Special statement: Proposed quality metrics to assess accuracy of prenatal detection of congenital heart defects

Patient Safety and Quality Committee, Society for Maternal-Fetal Medicine; C. Andrew Combs, MD, PhD; Afshan B. Hameed, MD; Alexander M. Friedman, MD, MPH; Iffath Abbasi Hoskins, MD

The Problem: A high percentage of congenital heart defects (CHD) are not detected prenatally, resulting in suboptimal care and increased neonatal morbidity and mortality.

A Solution: We propose quality metrics for tracking CHD detection rates to address this quality gap.

Congenital heart defects are a leading cause of neonatal morbidity and mortality. Accurate prenatal diagnosis of congenital heart defects can reduce morbidity and mortality by improving prenatal care, facilitating predelivery pediatric cardiology consultation, and directing delivery to facilities with resources to manage the complex medical and surgical needs of newborns with congenital heart defects. Unfortunately, less than one half of congenital heart defect cases are detected prenatally, resulting in lost opportunities for counseling, shared decision-making, and delivery at an appropriate facility. Quality improvement initiatives to improve prenatal congenital heart defects detection depend on the ability to measure the rate of detection at the level of providers, facilities, or populations, but no standard metric exists for measuring the detection of congenital heart defects at any level. The need for such a metric was recognized at a Cooperative Workshop held at the 2016 Annual Meeting of the Society for Maternal-Fetal Medicine, which recommended the development of a quality metric to assess the rate of prenatal detection of clinically significant congenital heart defects. In this paper, we propose potential quality metrics to measure prenatal detection of critical congenital heart defects, defined as defects with a high rate of morbidity or mortality or that require surgery or tertiary follow-up. One metric is based on a retrospective approach, assessing whether postnatally diagnosed congenital heart defects had been identified prenatally. Other metrics are based on a prospective approach, assessing the sensitivity and specificity of prenatal diagnosis of congenital heart defects by comparing prenatal ultrasound findings with newborn findings. Potential applications, limitations, challenges, barriers, and value for both approaches are discussed. We conclude that future development of these metrics will depend on an expansion of the International Classification of Diseases system to include specific codes that distinguish fetal congenital heart defects from newborn congenital heart defects and on the development of record systems that facilitate the linkage of fetal records (in the maternal chart) with newborn records.

Key words: congenital heart defects, neonatal morbidity, neonatal mortality, quality metric

Congenital heart defects (CHD) are among the most common congenital anomalies, affecting 0.4–0.8% of newborns and are the leading cause of death among infants with birth defects. Unfortunately, less than one half of CHD are detected prenatally, although detection rates vary, depending on the type of defect, examiner skill, and specific population. Prenatal detection of CHD improves neonatal surgical outcomes and childhood developmental milestones and reduces neonatal mortality. Prenatal diagnosis facilitates counseling, shared decision-making, and, if appropriate, transfer for delivery at a facility with resources to manage a critically ill newborn.

Quality improvement (QI) initiatives can improve the prenatal detection of CHD. The QI catchphrase, “If you cannot measure it, you cannot improve it,” implies that improving CHD detection depends on the measurement of the rate of correct diagnoses. Currently there is no...
standardized quality metric to measure CHD detection performance.

In 2016, the Society for Maternal-Fetal Medicine (SMFM) hosted a multistakeholder Cooperative Workshop entitled “Quality Measures in High-Risk Pregnancies” to consider new quality metrics for high-risk pregnancies. The Workshop Executive Summary recommended the development of a quality metric for prenatal detection of clinically significant CHD. In this paper, we define clinically significant CHD as any of the diagnoses listed in Table 1, conditions that are considered critical and that typically require surgical repair in infancy to avoid death or major morbidity. These diagnoses comprise approximately 20–30% of all CHD. We propose 2 types of quality metrics to measure prenatal detection of critical CHD. The first type is retrospective, initially identifying newborns with critical CHD and retrospectively determining whether CHD was diagnosed prenatally. The second uses a prospective approach, starting with a cohort of prenatal ultrasound examinations, followed by measurement of the sensitivity and specificity of subsequent diagnoses of CHD. For both approaches, potential uses, limitations, and barriers are discussed.

Retrospective cohort metric: percentage of newborns with CHD who had a prenatal diagnosis

This metric begins with a cohort of newborns with CHD and retrospectively evaluates the proportion with a prenatal diagnosis of CHD. The simplified measure specification is:

**Denominator:** Number of newborns with a critical CHD diagnosis (Table 1).

**Numerator:** Number of newborns in the denominator who had CHD identified prenatally.

**Metric:** Percentage (numerator divided by denominator). Theoretical ideal detection is 100%.

This metric is similar to the one proposed at the 2016 Cooperative Workshop. It would most readily be applied at the hospital or hospital-system level. Although it initially appears straightforward, the metric is complicated by several barriers and limitations, as detailed in the critique in Table 2. First, it cannot be tracked using claims or administrative data because there is no specific code for fetal CHD in the International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM). Metrics that can be tracked via ICD-10-CM codes or other structured data are generally preferred because they minimize the burden of data collection. Without a specific code, a manual review of maternal charts will be required to determine whether each case of newborn CHD was diagnosed prenatally.

Another limitation is that prenatal diagnoses are documented in maternal records, but postnatal diagnoses appear in neonatal records. Electronic health record systems rarely provide linkage between the maternal and newborn records to facilitate automated tracking of these diagnoses, so manual linkage will likely be required. Moreover, hospitals may not have access to prenatal ultrasound reports and would thus need to rely on admitting diagnoses, which may be incomplete.

The likely value of this metric would be to encourage hospitals to develop QI programs to help improve the performance of local practices that provide prenatal diagnostic services for patients who deliver there, including prenatal care clinics and maternal-fetal medicine, radiology, and pediatric cardiology services. This effort could start with a quality-assurance review of cases in which the CHD diagnosis was missed. A focused review may reveal system issues, such as lack of access to prenatal ultrasound, lack of referral for fetal echocardiogram in high-risk patients, or lack of prenatal care (Figure). For patients who had prenatal ultrasound examinations that missed the diagnosis of CHD, the review process could include feedback to all relevant ultrasound providers and a request that images be reviewed for technical quality and accuracy. This feedback alone may help to improve care because many prenatal ultrasound providers receive no systematic information about missed diagnoses and thus have no knowledge about potential quality issues.

Although this metric could be appropriate for hospital-level measurement, it would not provide valid between-

---

**TABLE 1**

Critical congenital heart defects considered for quality metrics

<table>
<thead>
<tr>
<th>Congenital heart defect</th>
<th>ICD-10-CM code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anomalies of the great veins</td>
<td>Q26^a</td>
</tr>
<tr>
<td>Pulmonary artery atresia</td>
<td>Q25.5</td>
</tr>
<tr>
<td>Coarctation or atresia of the aorta</td>
<td>Q25.1, Q25.2, Q25.21, Q25.29</td>
</tr>
<tr>
<td>Aortic stenosis, valvular, subaortic, or supravalvular</td>
<td>Q23.0, Q24.4, Q25.3</td>
</tr>
<tr>
<td>Ebstein anomaly</td>
<td>Q22.5</td>
</tr>
<tr>
<td>Hypoplastic right heart syndrome</td>
<td>Q22.6</td>
</tr>
<tr>
<td>Hypoplastic left heart syndrome</td>
<td>Q23.4</td>
</tr>
<tr>
<td>Tricuspid atresia or stenosis</td>
<td>Q22.4, Q22.8, Q22.9, I36.0</td>
</tr>
<tr>
<td>Pulmonary valve atresia, stenosis, insufficiency</td>
<td>Q22.0, Q22.1, Q22.3</td>
</tr>
<tr>
<td>Endocardial cushion defect</td>
<td>Q21.2</td>
</tr>
<tr>
<td>Tetralogy of Fallot</td>
<td>Q21.3</td>
</tr>
<tr>
<td>Common ventricle</td>
<td>Q20.4</td>
</tr>
<tr>
<td>Transposition of the great arteries</td>
<td>Q20.3</td>
</tr>
<tr>
<td>Common truncus arteriosus</td>
<td>Q20.0</td>
</tr>
</tbody>
</table>

ICD-10-CM: International Classification of Diseases, 10th Revision, Clinical Modification.

Adapted from Jelliffe-Pawlowski et al. Any digit 0–9.

hospital comparisons for several reasons. First, it does not necessarily reflect prenatal diagnostic services rendered by the maternity hospital but likely includes services by various community providers. Second, for hospitals with small numbers of cases, the metric will have poor precision (wide confidence intervals). Third, as noted in Table 2, the metric can be biased by referral patterns; a community hospital with a high detection rate that appropriately refers patients to deliver at a tertiary center will have an artifactualy low rate on this metric and the receiving hospital will have an artifactualy inflated rate. A similar bias would result from pregnancy terminations by women diagnosed with fetal CHD, producing artifactualy low apparent detection rates. **Prospective cohort metrics: sensitivity and specificity of prenatal ultrasound examinations**

The prospective approach defines a cohort of patients who had a prenatal ultrasound examination during a specified period. **TABLE 2**

<table>
<thead>
<tr>
<th>Issue</th>
<th>Explanation</th>
<th>Limitations, barriers, disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population studied</td>
<td>All newborns delivered at a given hospital during a specified time period</td>
<td>A small number of cases will result in rates with wide confidence intervals. (For example, a hospital with 1000 births per year may have only 10 CHD cases per year. If 5 are detected, the rate is 50%, but the 95% confidence interval ranges from 20% to 80%).</td>
</tr>
<tr>
<td>Likely time period of analysis</td>
<td>Calendar year</td>
<td></td>
</tr>
<tr>
<td>Method of capturing denominator</td>
<td>ICD-10-CM code list (Table 1) from newborn record</td>
<td>● Denominator does not include CHD cases that are not detected by the time of newborn discharge. If eligibility time period is extended to capture diagnoses after discharge, the hospital may not have access to infant outpatient records. ● Denominator does not include CHD cases that do not result in a newborn diagnosis (eg, miscarriage, termination, transfer for delivery elsewhere).</td>
</tr>
<tr>
<td>Method of capturing numerator</td>
<td>Manual review of admitting diagnoses in maternal admitting record and prenatal record</td>
<td>● There is no specific ICD-10-CM for fetal CHD; therefore, manual review is required. ● Numerator is obtained from maternal record, but denominator is defined by newborn record. Health systems may not have effective methods to link maternal and newborn records.</td>
</tr>
<tr>
<td>Attribution</td>
<td>Metric tracked at the hospital level</td>
<td>● Metric is attributed to hospital, but source of prenatal detection may include a variety of community ultrasound providers. ● Hospitals have limited ability to influence these community sources and thereby improve quality. ● As defined, the metric does not require assessing rates for individual ultrasound facilities or providers. If individual rates are assessed, a small number of cases for each facility or provider will result in wide confidence intervals.</td>
</tr>
<tr>
<td>Referral bias</td>
<td>Referral of correctly diagnosed cases will affect rates of both the sending and the receiving hospitals</td>
<td>● For example, if 50 CHD cases occur within a community hospital’s catchment area and 40 are detected, the actual detection rate is 80%. If the 40 detected cases are referred to deliver at the regional tertiary hospital, this leaves 10 undetected CHD cases delivering at the local hospital. The metric would be 0 cases detected of 10 CHD births at the local hospital, or a rate of 0%, despite an actual detection rate of 80% in the community. ● Conversely, the delivery population at the regional tertiary hospital population will be enriched by cases sent from the surrounding community hospitals after prenatal detection of CHD. This will inflate the metric for the tertiary hospital.</td>
</tr>
<tr>
<td>Termination bias</td>
<td>Pregnancy termination in correctly diagnosed cases will result in artifactualy low rate</td>
<td>For example, if 50 CHD cases occur within a hospital’s catchment area and 40 cases are detected prenatally, the actual detection rate is 80%. If 30 of the detected cases terminate pregnancy and 10 continue pregnancy, the apparent metric will be based on the 10 births in continuing pregnancies plus the 10 births among the undetected cases. The metric would be 10 detected of 20 total CHD births, or 50% despite an actual detection rate of 80% in the community.</td>
</tr>
<tr>
<td>False-positive diagnosis</td>
<td>Not captured</td>
<td>● A false-positive prenatal diagnosis of CHD has no bearing on the numerator or denominator of this metric. ● The metric rate can be increased by increasing the number of diagnoses of CHD, whether accurate or not. ● A balancing metric would be desirable to detect the extent of over-diagnosis.</td>
</tr>
</tbody>
</table>

time period and then determines the rate of correct prenatal CHD diagnosis. The simplified specification for the main proposed metric is:

**Cohort:** All pregnant women who had 1 or more obstetric ultrasound examinations during the measurement period using 1 or more of the following Current Procedural Terminology (CPT) codes: 76805 (basic second- or third-trimester ultrasound), 76810 (basic second- or third-trimester ultrasound, additional fetus), 76811 (detailed fetal anatomy examination), 76812 (detailed fetal anatomy examination, additional fetus), or 76825 (fetal echocardiogram).

**Denominator:** Number of patients in the cohort who delivered a newborn without CHD.

**Numerator:** Number of patients in the cohort who delivered a newborn without CHD after prenatal diagnosis of CHD (ICD-10 code Z33.2 or O04.xx) after prenatal diagnosis of CHD.

**Metric (sensitivity):** Percentage (numerator divided by denominator). Theoretical ideal detection is 100%.

This metric is readily recognizable as the sensitivity of ultrasonography as a screening test for CHD. Other test characteristics, such as specificity and predictive values, can be calculated from the $2 \times 2$ table shown in Table 3.

A second metric reflecting specificity can be defined as:

**Denominator:** Number of patients in the cohort who delivered a newborn without CHD.

**Numerator:** Number of patients in the cohort without a prenatal diagnosis of CHD.
TABLE 3
Prospective cohort approach to a quality metric for detection of CHD

<table>
<thead>
<tr>
<th>Postnatal findings</th>
<th>Prenatal findings</th>
<th>No suspicion of CHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHD</td>
<td>True positive</td>
<td>False negative</td>
</tr>
<tr>
<td>No CHD</td>
<td>False positive</td>
<td>True negative</td>
</tr>
</tbody>
</table>

CHD, congenital heart defects.

Sensitivity metric is true positive/(true positive plus false negative), and specificity metric is true negative/(true negative plus false positive). Positive predictive value is true positive/(true positive plus false positive), and negative predictive value is true negative/(true negative plus false negative).

Metric (specificity): Percentage (numerator divided by denominator). Theoretical ideal specificity is 100%.

Specificity is a key important balancing metric to guard against the overdiagnosis of fetal CHD. False-positive diagnoses have major medical, emotional, and financial implications, including recommendations for potentially unnecessary tests and procedures such as fetal echocardiography, genetic testing, consultation with pediatric cardiologists and neonatologists, transfer to tertiary facilities, or even pregnancy termination.

Limitations to the collection and use of these metrics are outlined in Table 4. As with the retrospective metric, the lack of specific ICD-10-CM code(s) for fetal CHD is a major barrier that prevents the use of claims data or administrative data for measure tracking, necessitating burdensome manual record review. As with the retrospective metric, another critical limitation is the general lack of a systematic linkage between fetal diagnoses (which appear in maternal records) and newborn diagnoses (in newborn records). Another limitation is the potential for bias as a result of pregnancy terminations.

These metrics could potentially be used by provider groups or health systems to evaluate the diagnostic accuracy of individual ultrasound providers (sonographers or physicians). If significant between-provider variation is found, QI efforts can be focused on providers with low values of either metric. Another potential use might be for a payer to evaluate the diagnostic accuracy of individual providers or provider groups.

Next steps and actions needed
Many stakeholders have an interest in knowing about the accuracy of prenatal diagnosis of CHD. Patients want to know whether ultrasound findings can be relied upon. Ultrasound providers want a reputation for high quality because continued referrals depend on that reputation. Payers want assurance that their payments for ultrasound services have value, which depends on accuracy. Thus, we agree with the Cooperative Workshop recommendation that metric(s) reflecting the accuracy of prenatal diagnosis of CHD should be developed.41

The proposed metrics, in principle, can provide an assessment of the accuracy of detection of CHD. However, they are fraught with practical limitations that hobble the ability of hospitals, providers, or payers to actually track them. To minimize the burden of data collection, it should be possible to track metrics using data that are already in the electronic records, administrative records, or claims databases. Claims data based on CPT and ICD-10-CM codes are frequently used for such purposes. Unfortunately, such automation is not currently possible for the proposed metrics because there is no specific ICD-10-CM code for fetal CHD and because it is difficult to link fetal records (in the maternal chart) with newborn records. QI initiatives to improve CHD detection can be done despite these limitations,35-39 but these efforts require a significant investment of resources to perform manual record review. Providers and hospitals are unlikely to invest these resources without external motivation to do so. Payers typically have access only to claims data and so cannot evaluate the metrics as currently proposed.

A critical first step is to develop specific ICD codes for fetal CHD and indeed for all congenital fetal anomalies. As of 2020, ICD-10-CM has only 3 codes for fetal anomalies (O35.0XXn, O35.8XXn, and O35.9XXn, where n = fetus number). These codes reflect “maternal care for (suspected) fetal central nervous system malformation,” “maternal care for (suspected) other fetal abnormality and damage,” and “maternal care for (suspected) fetal abnormality and damage unspecified,” respectively.

This lack of specificity is startling in a classification system that is renowned for specificity (eg, codes that distinguish trauma from impact with a parrot [W61.02], macaw [W61.12], chicken [W61.32], goose [W61.52], or duck [W61.62]). There is no such specificity for fetal diagnoses. The same ICD-10 code (O35.8XXn) is used for critical fetal anomalies (eg, tricuspid stenosis) and minor or incidental findings (eg, supernumerary digit). We believe that each anomaly listed in Table 1 should have a specific ICD code reflecting its occurrence in a fetus. More generally, having specific ICD codes for all common fetal anomalies (central nervous system, gastrointestinal, genitourinary, musculoskeletal, orofacial) may facilitate a variety of research and QI projects. In the future, the SMFM Coding Committee will recommend specific fetal anomalies that should be considered for addition to the ICD.

Another critical step is to develop record systems that facilitate the linkage of fetal and newborn findings. Without such linkage, there is no systematic way for ultrasound providers to know whether prenatal findings were confirmed or refuted by newborn examination findings, for newborn care providers to directly review
prenatal findings, or for payers to match maternal claims (which include fetal findings) with newborn claims. Such a linkage will require substantial development by vendors of electronic health record systems. The SMFM Informatics Committee has liaisons with several electronic health record vendors and can encourage such development.

Overcoming these hurdles will allow these metrics to satisfy the Measure Evaluation Criteria of the National Quality Forum. Future development would include pilot testing with interested health systems and subsequent application for endorsement of the metrics by the National Quality Forum.

**REFERENCES**

Reprints will not be available.

All authors and Committee members have filed a conflict of interest disclosure delineating personal, professional, and/or business interests that might be perceived as a real or potential conflict of interest in relation to this publication. Any conflicts have been resolved through a process approved by the Executive Board. The Society for Maternal-Fetal Medicine (SMFM) has neither solicited nor accepted any commercial involvement in the development of the content of this publication.

This document has undergone an internal peer review through a multilevel committee process within SMFM. This review involves critique and feedback from the SMFM Patient Safety and Quality and Document Review Committees and final approval by the SMFM Executive Committee. SMFM accepts sole responsibility for document content. SMFM publications do not undergo editorial and peer review by the American Journal of Obstetrics & Gynecology. The SMFM Patient Safety & Quality Committee reviews publications every 36-48 months and issues updates as needed. Further details regarding SMFM Publications can be found at www.smfm.org/publications.

SMFM has adopted the use of the word “woman” (and the pronouns “she” and “her”) to apply to individuals who are assigned female sex at birth, including individuals who identify as men as well as nonbinary individuals who identify as both genders or neither gender. As gender-neutral language continues to evolve in the scientific and medical communities, SMFM will reassess this usage and make appropriate adjustments as necessary.

All questions or comments regarding the document should be referred to the SMFM Patient Safety and Quality Committee at smfm@smfm.org.