NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

This form contains the measure information submitted by stewards. Blank fields indicate no information was provided. Attachments also may have been submitted and are provided to reviewers. The subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. Hyperlinks to the evaluation criteria and ratings are provided in each section.

TAP/Workgroup (if utilized)**:** Complete all yellow highlighted areas of the form. Evaluate the extent to which each subcriterion is met. Based on your evaluation, summarize the strengths and weaknesses in each section.

Note: *If there is no TAP or workgroup, the SC also evaluates the subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, noting any areas of disagreement; then evaluate the extent to which each major criterion is met; and finally, indicate your recommendation for the endorsement. Provide the rationale for your ratings.

Evaluation ratings of the extent to which the criteria are met
C = Completely (unquestionably demonstrated to meet the criterion)
P = Partially (demonstrated to partially meet the criterion)
M = Minimally (addressed BUT demonstrated to only minimally meet the criterion)
N = Not at all (NOT addressed; OR incorrectly addressed; OR demonstrated to NOT meet the criterion)
NA = Not applicable (only an option for a few subcriteria as indicated)

(for NQF staff use) **NQF Review #:** ACP-002-10 **NQF Project:** Ambulatory Care - Additional Outpatient Measures 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Ultrasound determination of pregnancy location for pregnant patients with abdominal pain
De.2 **Brief description of measure**: Percentage of pregnant patients who present to the ED with a chief complaint of abdominal pain and or vaginal bleeding who receive a trans-abdominal or trans-vaginal ultrasound.
1.1-2 **Type of Measure:** Process De.3 **If included in a composite or paired with another measure, please identify composite or paired measure**
De.4 **National Priority Partners Priority Area:** Safety De.5 **IOM Quality Domain**: Patient-centered, Safety De.6 **Consumer Care Need:**

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** Proprietary measure A.3 **Measure Steward Agreement:** Agreement will be signed and submitted prior to or at the time of measure submission A.4 **Measure Steward Agreement attached:** acep agreement-634025703247393099.doc	**A** Y☐ N☐
B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and	**B**

update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	Y☐ N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement Accountability, Payment incentive, Accreditation	C Y☐ N☐
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1**Testing:** No, testing will be completed within 24 months D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	D Y☐ N☐
(for NQF staff use) **Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	**Met** Y☐ N☐
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:**	
Staff Reviewer Name(s):	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	Eval Ratin g
(for NQF staff use) Specific NPP goal:	
1a.1 **Demonstrated High Impact Aspect of Healthcare:** Leading cause of morbidity/mortality, Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** Ectopic Pregnancy is a relatively common condition which can result in morbidity or mortality if misdiagnosed resulting in a delay to appropriate treatment. Abdominal pain is a frequent presenting complaint of women with ruptured ectopic pregnancy. Pelvic ultrasound can establish a pregnancy as intrauterine and identify high risk features for ectopic pregnancy (pelvic free fluid, complex adnexal mass). Early ultrasound can shorten the time to diagnosis of ectopic pregnancy and can help risk stratify pregnant patients with the complaint of abdominal pain or vaginal bleeding for discharge with routine follow-up, discharge with early follow-up or admission. 1a.4 **Citations for Evidence of High Impact:** 1. Adhikari S. Blaivas M. Lyon M. Diagnosis and management of ectopic pregnancy using bedside transvaginal ultrasonography in the ED: a 2-year experience. American Journal of Emergency Medicine. 25(6):591-6, 2007 Jul. 2. Dart RG. Burke G. Dart L. Subclassification of indeterminate pelvic ultrasonography: prospective evaluation of the risk of ectopic pregnancy. Annals of Emergency Medicine. 39(4):382-8, 2002 Apr. 3. Choi H, Blaivas M, Lambert MJ. Gestational outcome in patients with first-trimester pregnancy complications and ultrasound-confirmed live intrauterine pregnancy. Acad Emerg Med. 2000;7(2):200-3. 4. Pandya PP, et al. The prevalence of non-viable pregnancy at 10–13 weeks of gestation. Ultrasound	**1a** C☐ P☐ M☐ N☐

Comment [KP1]: 1a. The measure focus addresses:
•a specific national health goal/priority identified by NQF's National Priorities Partners; OR
•a demonstrated high impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use (current and/or future), severity of illness, and patient/societal consequences of poor quality).

Obstet Gynecol 1996;7:170.
5. Hals G, Tolbert A. Vaginal bleeding during the first 20 weeks of pregnancy: guidelines for emergency department evaluation and management. Part 2: differential diagnosis and management of ectopic pregnancy and spontaneous miscarriage. Emerg Med Rep 2000;21:153–64.
6. Levi CS, Lyons EA, Lindsay DJ. Early diagnosis of nonviable pregnancy with endovaginal US. Radiology 1988; 167(2):383-385.

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: The measure will aid in both an access to care as well as raise the awareness to the importance to exclude ectopic pregnancy in patients presenting with the highest risk complaints associated with ectopic pregnancy.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
Mateer et al demonstrated a reduction in the rate of ruptured ectopic pregnancy from 50% (12/24) in historical controls to 9% (1/11) when using emergency ultrasound for pregnant patients presenting with abdominal pain or vaginal bleeding.

Mateer JR, Valley VT, Aiman EJ, Phelan MB, Thoma ME, Kefer MP. Outcome analysis of a protocol including bedside endovaginal sonography in patients at risk for ectopic pregnancy. Ann Emerg Med. Mar 1996;27(3):283-9. [Medline].

1b.3 Citations for data on performance gap:
1. Tayal VS, Cohen H, Norton HJ. Outcome of patients with an indeterminate emergency department first-trimester pelvic ultrasound to rule out ectopic pregnancy. Acad Emerg Med 2004 Sep;11(9):912-7
2. Gracia CR, Barnhart KT. Diagnosing ectopic pregnancy: decision analysis comparing six strategies. Obstet Gynecol 2001; 97:464-70.

1b.4 Summary of Data on disparities by population group:
None.

1b.5 Citations for data on Disparities:
None.

1b C☐ P☐ M☐ N☐

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** The incidence of ectopic pregnancy in women presenting to the emergency department with vaginal bleeding or pain in the first trimester has been estimated at approximately 10%, and it remains the leading cause of maternal death in the first trimester. Ultrasound identification effectively rules out an ectopic pregnancy, thereby reducing the maternal risk.

Albayram F, Hamper UM. First-trimester obstetric emergencies: spectrum of sonographic findings. J Clin Ultrasound 2002;30:161-77.

1c.2-3. Type of Evidence: Evidence-based guideline

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
Maternal Morbidity and Mortality - study group statement The Royal College of Obstetricians and Gynecologists. Clinicians and medical students must be made more aware of atypical clinical presentations of ectopic pregnancy and the option of beta human chorionic gonadotrophin testing in women with unexplained abdominal pain of recent onset. (Grade C)

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom*)**:**
Grade C

1c C☐ P☐ M☐ N☐

Comment [KP2]: 1b. Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating considerable variation, or overall poor performance, in the quality of care across providers and/or population groups (disparities in care).

Comment [k3]: 1 Examples of data on opportunity for improvement include, but are not limited to: prior studies, epidemiologic data, measure data from pilot testing or implementation. If data are not available, the measure focus is systematically assessed (e.g., expert panel rating) and judged to be a quality problem.

Comment [k4]: 1c. The measure focus is:
•an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;
OR
•if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
oIntermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
oProcess – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
oStructure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
oPatient experience – evidence that an association exists between the measure … [1]

Comment [k5]: 4 Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status – patients must be vaccinated to achieve … [2]

Comment [k6]: 3 The strength of the body of evidence for the specific measure focus should be systematically assessed and rated (e.g., USPSTF grading system http://www.ahrq.gov/clinic/uspstf07/methods/benefit.htm). If the USPSTF grading system was not used, the grading system is explained including how it relates to the USPSTF grades or why it does not. However, evidence is not limited to quantitative studies and the best type of evidence depends upon the question being studied (e.g., randomized controlled trials appropriate for studying drug efficacy are not well suited for complex system … [3]

1c.6 Method for rating evidence: Evidence Grading Recommendation - A: Requires at least one randomized controlled trial as part of the body of literature of overall good quality and consistency addressing the specific recommendation. B: Requires availability of well-conducted clinical studies but no randomised clinical trials on the topic of recommendation. C: Requires evidence from expert committee reports or opinions and/or clinical experience of respected authorities. Indicates absence of directly applicable studies of good quality.

1c.7 Summary of Controversy/Contradictory Evidence: None indicated.

1c.8 Citations for Evidence (*other than guidelines***):** N/A

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
"Use of emergency ultrasound in pelvic disorders centers on the detection of intrauterine pregnancy (IUP), detection of ectopic pregnancy, detection of fetal heart rate in all stages of pregnancy, dating of the pregnancy, and detection of significant free fluid. Bedside pelvic ultrasound during the first trimester of pregnancy can be used to exclude ectopic pregnancy by demonstrating an intrauterine pregnancy. Studies of EP-performed ultrasound in this setting have demonstrated sensitivity of 76-90% and specificity of 88- 92% for the detection of ectopic pregnancy.87,88 In one study, EPs were able to detect an intrauterine pregnancy in 70% of patients with suspected ectopic pregnancy (first trimester pregnancy with abdominal pain or vaginal bleeding).87 When intrauterine fetal anatomy was visualized at the bedside, ectopic pregnancy was ruled out with a negative predictive value of essentially 100%. When bedside ultrasound evaluation was incorporated into a clinical algorithm for the evaluation of patients with suspected ectopic pregnancy, the incidence of discharged patients returning with ruptured ectopic pregnancy was significantly reduced."

1c.10 Clinical Practice Guideline Citation: Emergency Ultrasound Guidelines. American College of Emergency Physicians. Ann Emerg Med. 2009; 53:562-563.
1c.11 National Guideline Clearinghouse or other URL:
http://www.guideline.gov/summary/summary.aspx?doc_id=8318&nbr=004650&string=first+AND+trimester+AND+vaginal+AND+bleeding; http://www.annemergmed.com/

1c.12 Rating of strength of recommendation (*also provide narrative description of the rating and by whom***):**
Published literature for clinical sonographic evaluation of early intrauterine pregnancy confirmation is Class II (both A & B) and Class III evidence.

1c.13 Method for rating strength of recommendation (*If different from* USPSTF system*, also describe rating and how it relates to USPSTF)***:**
1. Literature review in which emergency ultrasound applications graded in the Fryback-Pearl hierarchical model of effectiveness assessment.

2. Assessment of scientific evidence traditionally presented as classes:
- Class I Evidence: Randomized controlled trials (RCTs) are the gold standard
- Class II Evidence:
A. Data collected prospectively
B. Retrospective analyses from clearly reliable data
- Class III Evidence: Most studies based on retrospectively collected data

3. After completing an assessment of the scientific evidence, the confidence in recommending the use of clinical sonography can be rendered and presented as levels:
- Level 1: Convincingly justifiable based on the available scientific information alone
- Level 2: Reasonably justifiable by available scientific evidence and strongly supported by expert opinion
- Level 3: Supported by available data but adequate scientific evidence is lacking.

1c.14 Rationale for using this guideline over others:
Strength of evidence.

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report?* **1**

Comment [k7]: USPSTF grading system **http://www.ahrq.gov/clinic/uspstf/grades.htm**: **A -** The USPSTF recommends the service. There is high certainty that the net benefit is substantial. **B -** The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. **C -** The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is at least moderate certainty that the net benefit is small. Offer or provide this service only if other considerations support the offering or providing the service in an individual patient. **D -** The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. **I -** The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

Steering Committee: Was the threshold criterion, *Importance to Measure and Report*, **met?** **Rationale:**	**1** **Y☐** **N☐**
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	Eval Rating
2a. MEASURE SPECIFICATIONS	
S.1 Do you have a web page where current detailed measure specifications can be obtained? **S.2 If yes, provide web page URL:** **2a. Precisely Specified**	
2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):** Number of appropriate patients who receive a trans-abdominal or trans-vaginal ultrasound. **2a.2 Numerator Time Window (***The time period in which cases are eligible for inclusion in the numerator***):** None. **2a.3 Numerator Details (***All information required to collect/calculate the numerator, including all codes, logic, and definitions***):** • CPT E/M Service Codes: 99281, 99282, 99283, 99284, 99285, 99291 and • patients who receive a trans-abdominal or trans-vaginal ultrasound during their ED visit (Recommend new CPT2 or G codes be created).	
2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):** All pregnant patients who present to the ED with a chief complaint of lower abdominal pain, and or vaginal bleeding. **2a.5 Target population gender:** Female **2a.6 Target population age range:** 14 to 50 **2a.7 Denominator Time Window (***The time period in which cases are eligible for inclusion in the denominator***):** None. **2a.8 Denominator Details (***All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):** CPT E/M Service Codes: 99281, 99282, 99283, 99284, 99285, 99291 Additional codes: ICD-9-CM 789.0, 789.1, 789.2, 789.3, 789.4, 789.5, 789.6, 789.7, 789.9; ICD-9-CM 623.8.	
2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):** 1. Women for whom location of pregnancy is already documented or reported as intra-uterine 2. Patient refusal 3. Ultrasound is not feasible (facility reason) 4. US machine not available • ED does not have access to ultrasound 5. Licensed independent provider not credentialed in ultrasound. **2a.10 Denominator Exclusion Details (***All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):** Chart review evidence of trans-abdominal or trans-vaginal ultrasound.	**2a-specs** **C☐** **P☐** **M☐** **N☐**
2a.11 Stratification Details/Variables (*All information required to stratify the measure including the*	

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable

Comment [KP8]: 2a. The measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability. The required data elements are of high quality as defined by NQF's Health Information Technology Expert Panel (HITEP) .

Comment [k9]: 11 Risk factors that influence outcomes should not be specified as exclusions.
12 Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

stratification variables, all codes, logic, and definitions)**:** None.	
2a.12-13 **Risk Adjustment Type:** No risk adjustment necessary 2a.14 **Risk Adjustment Methodology/Variables (***List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):** N/A 2a.15-17 **Detailed risk model available Web page URL or attachment:**	
2a.18-19 **Type of Score:** Count 2a.20 **Interpretation of Score:** 2a.21 **Calculation Algorithm (***Describe the calculation of the measure as a flowchart or series of steps***):** N/A	
2a.22 **Describe the method for discriminating performance** *(e.g., significance testing)***:** N/A	
2a.23 **Sampling (Survey) Methodology** *If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate)*: N/A	
2a.24 **Data Source (***Check the source(s) for which the measure is specified and tested)* Paper medical record/flow-sheet, Electronic administrative data/claims, Electronic clinical data, Electronic Health/Medical Record 2a.25 **Data source/data collection instrument (***Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.***):** Data will be collected from the medical record. These can be easily recorded either electronically or on paper using institution-specific instruments. 2a.26-28 **Data source/data collection instrument reference web page URL or attachment:** 2a.29-31 **Data dictionary/code table web page URL or attachment:** 2a.32-35 **Level of Measurement/Analysis** (*Check the level(s) for which the measure is specified and tested)* Clinicians: Individual, Clinicians: Group, Can be measured at all levels 2a.36-37 **Care Settings (***Check the setting(s) for which the measure is specified and tested)* Ambulatory Care: Emergency Dept 2a.38-41 **Clinical Services** (*Healthcare services being measured, check all that apply*) Clinicians: Physicians (MD/DO)	
TESTING/ANALYSIS	
2b. Reliability testing 2b.1 **Data/sample** *(description of data/sample and size)***:** ACEP has not conducted testing, and the measure is not in current use. 2b.2 **Analytic Method** *(type of reliability & rationale, method for testing)***:** N/A 2b.3 **Testing Results** *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:** N/A	**2b** C☐ P☐ M☐ N☐
2c. Validity testing	**2c** C☐

Comment [KP10]: 2b. Reliability testing demonstrates the measure results are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period.

Comment [k11]: 8 Examples of reliability testing include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing may address the data items or final measure score.

Comment [KP12]: 2c. Validity testing demonstrates that the measure reflects the quality of care provided, adequately distinguishing good and poor quality. If face validity is the only validity addressed, it is systematically assessed.

2c.1 Data/sample *(description of data/sample and size)*: N/A

2c.2 Analytic Method *(type of validity & rationale, method for testing)*:
N/A

2c.3 Testing Results *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
N/A

P☐ M☐ N☐

2d. Exclusions Justified

2d.1 Summary of Evidence supporting exclusion(s):
N/A

2d.2 Citations for Evidence:
N/A

2d.3 Data/sample *(description of data/sample and size)*: N/A

2d.4 Analytic Method *(type analysis & rationale)*:
N/A

2d.5 Testing Results *(e.g., frequency, variability, sensitivity analyses)*:
N/A

2d C☐ P☐ M☐ N☐ NA☐

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample *(description of data/sample and size)*: N/A

2e.2 Analytic Method *(type of risk adjustment, analysis, & rationale)*:
N/A

2e.3 Testing Results *(risk model performance metrics)*:
N/A

2e.4 If outcome or resource use measure is not risk adjusted, provide rationale: N/A

2e C☐ P☐ M☐ N☐ NA☐

2f. Identification of Meaningful Differences in Performance

2f.1 Data/sample from Testing or Current Use *(description of data/sample and size)*: N/A

2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance *(type of analysis & rationale)*:
N/A

2f.3 Provide Measure Scores from Testing or Current Use *(description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:
N/A

2f C☐ P☐ M☐ N☐

2g. Comparability of Multiple Data Sources/Methods

2g.1 Data/sample *(description of data/sample and size)*: N/A

2g.2 Analytic Method *(type of analysis & rationale)*:
N/A

2g.3 Testing Results *(e.g., correlation statistics, comparison of rankings)*:
N/A

2g C☐ P☐ M☐ N☐ NA☐

2h. Disparities in Care

2h

Comment [k13]: 9 Examples of validity testing include, but are not limited to: determining if measure scores adequately distinguish between providers known to have good or poor quality assessed by another valid method; correlation of measure scores with another valid indicator of quality for the specific topic; ability of measure scores to predict scores on some other related valid measure; content validity for multi-item scales/tests. Face validity is a subjective assessment by experts of whether the measure reflects the quality of care (e.g., whether the proportion of patients with BP < 140/90 is a marker of quality). If face validity is the only validity addressed, it is systematically assessed (e.g., ratings by relevant stakeholders) and the measure is judged to represent quality care for the specific topic and that the measure ... [4]

Comment [KP14]: 2d. Clinically necessary measure exclusions are identified and must be:
•supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;
AND ... [5]

Comment [k15]: 10 Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, sensitivity analyses with and without the exclusion, and variability of exclusions across providers.

Comment [KP16]: 2e. For outcome measures and other measures (e.g., resource use) when indicated:
•an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outc ... [6]

Comment [k17]: 13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in trea ... [7]

Comment [KP18]: 2f. Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.

Comment [k19]: 14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage ... [8]

Comment [KP20]: 2g. If multiple data sources/methods are allowed, there is demonstration they produce comparable results.

Comment [KP21]: 2h. If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender);OR rationale/data justifies why stratification is not necessary or not feasible.

2h.1 If measure is stratified, provide stratified results *(scores by stratified categories/cohorts)***:** N/A **2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:** N/A	C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties***, met?** **Rationale:**	*2* C☐ P☐ M☐ N☐
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	Eval Rating
3a. Meaningful, Understandable, and Useful Information **3a.1 Current Use:** Testing not yet completed **3a.2 Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:** This measure is not in use. **3a.3 If used in other programs/initiatives (***If used in quality improvement or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years***):** N/A **Testing of Interpretability** (*Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement*) **3a.4 Data/sample** *(description of data/sample and size)***:** N/A **3a.5 Methods** *(e.g., focus group, survey, QI project)***:** N/A **3a.6 Results** *(qualitative and/or quantitative results and conclusions)***:** N/A	**3a** C☐ P☐ M☐ N☐
3b/3c. Relation to other NQF-endorsed measures **3b.1 NQF # and Title of similar or related measures:**	
(for NQF staff use) **Notes on similar/related** endorsed **or submitted measures:**	
3b. Harmonization If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population): **3b.2 Are the measure specifications harmonized? If not, why?**	**3b** C☐ P☐ M☐ N☐ NA☐
3c. Distinctive or Additive Value **3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:**	**3c** C☐ P☐

Comment [KP22]: 3a. Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audience(s) for both public reporting (e.g., focus group, cognitive testing) and informing quality improvement (e.g., quality improvement initiatives). An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement.

Comment [KP23]: 3b. The measure specifications are harmonized with other measures, and are applicable to multiple levels and settings.

Comment [k24]: 16 Measure harmonization refers to the standardization of specifications for similar measures on the same topic (e.g., *influenza immunization* of patients in hospitals or nursing homes), or related measures for the same target population (e.g., eye exam and HbA1c for *patients with diabetes*), or definitions applicable to many measures (e.g., age designation for children) so that they are uniform or compatible, unless differences are dictated by the evidence. The dimensions of harmonization can include numerator, denominator, exclusions, and data source and collection instructions. The extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources.

Comment [KP25]: 3c. Review of existing endorsed measures and measure sets demonstrates that the measure provides a distinctive or additive value to existing NQF-endorsed measures (e.g., provides a more complete picture of quality for a particular condition or aspect of healthcare, is a more valid or efficient way to measure).

5.1 If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality: N/A	**M☐ N☐ NA☐**
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability*, met? **Rationale:**	**3** **C☐ P☐ M☐ N☐**
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)	Eval Rating
4a. Data Generated as a Byproduct of Care Processes **4a.1-2 How are the data elements that are needed to compute measure scores generated?** Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)	**4a** **C☐ P☐ M☐ N☐**
4b. Electronic Sources **4b.1 Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*) No **4b.2 If not, specify the near-term path to achieve electronic capture by most providers.** In EDs where EMR is present data elements will be available electronically, as adoption improves, electronic capture will improve.	**4b** **C☐ P☐ M☐ N☐**
4c. Exclusions **4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** No **4c.2 If yes, provide justification.**	**4c** **C☐ P☐ M☐ N☐ NA☐**
4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences **4d.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results.** N/A	**4d** **C☐ P☐ M☐ N☐**
4e. Data Collection Strategy/Implementation **4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** Measure has not been tested by ACEP. **4e.2 Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** The costs to implement this measure will depend on the method used to collect data. Personnel time will be needed if paper medical records are to be reviewed in order to determine whether patients received	**4e** **C☐ P☐ M☐ N☐**

Comment [KP26]: 4a. For clinical measures, required data elements are routinely generated concurrent with and as a byproduct of care processes during care delivery. (e.g., BP recorded in the electronic record, not abstracted from the record later by other personnel; patient self-assessment tools, e.g., depression scale; lab values, meds, etc.)

Comment [KP27]: 4b. The required data elements are available in electronic sources. If the required data are not in existing electronic sources, a credible, near-term path to electronic collection by most providers is specified and clinical data elements are specified for transition to the electronic health record.

Comment [KP28]: 4c. Exclusions should not require additional data sources beyond what is required for scoring the measure (e.g., numerator and denominator) unless justified as supporting measure validity.

Comment [KP29]: 4d. Susceptibility to inaccuracies, errors, or unintended consequences and the ability to audit the data items to detect such problems are identified.

Comment [KP30]: 4e. Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, etc.) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use).

trans-abdominal or trans-vaginal ultrasound in the ED. 4e.3 **Evidence for costs:** Not available. 4e.4 **Business case documentation:** Not available.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?	**4**
Steering Committee: Overall, to what extent was the criterion, *Feasibility*, **met?** **Rationale:**	**4** **C☐** **P☐** **M☐** **N☐**
RECOMMENDATION	
(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**	**Time-limited ☐**
Steering Committee: Do you recommend for endorsement? **Comments:**	**Y☐** **N☐** **A☐**

CONTACT INFORMATION

Co.1 **Measure Steward (Intellectual Property Owner)**
Co.1 **Organization**
American College of Emergency Physicians, 2121 K Street NW, Suite 325, Washington, District Of Columbia, 20037

Co.2 **Point of Contact**
Angela , Franklin, JD, afranklin@acep.org, 202-728-0610-3014

Measure Developer If different from Measure Steward
Co.3 **Organization**
American College of Emergency Physicians, 2121 K Street NW, Suite 325, Washington, District Of Columbia, 20037

Co.4 **Point of Contact**
Angela , Franklin, JD, afranklin@acep.org, 202-728-0610-3014

Co.5 **Submitter If different from Measure Steward POC**
Angela , Franklin, JD, afranklin@acep.org, 202-728-0610-3014, American College of Emergency Physicians

Co.6 **Additional organizations that sponsored/participated in measure development**

ADDITIONAL INFORMATION

Workgroup/Expert Panel involved in measure development
Ad.1 **Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.**
The following workgroup developed the measure:

Co-CHAIR
Stephen V. Cantrill, MD FACEP
937 S. Emporia Street
Denver, CO 80247-1900
(W) 303.436.7174
(W-Fax) 303.436.7541
stephen.cantrill@dhha.org

Co-CHAIR

Jeremiah Schuur, MD
Brigham & Women's Hospital
75 Francis Street
Boston, MA 02115
Phone: 617.732-5636
(C) 401.480.7468
(Fax) 617.264.6848
jschuur@partners.org

Brent R. Asplin, MD MPH FACEP
Chair, Dept of Emergency Medicine
Mayo Clinic/GE GR G-410
200 First Street SW
Rochester, MN 55905
(W): (507) 255-6501 (Sue Kirk)
(C) 651.261.7939
Asplin.Brent@mayo.edu

Christopher Baugh, MD (EMRA)
1163 Beacon Street, Apt. 4
Brookline, MA 02446-5512
(H) 617.935.3589
(W) 617.732.8070
(W-Fax) 617.264.6848
cbaugh@partners.org

Robert I. Broida, MD FACEP
P.O. Box 5404
Akron, OH 44334-0404
(W) 330.493.4443 ext. 1307
(W-Fax) 330.491-4088
rbroida@emp.com

Dickson S. Cheung, MD
10360 Bluffmont Dr
Lone Tree, CO 80124-5579
(H) 303.662.9999
(C) 303.956.7381
dscheung@alum.mit.edu

William C. Dalsey, MD MBA FACEP
945 Len Mar Drive
Blue Bell, PA 19422-2000
(H) 215.654.1190
(Fax) 215.643.8787
wcderdoc@aol.com

Enrique R. Enguidanos, MD FACEP
North Sound Emergency Medicine
1001 N. Broadway, Suite A11
Everett, WA 98201-1582
(W) 425.259.0212
(H) 206.522.5935
(W-Fax) 425.259.0209
enrique.enguidanos@providence.org

David P. John, MD FACEP

Caritas Carney Hospital
2100 Dorchester Ave
Dorchester, MA 02124-5666
(W) 617.506.4463
(C) 203.671.5972
David.John@caritaschristi.org

Helmut W. Meisl, MD FACEP
Good Samaritan Hospital
2425 Samaritan Drive
San Jose, CA 95124
(W) 408.559.2552
(C) 650.283.7345
hmeisl@earthlink.net

Neal P. O'Connor, MD FACEP
Medical Center of Aurora
1501 S. Potomac
Aurora, CO 80012-5411
(W) 303.436.2721
(C) 303.589.9172
no'connor@carepointpc.com

Shari J. Welch, MD FACEP
3822 Brockbank Drive
Salt Lake City, UT 84124-3954
(H) 801.943.3308
sjwelch56@aol.com

Richard T. Griffey MD MPH FACEP
Division of Emergency Medicine
Washington University School of Medicine
Campus Box 8072
660 S. Euclid
St. Louis MO 63110
(w) 314-747-4899
(h) 314-644-1799
griffeyr@wustl.edu

Ad.2 If adapted, provide name of original measure: N/A
Ad.3-5 If adapted, provide original specifications URL or attachment

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.6 Year the measure was first released:
Ad.7 Month and Year of most recent revision:
Ad.8 What is your frequency for review/update of this measure? This is a newly developed measure, however we expect to review at least every 3 years
Ad.9 When is the next scheduled review/update for this measure? 10, 2012

Ad.10 Copyright statement/disclaimers: N/A

Ad.11 -13 Additional Information web page URL or attachment:

Date of Submission (*MM/DD/YY*)**:** 07/20/2010

Page 3: [1] Comment [k4] **Karen Pace** **10/5/2009 8:59:00 AM**

1c. The measure focus is:

- an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;

OR

- if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
  - Intermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
  - Process – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
    if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
  - Structure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
  - Patient experience – evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
  - Access – evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
  - Efficiency – demonstration of an association between the measured resource use and level of performance with respect to one or more of the other five IOM aims of quality.

Page 3: [2] Comment [k5] **Karen Pace** **10/5/2009 8:59:00 AM**

4 Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status – patients must be vaccinated to achieve immunity. This does not preclude consideration of measures of preventive screening interventions where there is a strong link with desired outcomes (e.g., mammography) or measures for multiple care processes that affect a single outcome.

Page 3: [3] Comment [k6] **Karen Pace** **10/5/2009 8:59:00 AM**

3 The strength of the body of evidence for the specific measure focus should be systematically assessed and rated (e.g., USPSTF grading system http://www.ahrq.gov/clinic/uspstf07/methods/benefit.htm). If the USPSTF grading system was not used, the grading system is explained including how it relates to the USPSTF grades or why it does not. However, evidence is not limited to quantitative studies and the best type of evidence depends upon the question being studied (e.g., randomized controlled trials appropriate for studying drug efficacy are not well suited for complex system changes). When qualitative studies are used, appropriate qualitative research criteria are used to judge the strength of the evidence.

Page 7: [4] Comment [k13] **Karen Pace** **10/5/2009 8:59:00 AM**

9 Examples of validity testing include, but are not limited to: determining if measure scores adequately distinguish between providers known to have good or poor quality assessed by another valid method; correlation of measure scores with another valid indicator of quality for the specific topic; ability of measure scores to predict scores on some other related valid measure; content validity for multi-item scales/tests. Face validity is a subjective assessment by experts of whether the measure reflects the quality of care (e.g., whether the proportion of patients with BP < 140/90 is a marker of quality). If face validity is the only validity addressed, it is systematically assessed (e.g., ratings by relevant stakeholders) and the measure is judged to represent quality care for the specific topic and that the measure focus is the most important aspect of quality for the specific topic.

Page 7: [5] Comment [KP14] **Karen Pace** **10/5/2009 8:59:00 AM**

2d. Clinically necessary measure exclusions are identified and must be:

- supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;

AND

- a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;

AND

if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

Page 7: [6] Comment [KP16] **Karen Pace** **10/5/2009 8:59:00 AM**

2e. For outcome measures and other measures (e.g., resource use) when indicated:

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care;[Error! Bookmark not defined.] OR

rationale/data support no risk adjustment.

Page 7: [7] Comment [k17] **Karen Pace** **10/5/2009 8:59:00 AM**

13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than adjusting out differences.

Page 7: [8] Comment [k19] **Karen Pace** **10/5/2009 8:59:00 AM**

14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% v. 75%) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall poor performance may not demonstrate much variability across providers.