NATIONAL QUALITY FORUM

Measure Submission and Evaluation Worksheet 5.0

This form contains the information submitted by measure developers/stewards, organized according to NQF's measure evaluation criteria and process. The evaluation criteria, evaluation guidance documents, and a blank online submission form are available on the submitting standards web page.

NQF #: 0502   **NQF Project:** Perinatal and Reproductive Health Project
(for Endorsement Maintenance Review) **Original Endorsement Date:** Oct 24, 2008   **Most Recent Endorsement Date:** Oct 24, 2008

BRIEF MEASURE INFORMATION
De.1 Measure Title: Pregnancy test for female abdominal pain patients.
Co.1.1 Measure Steward: American College of Emergency Physicians
De.2 Brief Description of Measure: Percentage of female patients aged 14 to 50 who present to the emergency department (ED) with a chief complaint of abdominal pain for whom a pregnancy test ordered
2a1.1 Numerator Statement: Number of patients in the denominator who have a pregnancy test (urine or serum) ordered in the ED
2a1.4 Denominator Statement: All women, ages 14 – 50 years old, who present to the ED with a chief complaint of abdominal pain.
2a1.8 Denominator Exclusions: i. Females for whom pregnancy is already documented or reported (verbal report by patient is acceptable). ii. Females with documented or reported hysterectomy (verbal report by patient is acceptable). iii. Females documented or reported to be post-menopausal (verbal report by patient is acceptable). iv. Patient refusal v. Patients who do not complete their ED evaluation (Left before completion, Left AMA, etc.)
1.1 Measure Type: Process **2a1. 25-26 Data Source:** Administrative claims, Electronic Clinical Data : Electronic Health Record, Paper Records **2a1.33 Level of Analysis:** Clinician : Group/Practice, Clinician : Individual, Facility **1.2-1.4 Is this measure paired with another measure?** No **De.3 If included in a composite, please identify the composite measure** (*title and NQF number if endorsed*)**:**

STAFF NOTES (*issues or questions regarding any criteria*)
Comments on Conditions for Consideration:
Is the measure untested?   Yes☐   No☐   If untested, explain how it meets criteria for consideration for time-limited endorsement:
1a. Specific national health goal/priority identified by DHHS or NPP addressed by the measure (*check De.5*): 5. Similar/related endorsed or submitted measures (*check 5.1*): Other Criteria:
Staff Reviewer Name(s):

1. IMPACT, OPPORTUNITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.

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: H☐ M☐ L☐ I☐
(*The measure directly addresses a specific national health goal/priority identified by DHHS or NPP, or some other high impact aspect of healthcare.*)

De.4 Subject/Topic Areas (*Check all the areas that apply*): Perinatal
De.5 Cross Cutting Areas (*Check all the areas that apply*): Safety

1a.1 Demonstrated High Impact Aspect of Healthcare: Affects large numbers, Patient/societal consequences of poor quality

1a.2 If "Other," please describe:

1a.3 Summary of Evidence of High Impact (*Provide epidemiologic or resource use data*):
Ectopic Pregnancy is a 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. Consensus clinical guidelines recommend that women of reproductive age who present to an ED with atraumatic abdominal pain should receive a pregnancy test.

Pregnancy testing is recommended in teh Emergency Department for females who might be pregnant because clinical history is unreliable (Ann Emerg Med 1989). The impportance of pregnancy diagnosis is particularly true in pateints with abdominal pain and/or piror to radiologic procedures where failure to diagnose pregnancy is a risk to the woman and her unborn child.

1a.4 Citations for Evidence of High Impact cited in 1a.3: American College of Emergency Physicians. Clinical Policy: Critical Issues for the Initial Evaluation and Management of Patients Presenting With a Chief Complaint of Nontraumatic Acute Abdominal Pain. Annals of Emergency Medicine 36: 406-415l 2000 Oct.

Ramoska EA, Sacchetti AD, Nepp M, Reliability of patient history in determing the possibility of pregnancy. Ann Emerg Med 1989; 18:48-50.

1b. Opportunity for Improvement: H☐ M☐ L☐ I☐
(*There is a demonstrated performance gap - variability or overall less than optimal performance*)

1b.1 Briefly explain the benefits (improvements in quality) envisioned by use of this measure:
Use of the measure can eliminate the risk of the physician failing to diagnose a patient is pregnancy, thereby reducing the possibility that a patient with ectopic pregnancy is not identified.

1b.2 Summary of Data Demonstrating Performance Gap (*Variation or overall less than optimal performance across providers*):
[*For Maintenance – Descriptive statistics for performance results for this measure - distribution of scores for measured entities by quartile/decile, mean, median, SD, min, max, etc.*]
Two studies have addressed this recently:

1)	Schuur et al. found a large gap in a national database of US EDs (67%). In a chart review at 4 northeaster academic EDs, they found a smaller performance gap (6%)
STUDY OBJECTIVE: We assess performance and explore definitions for a new emergency department (ED) quality measure: the proportion of women aged 14 to 50 years who have abdominal pain and receive pregnancy testing (aimed at detecting ectopic pregnancy).
METHODS: We analyzed data from the National Hospital Ambulatory Medical Care Survey (NHAMCS) (2002 to 2006) to test trends and predictors of the new measure, using both restrictive and broad definitions from the International Classification of Diseases, Ninth Revision (ICD-9) and reason-for-visit codes, and determine the proportion of women with ectopic pregnancy who had undergone pregnancy testing. For comparison, we conducted a detailed chart review in 4 US hospitals among patients who visited the ED in 2006.
RESULTS: Using a broad ICD-9 definition for inclusion in NHAMCS, 2.13 million women aged 14 to 50 years with abdominal pain

visited an ED annually between 2002 and 2006. Of those, 33.0% (95% confidence interval [CI] 30.5% to 35.5%) received pregnancy testing. Testing rates were materially stable, regardless of the definition used (broad or restrictive ICD-9 or reason-for-visit code). Among women with an ICD-9 diagnosis of ectopic pregnancy, 55.6% (95% CI 43.7% to 67.6%) had a documented pregnancy test. In the chart review, among 200 women aged 14 to 50 years and with abdominal pain, 89.4% (95% CI 85.0% to 94.0%) were eligible for the measure; of those, 93.9% (95% CI 90.3% to 97.4%) received testing.
CONCLUSION: Analysis of national ED survey data demonstrated a large performance gap for a new pregnancy testing quality measure, whereas focused chart review at 4 sites showed a smaller gap. Given these discrepancies, additional study is recommended before the widespread implementation of the pregnancy testing measure as an assessment of ED performance.

2)   L. Graff evaluated the measure in a large community ED in the northeast and found a gap of ~20% in 2 pre-intervention years and a gap of ~4% in 2010 after a quality improvement intervention. (Pre-publication data)

NQF Quality Measure #502: rate of HCG testing among women age 14 through 50 with chief complaint of abdominal pain

Improved quality of patient care with feedback of NQF HCG Quality Measure #502 (rate of HCG testing among women age 14 through 50 with chief complaint of abdominal pain) and Adverse Events (CT scan imaging of the abdomen or chest with pregnancy testing)

Background: Pregnancy testing is recommended in the Emergency Department, (ED) particularly for patients with abdominal pain and/or prior to radiologic procedures, because clinical history is unreliable and failure to diagnose pregnancy is a risk to the woman and her unborn child.

Methods: Two community teaching hospital EDs with combined 105,000 patient visits per year participated in an initiative to improve the use of urine or serum human chorionic gonadotropin HCG testing in diagnosing pregnancy. CY 2008 served as a control period. During CY 2009, all ED physicians (MDs) and physician assistants (PAs) were given each month their HCG testing rate (rate of HCG testing among women age 14 through 50 years). During CY 2010 all ED MDs and PAs were given each month 1) their NQF HCG Quality Measure #502 rate (rate of HCG testing among women age 14 through 50 with chief complaint of abdominal pain), 2) their cases that failed the Quality Measure, and 3) their cases that had an adverse event (CT scan imaging of the abdomen or chest or back without pregnancy testing to confirm the patient not pregnant). Record abstraction was used to determine exclusions from the Quality Measure (patient known to be pregnant; status post hysterectomy; status post menopause; patient refusal) or the adverse event (CT scan imaging).

Results: During the three years examined, the number of women age 14 through 50 remained constant: 30,521 in 2008; 34,747 in 2009; 32,855 in 2010, of which 4172, 4717, and 3566 had chief complaint of abdominal pain in 2008, 2009, and 2010, respectively. The HCG testing rate was 31.7%(95% CI 30.3% 33.1%) during 2008 (control), and did not increase 2009 (feedback of HCG testing rates) 32.5% (95% CI 31.2% 33.9%), but did improve with feedback of the Quality Measure and adverse events in 2010 40.6% (95% CI 39.0 42.2). Similarly, only with feedback of the Quality Measure and adverse events did performance improve on the HCG Quality Measure from 80.2% (95% CI 78.9% 81.3%) in 2008 and 80.3 (95% CI 79.2% 81.5%) in 2009 to 96% (95% CI 95.4%, 96.6%) in 2010 as did performance improve on the Adverse Event rate from 3.2% (95% CI 2.7% 3.7%) in 2008, 3.5% (95% CI 2.9% 4.0%) in 2009 to 0.7% (95% CI 0.5%,1.0%) in 2010 .

Conclusion: Clinician performance did not improve with individual feedback of HCG testing rates but did improve with individual feedback of HCG Quality Measure and adverse events. The construction of clinical performance measures has important implications concerning how clinicians respond to feedback on their performance. Raw test utilization data was ineffective but a quality measure with 100% theoretical maximum performance and the ability to delineate adverse events appears actionable by clinicians.

1b.3 **Citations for Data on Performance Gap:** [*For Maintenance – Description of the data or sample for measure results reported in 1b.2 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included*]
1)   Schuur JD, Tibbetts SA, Pines JM. Pregnancy testing in women of reproductive age in US emergency departments, 2002 to 2006: assessment of a national quality measure. Ann Emerg Med. 2010 May;55(5):449-457.e2. Epub 2009 Nov 22.
2)   L. Graff, 2010, Pre-publication summary attached. NOT FO DISTRIBUTION. Available to comment at Louisgraff4@aol.com

1b.4 Summary of Data on Disparities by Population Group: [*For <u>Maintenance</u> –Descriptive statistics for performance results for this measure by population group*]

Schuur et al found that in the NHAMCS data, lower testing rates were associated with increasing age, admission to hospital, Medicare insurance, provider characteristics (no resident/intern or physician assistant/nurse practitioner involvement), and hospital location in the western region. They found no yearly trend and no difference by race/ethnicity. No disparities were identified in the chart review data.

1b.5 Citations for Data on Disparities Cited in 1b.4: [*For <u>Maintenance</u> – Description of the data or sample for measure results reported in 1b.4 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included*]

3)   Schuur JD, Tibbetts SA, Pines JM. Pregnancy testing in women of reproductive age in US emergency departments, 2002 to 2006: assessment of a national quality measure. Ann Emerg Med. 2010 May;55(5):449-457.e2. Epub 2009 Nov 22.

1c. Evidence (*Measure focus is a health outcome OR meets the criteria for quantity, quality, consistency of the body of evidence.*)
Is the measure focus a health outcome?   Yes☐ No☐   <u>If not a health outcome</u>, rate the body of evidence.

Quantity: H☐ M☐ L☐ I☐   Quality: H☐ M☐ L☐ I☐   Consistency: H☐ M☐ L☐ I☐

Quantity	Quality	Consistency	Does the measure pass subcriterion1c?
M-H	M-H	M-H	Yes☐
L	M-H	M	Yes☐ IF additional research unlikely to change conclusion that benefits to patients outweigh harms: otherwise **No**☐
M-H	L	M-H	Yes☐ IF potential benefits to patients clearly outweigh potential harms: otherwise **No**☐
L-M-H	L-M-H	L	No ☐

Health outcome – rationale supports relationship to at least one healthcare structure, process, intervention, or service	Does the measure pass subcriterion1c? Yes☐ IF rationale supports relationship

1c.1 Structure-Process-Outcome Relationship (*Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome*):
process-health outcome

1c.2-3 Type of Evidence (*Check all that apply*):
Clinical Practice Guideline, Selected individual studies (rather than entire body of evidence)

1c.4 Directness of Evidence to the Specified Measure (*State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population*):
L. Graff´s study was the exact measure on the general emergency department population at a large community teaching hospital and the outcomes clearly show the value of the quality measure.

1c.5 Quantity of Studies in the Body of Evidence (*Total number of studies, not articles*): 1 study

1c.6 Quality of Body of Evidence (Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events): a)   There were several limitations of this study. There is variability in the use of chief complaints between clinicians since many patients with abdominal pain present with multiple complaints that the clinician could choose instead such as vomiting or nausea or back pain or diarrhea. This was a pre post study and a prospective trial would be stronger evidence. Implementation of the cohort quality measure will be difficult at most hospitals. At the two study emergency departments only selective components of PQRI quality measure are defined fields with data dictionary in the electronic chart. Other components of the quality measure needed chart review to find the presence or

absence of documentation of exclusion criteria. This makes it a very labor intensive measure. B) the study was direct evidence of the value of the measure,   c) the confidence intervals are consistent with a large size study with approximately 4,000 patients each year for the 3 years of the study / performance improvement initiative.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* The findings of the study were consistent across the range of analyses in the study

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)***:**
The harm at base line was 3 ½ percent of the patients were having adverse effects (CT imaging of the abdomen or chest without pregnancy test to confirm they were not pregnant) and with the quality measure this was reduced to less than 1%

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? No

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias:

1c.11 System Used for Grading the Body of Evidence: Other

1c.12 If other, identify and describe the grading scale with definitions: n/a

1c.13 Grade Assigned to the Body of Evidence:

1c.14 Summary of Controversy/Contradictory Evidence:

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)***:**
4)   L. Graff, 2010, Pre-publication summary attached. NOT FOR DISTRIBUTION. Available to comment at Louisgraff4@aol.com

1c.16 Quote verbatim, <u>the specific guideline recommendation</u> *(Including guideline # and/or page #)***:**

1c.17 Clinical Practice Guideline Citation: American College of Emergency Physicians. Clinical Policy: Critical Issues for the Initial Evaluation and Management of Patients Presenting With a Chief Complaint of Nontraumatic Acute Abdominal Pain. Annals of Emergency Medicine 36: 406-415l 2000 Oct.

1c.18 National Guideline Clearinghouse or other URL:

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? No

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias:

1c.21 System Used for Grading the Strength of Guideline Recommendation: Other

1c.22 If other, identify and describe the grading scale with definitions: n/a

1c.23 Grade Assigned to the Recommendation:

1c.24 Rationale for Using this Guideline Over Others: This is the only available guideline that we found in a thorough literature review.

Based on the NQF descriptions for rating the evidence, what was the <u>developer's assessment</u> of the quantity, quality, and consistency of the body of evidence?
1c.25 Quantity: Low   **1c.26 Quality:** Moderate**1c.27 Consistency:** Moderate

Was the threshold criterion, *Importance to Measure and Report*, met?

Was the threshold criterion, *Importance to Measure and Report*, met?
(*1a & 1b must be rated moderate or high and 1c yes*)   Yes☐ No☐
Provide rationale based on specific subcriteria:

For a new measure if the Committee votes NO, then STOP.
For a measure undergoing endorsement maintenance, if the Committee votes NO because of 1b. (no opportunity for improvement), it may be considered for continued endorsement and all criteria need to be evaluated.

2. RELIABILITY & VALIDITY - SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES

Extent to which the measure, <u>as specified</u>, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (<u>evaluation criteria</u>)
Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See <u>guidance on measure testing</u>.

S.1 **Measure Web Page** *(In the future, NQF will require measure stewards to provide a URL link to a web page where current detailed specifications can be obtained)*. Do you have a web page where current detailed specifications for <u>this</u> measure can be obtained? No

S.2 **If yes, provide web page URL:**

2a. RELIABILITY. Precise Specifications and Reliability Testing: H☐ M☐ L☐ I☐

2a1. Precise Measure Specifications. (*The measure specifications precise and unambiguous.*)

2a1.1 **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, e.g., cases from the target population with the target process, condition, event, or outcome)*:
Number of patients in the denominator who have a pregnancy test (urine or serum) ordered in the ED

2a1.2 **Numerator Time Window** *(The time period in which the target process, condition, event, or outcome is eligible for inclusion)*:
This measure is to be reported each time a patient presents to the emergency department with a chief complaint of abdominal pain.

2a1.3 **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, codes with descriptors, and/or specific data collection items/responses*:

Numerator Coding:
CPT 84703, 81025

2a1.4 **Denominator Statement** *(Brief, narrative description of the target population being measured)*:
All women, ages 14 – 50 years old, who present to the ED with a chief complaint of abdominal pain.

2a1.5 **Target Population Category** *(Check all the populations for which the measure is specified and tested if any)*: Adult/Elderly Care, Maternal Care, Populations at Risk, Special Healthcare Needs

2a1.6 **Denominator Time Window** *(The time period in which cases are eligible for inclusion)*:
Female patients ages 14 through 50 years old on date of encounter.

2a1.7 **Denominator Details** (*All information required to identify and calculate the target population/denominator such as definitions, codes with descriptors, and/or specific data collection items/responses)*:
Denominator Coding:
CPT E/M service codes: 99281, 99282, 99283, 99284, 99285, 99291
ICD-9 diagnosis codes: 789.00, 789.01, 789.02, 789.03, 789.04, 789.05, 789.06, 789.07, 789.09, 789.30, 789.31, 789.32, 789.33, 789.34, 789.35, 789.36, 789.37, 789.39,789.40, 789.41, 789.42, 789.43, 789.44, 789.45, 789.46, 789.47, 789.49, 789.60, 789.61, 789.62, 789.63, 789.64, 789.65, 789.66, 789.67, 789.69

Place of service indicator must be 23 (Emergency Department)

2a1.8 Denominator Exclusions *(Brief narrative description of exclusions from the target population)*:
i. Females for whom pregnancy is already documented or reported (verbal report by patient is acceptable).
ii. Females with documented or reported hysterectomy (verbal report by patient is acceptable).
iii. Females documented or reported to be post-menopausal (verbal report by patient is acceptable).
iv. Patient refusal
v. Patients who do not complete their ED evaluation (Left before completion, Left AMA, etc.)

2a1.9 Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, codes with descriptors, and/or specific data collection items/responses)*:
n/a

2a1.10 Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, codes with descriptors, definitions, and/or specific data collection items/responses)*:
n/a

2a1.11 Risk Adjustment Type *(Select type. Provide specifications for risk stratification in 2a1.10 and for statistical model in 2a1.13)*: No risk adjustment or risk stratification   **2a1.12 If "Other," please describe:**

2a1.13 Statistical Risk Model and Variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development should be addressed in 2b4.)*:
n/a

2a1.14-16 Detailed Risk Model Available at Web page URL (or attachment). Include coefficients, equations, codes with descriptors, definitions, and/or specific data collection items/responses. Attach documents only if they are not available on a webpage and keep attached file to 5 MB or less. NQF strongly prefers you make documents available at a Web page URL. Please supply login/password if needed**:**

2a1.17-18. Type of Score: Rate/proportion

2a1.19 Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*: Better quality = Higher score

2a1.20 Calculation Algorithm/Measure Logic *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*:
n/a

2a1.21-23 Calculation Algorithm/Measure Logic Diagram URL or attachment:

2a1.24 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):
The measure can be calculated from a sample or a complete count. The recommended minimum sample is 25 and should be obtained from a complete list of all ED patients with applicable denominator inclusion codes.

2a1.25 Data Source *(Check all the sources for which the measure is specified and tested)*. If other, please describe:
 Administrative claims, Electronic Clinical Data : Electronic Health Record, Paper Records

2a1.26 Data Source/Data Collection Instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.*)**:** A data collection instrument has been developed and is available.

2a1.27-29 Data Source/data Collection Instrument Reference Web Page URL or Attachment: URL
http://www.brighamandwomens.org/Departments_and_Services/emergencymedicine/Quality_Improvement.aspx?sub=0

2a1.30-32 Data Dictionary/Code Table Web Page URL or Attachment:

2a1.33 Level of Analysis (*Check the levels of analysis for which the measure is specified and tested*)**:** Clinician : Group/Practice, Clinician : Individual, Facility

2a1.34-35 Care Setting (*Check all the settings for which the measure is specified and tested*)**:** Ambulatory Care : Clinic/Urgent Care, Hospital/Acute Care Facility

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability.*)

2a2.1 Data/Sample (*Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included*)**:**
Neither study formally tested reliability, but they contain reliability data.

Schuur et al tested the reliability of the measure by reviewing charts of patients with a diagnosis of abdominal pain. Of 200 charts reviewed, all had abdominal pain as a preseting complaint and were eligible to be considered fo the measure. Among 200 women aged 14 to 50 years and with a final diagnosis of abdominal pain (789.xx) in 4 EDs in 2006, 89.4% (95% CI 85.0% to 94.0%) were eligible for the pregnancy quality measure (Table 3). Reasons for ineligibility were 7 patients with a previous hysterectomy, 10 who had documented infertility, and 4 who had clinical evidence of pregnancy or reported a positive pregnancy test before their ED visit. Pregnancy testing was performed in the ED for 93.9% (95% CI 90.3% to 97.4%) of eligible women, with rates ranging from 88.1% to 97.9% across the EDs.
Graff used EHR data, based on the patient's chief complaint and did not formally test reliability.

2a2.2 Analytic Method (*Describe method of reliability testing & rationale*)**:**

2a2.3 Testing Results (*Reliability statistics, assessment of adequacy in the context of norms for the test conducted*)**:**

2b. VALIDITY. Validity, Testing, <u>including all Threats to Validity</u>:   H☐ M☐ L☐ I☐

2b1.1 Describe how the measure specifications (*measure focus, target population, and exclusions*) **are consistent with the evidence cited in support of the measure focus (***criterion 1c)* **and identify any differences from the evidence:**
The measure directly addresses the recommendations of the ACEP guideline.

2b2. Validity Testing. (*Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.*)

2b2.1 Data/Sample (*Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included*)**:**
Survey of experts in emergency medicine. Details below in methods.

2b2.2 Analytic Method (*Describe method of validity testing and rationale; if face validity, describe systematic assessment*)**:**
Expert Panel established face validity:

Overview of Methods
In committee year 2007-2008, the ACEP Board of Directors (BOD) charged the Quality and Performance Committee (QPC) with

developing a set of performance measures for emergency medicine. With national bodies such as the National Quality Forum and Center for Medicare and Medicaid Services increasingly calling for performance measures from specialty societies, ACEP wanted to have a set of potential performance measures ready to submit to such programs. The aim of this process was to develop measures that would be appropriate for inclusion in national quality programs such as CMS PQRI, and ultimately used for public-reporting or pay-for-performance. The aim was not to develop measures to help EDs guide internal improvement.

The full QPC discussed possible measurement topics and reviewed available emergency medicine measures under development. In particular the Chief Complaint Based Quality indicators developed by ACEP Quality and Performance Section were reviewed. Candidate measure ideas were ranked through a web-based survey by all members of the QPC. The highest ranking measures were fully specified into performance measures by QPC Workgroup 2 and re-circulated to the full QPC for comments.

The fully specified measures were evaluated across the domains of quality performance measurement using a web based survey tool. All members of the QPC and the Clinical Policies committee were invited to respond. We received a>60% response rate.

Workgroup 2 reviewed the collated comments, available literature and developed recommendations for the QPC. The following recommendations were considered:

- Reject – definitive evaluation of measure is possible and it is not appropriate for submission to national quality organizations
- Table – conceptually sound measure, but insufficient evidence exists to reject or endorse as a performance measure for submission to national quality organizations. Re-evaluate in one year
- Approve – conceptually sound measure with sufficient evidence to consider for endorsement as a performance measure for submission to national quality organizations. Field testing of measure reliability and feasibility are required before implementation.

2b2.3 Testing Results *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)*:
Results: This measure scored highest of all measures evaluated out of the ACEP TEP

Question	Median	Average	Standard Deviation	
Useful: Evidence-based	4	3.9	0.87	
Useful: Interpretable	5	4.7	0.48	
Useful: Actionable	5	4.6	0.49	
Measure Design: Denominator	5	4.5	0.51	
Measure Design: Numerator	4	4.4	0.57	
Measure Design: Validity	4	4.1	0.72	
Measure Design: Reliability	4	4.5	0.51	
Measure Implementation: Feasibility		4	4.1	0.82
Overall Assessment	4	4.2	0.85	

POTENTIAL THREATS TO VALIDITY. (*All potential threats to validity were appropriately tested with adequate results.*)

2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 Data/Sample for analysis of exclusions *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
Exclusions were drawn up by the Expert Panel. Schuur et al tested them in chart review – details above.

2b3.2 Analytic Method (*Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference*):

2b3.3 Results *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*

2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:

2b4.2 Analytic Method (*Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables)*:

2b4.3 Testing Results *(*<u>Statistical risk model</u>*: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models.* <u>Risk stratification</u>*: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata)*:

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.*)

2b5.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)*:

2b5.3 Results *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches result in comparable scores.*)

2b6.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
See schuur et al

2b6.2 Analytic Method *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure)*:

2b6.3 Testing Results *(Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted)*:
Schuur et al faound a large gap between the NHAMCS data and their chart review. This speaks to the difference between the quality of chart review, as NHAMCS is a large dataset based on chart review.

2c. Disparities in Care: H☐ M☐ L☐ I☐ NA☐ (*If applicable, the measure specifications allow identification of disparities.*)

2c.1 If measure is stratified for disparities, provide stratified results *(Scores by stratified categories/cohorts)*:

2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:

2.1-2.3 Supplemental Testing Methodology Information:
Attachment
QPC ACEP Emergency Medicine Performance Measures Technical Report 06-2009 (af 09-2009 edit).doc

Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met?
(*Reliability and Validity must be rated moderate or high*) Yes☐ No☐
Provide rationale based on specific subcriteria:

If the Committee votes No, STOP

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. (<u>evaluation criteria</u>)

C.1 **Intended Purpose/ Use** (*Check all the purposes and/or uses for which the measure is intended*): Public Reporting, Quality Improvement (Internal to the specific organization)

3.1 **Current Use** (*Check all that apply; for any that are checked, provide the specific program information in the following questions*): Public Reporting, Quality Improvement (Internal to the specific organization)

3a. Usefulness for Public Reporting: H☐ M☐ L☐ I☐
(*The measure is meaningful, understandable and useful for public reporting.*)

3a.1. Use in Public Reporting - disclosure of performance results to the public at large (*If used in a public reporting program, provide name of program(s), locations, Web page URL(s)*). <u>If not publicly reported in a national or community program</u>, state the reason AND plans to achieve public reporting, potential reporting programs or commitments, and timeline, e.g., within 3 years of endorsement: [*For <u>Maintenance</u> – If not publicly reported, describe progress made toward achieving disclosure of performance results to the public at large and expected date for public reporting; provide rationale why continued endorsement should be considered.*]
This measure is included in the CMS Physician Quality Reporting System program.

3a.2.Provide a rationale for why the measure performance results are meaningful, understandable, and useful for public reporting. <u>If usefulness was demonstrated</u> (e.g., focus group, cognitive testing), describe the data, method, and results:

3.2 Use for other Accountability Functions (payment, certification, accreditation). If used in a public accountability program, provide name of program(s), locations, Web page URL(s):

3b. Usefulness for Quality Improvement: H☐ M☐ L☐ I☐
(*The measure is meaningful, understandable and useful for quality improvement.*)

3b.1. Use in QI. If used in quality improvement program, provide name of program(s), locations, Web page URL(s):
[*For <u>Maintenance</u> – If not used for QI, indicate the reasons and describe progress toward using performance results for improvement*].
The results of the quality measure are very understandable to the public, i.e. whether the physicians at the ED for which the results are published are reliably confirming pregnancy status for women with abdominal pain. The risk of ectopic pregnancy and receiving CT scan imaging radiation when they are pregnant are easily understandable.
The measure is in use for QI at Dr. Graff's hospital and at EDs staffed by Emergency Medicine Physicians' (Dr. Kevin Klauer can further comment kklauer@emp.com)

3b.2. Provide rationale for why the measure performance results are meaningful, understandable, and useful for quality improvement. <u>If usefulness was demonstrated</u> (e.g., QI initiative), describe the data, method and results:
Year 3 of the study clearly showed that this quality measure changed physician performance so they reliably ensured the women of fertile age were not pregnant. This is in contrast to general education of the value of pregnancy testing and the physician specific

feedback of pregnancy testing rates which did not change physician performance to the high performance that the public would expect.

Overall, to what extent was the criterion, *Usability*, met? H☐ M☐ L☐ I☐
Provide rationale based on specific subcriteria:

4. FEASIBILITY

Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (**evaluation criteria**)

4a. Data Generated as a Byproduct of Care Processes: H☐ M☐ L☐ I☐

4a.1-2 How are the data elements needed to compute measure scores generated? *(Check all that apply).*
Data used in the measure are:
generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition

4b. Electronic Sources: H☐ M☐ L☐ I☐

4b.1 Are the data elements needed for the measure as specified available electronically *(Elements that are needed to compute measure scores are in defined, computer-readable fields)*: Some data elements are in electronic sources

4b.2 If ALL data elements are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources: ED's that do not have electronic charts will not be able to calculate the measure exclusions without manual chart review.

4c. Susceptibility to Inaccuracies, Errors, or Unintended Consequences: H☐ M☐ L☐ I☐

4c.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measurement identified during testing and/or operational use and strategies to prevent, minimize, or detect. If audited, provide results:
Graff found that with an electronic health record in the emergency department the data are readily available and the quality staff in the emergency department can perform the measure on each physician each month. Schuur found that with paper based charting, the measure can be collected quickly (<10 minutes per chart total).

The Data Elements are Generated as Byproduct of Care Processes and easily collated and analyzed and report formed if there is an electronic health record

4d. Data Collection Strategy/Implementation: H☐ M☐ L☐ I☐

A.2 Please check if either of the following apply (*regarding proprietary measures*):
4d.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 and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues *(e.g., fees for use of proprietary measures)*:
We have learned that this measure which involves the ordering of 1 test based on 3 factors (gender, age, chief complaint of the patient) can be automatically created by an electronic health record that has cpoe (computerized physician order entry) with required 4 fields in the physician order entry (presence or not of the exclusions of hysterectomy, known pregnant, patient refuses, menopause)

Overall, to what extent was the criterion, *Feasibility*, met? H☐ M☐ L☐ I☐
Provide rationale based on specific subcriteria:

OVERALL SUITABILITY FOR ENDORSEMENT

Does the measure meet all the NQF criteria for endorsement? Yes☐ No☐
Rationale:

If the Committee votes No, STOP.
If the Committee votes Yes, the final recommendation is contingent on comparison to related and competing measures.

5. COMPARISON TO RELATED AND COMPETING MEASURES
If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure before a final recommendation is made.
5.1 If there are related measures *(either same measure focus or target population)* or competing measures *(both the same measure focus and same target population)*, list the NQF # and title of all related and/or competing measures:

5a. Harmonization
5a.1 If this measure has EITHER the same measure focus OR the same target population as NQF-endorsed measure(s): Are the measure specifications completely harmonized?   **5a.2** If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden:

5b. Competing Measure(s)
5b.1 If this measure has both the same measure focus and the same target population as NQF-endorsed measure(s): Describe why this measure is superior to competing measures *(e.g., a more valid or efficient way to measure quality)*; OR provide a rationale for the additive value of endorsing an additional measure. *(Provide analyses when possible)*:

CONTACT INFORMATION
Co.1 Measure Steward (Intellectual Property Owner): American College of Emergency Physicians, 1125 Executive Circle, Irving, Texas, 75038
Co.2 Point of Contact: Dainsworth, Chambers, dchambers@acep.org, 202-728-0610-3014
Co.3 Measure Developer if different from Measure Steward: American College of Emergency Physicians, 1125 Executive Circle, Irving, Texas, 75038   **Co.4** Point of Contact: Dainsworth, Chambers, dchambers@acep.org, 202-728-0610-3014
Co.5 Submitter: Dainsworth, Chambers, dchambers@acep.org, 202-728-0610-3014, American College of Emergency Physicians
Co.6 Additional organizations that sponsored/participated in measure development:
Co.7 Public Contact: Dainsworth, Chambers, dchambers@acep.org, 202-728-0610-3014, American College of Emergency Physicians

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.
Ad.2 If adapted, provide title of original measure, NQF # if endorsed, and measure steward. Briefly describe the reasons for adapting the original measure and any work with the original measure steward:
Measure Developer/Steward Updates and Ongoing Maintenance   **Ad.3** Year the measure was first released: 2009

Ad.4 Month and Year of most recent revision: 09, 2011 **Ad.5 What is your frequency for review/update of this measure?** At least yearly **Ad.6 When is the next scheduled review/update for this measure?** 09, 2012
Ad.7 Copyright statement:
Ad.8 Disclaimers:
Ad.9 Additional Information/Comments: •2a1.34. Care Setting: The measure is specified for the emergency department, but is equally applicable to the "Ambulatory Care : Clinic/Urgent Care" setting. The NQF should consider the measure for this care setting.
Date of Submission (*MM/DD/YY*): 10/17/2011