Quality Measures (for Accountability): What Should We Measure, and Why?

Munish Gupta, MD MMSc

MidCaN Webinar Series

March 17, 2022

American Academy of Pediatrics
DEDICATED TO THE HEALTH OF ALL CHILDREN®

Section on Neonatal-Perinatal Medicine

A little context…

Outline

- A little background on quality measures
- Measures for improvement
- Measures for accountability
- Measures for benchmarking
- SONPM Quality Measures Project

Donabedian Framework for Evaluating Quality

Donabedian, Evaluating the Quality of Medical Care, Milibank Quarterly, 1966

Why measure?

- Measures drive improvement
- Measures inform consumers
- Measures influence payment

National Quality Forum, ABCs of Measurement

https://www.qualityforum.org/Measuring_Performance/ABCs/The_Difference_a_Good_Measure_Can_Make.aspx

Measurement for Research vs. Improvement

IHI, “Science of Improvement: Establishing Measures”

	Measurement for Research	Measurement for Improvement
Purpose	To discover new knowledge	To bring new knowledge into daily practice
Tests	One large "blind" test	Many sequential, observable tests
Biases	Control for as many biases as possible	Stabilize the biases from test to test
Data	Gather as much data as possible, "just in case"	Gather "just enough" data
Duration	Can take long periods of time to obtain results	"Small tests of significant changes"

http://www.ihi.org/resources/Pages/HowtoImprove/ScienceofImprovementEstablishingMeasures.aspx

JOURNAL ON QUALITY IMPROVEMENT

We are increasingly realizing not only how critical measurement is to the quality improvement we seek but also how counterproductive it can be to mix measurement for accountability or research with measurement for improvement.

PERFORMANCE MEASURES AND MEASUREMENT

The Three Faces of Performance Measurement:

Improvement, Accountability, and Research

LEIF I. SOLBERG, MD
GORDON MOSSER, MD
SHARON MCDONALD, RN, PHD

Solberg et al, Three Faces of Performance Measurement, Joint Commission Journal on Quality Improvement, 1997

Measurement for Improvement vs. Accountability

	Improvement	Accountability
Audience	Medical group, QI team Providers Administrators	Purchases Payers Patients
Measures	Few Easy to collect Approximate	Very few Complex collection Precise and valid
Time period	Short, current	Long, past
Confounders	Consider but rarely measure	Describe and try to measure
Sample Size	Small	Large
Collection	Simple, minimal cost and expertise Usually repeated	Complex, moderate effort and cost

Solberg et al, Three Faces of Performance Measurement, Joint Commission Journal on Quality Improvement, 1997

Measurement for Improvement vs. Accountability

- Measures for improvement:
  - Used by individual units or groups
  - Seek to drive improvement for specific goals
  - Can have variability in definition or measures across institutions
  - Definitions should be consistent within the institution

- Measures for accountability:
  - Used by organizations that regulate, pay, or report to public
  - Seek to compare performance
  - Require uniform and reliable definition across institutions

Table 3. Recommendations on How to Make More Use of Measurement in Improvement Efforts

1. Limit the number of measurements.
2. Pick measurements that are important to clinicians (and patients), ideally by having the users do the selection.
3. Make the data collection easy enough and the time frames short enough so that data collection can be repeated frequently to allow for trending changes over time.
4. Do not try to have the measures serve accountability or research purposes at the same time as improvement.
5. Build in baseline measures before implementing any changes.
6. Provide training, tools, and examples to those in clinical settings who are not used to data and this type of measurement.

Solberg et al, Three Faces of Performance Measurement, Joint Commission Journal on Quality Improvement, 1997

MEASURES FOR IMPROVEMENT

Thoughts on Measures for Improvement

- They can be anything your team thinks is important
- Ideally tied to your aims, drivers, and changes
- They should be clear and precise (operational definitions)
- Try to minimize data collection burden (can be tough)
- Try to collect data in real time (can be tough)
- Collect baseline data if you can (can be tough)
- Collect data over time and use time-series data analysis
- If you can benchmark, even better

Examples of Measures for Improvement

Reducing Vancomycin Use in a Level IV NICU

Rana F. Hamdy, MD, MPH, MSCE,[a,b] Sopnil Bhattarai, BS,[c] Sudeepta K. Basu, MD,[b,d] Andrea Hahn, MD, MS,[a,b] Brian Stone, MD, MBA,[b,d] Eleanor D. Sadler, PharmD,[e] Benjamin M. Hammer, PharmD,[e] John Galiote, MD,[b,d] Julie Slomkowski, PharmD,[f] Anne M. Casto, MSN, NNP-BC,[d,g] Katelyn P. Korzuch, MMS, PA-C,[d] Hannah Chase, BS,[h,i] Nneka Nzegwu, DO, MPH,[b,d] Isabella Greenberg, MPH,[c] Noelle Ortiz, MPH, MBS,[h,j] Carmen Blake, RN,[g] Jaeho Chang, MA,[c] James E. Bost, MS, PhD,[h,j] Asha S. Payne, MD, MPH,[d,j] Rahul K. Shah, MD, MBA,[c,k] Lamia Soghier, MD, MEd[b,d]

Hamdy et al, Pediatrics, 2020

Examples of Measures for Improvement

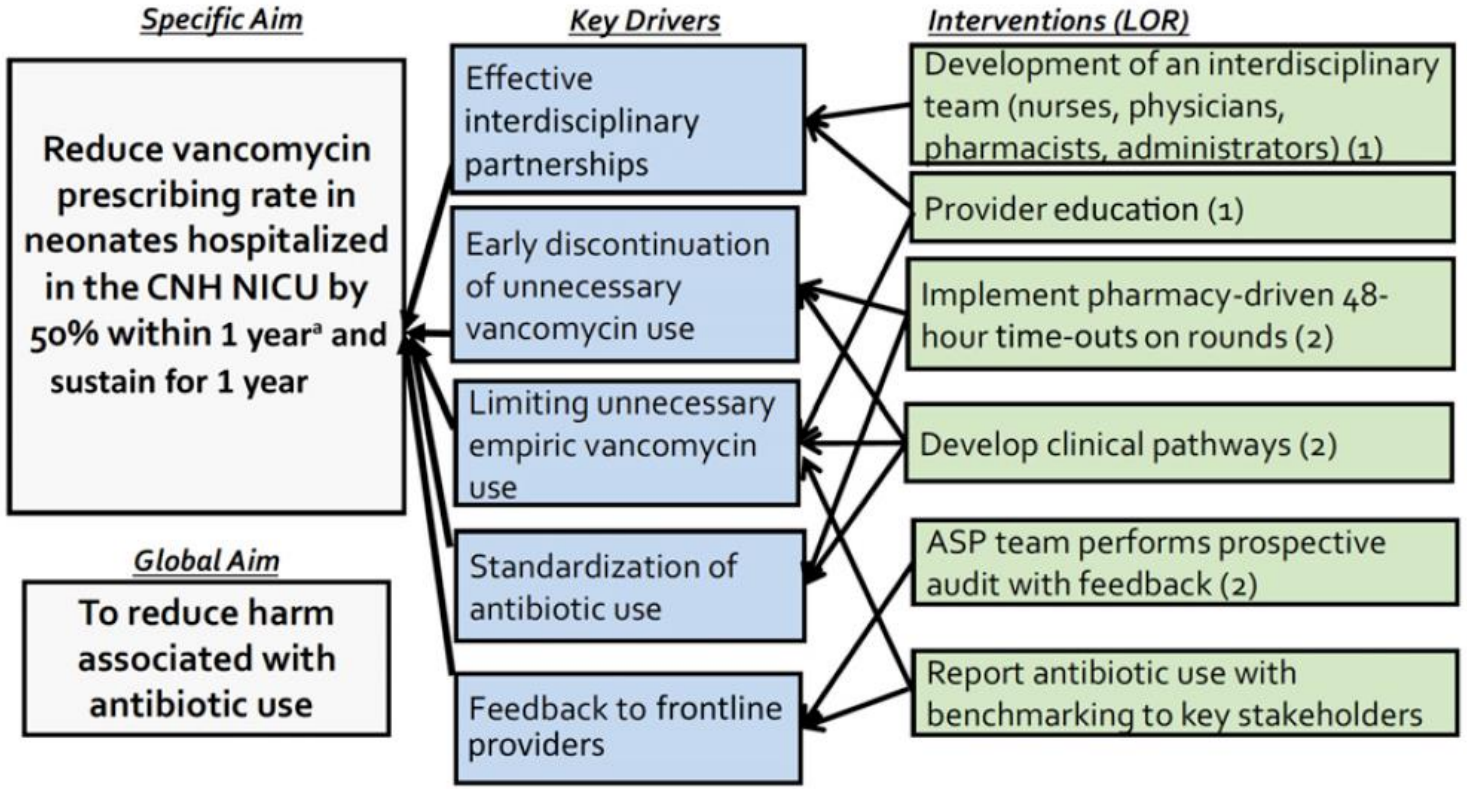


FIGURE 2

Key driver diagram for vancomycin reduction in the NICU. Levels of reliability (LOR) of 1 (helping teams develop a standardized workflow) and 2 (developing error-proof systems to prevent work-arounds) were implemented[20]; no level 3 interventions were used. [a] A 50% reduction by September 2018 as compared to September 2016 through August 2017.

Hamdy et al, Pediatrics, 2020

Examples of Measures for Improvement

Outcome Measure

The primary outcome measure was vancomycin DOTs per 1000 patient-days. A DOT was defined as ≥1 doses of intravenous vancomycin administered to the patient on a calendar day. Antibiotic data were obtained from the hospital pharmacy dispense database. NICU patient-days were obtained from the hospital daily census.

Process Measures

A third process measure, the number of vancomycin orders per month for which a 48-hour time-out was performed and documented by the unit-based pharmacist (and whether it was followed), was used to monitor pharmacist compliance with the 48-hour time-out policy.

Hamdy et al, Pediatrics, 2020

Examples of Measures for Improvement

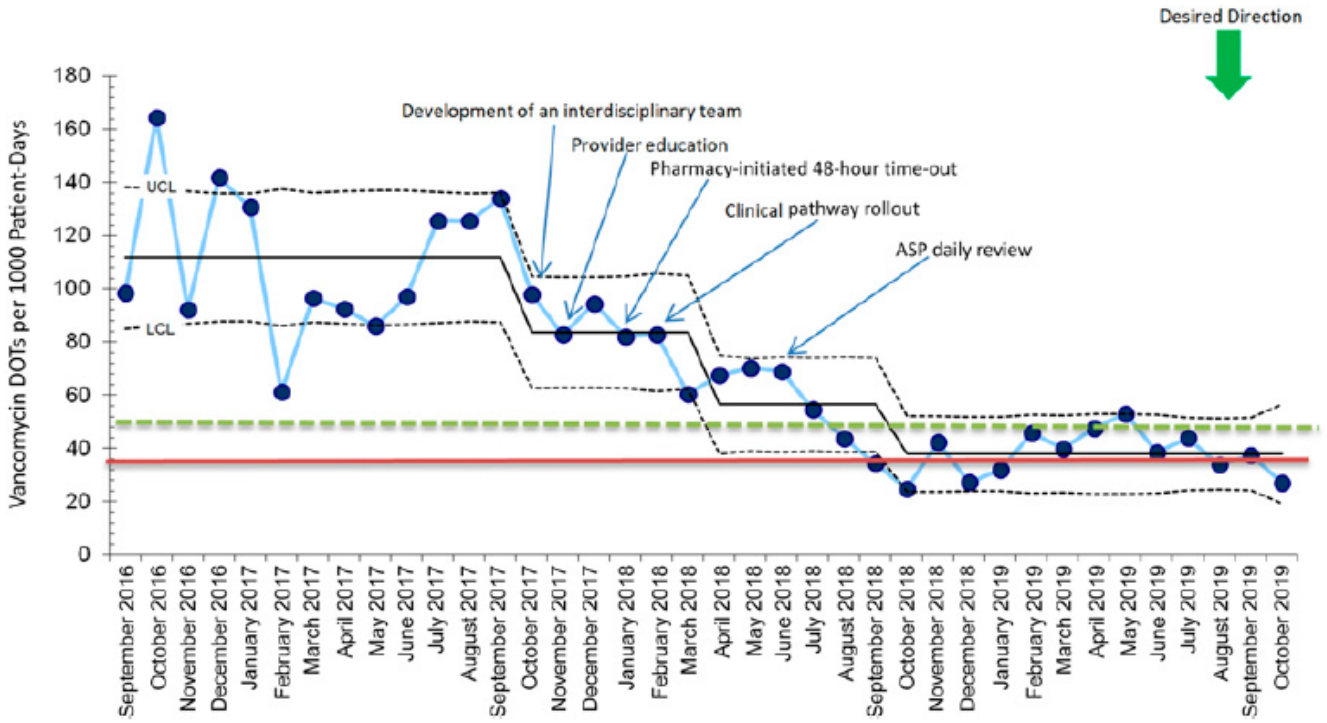


FIGURE 3

Vancomycin use in the Children's National NICU, 2016–2019: U-chart for vancomycin DOTs per 1000 patient-days. The green dashed line indicates target vancomycin DOTs per 1000 patient-days (50% of baseline; 56 DOTs per 1000 patient-days). The red solid line indicates mean vancomycin DOTs per 1000 patient-days for neonatal services (34 DOTs per 1000 patient-days) among 40 children's hospitals contributing data to Pediatric Health Information System. LCL, lower control limit; UCL, upper control limit.

Hamdy et al, Pediatrics, 2020

MEASURES FOR ACCOUNTABILITY

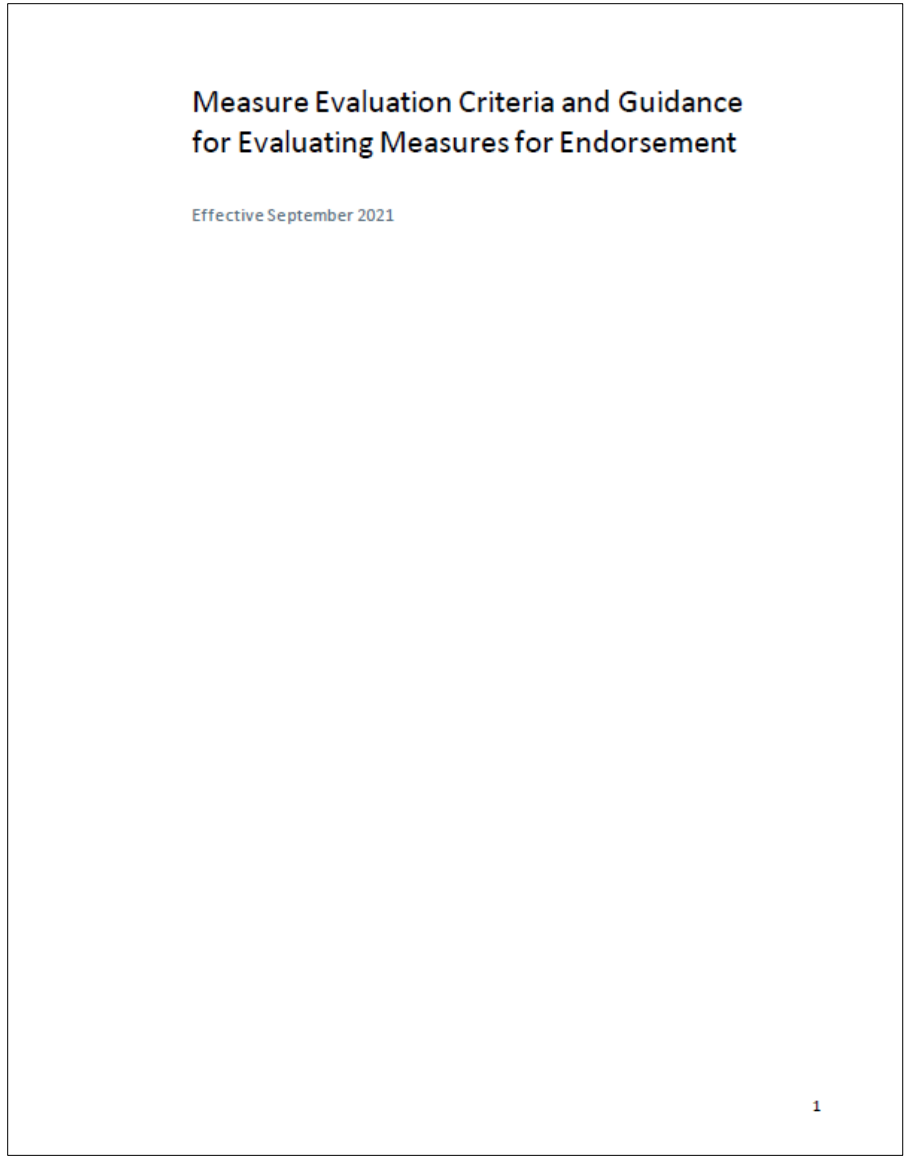
Some History of Measures for Accountability

- 1987: Health Care Financing Administration (HCFA) released hospital-specific mortality data
- 1989: New York state DPH released annual risk-adjusted CABG mortality by hospital and surgeon
- 1990s: Health Employer Data Information Set (HEDIS) from National Committee on Quality Assurance (NCQA)
- 1998: President Clinton Advisory Commission on quality
- 1999: National Quality Forum (NQF) established

National Quality Forum

- Not-for profit, non-partisan organization
- Coalition of private and public sector leaders
- Consensus-based methods
- Recommends measures for use in payment and public reporting programs
- Extensive process for reviewing measures

NQF Framework for Measure Evaluation



Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement

Effective September 2021

1

Measure Evaluation Criteria	
Importance	• Evidence-based • Performance gap or variation
Scientific Acceptability	• Reliability: consistent, repeatable • Valid: adequately measures quality, risk-adjusted, appropriate exclusions
Feasibility	• Clear to stakeholders • Data readily available, can be captured without undue burden
Usability and Use	• Contributes to improvement • actionable • Limited unintended consequences

NQF Neonatal Measures Currently Endorsed

- Exclusive breast milk feeding (PC-05)
- Proportion of infants 22 to 29 weeks screened for ROP
- Unexpected complications in term newborns

NQF Neonatal Measures NOT Endorsed

- Birth dose of Hep B vaccine and HBIG
- First NICU temp < 36 degrees C
- First temp measured within 1 hour of NICU admission
- Hep B vaccination prior to hospital discharge
- Late sepsis or meningitis in VLBW infants
- Neonatal blood stream infection rate
- Newborn hearing screening
- Health care associated bloodstream infections
- Surfactant within 2 hours of birth
- Under 1500 gm not delivered at appropriate level of care

Pay for Performance: Evidence is mixed

- 2017 systematic review: low-strength evidence for for improvement in short-term processes of care in ambulatory settings, some evidence for reducing readmissions, no evidence on improving health outcomes
  (Mendelson et al, Annals of Internal Medicine, 2017)
- 2019 Cochrane review: low certainty of evidence for impact on outcomes, quality, equity, or resource use; effects on patient outcomes in hospitals “at most small”
  (Mathes et al, Cochrane Database, 2019)

Pay for Performance: Unintended Consequences

- Hospital Readmissions Reduction Program (HRRP)
  - Gaming: triage in ER, observation rather than admission
  - Practice: delaying readmission beyond 30 days after discharge
  - Financial penalties mostly impacted safety-net hospitals as metrics not adjusted for socioeconomic factors
  - Potential association with increase in mortality in heart failure

Gupta A et al, “The Hospital Readmissions Reduction Program – Learning from Failure of a Healthcare Policy”, European Journal of Heart Failure, 2018

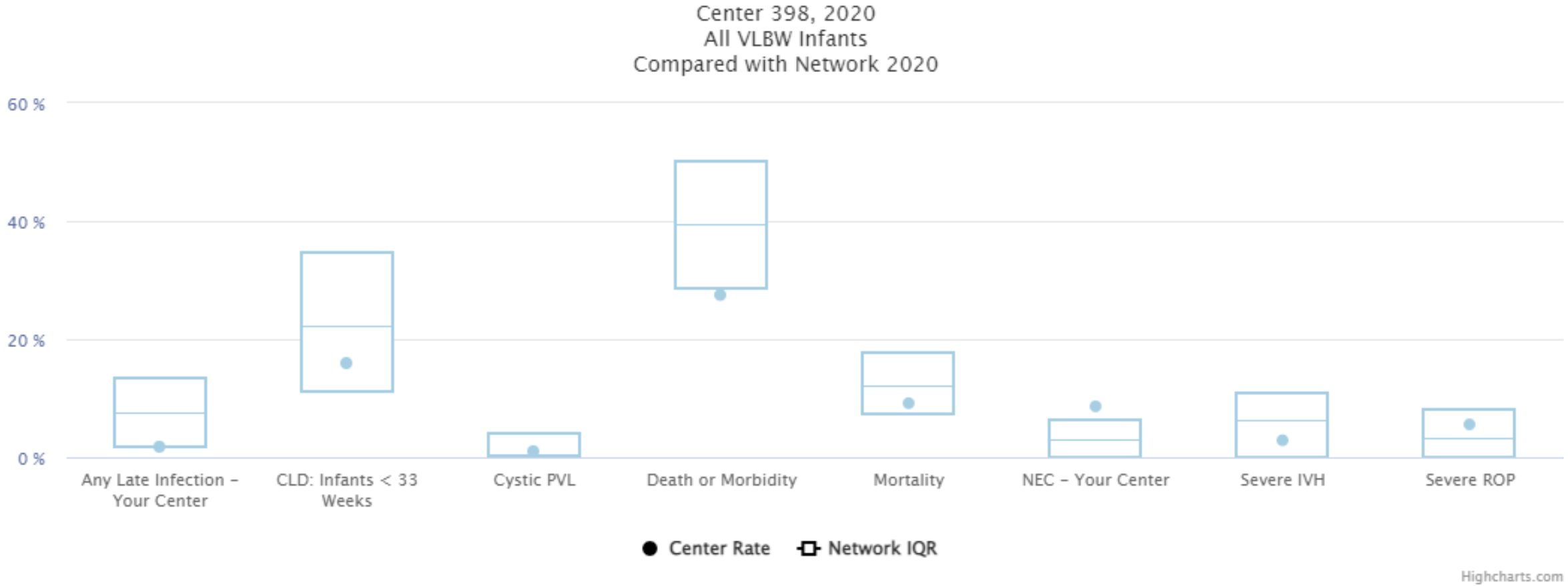
MEASURES FOR BENCHMARKING

Measures for Benchmarking?

- Comparative measures to drive local improvement
- VON, Mednax, CHNC
- INTERNAL accountability vs. external
- Somewhat in-between measures for improvement and measures for accountability
- Need clear, precise definitions
- If used for internal accountability, may not need as thorough evidence base and risk adjustment

Measures for Benchmarking: Potential Problems

- Small denominator
- Different populations
- Imprecise definitions (despite best efforts)
- Impact of practice differences


Center 398, 2020
All VLBW Infants
Compared with Network 2020
60 %
40 %
20 %
0 %
Any Late Infection – Your Center
CLD: Infants < 33 Weeks
Cystic PVL
Death or Morbidity
Mortality
NEC – Your Center
Severe IVH
Severe ROP
Center Rate
Network IQR
Highcharts.com

SONPM QUALITY MEASURES PROJECT

SONPM Quality Measures Project Goal

- Context: increasing awareness that accountability (and pay-for-performance) would impact neonatology
- Initially: review universe of quality measures in neonatal care to identify those most appropriate for accountability
- Final: systematically reviewing published measures, with goal of identifying measures that are ok for external quality assessment

Quick poll!

- Pollev.com/midcan
- Or text MIDCAN to 37607

Have quality measures related to NEWBORN CARE been formally linked to payment in your NICU/hospital?

Other examples: American College of Physicians

The NEW ENGLAND JOURNAL of MEDICINE

Perspective

Time Out — Charting a Path for Improving Performance Measurement

Catherine H. MacLean, M.D., Ph.D., Eve A. Kerr, M.D., M.P.H., and Amir Qaseem, M.D., Ph.D., M.H.A.

MacLean et al, NEJM, 2018

Other examples: American College of Physicians

- Reviewed 86 measures from Medicare Merit-Based Incentive Payment System (MIPS) relevant to ambulatory general internal medicine
- Reviewed on 5 domains: importance, appropriate care, evidence base, measure specifications, feasibility
- 37% valid, 35% not valid, 24% uncertain validity

MacLean et al, NEJM, 2018

Other examples: American College of Physicians

Ratings for a Sample of Measures.*

Rating	NQF-Endorsed	Steward	Measure	Importance	Appropriateness	Clinical Evidence	Specifications	Feasibility	Rationale
Valid	Yes	NCQA	Avoidance of Antibiotic Treatment in Adults with Acute Bronchitis Numerator: MIPS 116 (NQF 0058)	+	+	+	+	+	Based on appropriate evidence; specifications include appropriate exclusion criteria for patients with COPD or immunocompromised patients.

MacLean et al, NEJM, 2018

Other examples: SMFM

Society for Maternal-Fetal Medicine
High-risk pregnancy experts

SMFM Special Report

smfm.org

Quality measures in high-risk pregnancies: Executive Summary of a Cooperative Workshop of the Society for Maternal-Fetal Medicine, National Institute of Child Health and Human Development, and the American College of Obstetricians and Gynecologists

Brian K. Iriye, MD; Kimberly D. Gregory, MD; George R. Saade, MD; William A. Grobman, MD; Haywood L. Brown, MD

Iriye et al, AJOG, 2017

Other examples: SMFM

- Reviewed 34 measures in 5 clinical areas
- Used NQF domains: importance, scientific acceptability, usability, feasibility
- 15 out of 34 recommended for further consideration or development

Iriye et al, AJOG, 2017

Other examples: SMFM

TABLE 1
Principal workshop-proposed measures for participant consideration as a quality measure: rationale for inclusion or exclusion

Proposed measures	Importance	Scientific acceptability	Usability	Feasibility	Recommended for further consideration or development
Preterm birth prevention (outpatient)					
Universal TVCL screening	Yes Preterm birth occurs in 10% of pregnancies and is the leading cause of neonatal morbidity and death.	Yes Identification of patients with short cervix ≤2.0 cm would provide opportunities for treatment.	Yes Assists with medical decision-making.	Unclear Difficult to abstract data from outpatient electronic health records	No

Iriye et al, AJOG, 2017

SONPM Quality Measures Project Timeline

- A long time ago: SONPM chair (David Burchfield) proposed SONPM convenes stakeholders to review quality measures in neonatology
- A fairly long time ago: several informal discussions between SONPM leadership and quality organizations
- Several years ago: SONPM quality measures task force launched

SONPM Quality Measures Task Force

- Heather Kaplan
- Michael Prendergast
- Mark Hudak

SONPM Quality Measures Task Force Process

- Systematically reviewed neonatal measures promoted by regulatory and quality organizations with significant state or national-level impact
- High-level review of measures to identify those thought to be most important and commonly used
- Developed framework for evaluation based on NQF
- Recruited volunteer teams of topic leads, QI experts, and clinical experts that performed deep dives into high priority measures
- Consensus conference of volunteer teams and key partners to review initial measure evaluations and provide feedback
- Revised evaluation framework based on feedback
- Reviewing measure recommendations across teams to insure consistency

Categorization of Neonatal Quality Assessment Metrics

Type	Organization	# of Measures in Database	# of Measures in Prioritized List
National Agencies	National Quality Forum (NQF)	24	12
	Joint Commission	5	5
	National Healthcare Safety Network (NHSN)	5	5
	Agency for Healthcare Research and Quality (AHRQ)	11	5
	Health Resources and Services Administration (HRSA)	22	4
National Reporting Groups	Leapfrog	5	5
	U.S News and World Report	142	21
Payers	Centers for Medicare and Medicaid Services (CMS)	0	0
	Blue Cross Blue Shield	0	0
	State Medicaid agencies	0	0
Professional Organizations	Vermont Oxford Network (VON)	18	18
	California Perinatal Quality Care Collaborative (CPQCC)	20	18
	Children's Hospitals Neonatal Consortium (CHNC)	18	7
	Pediatric National Surgical Quality Improvement Program (NSQIP)	17	9
	Mednax	8	8
Others	Vizient	0	0
	Premier	0	0
Totals		295	117

Clinical Categories, Unified Measures, and Volunteer Teams

Measure Category	# Individual Measures	# Unified Measures	Team	Topic leads	QI Advisor	Clinical Advisor
Health Care Maintenance	6	5	1	Barry Weinberger Meredith Mowitz Nina Menda	Scott Lorch	DeWayne Pursley Mark Hudak
Safety	2	2				
Post-Discharge/FU/Readmissions	4	3				
Mortality	7	1	2	Dmitry Dukhovny Mike Posencheg Josh Petrikin	Jochen Profit	Theresa Grover John Zupancic
Miscellaneous	5	2				
Structural Measures	5	7				
Infection Measures	25	6	3	Jessica Davidson Alan Picarillo Maya Balakrishnan	Ravi Patel	Karen Puopolo Athena Kourtis
Neurology/Ophthalmology	11	6	4	Ulrike Mietzsch Dena Hubbard	Stephen Pearlman	Sonia Bonifacio
Nutrition and Growth	13	5	5	Jim Barry Krithika Lingappan Colby Day	Henry Lee	Brenda Poindexter Jay Goldsmith
Respiratory	11	5				
Surgery	12	12	6	Alexis Davis Rebecca Vartanian Amy Nathan	Robert Ursprung	Mike Padula Shawn Rangel
Transport	2	2				
Obstetric/Perinatal Measures	7	3		Not pursued		
Composite Measures	5	3		Not pursued		
TOTAL	115	62				

Team 7: Family-Centered Measures

- Goal: identify potential family-centered measures using same evaluation criteria
- Team: Danielle Ehret, Mark Hudak, Jeffrey Horbar, Jochen Profit, Carl Bose, Lelis Vernon

Key Partners at Consensus Conference

Key Partners	Status for Conference
AHA	Confirmed
AWHONN	Confirmed
BCBS	Confirmed
CDC	Confirmed
CHNC	Confirmed
HCA	Confirmed
Joint Commission	Confirmed
Kaiser	Confirmed
March of Dimes	Confirmed
MEDNAX	Confirmed
NANN	Confirmed
NQF	Confirmed
OPQC	Confirmed
COFN	On task force
CPQCC	On task force
Family Partners	On task force
VON	On task force

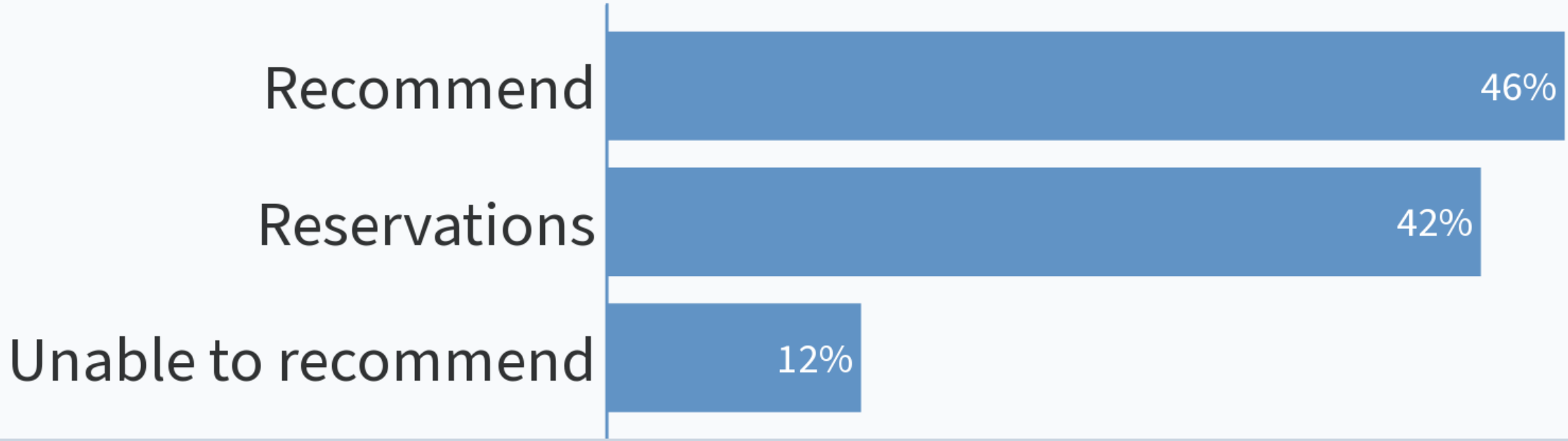
SONPM Evaluation Framework*

CATEGORY	DEFINITIONS	OPTIONS
Importance	Priority Performance gap	High Moderate Low
Scientific Value	Evidence base Structure-process-outcome link Modifiable	Adequate Limited or with face validity Absent without face validity
Measurement	Reliability Validity Feasibility Usability	Minimal issues Significant issues but addressable Significant issues hard to address

OVERALL RECOMMENDATION	
Highest ratings for all three categories:	**Recommend for further consideration**
All else:	**Reservations**
Lowest ratings in any category:	**Unable to recommend**

** Multiple earlier versions*

Based on what you know, how would you rate a measure of HAND HYGIENE RATES in your NICU for use for accountability?



Example of final measure evaluation

Measure		Hand Hygiene
Type		Process
Importance	**Rating**	**High**
	Notes	WHO, AAP priority; variation exists
Scientific Value	**Rating**	**Limited or with face validity**
	Notes	Modest evidence linking improvement to outcomes
Measurement	**Rating**	**Significant issues hard to address**
	Notes	Inconsistent definition; measurement can be time-consuming
Recommendation		**Unable to recommend**

Preliminary Final Results

Recommendation	Number of Measures
Recommend for further consideration	9
Reservations	23
Unable to recommend	32

Preliminary Final Results

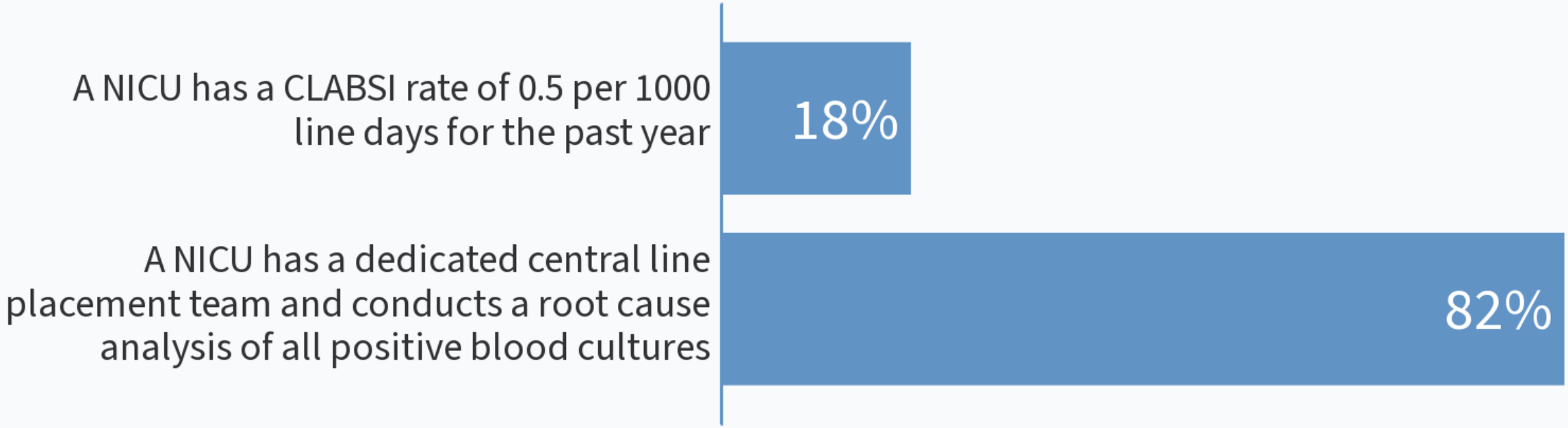
- Among 23 rated as “Reservations”
  - 14 were high importance, 9 were moderate
  - 12 had adequate validity, 11 had limited
  - 1 had minimal measurement issues, 22 had significant but addressable
- Among 32 rated as “Unable to recommend”
  - 13 were high importance, 7 moderate, and 12 low
  - 9 had adequate validity, 18 limited, and 5 absent
  - 4 had minimal measurement issues, 6 had significant but addressable, 22 had hard to address

WHERE TO FROM HERE?

Thoughts

- Current quality measures for external accountability in neonatology are not that great
- Recommended substantially lower percentage of measures than ACP and SMFM
- Measurement issues (including risk-adjustment and feasibility) are most common challenge
- Process measures are generally more appealing for accountability than outcome measures

Which do you think is a better indicator of quality of care?



New “Structural” Measures for accountability?

- Is performance on clinical measures more important than robustness of quality and safety processes?
- Could structural measures be developed that examine quality and safety infrastructure or processes within units?
- Some basis in literature: NICU volume, culture, staffing
- Potential link with SONPM NICU Verification Project and development of national standards

Take home points?

- You SHOULD be doing local QI, and using well-defined measures that are important to you.
- You SHOULD be part of a collaborative that allows you to benchmark key measures, and while a grain of salt is ok, at some point, you should trust those comparisons.
- While externally reported measures are important, we need to be careful about current ones.
- We really need to find better measures for accountability.

References

- Donabedian, Evaluating the Quality of Medical Care, Milibank Quarterly, 1966
- Ellsbury et al, A Multifaceted Approach to Improving Outcomes in the NICU, Pediatrics, 2016
- Gupta A et al, “The Hospital Readmissions Reduction Program – Learning from Failure of a Healthcare Policy”, European Journal of Heart Failure, 2018
- Iriye et al, Quality Measures in High-Risk Pregnancies, AJOG, 2017
- MacLean et al, Time Out – Charting a Path for Improving Performance Measurement, NEJM, 2018
- Solberg et al, Three Faces of Performance Measurement, Joint Commission Journal on Quality Improvement, 1997