious adverse event(s) <sup>a</sup> associated with antibiotic treatment, <i>including but not limited to:</i> (7) <ul style="list-style-type: none"> <li>a. anaphylaxis,</li> <li>b. thrombocytopenia,</li> <li>c. hemolytic anemia,</li> <li>d. fetal abnormalities; and,</li> </ul> 2. non-serious adverse event(s) associated with treatment, <i>including but not limited to:</i> (4) <ul style="list-style-type: none"> <li>a. alterations in vaginal/perineal microbiome (e.g., candidiasis, vaginitis),</li> <li>b. antibiotic-induced diarrhea,</li> <li>c. rash,</li> <li>d. vomiting</li> </ul><br><u>Exclude:</u> screening for group B streptococcus near delivery or at time of rupture of membranes for the prevention or treatment of chorioamnionitis or neonatal GBS (without other outcomes of interest listed above) |
| <b>Study Designs</b> | RCTs  |
| <b>Language</b>      | English and French  |
| <b>Setting</b>       | Primary care and clinical settings (e.g., prisons, remote stations, community centers, midwifery practices)   |
| <b>Time frame</b>    | No publication date limits  |

KQ: key question; RCT: randomized controlled trial

\*Serious adverse event (experience) or reaction is any untoward medical occurrence that: a) results in death, b) is life-threatening, c) requires in-patient hospitalisation or prolongation of existing hospitalisation, d) results in persistent or significant disability/incapacity, or e) is a congenital anomaly/birth defect (Health Canada, 2011)

**Table 5. KQ5: Diagnostic accuracy of screening tests**

|                   |   |
|-------------------|---|
| <b>Population</b> | Asymptomatic pregnant women at any stage of pregnancy who are not at high risk for bacteriuria.<br><br><u>Patient subgroups:</u> women with kidney infection, urogenital anomalies, polycystic kidneys, recurrent urinary tract infection [UTI], diabetes, sickle cell disease, socioeconomic status, ethnicity, urban/rural<br><br><u>Exclude:</u> studies <i>exclusively</i> including women with conditions that place them at substantially higher than average risk of bacteriuria (kidney infection, urogenital anomalies, polycystic kidneys, recurrent urinary tract infection [UTI], diabetes, and sickle cell disease), or with symptoms of UTI |
|-------------------|---|

|                                      |  |
|--------------------------------------|--|
| <b>Interventions/Index Test</b>      | <p>Any index test</p> <p><u>Screening subgroups/algorithm, including:</u> urine collection method, frequency of testing, number of samples in one collection, criteria for a positive test (including number of consecutive positive specimens, bacterial colony count, and specified pathogen(s)), follow-up testing during pregnancy, timing during pregnancy</p> <p><u>Exclude:</u> urine screening is done for other conditions (e.g., proteinuria, glycosuria, Chlamydia), non-urine screening test (e.g., vaginal/rectal swab culture for GBS testing)</p> |
| <b>Comparator/Reference Standard</b> | <p>A urine culture</p> <p>Screening subgroups/algorithm, including: urine collection method, frequency of testing, number of samples in one collection, criteria for a positive test (including number of consecutive positive specimens, bacterial colony count, and specified pathogen(s)), follow-up testing during pregnancy, timing during pregnancy</p>  |
|                                      | <u>Exclude:</u> urine screening is done for other conditions (e.g., proteinuria, glycosuria, Chlamydia), non-urine screening test (e.g., vaginal/rectal swab culture for GBS testing)  |
| <b>Outcomes</b>                      | Sensitivity, specificity, false positives, true positive, false negatives, true negatives, positive and negative likelihood ratios, prevalence/pre-test probability (true positive + false positive)/total number of people)   |
| <b>Study Designs</b>                 | <p>Prospective and retrospective studies where a consecutive or random sample of participants receive both the index test(s) and the reference standard, or where participants are randomized to different index tests but all receive the reference standard, and assessment in a cross-sectional manner</p> <p><u>Exclude:</u> case-control studies and studies with longitudinal assessment of the reference standard</p>   |
| <b>Language</b>                      | English and French   |

## Screening and Selecting Studies for Inclusion

For the database searches, two reviewers will independently screen the titles and abstracts (when available) using broad inclusion/exclusion criteria. Citations will be classified as “include/unsure,” “exclude,” or “reference” (i.e., conference abstracts, protocols, and systematic reviews). One reviewer will review the “reference” group and will conduct all other searching as outlined in the above section. The full text of all studies classified as “include/unsure” or identified after reviewing the reference citations will be retrieved for full review; two reviewers will independently assess eligibility using a standard form that outlines the inclusion and exclusion criteria. Disagreements on final inclusion of all studies will be resolved through consensus or third party adjudication. For KQs 4 and 5, any existing systematic review(s) identified as relevant will be assessed for eligibility based on whether the authors: i) searched more than one database, ii) report their selection criteria, and iii) use PICOTS criteria that are a close match to that for the relevant KQ. In cases where there is more than one possible review providing results for the same intervention-outcome pair, we will choose one based on: AMSTAR (23) rating (score 8 or higher preferred), comprehensiveness of search (i.e., reports on most or more papers included by other existing reviews), closest match to our PICOTS, most recent date of study inclusion/search, and the quality and extent of reporting on individual study characteristics, data, and quality assessments. All decisions to exclude a study at full text review will be provided. The title/abstract screening and full-text selection processes will be conducted and documented in DistillerSR. The flow of literature and reasons for full text exclusions will be recorded in a PRISMA Flow Chart.

## Data Extraction & Reporting

One reviewer will independently extract data from each included study or systematic review into DistillerSR; a second reviewer will verify all data. Disagreements will be resolved through



discussion or third-party consultation until consensus is reached.

When using individual studies for a KQ, a narrative summary (with accompanying tables) will be provided to report on all studies by design, country of origin, sample sizes, population(s) (including subgroups), intervention(s)/index tests (including data on thresholds and for subgroup questions), comparator(s)/reference test, setting, and outcome measures, as reported by studies. When there are multiple publications associated with a study we will consider the earliest report of the main (primary) outcome data to be the primary data source. We will extract data from the primary source first and then add outcome data reported in the secondary/associated publications and data sources. We will reference the primary source throughout the evidence report; all associated literature will be tabulated for reference.

When relying on systematic reviews for KQs 4 (treatment) and 5 (test accuracy), we will extract data on the characteristics of the systematic review (PICOTS), the included studies with specifics related to the population (size and characteristics), outcomes evaluated (including definitions and timing of assessment), quality/risk of bias (by domain/construct if available), the methods of analysis (meta-analytical approach and its findings in relation to heterogeneity, if applicable), findings from their syntheses including subgroup analysis and GRADE or other quality assessments if performed across studies, and any limitations noted by the systematic review authors. For KQs 4 and 5, data verification will be completed on 5 to 10% of included studies in any existing systematic review(s), and if satisfied with concordance, we will consider incorporating the reported data on study and participant characteristics without returning to the primary studies. If additional studies are included (e.g., new studies from updated search [KQ4] or excluded studies in the identified systematic review that is subsequently included for current review to ensure coverage of scope [KQ5]), these will be clearly identified and presented.

When using individual studies, we will record intention-to-treat results, if possible. For continuous outcomes measures, we will extract (by arm) the mean baseline and endpoint or change scores, standard deviations (SD) or other measure of variability, and number analyzed. We will not include outcome data from studies that did not provide a follow up change or endpoint mean or data that could be used to calculate follow up scores. If necessary, we will approximate means by medians. If standard deviations are not given, they will be computed from p-values, 95% confidence intervals (95% CIs), standard errors, z-statistics, or t-statistics. If computation is not possible they will be estimated from upper bound p-values, ranges, inter-quartile ranges, or (as a last resort) by imputation using the largest reported SD from the other studies in the same meta-analysis. When computing SDs for change from baseline values, we will assume a correlation of 0.5, unless other information is present in the study that allows us to compute it more precisely. For dichotomous outcomes, we will report counts or proportions, and sample size, by study arm.

For dichotomous data on harms, each adverse event (AE) will be counted as if it represents a unique individual; because a single individual might experience more than one AE, this assumption may overestimate the number of people having an AE. Only numerical data for AEs will be extracted; that is, we will make no assumptions on lack or presence of an AE if this is not reported; authors that report only p-values or that one arm had fewer events than another (but where it is explicit that the outcome was captured in the study) will be contacted (3 times via email) to provide the data.

Data on within-study subgroup analysis will be collected, including: subgroups (independent variables), the type of analysis (e.g., subgroup/stratified or regression analysis), the outcomes assessed (dependent variables), and the authors' conclusions. We will collect data suitable for all patient and intervention subgroups (see Table 1) for performing our own subgroup analyses (e.g., stratified analysis, meta-regression) based on study-level data.

### **Risk of Bias/Methodological Quality Assessment**

Two reviewers will independently assess the risk of bias (ROB) of each included study (KQs 1-3), with disagreements resolved through discussion or third-party consultation to reach consensus. The results for each study and across studies will be reported by each domain and for the overall ROB score. The ROB for each study will be assessed on an outcome basis where needed, particularly when different outcomes are assumed to have different susceptibilities to bias; for example, subjective outcomes and expected harms are more prone to bias from non-blinding than objective outcomes and unexpected/rare harms.

RCTs and CCTs (theoretically only differing from RCTs by lack of random sequence generation and not by other ROB domains) will be appraised using the Cochrane Risk of Bias tool (24). This tool consists of six domains (sequence generation, allocation concealment, blinding, incomplete outcome data, selective outcome reporting, and "other" sources of bias) and a categorization of the overall risk of bias. Blinding will be assessed separately for patients/providers and outcome assessors taking into account the type of outcome that may be affected (e.g. subjective vs. objective). To assist with outcome reporting bias assessments, we will seek study protocols and studies/data from registries. The overall assessment is based on the responses to individual domains. If one or more individual domains are assessed as having a high risk of bias, the overall score will be rated as high risk of bias. If at least one domain is assessed as unclear, and no domains are assessed as high, the overall score will be rated as unclear risk of bias. The overall risk of bias will be considered low only if all components are rated as having a low risk of bias.

Controlled observational studies will be appraised using the Newcastle-Ottawa Quality Assessment Scale (25); three domains (sample selection, comparability of cohorts, and assessment of outcomes) are evaluated. Each item that is adequately addressed is awarded one star, except for the "comparability of cohorts" item, for which a maximum of two stars can be given. The overall score is calculated by tallying the stars. We will consider a total score of 6 to 8 stars to indicate low ROB, 4 or 5 stars to indicate moderate ROB, and 3 or fewer stars to indicate high ROB.

For diagnostic accuracy studies (KQ5), we will rely on the Quality of Diagnostic Accuracy Studies (QUADAS-2) (26) used to assess ROB. This tool assesses concerns of risk of bias among four domains (patient selection, index test, reference standard, and flow and timing) and concerns of applicability across the first three domains.

If one or more systematic review(s) is used to provide evidence for KQ4 (treatment) or KQ5 (accuracy), we will assess if the review used an explicit tool (e.g., Cochrane ROB [KQ4], QUADAS [KQ5]) for assessing the main sources of potential bias. If so, we will complete assessments on 5 to 10% of included studies to establish concordance before considering the use of assessments reported by each review.

Studies answering KQ2 (outcome valuation) will be evaluated by tools appropriate to their study design: for surveys and qualitative studies we will use tools developed by the Center for Evidence-based Management (<http://www.cebma.org/resources-and-tools/what-is-critical-appraisal/>) The quality of economic evaluation studies (KQ3) will be assessed using Drummond's checklist for economic evaluation studies (27).

## **Data Analysis & Synthesis**

We will provide summaries of intervention effects for each study by calculating the appropriate statistics based on types of outcomes.

### ***Key Question 1***

For pair-wise meta-analysis in KQ1 (screening effectiveness), we will employ a random effects model. For continuous outcomes, we will report a pooled mean difference (MD) when one measurement tool is used, or other options that exist for communicating results when combining two or more outcome scales measuring similar constructs (28, 29). For dichotomous outcomes, we will report relative risks (RR) and risk differences (RD) between groups with corresponding 95% CIs. For those outcomes (e.g. serious adverse events) where at least one intervention group contains zero events, only the risk difference will be used. For calculating the RD, we will use the median baseline risk for the control group in the included studies, although may perform sensitivity analysis using differing baseline risks if thought suitable (30, 31). The decision to pool studies will not be based on the statistical heterogeneity ( $I^2$  statistic will be reported), but rather on interpretation of the clinical and methodological differences between studies. When substantial heterogeneity is suspected, we will conduct sensitivity analyses if appropriate (e.g., in the presence of studies with outlying effect sizes, for studies rated as high risk of bias in some domains such as incomplete outcome data [ $<80$  percent] or lack of allocation concealment, parallel versus cross-over designs). Heterogeneity will also be examined during our planned subgroup analyses for important patient and intervention variables (see Table 1). Where there are at least eight studies in a meta-analysis, we will analyze publication bias both visually using the funnel plot and quantitatively using Egger's test (32). We will not combine results from RCTs with CCTs or controlled observational studies (if used via staging approach for KQ on screening); rather, the latter two will be used to support or provide context for the evidence from RCTs.

### ***Key Questions 2 & 3***

For KQs 2 (outcome valuation) and 3 (cost-effectiveness), results will be narratively described in most cases. If more than one study is identified providing numerical values for ranking benefits and/or harms (KQ2) or similar outcomes (KQ3) these will be summarized descriptively and results across studies compared. Thematic analysis may be undertaken for KQ2, including coding data (meaning and context) into descriptive themes that accurately reflect the data and then summarizing this in a narrative format.

### ***Key Questions 4 and 5***

When using systematic reviews for stages 2 and 3, any meta-analysis will be reconstructed if possible to provide graphical representation of the findings to support our interpretations. Meta-analysis may be recalculated, if possible, when new studies are found in search updates (KQ4), analysis methods are not thought appropriate (e.g., use of random rather than fixed effects

models, ability but no use of HSROC models [see below]) or if further analysis (e.g. between-study stratification) may be possible for subgroups of interest. When substantial methodological heterogeneity was found, we may conduct sensitivity analyses if appropriate and able (e.g., for studies rated as high risk of bias, different study designs) or decide to not use the pooled/combined estimate. If not conducted by the authors and when there are at least eight studies in a meta-analysis, we will if possible analyze publication bias both visually using the funnel plot and quantitatively using Egger's test (32). If meta-analysis was not performed, we will summarize the findings of the systematic review authors.

For KQ5 (diagnostic accuracy), if individual studies are incorporated we will construct 2 x 2 tables and calculate sensitivity, specificity, and positive and negative likelihood ratios (LR+, LR-). Sensitivity and specificity are measures of test accuracy. Likelihood ratios are used to estimate the increased or decreased probability of disease (i.e., ASB) for a patient and can be used to refine clinical judgement based on varying pre-test probabilities. The larger the LR+, the more accurate the test is and the greater the likelihood of disease following a positive test; the smaller the LR-, the more accurate the test is, the lesser the likelihood of disease following a negative test (33). A LR+ that is >10 indicates a large and often conclusive probability that the condition is present; a LR- that is <0.10 suggests a large and often conclusive probability that the condition is not present. A likelihood ratio of one means that a positive or negative result is equally probable in a patient with and without the disease/condition.

If there are more than three studies and they are clinically homogenous (i.e., timing in pregnancy, thresholds, diagnostic criteria), we will pool data using a hierarchical summary receiver-operator curve (HSROC) and bivariate analysis of sensitivity and specificity (34). The HSROC simultaneously compares the sensitivity and specificity (taking their correlation into account) for all studies comparing a particular screening test with ASB diagnostic criteria. We will use Review Manager Version 5.0 (The Cochrane Collaboration, Copenhagen, Denmark) to perform meta-analyses, and Stata 11.0 (metandi program; StataCorp LP, College Station, TX, USA) to fit the bivariate and HSROC models and produce the pooled estimates of sensitivity, specificity, and likelihood ratios.

The results will be organized by type of screening test. If possible, we will examine the impact of screening before and after 12-16 weeks' gestation and in relation to other intervention subgroups described in Table 5. Sensitivities, specificities, and likelihood ratios and their 95% confidence intervals (CI) will be presented in summary tables that include all screening tests and diagnostic criteria. Based on the findings for sensitivity and specificity and estimates of one or more relevant baseline prevalence, an evidence profile will be generated for the outcomes FN, FP, TN, and TP (30).

### **Subgroup Analyses**

Our primary approach for evaluating differential effect for subgroups will be to record any within-study subgroup analyses performed by study investigators using individual patient data; these results preserve the within-study randomization. Because these results are often based on diverse methodology and may be difficult to interpret across the body of evidence, we will also perform our own subgroup analyses using study-level data, as possible, using formal statistical approaches (e.g., meta-regressions) or by stratifying the results of the pairwise meta-analyses by subgroup variables. When determining whether entire studies fall into a particular subgroup

category (e.g., recurrent UTI), we will consider  $\geq 80$  percent of the study population meeting the criteria as sufficient. We will employ regression analyses when: for continuous variables (e.g., timing during pregnancy) there are at least six to ten studies reporting on the outcome within a specific subgroup, and for categorical variables (e.g., history of recurrent UTI) there are at least three studies for each category level. The number of sufficient studies serves as a rule of thumb for the lower bound that investigators can consider for a meta-regression, but power will vary according to the size and variability of the effect. These analyses would rely on study-level data, such that the results would be considered observational in nature.

## Assessment of the Overall Quality of the Evidence using GRADE

Two reviewers will independently assess the quality of the body of evidence or confidence in the effect for each outcome of interest (see Table 1) using the GRADE methodology. Discrepancies will be resolved through discussion or third-party consultation to reach consensus. Assessments will be entered into the GRADEPro software and summarized in GRADE evidence profiles, Summary of Findings tables and Evidence to Decision Tables. Footnotes to the tables will explain all decisions. The CTFPHC will then use this evidence on each outcome, to assess the net benefits and harms of each service, consider patient preferences and values, and other elements of the GRADE methodology to develop the recommendations on screening for bacteriuria (feasibility, acceptability and equity).

The general approach is outlined here although methods will align with GRADE guidance (30, 35). When using systematic reviews, GRADE assessments will be based on the individual studies and reporting by review authors (e.g., on ROB assessments and PICOTS characteristics) and upon validation of a sample by the review team. For evidence on the benefits and harms of screening (KQ1), as a starting point the quality is assigned as high for evidence from RCTs and low for evidence from observational studies, when used. For accuracy studies, cross-sectional or cohort studies in patients with diagnostic uncertainty and direct comparison of test results with an appropriate reference standard will be considered high quality. Thereafter, we will examine and potentially downgrade the quality based on five core domains: study limitations/ROB, inconsistency, indirectness, imprecision, and publication/reporting bias. For outcomes where there is evidence from observational studies and no other reason to downgrade the evidence, we will also consider the additional domains of dose-response association, plausible confounding, and strength of association (i.e., large magnitude of effect [i.e., large  $\leq 0.5$  or  $\geq 2.0$  or very large  $RR \leq 0.2$  or  $\geq 5.0$ ]), to potentially upgrade the quality (36).

For the *study limitations (risk of bias)* domain RCTs and CCTs may be downgraded one or two levels depending on the proportion of trials (e.g., one very large trial may outweigh two very small trials) assessed as having high ROB for the particular outcome under consideration (37). Evidence from observational studies will be downgraded when most studies have moderate or high ROB. For *inconsistency* (consistent, inconsistent) we will assess the magnitude of the effects of the included studies (e.g., inconsistent when lack of overlap in 95% CIs for some studies) (38). *Indirectness* of the evidence (direct or indirect) is based on evaluating the relevance of the study's PICO compared to ours for our primary KQ1 (effectiveness of screening); when relying on test accuracy and treatment studies there will be downgrading by at least one level for this domain (36). We will assess *imprecision* (precise or imprecise) on the basis of clinical thresholds and Optimal Information Size (39). For outcomes where clinical

thresholds are used/determined, we will typically downgrade this domain once if the entire pooled 95% CI does not cross the threshold (i.e. only one limit of the CI crosses), and downgrade twice if the 95% CI crosses the threshold and no difference (0 MD or 1.0 RR) or does not cross the threshold at all. Thresholds may be determined a priori (prior to viewing results from studies) but may also be revised post hoc based on careful benefit-harm considerations when considering all outcomes together (e.g., lower benefit threshold in cases of few and minor harms). A precise estimate is one that allows for a clinically useful conclusion. *Reporting bias* (suspected or undetected) will be evaluated with respect to publication bias.

Interpreting these domains when relying on evidence from diagnostic test (KQ5) data has certain considerations, including how certain the CTFPHC is about the consequences of each outcome (FP, FN, TP, TN) in relation to the main outcomes of interest for KQs 1, 2 & 4 (30).

## External Review

The evidence review will be peer-reviewed by external content experts (minimum 3) and invited stakeholder organizations (minimum 10), with response to all comments shared with all reviewers approximately two months after posting of the final review.

## Planned Schedule and Timeline

Draft protocol approved by CTFPHC members: July 29, 2016

External peer review: August 1-10, 2016

Final protocol: November 30, 2016

Draft evidence review: January 31, 2017

Final evidence review: March 31, 2017

## Conflict of Interest Statement

None of the study team members have any known actual or perceived conflicts of interest related to this review.

## References

1. Glaser AP, Schaeffer AJ. Urinary Tract Infection and Bacteriuria in Pregnancy. The Urologic clinics of North America. 2015;42(4):547-60.
2. Schnarr J, Smaill F. Asymptomatic bacteriuria and symptomatic urinary tract infections in pregnancy. European journal of clinical investigation. 2008;38 Suppl 2:50-7.
3. Ipe DS, Sundac L, Benjamin WH, Jr., Moore KH, Ulett GC. Asymptomatic bacteriuria: prevalence rates of causal microorganisms, etiology of infection in different patient populations, and recent advances in molecular detection. FEMS microbiology letters. 2013;346(1):1-10.
4. Smaill FM, Vazquez JC. Antibiotics for asymptomatic bacteriuria in pregnancy. The Cochrane database of systematic reviews. 2015;8:CD000490.
5. Connolly A, Thorp JM, Jr. Urinary tract infections in pregnancy. The Urologic clinics of

North America. 1999;26(4):779-87.

6. Nicolle LE, Bradley S, Colgan R, Rice JC, Schaeffer A, Hooton TM. Infectious Diseases Society of America guidelines for the diagnosis and treatment of asymptomatic bacteriuria in adults. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2005;40(5):643-54.
7. Kass EH. Bacteriuria and the pathogenesis of pyelonephritis. *Laboratory investigation; a journal of technical methods and pathology*. 1960;9:110-6.
8. Wing DA, Fassett MJ, Getahun D. Acute pyelonephritis in pregnancy: an 18-year retrospective analysis. *American journal of obstetrics and gynecology*. 2014;210(3):219 e1-6.
9. Kazemier BM, Koningstein FN, Schneeberger C, Ott A, Bossuyt PM, de Miranda E, et al. Maternal and neonatal consequences of treated and untreated asymptomatic bacteriuria in pregnancy: a prospective cohort study with an embedded randomised controlled trial. *The Lancet Infectious diseases*. 2015;15(11):1324-33.
10. Rogozinska E, Formina S, Zamora J, Mignini L, Khan KS. Accuracy of Onsite Tests to Detect Asymptomatic Bacteriuria in Pregnancy: A Systematic Review and Meta-analysis. *Obstet Gynecol*. 2016;128(3):495-503. Epub 2016/08/09.
11. Hilt EE, McKinley K, Pearce MM, Rosenfeld AB, Zilliox MJ, Mueller ER, et al. Urine Is Not Sterile: Use of Enhanced Urine Culture Techniques To Detect Resident Bacterial Flora in the Adult Female Bladder. *Journal of Clinical Microbiology*. 2014;52(3):871-6.
12. Price TK, Dune T, Hilt EE, Thomas-White KJ, Kliethermes S, Brincat C, et al. The Clinical Urine Culture: Enhanced Techniques Improve Detection of Clinically Relevant Microorganisms. *Journal of Clinical Microbiology*. 2016;54(5):1216-22.
13. Rouse DJ, Andrews WW, Goldenber RL, Owen J. Screening and Treatment of Asymptomatic Bacteriuria of Pregnancy to Prevent Pyelonephritis - a Cost-Effectiveness and Cost-Benefit-Analysis. *Obstet Gynecol*. 1995;86(1):119-23.
14. Guinto VT, De Guia B, Festin MR, Dowswell T. Different antibiotic regimens for treating asymptomatic bacteriuria in pregnancy. *The Cochrane database of systematic reviews*. 2010(9):CD007855.
15. Allen VM, Yudin MH, Bouchard C, Boucher M, Caddy S, Castillo E, et al. Management of group B streptococcal bacteriuria in pregnancy. *Journal of obstetrics and gynaecology Canada : JOGC = Journal d'obstetrique et gynecologie du Canada : JOGC*. 2012;34(5):482-6.
16. United States Preventive Services Task Force. Screening for asymptomatic bacteriuria in adults: U.S. Preventive Services Task Force reaffirmation recommendation statement. *Annals of internal medicine*. 2008;149(1):43-7.

17. Smaill F, Vazquez JC. Antibiotics for asymptomatic bacteriuria in pregnancy. The Cochrane database of systematic reviews. 2007(2):CD000490.
18. American Academy of Family Physicians. Summary of Recommendations for Clinical Preventive Services. Leawood, KS: 2015.
19. American Academy of Pediatrics, American College of Obstetricians and Gynecologists. Guidelines for perinatal care. 7 ed: Amer Academy of Pediatrics; 2012.
20. NICE (National Institute for Health and Care Excellence). Antenatal care for uncomplicated pregnancies. Clinical guideline 62. United Kingdom: NICE; 2016. Available from: <https://www.nice.org.uk/guidance/cg62>.
21. SIGN (Scottish Intercollegiate Guidelines Network). SIGN 88. Management of suspected bacterial urinary tract infections in adults. Edinburgh, Scotland: SIGN; 2012. Available from: <http://www.sign.ac.uk/guidelines/fulltext/88/index.html>.
22. Malik K, Human Development Report 2013 Team. Human Development Report 2013. The Rise of the South: Human Progress in a Diverse World. N.Y., New York. : United Nations Development Programme, 2013.
23. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Medical Research Methodology. 2007;7(1):10.
24. Higgins JP, Green S. Section 8. Assessing risk of bias in included studies.: The Cochrane Collaboration; 2011. Available from: <http://handbook.cochrane.org>
25. Wells GA, Shea B, O'connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2000.
26. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine. 2011;155(8):529-36.
27. Drummond MF SM, Torrance G, O'Brien B, Stoddart G. Methods for the Economic Evaluation of Health Care Programmes. 3rd ed. Oxford, UK: Oxford University Press; 2005.
28. Guyatt GH, Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA, et al. GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles-continuous outcomes. Journal of clinical epidemiology. 2013;66(2):173-83.
29. Johnston BC, Patrick DL, Thorlund K, Busse JW, da Costa BR, Schunemann HJ, et al. Patient-reported outcomes in meta-analyses-part 2: methods for improving interpretability for decision-makers. Health and quality of life outcomes. 2013;11:211.
30. Hsu J, Brozek JL, Terracciano L, Kreis J, Compalati E, Stein AT, et al. Application of GRADE: making evidence-based recommendations about diagnostic tests in clinical practice



guidelines. Implementation science : IS. 2011;6:62.

31. Spencer FA, Iorio A, You J, Murad MH, Schunemann HJ, Vandvik PO, et al. Uncertainties in baseline risk estimates and confidence in treatment effects. *BMJ*. 2012;345:e7401.
32. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ (Clinical research ed)*. 1997;315(7109):629-34.
33. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *Jama*. 1994;271(9):703-7.
34. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of clinical epidemiology*. 2005;58(10):982-90.
35. Guyatt GH, Oxman AD, Schunemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. *Journal of clinical epidemiology*. 2011;64(4):380-2.
36. Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. *Journal of clinical epidemiology*. 2011;64(12):1311-6.
37. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *Journal of clinical epidemiology*. 2011;64(4):407-15.
38. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. *Journal of clinical epidemiology*. 2011;64(12):1294-302.
39. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. *Journal of clinical epidemiology*. 2011;64(12):1283-93.

## **Appendix 1. MEDLINE Search Strategy (KQ1 [screening effectiveness])**

Database: Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present

Search Title: PHTF Bacteriuria Screening in Pregnancy

Strategy:

1. Asymptomatic Infections/ and (bacteriuria\* or bladder\* or cystitis\* or kidney\* or pyelo-

cystiti\* or pyelocystiti\* or pyelo-nephriti\* or pyelonephriti\* or urin\* or UTI\*).mp.

2. Bacteriuria/
3. exp Cystitis/
4. Dysuria/
5. Pyelonephritis/
6. Urinary Tract Infections/
7. bacilluria\*.tw,kf.
8. bacteriuria\*.tw,kf.
9. cystiti\*.tw,kf.
10. (cysto-pyeliti\* or cystopyeliti\*).tw,kf.
11. dysuria\*.tw,kf.
12. (infection\* adj2 (bladder\* or genitourin\* or kidney\* or urin\* or urogenita\*)).tw,kf.
13. (pyelo-cystiti\* or pyelocystiti\*).tw,kf.
14. (pyelo-nephriti\* or pyelonephriti\*).tw,kf.
15. (UTI or UTIs).tw,kf.
16. or/1-15 [Combined MeSH & text words for bacteriuria]
17. Antibody-Coated Bacteria Test, Urinary/
18. \*Bacteriuria/di, pc, mi, ur
19. exp \*Cystitis/di, pc, mi, ur
20. Mass Screening/
21. Microbial Sensitivity Tests/
22. Microscopy/
23. Predictive Value of Tests/
24. \*Pyelonephritis/di, pc, mi, ur
25. Reagent Kits, Diagnostic/
26. Reagent Strips/
27. "Sensitivity and Specificity"/
28. Urinalysis/
29. \*Urinary Tract Infections/di, pc, mi, ur
30. ((accurac\* or diagnostic) adj5 (algorithm\* or test\*)).tw,kf.
31. diagnostic accurac\*.tw,kf.
32. culture\*.tw,kf.
33. (detect\* or predict\* or screen\*).tw,kf.
34. (dip slide\* or dipslide\* or dip stick\* or dipstick\*).tw,kf.
35. (micro-scopy or microscopy).tw,kf.
36. (microb\* adj2 test\*).tw,kf.
37. ((re-agent\* or reagent) adj3 (strip\* or test\*)).tw,kf.
38. strip\* test\*.tw,kf.
39. urine test\*.tw,kf.
40. (urinalys\* or urine analys\*).tw,kf.
41. uriscreen.tw,kf.
42. or/17-41 [Combined MeSH & text words for screening]
43. exp Pregnancy/
44. Pregnancy Complications, Infectious/
45. Pregnant Women/
46. Prenatal Care/
47. Prenatal Diagnosis/

48. (antenatal\* or pre-natal\* or prenatal\*).mp.
49. (expect\* adj (female? or mother? or wom#n)).tw,kf.
50. pregnan\*.mp.
51. or/43-50 [Combined MeSH & text words for pregnancy]
52. and/16,42,51 [Combined searches for bacteriuria, screening & pregnancy]
53. Male/ not (Female/ and Male/)
54. 52 not 53 [Male only records excluded]
55. exp Animals/ not (exp Animals/ and Humans/)
56. 54 not 55 [Animal only records excluded]
57. (comment or editorial or news or newspaper article).pt.
58. (letter not (letter and randomized controlled trial)).pt.
59. 56 not (57 or 58) [Opinion pieces excluded]
60. case reports.pt.
61. 59 not 60 [Case reports excluded]
62. limit 61 to (english or french)
63. remove duplicates from 62

## **Appendix 2. Methods for Integrating Existing Systematic Reviews into New Reviews**

One or more systematic reviews may exist that align with one or more key questions (KQs) of the reviews undertaken to inform CTFPHC guidelines. The CTFPHC and ERSCs have considered the manner in which new reviews conducted for CTFPHC guidelines can benefit from efficiencies by incorporating existing systematic reviews, while maintaining methodological rigor in their own systematic review conduct, closely aligning existing reviews within their review scope (i.e., inclusion/exclusion criteria), and maintaining consistency with other CTFPHC Methods. They have based their approach on work conducted by a methods working group composed of investigators from the Evidence-based Practice Center Program funded by the U.S. Agency for Healthcare Research and Quality.<sup>1,2</sup> A summary of the way the ERSCs will operationalize the 12 AHRQ recommendations (Box 1) to meet their needs is outlined below. This approach differs from situations when “updating” a single existing systematic review is deemed suitable, that is, in some cases a high-quality review will be used to answer one or more of the CTFPHC KQs in entirety, usually without revisions to the review’s scope, search for evidence (apart from updating to present), methodological quality/risk of bias assessments, data extraction, or data analysis.

### *Summary of CTFPHC Approach*

The recommendations developed by AHRQ (Box 1) will serve as an overall framework for ERSC reviews, although in most cases existing systematic reviews will be used to build efficiencies in discrete steps within the review process—mainly search and selection of literature, and data extraction—which will not generally include refinement of the scope or data analysis and interpretation. Moreover, we will not in most circumstances include a systematic review itself as a study design for inclusion (unless the intention is to specifically conduct an overview of reviews). The ability to use any given systematic review will largely depend on how it aligns with the CTFPHC review’s scope (PICOTS). A further primary consideration will be the comprehensiveness of its search strategy and reporting of literature flow. It is important to note that some CTFPHC reviews need to be complex with multiple stages (e.g., a review of screening effectiveness for patient-important benefits and harms may require including evidence on indirect evidence of test accuracy and treatment) such that existing systematic reviews may

exist for one or more discrete stages but not for others. Some key points on the operationalization, and minor revision, by the ERSCs of these recommendations are provided below.

1. **Choosing systematic reviews:** Following the identification of relevant reviews (a search for systematic reviews may be undertaken for some topics), the evidence for each will be mapped to the PICOTS elements and the quality of the review will be assessed (e.g., using the AMSTAR tool which has been evaluated and found effective to discriminate reviews with high and low quality of methods and reporting).<sup>3</sup> Some of the CTFPHC KQs may only have a single existing systematic review for possible incorporation, while others may have more than one; if suitable, a decision between systematic reviews will be based on methodological quality, comprehensiveness and quality of its literature search and reporting (e.g., assessed using PRESS checklist), comprehensiveness of reporting on included studies, and the best fit within the CTFPHC scope and methods. In some cases two or more reviews may be integrated because, together, they capture the full scope of the CTFPHC KQ(s). Rationale will be provided for choices made.

**Note:** If no review is deemed a good fit for purpose for integration (i.e., de novo process all together appears to be best option) we will at minimum examine available reviews for their search strategies (to ensure that our search strategies are comprehensive) and review their reference lists for identification of studies.

2. **Searching:** Various strategies will be considered. If one or more reviews are fit for purpose (but do not meet criteria for classification as a systematic review update) and cover a scope that is *very similar or broader* than the CTFPHC topic, we may update the search(es) if the last search date was prior to 6 months before commencing our review. When there are multiple reviews being considered, updating the literature to present may involve a new comprehensive search strategy to identify studies published after the date of the earliest existing review; this may reduce complexities when trying to implement, document, and remove duplicates from multiple searches. Alternatively, if the scope of the existing review(s) is *narrower* (e.g., missing an element in PICOTS) or the search *deemed sub-optimal in some manner* (e.g., missing key terms, additional database viewed as highly relevant) we may re-run the existing review's search concurrent with an original (e.g., broader) search and remove the citations previously screened for the other review. If more appropriate, we may update the other review's search and use a new search for the missing PICO element(s) (e.g., one additional intervention) for a longer time period to meet our timeframe. In cases where we feel screening excluded studies lists is appropriate we will also undertake this. Careful consideration will be used to ensure a comprehensive search is conducted regardless of approach taken; moreover, the ERSC librarians will help determine on a case-by-case basis what approach would be feasible for implementation to ensure aims of building efficiencies are possible.
3. **Screening and selection:** We will assess articles included in all relevant reviews (based on full text if necessary) to determine if they meet our inclusion criteria.
4. **Data extraction and methodological quality assessments:** We will consider incorporating the data on study and participant characteristics rather than extracting these data anew; we may also use the review author's risk of bias assessments if the tools/methods are consistent with CTFPHC methods. These steps will create efficiencies but because they are dependent on the quality of the systematic review and extent of reporting, the ERSC staff will verify the data on at least 5 to 10% of studies.<sup>1</sup>
5. **Data analysis:** We will consider using quantitative outcome data from reviews (with

verification), but will not typically use meta-analyses or quality (GRADE) assessments of existing reviews.

6. **Reporting:** Transparent reporting of all integration steps used will be included in the evidence review report.

### **Box 1. Recommendations developed by AHRQ EPCs\*<sup>1,2</sup>**

\*Strength of evidence refers to AHRQ's slightly modified approach to the GRADE quality of evidence approach

1. Existing reviews should be confirmed as systematic reviews through the application of a minimum set of eligibility criteria. We propose that the minimum eligibility criteria for systematic reviews include an explicit and adequate search, application of predefined eligibility criteria to select studies, risk of bias assessment for included studies, and synthesis of results.
2. Criteria to assess the relevance, in terms of question elements and currency, and quality of existing systematic reviews under consideration for inclusion in reviews should be predefined.
3. The quality of relevant existing systematic reviews should be assessed in an explicit manner with a minimum set of quality criteria that include search of multiple sources, use of a generally accepted tool for risk of bias assessment, and sufficient information to assess the strength of the body of evidence that includes the major domains of risk of bias, directness, consistency, precision, and reporting bias.
4. The risk of bias assessments from the existing systematic review may be used when the review described an explicit process, including the use of a tool or method that is compatible with the approach of the current review and that assessed the key sources of potential bias.
5. We suggest that risk of bias assessment be repeated in a sample of studies from an existing review under consideration for inclusion in a new review to confirm concordance with current review team approach.
6. We recommend that at a minimum, reviews should narratively describe findings of the prior review(s), including the number and types of studies included, and the overall findings.
7. We recommend that newly identified studies be clearly distinguished from studies in the existing review(s) when presented in the narrative and any tables (eg, separate tables).
8. Summary tables should include sufficient information to support ratings for overall strength of evidence, including ratings for individual strength of evidence domains (study limitations, consistency, precision, directness, reporting bias). The strength of evidence ratings should be based on the underlying primary evidence, not the number or quality of existing systematic reviews.
9. Using strength of evidence domains as a framework (study limitations, consistency, precision, directness, and reporting bias), review authors should consider how new evidence would change estimates of effect or ratings for strength of evidence. A new quantitative synthesis (ie, pooled estimate) is needed if new studies would change conclusions or strength of evidence judgements, or to obtain a more precise or more up-to-date estimate.
10. In cases where the existing systematic review(s) did not complete strength of evidence grading for a comparison and outcome of interest, the strength of evidence should be assessed for the body of evidence, considering primary studies from prior review(s) and any new studies identified.
11. In cases where no new studies are added to the body of evidence, the strength of evidence assessment from the existing systematic review may be used if conducted using an acceptable grading approach consistent with current review context. In these cases, we suggest that the overall strength of evidence assessment be reviewed, considering the strength of evidence domains, to confirm consistency with current review team assessments.

12. In cases where new studies are added to the body of evidence, the strength of evidence may need to be reassessed on the basis of all studies/evidence.

## **Appendix 2 References**

1. Robinson KA, Chou R, Berkman ND, et al. Integrating bodies of evidence: existing systematic reviews and primary studies. Methods Guide for Effectiveness and Comparative Effectiveness Reviews [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2015 Feb.
2. Robinson KA, Chou R, Berkman ND, et al. Twelve recommendations for integrating existing systematic reviews into new reviews: EPC guidance. J Clin Epidemiol. 2016 Feb;70:38-44. PMID: 26261004.
3. Foisy M, Hartling L. Challenges and considerations involved in using AMSTAR in overviews of reviews. 22<sup>nd</sup> Cochrane Colloquium. Hyderabad (India); 2014 Sept 21-26.