American Journal of Obstetrics and Gynecology (2004) **190**, 1747–58

ELSEVIER

AMERICAN JOURNAL of OBSTETRICS and GYNECOLOGY

www.elsevier.com/locate/ajog

Development of a large-scale obstetric quality-improvement program that focused on the nulliparous patient at term

Elliott K. Main, MD,[a,b,*] Lori Bloomfield, MS,[a] Gordon Hunt, MD[a] for the Sutter Health, First Pregnancy and Delivery Clinical Initiative Committee (complete list at the end of the article)

Division of Clinical Integration, Sutter Health, Sacramento, Calif[a]*; and Department of Obstetrics and Gynecology, California Pacific Medical Center, San Francisco, Calif*[b]

KEY WORDS
Quality improvement
Cesarean delivery
Perineal laceration
Induction
Nulliparity

Objective: The purpose of this study was to identify an appropriate population and a balanced set of maternal and neonatal measures to drive a hospital network obstetric quality improvement program.

Study design: Sutter Health, a large Northern California health care system with >40,000 births annually, served as the site for this project. We chose to focus on the standardized nulliparous patients: term, singleton, and vertex. A multidisciplinary task force evaluated and selected perinatal outcome and process measures. Data from every hospital were collected prospectively electronically and analyzed centrally.

Results: Outcome measures that were selected included term, singleton, and vertex rates of 3rd/4th-degree laceration, cesarean birth, 5-minute Apgar score of <7, and patient satisfaction. The process measures included episiotomy, induction (37-41 weeks), and admittance with cervical dilation of ≥3 cm. Data collection completeness improved each quarter; by the end of 2002, the data collection completeness rate had reached 99.7%. Every measure demonstrated a large variation among our hospitals, which indicates opportunities for improvement.

Conclusion: This balanced set of measures for term, singleton, and vertex patients has been straightforward to collect over a large and diverse hospital system and has engaged all participants successfully.

Over the last 5 years, we have seen a reappraisal of traditional quality indicators in obstetrics. The value of the measures (total cesarean birth rate and the rate of vaginal birth after cesarean [VBAC]) to accurately reflect a birthing unit's quality of care has been challenged by both clinical leaders and new data.[1,2] We describe here an alternative approach to obstetric quality assessment and improvement that focuses on treatment of a standardized population of women who are having their first delivery and the implementation of this method in a large hospital network.

Presented at the Seventieth Annual Meeting of the Pacific Coast Obstetrical and Gynecological Society, September 16-21, 2003, Anchorage, Alaska.

* Reprint requests: Elliott K. Main, MD, Department of Obstetric and Gynecology, California Pacific Medical Center, 3700 California St, San Francisco, CA 94118.

0002-9378/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ajog.2004.02.055

The laboratory we were using to test this model is Sutter Health, an integrated network of 25 hospitals and 7 medical groups in Northern California. The 20 birthing units in Sutter Health provide obstetric care for >40,000 births each year. This delivery total represents a full 1% of all US births (4,019,280 births in 2002).[3] The birthing units are remarkably diverse: large and small, urban and rural, and ethnically and racially varied and range in neonatal intensive care support from level I to level III. Thus, in breadth and scope, the efforts reported here could serve as a model for obstetric quality assessment and improvement at a regional, state, or national level.

From the outset, we recognized that an important lesson for obstetric quality improvement from the 1990s was that focusing on a single outcome (eg, cesarean birth) dismisses the complexity of desired outcomes: mother, baby, and satisfaction (each with multiple dimensions). Equally important, a single focus (eg, total cesarean birth rate) can lead to practice changes that potentially can put one set of outcomes (eg, baby) at risk for the sake of another (eg, mother).[1]

Material and methods

In 1999, Sutter Health Women's Advisory Group held a series of multidisciplinary meetings to determine the direction of quality improvement in the area of obstetric care. The charge was 3-fold: to identify an appropriate population/issue for intervention that would be relevant for all birthing units, to develop measures and data collection systems, and to establish a comprehensive quality program that creates both the appropriate "mind set" and the clinical tools to help accomplish improvement. This was done within the context of the Sutter Health system-wide Clinical Initiative program. The Sutter Health Clinical Initiative program provides central support and organization for carefully structured long-term (7+ year) quality improvement efforts that are standardized throughout the system.

Nulliparous birth was selected as our quality improvement focus on the basis of the following criteria: (1) relevancy for all facilities (large/small, neonatal intensive care unit [NICU] levels I through III), (2) a complex topic area that could engage staff for an extended period and provide a range of opportunities for improvement activities with spin-off for other patient populations, (3) a topic area that is recognized easily as important by both medical and nonmedical members of the community, and (4) a population in which we could have significant impact.

Nulliparous birth, when compared with multiparous births, is associated with longer, more difficult labors and with much higher morbidities, whether measured by rates of cesarean birth, perineal injury, or NICU admission. A similar approach to has been taken in the United Kingdom where the "Standardized Primip" (defined in Britain as nulliparous, term, singleton, vertex, white, aged 20-34 years, over 155 cm in height, and excluding medical complications of pregnancy) has been shown to be a useful interhospital quality improvement target group.[4] Nulliparous women are both very common (currently 45% of our hospitals' obstetric populations) and raise important management issues at all of our hospitals, large or small. Importantly, a focus on "first time birth" could engage all obstetric providers, prenatal educators, labor and delivery staff, and women and their partners. In recognition of these advantages, in March 2000, Sutter Health formally launched its system-wide obstetric quality improvement program as the "First Pregnancy and Delivery (FPAD) Clinical Initiative."

Potential measures were identified either from the National Quality Measures Clearinghouse, the Agency for Healthcare Research and Quality,[5] or by a search of the literature. Initial selection focused on outcome measures for mothers and term infants and particularly on those measures that would be effective for a hospital system–based quality improvement program for maternity services. The measures were then assessed for usefulness with the parameters that are described in Table I.[6,7]

One of the first priorities of the clinical initiative was for a simple comprehensive data collection scheme. Our goal was that data collection would be complete, accurate, and minimally demanding of staff time. Six of hospitals that accounted for >60% of the births used The Perinatal Data Center software (Site of Care Systems, San Francisco, Calif), a comprehensive maternal-neonatal charting system and data analysis/reporting tool. This product identifies and abstracts key data from the medical record as it is being entered and creates efficiencies for accurate data collection. This use of primary data is a distinct advantage over many approaches. By 2005, we anticipate that 90% of our data will be collected in this manner. The alternative, used largely at some of our smaller institutions, was a web-based limited electronic logbook that was copied from the unit's labor and delivery logbook after the woman's episode of care. For this project, a restricted de-identified set of 20 data elements for each patient was forwarded electronically to Sutter Health Clinical Integration data management center. There it was evaluated and assembled into 1 large database. Data definitions were widely distributed, and education sessions were held for data entry personnel. An analysis of data that were collected during the first year (2001) identified issues with completeness; therefore, we instituted a comprehensive data collection quality improvement plan that included specific tools for

Table I Criteria used to assess hospital-based obstetric quality measures

Criteria	Rationale
Frequency	The outcome should have a reasonable rate of occurrence per year; an infrequent outcome will not be useful for comparison purposes (eg, maternal mortality rate).
Collectability	The ability to collect data accurately is critically important; some outcomes are either hard to obtain from the record or the definition is not standardized.
Relevance	The measure must be relevant to current practice, which means that it should be an important, noncontroversial, and engaging current issue (eg, VBAC is not).
Controllability	The outcome should be amenable to intervention and under provider "control"; some obstetric outcomes are beyond the control of the provider (eg, cerebral palsy).
Evidence-based	The measure should have good supporting literature because good evidence is required to convince providers to change current practices.
Focused	The measure should be focused narrowly to allow for change intervention; some outcomes are too broad to allow hospital-based quality improvement activities (eg, perinatal mortality rates).
Universal	The measure should be applicable widely to all units; an outcome that is relevant for only some units (eg, tertiary care centers) will have less appeal.
Variable	The outcome should show variation that could indicate the effects of a different practice style; an outcome with the same rate everywhere is unlikely to be changeable.

individual sites to monitor data collection completeness and accuracy.

Results

Outcome and process measure selection

The analysis of maternal measures is given in Table II. The committee noted that the recent controversies around the use of total cesarean birth or even primary cesarean birth rates as quality indicators. (Who can argue against a physician/patient decision for a cesarean delivery for a non−vertex presentation or for a number of other indications?) Furthermore, the committee observed that primary cesarean birth rates and VBAC rates would be non-starters among the obstetric staff and deflect engagement in the project as a whole. However, there was considerable interest in the concept of cesarean birth rate for term nulliparous women in labor (largely driven by cephalopelvic disproportion/failure to progress and fetal distress) that is captured by the standardized nulliparous, term, singleton, vertex (NTSV) cesarean birth rate.[8] Although cesarean birth rate can be viewed as a process measure (rate of performance of a procedure), it is also an outcome in its own right. Cesarean birth has both immediate complications and more importantly creates the state of "previous cesarean birth," which bestows increased risks for subsequent pregnancies, however they are delivered. NTSV cesarean birth rate was also the measure chosen by the American College of Obstetricians and Gynecologists[9] and by the US Department of Health and Human Services for Healthy Person 2010 goals.[10] Indeed, in previous studies, the NTSV cesarean birth rate accounts for much of the inter-hospital and interprovider variation in the cesarean birth rate.[8,9,11]

The recent recognition of short- and long-term pelvic floor morbidity from difficult vaginal deliveries prompted the FPAD group to focus on a related morbidity: the rate of 3rd/4th-degree perineal laceration. There is a considerable literature on this topic and a clear understanding of the antecedent risk factors, a number of which are under provider control.[12,13] The greatest patient risk factor for 3rd/4th-degree perineal laceration is nulliparity (6- to 8-fold increased risk), so it fits well into the concept of a "first pregnancy" focus. NTSV cesarean births and 3rd/4th-degree lacerations are relevant to all birthing units and have a high degree of variability among centers.

Intrapartum hemorrhage and infection are important maternal morbidities, but as quality measures, they suffer from a limited literature and, most importantly, from the uncertainty that they actually result from provider actions. Furthermore, their incidence is relatively low at many centers, and their coding is incomplete at best. Prolonged maternal postpartum length of stay includes the aforementioned 2 morbidities, with the addition of medical complications. The literature does not well resolve whether differences in these measures truly reflect differences in the quality of medical care.

We felt that it was important to balance maternal outcome measures with at least 1 indicator for neonatal outcome that is relevant for term infants. Of the group of neonatal measures that we analyzed (Table III), there was not one that occurs with any large frequency. Two issues clouded many of the measures. First was applicability to all units; all the measures that involve NICU admission or care have no meaning to the large number of hospitals without a NICU; second was the degree to

Table II Assessment of maternal quality measures

Measure	Frequency*	Collectability†	Relevance‡	Controllability‡	Evidence-based‡	Focused‡	Universal‡	Variable‡
Total cesarean birth rate	+++	+++	+	+	++	+	+++	+++
Primary cesarean birth rate	+++	+++	+	+	++	+	+++	+++
NTSV cesarean birth rate§	+++	++	++	++	+++	+++	+++	+++
3rd/4th-Degree laceration rate	++	+++	+++	++	+++	++	+++	+++
NTSV 3rd/4th-degree laceration rate§	++	++	+++	++	+++	+++	+++	+++
Intrapartum hemorrhage > 1000 mL	+	+	++	?	+	+	+++	?
Intrapartum hemorrhage with transfusion	+	+	++	?	+	+	+++	?
Postpartum endometritis	++	+	+	?	+	+	+++	++
Prolonged postpartum LOS	+	+++	+	?	+	+	++	?

* Rating: +++ (>10%), ++ (2%-10%), + (<2%).

† Rating: +++ (obtainable from administrative data with good accuracy), ++ (easily obtainable from logbooks or standard databases with good accuracy), + (significant concern about definition and charting accuracy).

‡ Scaled by the FPAD committee from the literature and the preliminary studies at our institutions.

§ The highlighted rows indicate the highest scoring measures and were selected for our FPAD Clinical Initiative. All measures were ranked up to +++.

Table III Assessment of term infant quality measures

Measure	Frequency*	Collectability†	Relevance‡	Controllability‡	Evidence‡	Focused‡	Universal‡	Variable‡
Perinatal mortality rate	+	+	+++	+	++	+	++	+++
NICU admission rate	++	+++	+	+	+	+	0	++
NICU days per 1000 births	++	+++	+	+	+	+	0	++
Newborn infant length of stay > 4 d	++	+++	+	+	+	+	++	?
JCAHO/outcome research yields excellence "term sick infant" rate§	++	+	++	++	++	+	0	++
Neonatal ventilator use (d)	+	+	++	+	+	++	0	?
5-Minute Apgar score <7‖	+/++	++	++	++	++	++	+++	+++
5-Minute Apgar score <4	0/+	++	++	+	++	++	+++	++

JCAHO, Joint Commission on Accreditation of Healthcare Organizations.

* Rating: +++ (>10%), ++ (2%-10%), + (<2%).

† Rating: +++ (obtainable from administrative data with good accuracy), ++ (easily obtainable from logbooks or standard databases with good accuracy), + (significant concern about definition and charting accuracy or difficult to ascribe to the correct hospital [perinatal mortality rate]).

‡ Scaled by the FPAD committee from the literature and the preliminary studies at our institutions.

‖ Indicates the highest scoring measures and was selected for our FPAD Clinical Initiative; all measures were ranked up to +++.

§ Infants > 2500 g at birth with birth injury, seizure, 5-minute Apgar score <3, or NICU admission.

which these outcomes were under the control of the obstetric providers. Term 5- minute Apgar score of <7 stood out as relatively frequent (1%-2%) and, in many cases, reflective of labor care and resuscitation capabilities. Furthermore, there are recent studies that indicate both short-term morbidities and longer term associations with low 5-minute Apgar scores in term infants.[14,15]

Last, in most effective quality improvement programs, there should be a mix of outcome measures with process measures.[5] Process measures are typically the rate of doing a procedure that is linked to (or is a "driver" of) an outcome. It is easier to focus quality improvement activities and to effect quicker change with process measures than with outcome measures. Familiar examples of process measures include Papanicolaou test rate (for cervical cancer prevention), mammogram rate (for breast cancer early detection), and hemoglobin A1C rate in persons with diabetes mellitus (for attention to tight control of blood sugars). None of these represent an outcome in itself, but each is an important part of the care that leads to an outcome. We strove to identify several process measures that would help focus our quality improvement efforts in this manner. Table IV displays our considerations. The literature is quite strong for several of these: routine episiotomies linked to 3rd/4th-degree lacerations,[16,17] nulliparous inductions,[18-20]

Table IV Assessment of obstetric process measures

Measure	Frequency*	Collectability†	Relevance‡	Controllability‡	Evidence‡	Focused‡	Universal‡	Variable‡
NTSV episiotomy rate (drives 3rd/4th-degree lacerations)§	+++	+++	+++	+++	+++	+++	+++	+++
Operative vaginal delivery rate (drives 3rd/4th-degree lacerations)	+++	+++	+/++	++	++	++	+++	+++
NTSV induction rate (<41wk; drives NTSV cesarean birth rate)§	+++	++	+++	++	+++	+++	+++	+++
NTSV cervical dilation ≥3 cm at admission (drives NTSV cesarean birth rate)§	+++	++	+++	++	++	+++	+++	++
Hours from admission to delivery; drives NTSV cesarean birth rate and unit costs	++	+	+	++	+	+	+++	++
Ability to move legs during second-stage labor (drives operative vaginal deliveries and 3rd/4th-degree lacerations)	+++	+	++	++	+	++	++	?

* Rating: +++ (>10%), ++ (2%-10%), + (<2%).

† Rating: +++ (obtainable from administrative data with good accuracy), ++ (easily obtainable from logbooks or standard databases with good accuracy), + (significant concern about definition and charting accuracy).

‡ Scaled by the FPAD committee from the literature and the preliminary studies at our institutions.

§ Indicates the highest scoring measures and was selected for our FPAD Clinical Initiative; all measures were ranked up to +++.

and early admission in labor[21] that are linked to development of dystocia-related cesarean birth. Furthermore, all 3 proposed process measures appear to have a fair amount of provider control.

We also chose to use a patient satisfaction measure as an important outcome parameter.[5] Some of the recommended changes (eg, fewer elective nulliparous inductions, later labor admissions) could require considerable patient education not to decrease patient satisfaction. Having a measure of patient satisfaction would keep each hospital's focus on a more holistic view of obstetric care. We chose a single summary satisfaction measure that was defined as the product of the overall patient satisfaction score (which has little variability) by the score for "Would you recommend this hospital to your family or friends?" Unfortunately, the vendor for patient satisfaction surveys in our hospital system changed in the middle of this first phase, so results will not be presented here.

Baseline data collection

Data were collected from all 20 sites, beginning the first quarter of 2001. Time for project-related data entry and transmission ranged from 2 to 3 hours per quarter for the manual systems to 15 to 30 minutes per quarter for the integrated electronic system. A data collection quality improvement plan was introduced systemwide in January 2002 to address missing data. Figure 1 shows the success of data quality improvement efforts over the first year. Currently, we have <0.3% of data points missing of nearly 30,000 data points that were collected each quarter.

Individual hospital results for 2002 are shown in Figures 2 and 3 (each hospital is represented by a letter). There is considerable variation among the hospitals in all the measures, not explicable by patient variables. The outcome measures (NTSV 3rd/4th-degree laceration and NTSV cesarean birth rate) vary 2- to 3-fold (Figure 2). The process measures have even greater variability; among hospitals, the measures vary by 2.5- to 10-fold (Figure 3). Also of interest, there is considerable shuffling of hospital order among the measures so that an individual hospital's position varies greatly from one measure to another. Indeed, each hospital can claim credit for being near the top in ≥1 measures; at the same time, every hospital has measures that indicate an "opportunity for improvement."

Comment

In an era in which both physicians and patients are uninspired or even dismissive of a quality goal of low/lower/lowest total cesarean birth rate, we strove to

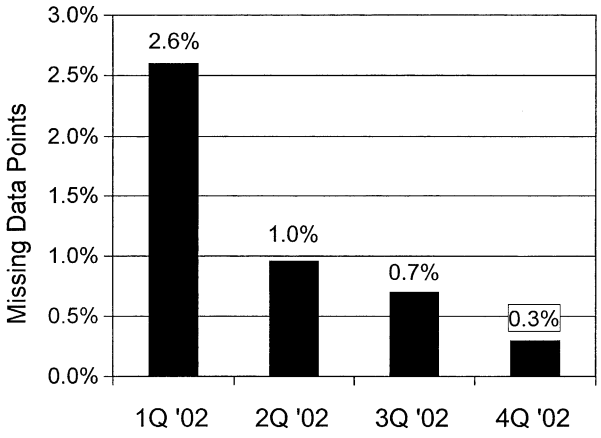

Figure 1 Data completeness, as measured by missing data points, during year-long data quality improvement effort (>28,000 data points were collected each quarter).

identify a meaningful alternative approach to quality improvement in obstetrics. First, we recognized that a hospital's entire maternity population was too heterogeneous in terms of diagnoses and clinical issues to be considered as the appropriate quality improvement focus (denominator). Prematurity (10%-15% of a hospital's births) is a major driver of interventions and determinant of poor outcomes but is often outside of the control of the provider and is distributed very unevenly among hospitals. Likewise, indicated cesarean delivery largely has replaced vaginal breech delivery for most singleton infants and for many twin infants (together totaling 3%-5% of births). Finally, the national reassessment of VBAC delivery should remove VBAC/repeat Cesarean birth (10%-18% of births) as a quality indicator. Therefore, we believe that a more limited but more homogeneous population is needed for quality assessment purposes. We embraced the principle of "high volume/high risk" and chose to focus on the NTSV population. This is similar to the "standardized primip" that is used widely in outcome studies and for unit comparisons in the United Kingdom. It is also the definition that we have used in Northern California to analyze cesarean birth rates for >10 years[8] and is now embraced by the American College of Obstetricians and Gynecologists[9] and by the US Department of Health and Human Services in their Healthy Person 2010 measures as the most appropriate measure for cesarean birth.[10] Here we apply this population definition to a broader range of pregnancy outcome measures. Women who are having their first birth are at much higher risk (often ≥5-fold) not only for cesarean birth, but also for 3rd/4th-degree lacerations and for episiotomies, when compared with women who have had a previous successful vaginal birth.[12,13,16,17] Likewise, women who are having their first birth are at higher risk of complications from an induction[18-20] and for babies with lower 5-minute Apgar scores and babies who require care in the NICU (data from this project). The NTSV group is also "high volume," comprising>40% of most modern obstetric unit births, which reflects the demographic shift in the United States towards smaller family size. This population is distributed relatively uniformly among all sizes and types of obstetric units, which makes it a good choice for interunit comparisons and relevancy.

We also recognized that a single outcome measure would not do justice to the complex balance of outcomes that are represented by birth. Therefore, we replaced the current "standard" quality measure, cesarean birth rate, with a "basket" of quality measures that attempt to balance the maternal risks of vaginal birth (3rd/4th-degree laceration rate) with cesarean birth (risk-adjusted cesarean birth rate); both are balanced against the infant's interests (5-minute Apgar score of <7 rate) and the mother's satisfaction with her care (patient satisfaction index). An additional set of measures (NTSV inductions at <41weeks of gestation, NTSV labor admissions ≥3 cm dilation, and NTSV episiotomies) that act as "drivers" of the aforementioned outcome measures are also collected to help focus the quality improvement efforts. This balanced set of measures (mother/baby, vaginal/cesarean birth, medical outcomes/patient satisfaction) gives a better appreciation for the complexity of the care that we provide and should help us to avoid overemphasizing change in 1 dimension (eg, reducing cesarean rates) that might be at the expense of the other outcomes of importance (eg, infant and perineal morbidities). In this model, the ideal obstetric provider is not the one with the lowest total cesarean rate but rather the one with the best mix of low risk-adjusted (NTSV) 3rd/4th-degree laceration rate, low NTSV cesarean birth rate, and low 5-minute Apgar score of <7 rate that also has the best patient satisfaction scores.

The use of nulliparity, term gestational age, singleton plurality, and vertex presentation as the major criteria to define a population to focus quality improvement efforts appears to be a reasonable start. Other models have included race as an adjustor (the effects are generally quite modest for these outcomes), but there are considerable arguments that to include race in the final paradigm (and therefore factor out race or social factors) could lead to reduced attention to race-related morbidities. Maternal age is another risk factor than is used in many models of cesarean birth. Indeed, after the 4 risk factors (NTSV) used here, maternal age follows as the next most important (but considerably less strong) patient-related contributor to cesarean delivery rates.[11] The average age of nulliparous women in our Northern California hospitals ranged widely from 22.5 years in many of the hospitals in the Central Valley to 31.7 years in hospitals in the San Francisco Bay region. Interestingly, however, when we excluded patients who were >35 years old from the analysis, there was very little effect on the hospital scores or rankings. Therefore, for simplicity in the model

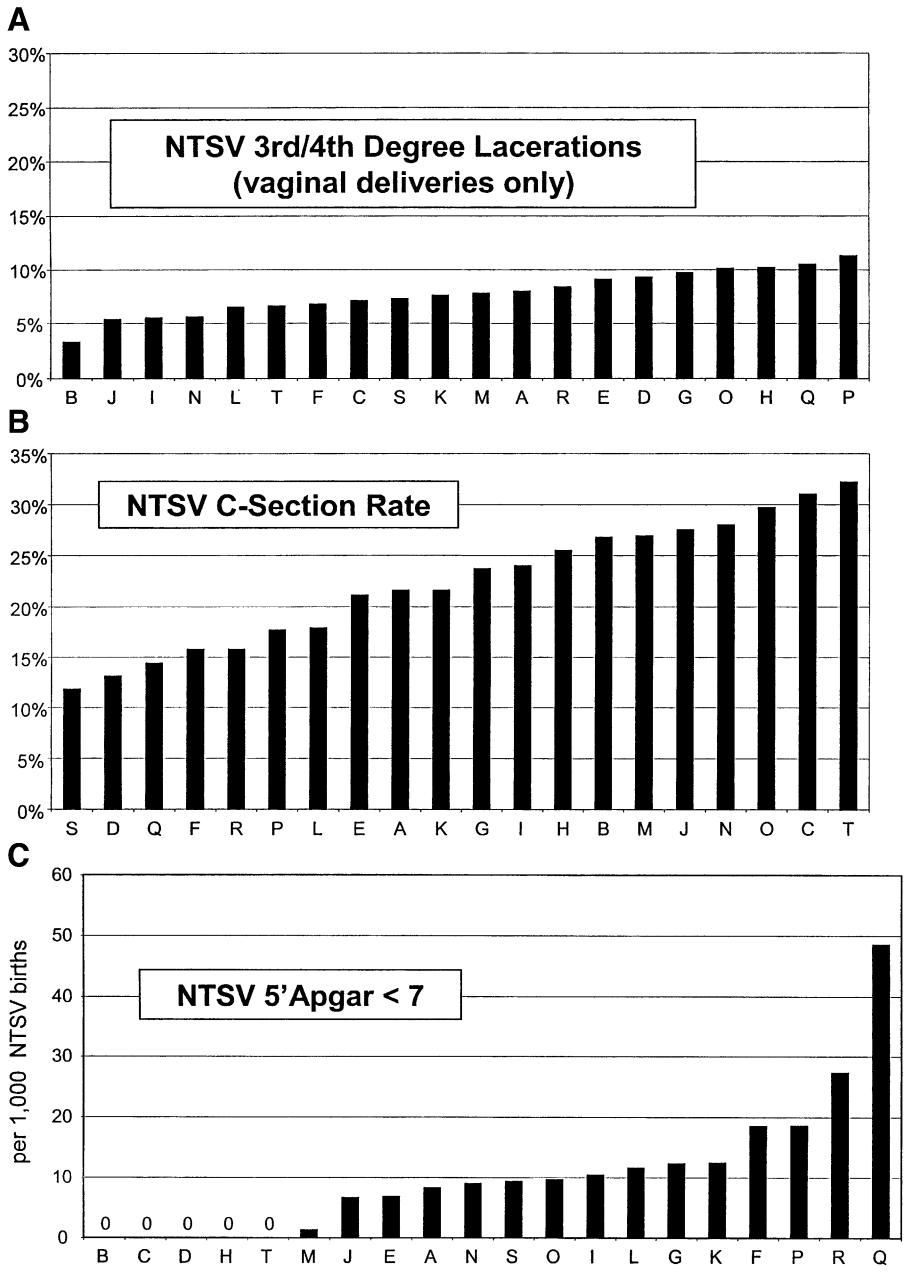

Figure 2 Outcome measures by hospital for the entire year of 2002 (NTSV 3rd/4th-degree laceration rate [**A**], NTSV cesarean birth rate [**B**], and NTSV 5-minute Apgar score <7 rate [**C**]). Each individual hospital within Sutter Health is marked by a letter that is consistent through each panel. The proportion of relevant patients with the measure is noted in the Y-axis (except the 5-minute Apgar score, which is the number per 1000 NTSV births).

construct and to be able to compare our results to state and national NTSV rates, we have chosen to not include age in our population definition. We intend to monitor these and other potentially useful risk adjustors on an ongoing basis. Indeed, we should emphasize that our intent is not to form a perfect risk-adjustment model but to create a risk-homogeneous subpopulation for quality improvement efforts that is both reasonable and easily grasped by providers and patients alike. Of interest, the variable that we did not include that has been demonstrated to have the largest effect on these results (beyond the ones controlled for) is the individual provider.[22] This would confirm that both the hospital and the individual obstetric provider are appropriate targets for quality improvement activities.

Parity is a powerful risk factor for nearly every labor outcome but lacks an ICD-9 code. This omission makes administrative (discharge diagnosis) data sets significantly reduced in value for doing many obstetric outcome analyses. Some researchers have compensated for this major deficiency by carefully matching (a time-consuming process) hospital discharge diagnosis

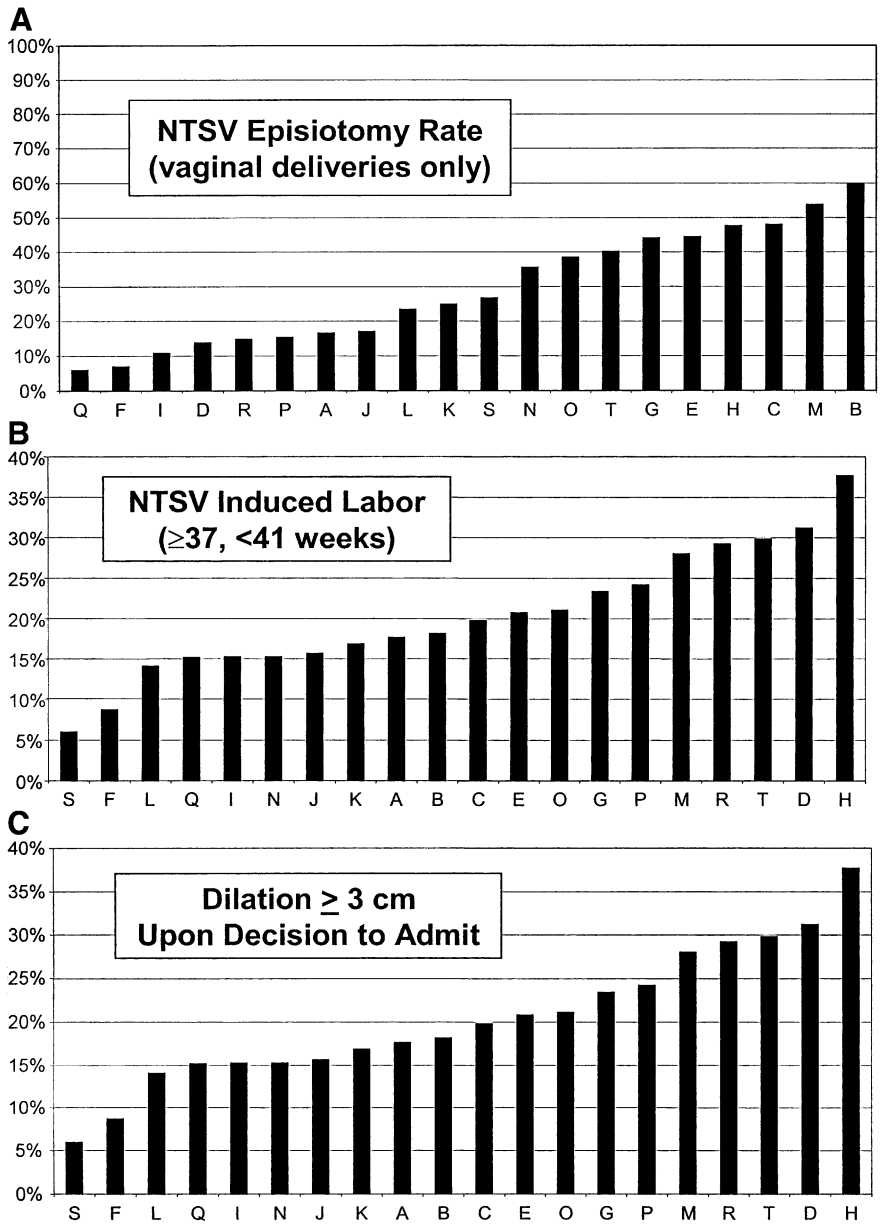

Figure 3 Process measures by hospital for the entire year of 2002 (NTSV episiotomy rate [**A**], NTSV <41 week induction rate [**B**], and NTSV labor admission ≥3 cm dilation rate [**C**]). Each individual hospital within Sutter Health is marked by a letter that is consistent through each panel. The proportion of relevant patients with the measure is noted in the Y-axis.

information with birth certificate data. Hopefully, for the sake of quality improvement in our specialty, a code for nulliparity can be added in the future.

Data collection efforts for this project indicate that large quality improvement efforts need not depend on administrative data. An integrated mother-baby database and report-generating system deployed at all the larger hospitals and simple electronic logs at many of the smaller centers allowed us to collect complete and meaningful data successfully. Indeed, the completeness of this data set (missing data rate, <0.3%) is considerably better than many states' birth certificate data. Data completeness and accuracy form a critical foundation for quality improvement activities, because the first defense of many providers is to "attack the data." In fact, the process of collecting data can be turned into an educational and quality improvement effort in its own right. Our most difficult measure to collect accurately, which was NTSV cervical dilation on admission, is now the focus of process changes and education on our obstetric units. This, in turn, necessarily focuses the unit's attention on the issue of latent phase labor admissions.

To be useful for quality improvement, a measure should be collected easily, be applicable to a wide range of facilities, be under provider control, be well supported

by literature (in other words be ''evidence-based''), and be associated with a relevant clinical outcome. In addition, the measure should show considerable variation among centers or providers. Significant variation in rates (especially within risk-adjusted populations) generally indicates that practice style rather than clinical conditions is playing an important role in determining the rates of both procedures and outcomes.[23,24] All of our measures exhibited significant variation among centers and, when examined at larger single centers, among individual practitioners as well (data not shown). This variation can be harnessed as an important tool for change.[8]

The approach of focusing on nulliparous women and of addressing a larger spectrum of balanced outcomes other than cesarean birth rate alone has re-engaged obstetricians, midwives, labor nurses, administrators, and pregnant women and their partners in quality improvement efforts. All parties recognize that the first delivery is the hardest, and all parties see the value in multiple dimensions of outcome measurement. In the past, if a birthing unit did poorly in its single quality measure (eg, total cesarean delivery rate), the professional members of that unit often withdrew and criticized the value of the goal. In this study, every birthing unit has done well on some of the measures (which creates pride in one's unit and job satisfaction) and less well on others (which provides each unit with ''opportunities for improvement''). Indeed, as we move into the next stage of the program, the introduction of clinical topic-specific quality improvement activities with focused measure-specific ''tool kits,'' we are seeing critical discussion of practice styles in each hospital and a comparison of outcomes at the level of the practitioner. Current literature is being reviewed in the context of one's own daily practices, and there are lively debates about widely held obstetric beliefs. We hypothesize that these activities and attitudes are the harbingers of real and sustained quality improvement.

Other members of the Sutter Health, FPAD Clinical Initiative Committee include: Carol Abrahams, RN, California Pacific Medical Center, San Francisco; Robert Altman, MD, Gould Medical Group Modesto; Kathy Bunger, RN, Marin General Hospital Greenbrae; Mary Campbell Bliss, RN, Sutter Memorial Hospital Sacramento; Bill Farrell, PhD, Sutter Health Support Services, Sacramento; Gail Fujimoto, RN, Sutter Medical Center Santa Rosa; Herman Hedriana, MD, Sutter Medical Center, Sacramento; Brian Jones, MD, Sutter Medical Center, Sacramento; Lizellen La Follette, MD, Marin General Hospital, Greenbrae; Kristin Ostrom, CNM, Sutter Lakeside Hospital, Lakeport; Laurie Polivy, RN, Mills-Peninsula Health Services, Burlingame; Anita Rushworth, RN, Sutter Auburn Faith Hospital, Auburn; Howard Salvay, MD, Sutter Maternity & Surgery Center, Santa Cruz; Nory Sargeant, RN, Sutter Roseville Medical Center, Roseville; Leon Schimmel, MD, Sutter Davis Hospital, Davis; Jacci Sterling, RN, Sutter Maternity & Surgery Center, Santa Cruz; John Whitelaw, MD, Sutter Health Support Services, Sacramento.

References

1. Sachs BP, Kobelin C, Castro MA, Frigoletto F. The risks of lowering the cesarean-delivery rate. N Engl J Med 1999;340:54-7.
2. Greene MF. Vaginal delivery after cesarean section: Is the risk acceptable?. N Engl J Med 2001;345:54-5.
3. Births: preliminary data for 2002. NVSR Vol. 51, No. 11. 20 pp. (PHS) 2003-1120. (June 25, 2003. Available at: http://www.cdc.gov/nchs/data/nvsr/nvsr51/nvsr51_11.pdf
4. Cleary R, Beard RW, Chapple J, Coles J, Griffin M, Joffe M, et al. The standard primipara as a basis for inter-unit comparisons of maternity care. BJOG 1996;103:223-9.
5. National Quality Measures Clearinghouse, Agency for Healthcare Research and Quality (AHRQ), US Department of Health and Human Services. Available at: http://www.qualitymeasures.ahrq.gov/
6. Brook RH, McGlynn EA, Cleary PD. Measuring quality of care. N Engl J Med 1996;335:966-70.
7. Kassirer JP. The quality of care and the quality of measuring it. N Engl J Med 1993;329:1263-5.
8. Main EK. Reducing cesarean birth rates with data-driven quality improvement activities. Pediatrics 1999;103:374-83.
9. Evaluation of cesarean delivery. American College of Obstetricians and Gynecologists. 2000, p. 34.
10. Healthy people 2010: Chapter 16: Maternal, infant and child health, (measure definition 16-9a). Office of Disease Prevention and Health Promotion, US Department of Health and Human Services Available at: http://www.healthypeople.gov/document/pdf/tracking/od16.pdf
11. Bailit JL, Dooley SL, Peaceman AN. Risk adjustment for interhospital comparison of primary cesarean rates. Obstet Gynecol 1999;3:1025-30.
12. Zetterstrom J, Lopez A, Anzen B, Norman M, Holmstrom B, Mellgren A. Anal sphincter tears at vaginal delivery: risk factors and clinical outcome of primary repair. Obstet Gynecol 1999;94:21-8.
13. Eason E, Labreque M, Wells G, Feldman P. Preventing perineal trauma during childbirth: a systematic review. Obstet Gynecol 2000;95:464-71.
14. Thorngren-Jerneck K, Herbst A. Low 5-minute Apgar score: a population-based register study of 1 million term births. Obstet Gynecol 2001;98:65-70.
15. Casey BM, McIntire DD, Leveno KJ. The continuing value of the Apgar score for the assessment of newborn infants. N Engl J Med 2001;344:467-71.
16. Myers-Helfgott MG, Helfgott AW. Routine use of episiotomy in modern obstetrics: Should it be performed? Obstet Gynecol Clin 1999;26:305-25.
17. Webb DA, Culhane J. Hospital variation in episiotomy use and the risk of perineal trauma during childbirth. Birth 2002;29:132-6.
18. Dublin S, Lydon-Rochelle M, Kaplan RC, Watts DH, Critchlow CW. Maternal and neonatal outcomes after induction of labor without an identified indication. Am J Obstet Gynecol 2000;183:986-94.
19. Seyb ST, Berka RJ, Socol ML, Dooley SL. Risk of cesarean delivery with elective induction of labor at term in nulliparous women. Obstet Gynecol 1999;94:600-7.
20. Johnson DP, Davis NR, Brown AJ. Risk of cesarean delivery after induction at term in nulliparous women with an unfavorable cervix. Am J Obstet Gynecol 2002;188:1565-72.

21. Lee L, Hitti J. Is there an association between early labor admission and the risk of cesarean section for dystocia in low risk, nulliparous women [abstract]? Am J Obstet Gynecol 2002;187:S134.
22. Luthy DA, Malmgren JA, Zingheim RW, Leninger CJ. Physician contribution to a cesarean delivery risk model. Am J Obstet Gynecol 2003;188: 1597-87.
23. O'Connor GT, Quinton HB, Traven ND, Ramunno LD, Dodds TA, Marciniak TA, et al. Geographic variation in the treatment of acute myocardial infarction: the cooperative cardiovascular project. JAMA 1999;281:627-33.
24. Wennberg JE, Cooper MM. The Dartmouth atlas of health care in the United States 1999. Chicago: American Publishing Inc; 1999.

Editor's note: This manuscript was revised after these discussions were presented.

Discussion

Dr Roger Rowles, Yakima, Wash. Over the last several years, this body has heard a number of reports detailing cesarean rates and VBAC rates as indicators of quality of care. This paper acknowledges those factors, but, I believe, takes us further down the road of quality assessment and assurance in obstetrics and leads us away from looking just at cesarean rates and VBAC rates as sole indicators of good obstetric care.

The laboratory in this assessment is comprised of a very large number of pregnant women who were delivered within the Sutter Health Network of Northern California. In an attempt to further define their population as a target for quality improvement, they choose the "standardized primipara," the woman who is having her first child at or near term, with a single baby, with vertex presentation. This is a patient who generally is considered to be at high risk for difficult labors and other morbidities, one who is prevalent in the obstetric population, and one who triggers a heightened state of awareness among obstetrics providers, labor and delivery staff, and, frequently, an extended and concerned family.

The authors looked at 20 measures that could potentially affect or reflect the quality of care, to include both process and outcome measures. They then submitted these measures to an examination as to their frequency, measurability, relevancy, controllability, universality, variability, and consistency with scientific evidence or general consensus. From this list of 20 measures, they arrived at 7 measures that they felt fulfilled all of the aforementioned criteria, which would apply to small and large hospitals alike, and which looked at the comprehensive maternal unit (mother, baby, and family). These 7 measures include outcome assessments (primary cesarean rate, 5-minute Apgar scores <7, and 3rd/4th-degree lacerations), and process measures that are thought to be linked to these outcomes (induction of labor between 37 and 41 weeks of gestation, latent labor admissions, and episiotomy) and, as a separate outcome measure, patient satisfaction.

Collection of this data on >40,000 patients per year, scattered over 20 birthing units, was initiated in January 2001 and was accomplished with a comprehensive charting system and data analysis\reporting tool. The first year of data collection was used to assess the quality of the data, and by the second year, an improved data collection plan was in place. By the end of that second year, <3% of data points were missing of nearly 30,000 data points that had been collected each quarter, which gives the authors assurances of the data accuracy and completeness. The numbers for each hospital were then presented, which allowed interhospital comparison and allowed hospitals to track their own results quarter to quarter.

The results were not surprising. As anticipated, there was considerable interhospital variation, which confirmed the suitability of these measures as a focus for improvement. These 7 measures were then analyzed as to the 2-year trend. There was a significant reduction in 3 measures: induction rates, episiotomy rates, and 3rd/4th-degree lacerations. Cesarean delivery rates and low Apgar scores did not change significantly, and latent labor admissions and patient satisfaction were ultimately difficult to assess for logistic reasons.

The idea of developing a standard clinic data set that is collected routinely across all units offers the chance to make a comparison between units with respect to both process and outcome and creates the potential for quality improvement. Historically, this has been difficult for 2 reasons: assuring accuracy and completeness of data and controlling for differences in case mix. Dr Main et al have successfully overcome the problems of data, and using the "standard primiparous" model, which was defined initially in the United Kingdom by Patterson et al[1] in 1991 and refined by Cleary et al[2] in 1994, have overcome the second difficulty. With the results at hand, they now launch into the next stage of the initiative, the introduction of topic-specific quality improvement tool kits.

There are several interesting observations on the data presented for each hospital that I would ask Dr Main to comment on. The hospital with the highest episiotomy rate had the lowest rate of 3rd/4th-degree lacerations, and the hospital with the lowest episiotomy rate had the second highest rate of lacerations. The hospital with the second highest induction rate had the second lowest cesarean rate. This all seems contradictory. How do you reconcile these data when approaching these hospitals with educational tools?

Did you consider further standardizing your "standard primiparous" by eliminating those with medical complications or those who had medical indications for induction of labor?

In your report in 1999,[3] you successfully reduced cesarean birth rates at 2 hospitals in San Francisco by unblinding the names of the practitioners and their individual cesarean birth rates. Could you expand on you quality improvement tool kits for this initiative, and whether results will also be unblinded?