


BMJ Open Quality Measuring practice preference variation for quality improvement: development of the Neonatology Survey of Interdisciplinary Groups in Healthcare Tool (NSIGHT)

Emily Whitesel ^{1,2}, Helen Healy^{1,2}, Wenyang Mao,¹ DeWayne M Pursley,^{1,2} John Zupancic,^{1,2} Munish Gupta^{1,2}

To cite: Whitesel E, Healy H, Mao W, *et al*. Measuring practice preference variation for quality improvement: development of the Neonatology Survey of Interdisciplinary Groups in Healthcare Tool (NSIGHT). *BMJ Open Quality* 2023;**12**:e002422. doi:10.1136/bmjog-2023-002422

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjog-2023-002422>).

Received 19 May 2023
Accepted 3 December 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Neonatology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

²Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA

Correspondence to

Dr Emily Whitesel;
ewhitesel@bidmc.harvard.edu

ABSTRACT

Background Understanding behavioural psychology and the human side of change are guiding principles for quality improvement (QI). Tools to measure these to guide improvement efforts are lacking.

Methods We created a clinical vignette-based survey to measure provider preferences for respiratory care in the neonatal intensive care unit. Fourteen vignettes were included, each vignette offering two reasonable practice choices. Responses were based on a 5-point Likert scale, ranging from neutral to strong preference for either choice. The survey was completed by physicians, nurses, advanced practice providers and respiratory therapists in 2017 and again in 2019. Net preference was measured as the median value of responses, and agreement was measured as the SD of responses. Net preference and agreement were assessed for all responses, by discipline, and by year.

Results Response rates were 51% of all staff in 2017 and 57% in 2019. Vignettes asking about non-invasive respiratory support showed more defined net preferences and higher agreement between years, coinciding with QI efforts and guideline implementation in this area during the interval time. Results on other areas of practice were consistent between years. Discipline comparisons showed nurses and physicians agreed the least often. Six response patterns were identified, ranging from net preference and high agreement to no net preference and low agreement.

Conclusion We propose this survey, called the Neonatology Survey of Interdisciplinary Groups in Healthcare Tool, is a novel method for measurement of hospital unit psychology and culture. Demonstrated improvement where QI efforts were focused and consistency in results in other areas support the validity of this tool. Measuring the human side of change may impact QI efforts.

INTRODUCTION

Modern quality improvement (QI) methods in healthcare are built on the System of Profound Knowledge, created by Deming.¹ Deming described four domains within the System of Profound Knowledge: appreciation

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Quality improvement (QI) provides a structured framework for examining and improving our delivery of healthcare, in which psychology is recognised as a key driver of system performance.

WHAT THIS STUDY ADDS

⇒ We currently lack an effective and simple means to measure and describe human factors behind improvement. Therefore, we developed a novel clinical vignette-based tool that can be used to measure individual practice preferences, providing levels of agreement between individual providers as well as between disciplines around specific clinical topics.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ We propose that this tool be used as a QI metric to help identify barriers to change, inform improvement efforts about unit culture and facilitate the development of shared mental models of practice.

for a system, understanding variation, theory of knowledge, and psychology.^{1 2} Numerous commonly used tools can help define and measure the first three domains: driver diagrams, process maps, fishbone diagrams and Pareto charts can help examine systems; time-series data analysis and statistical process control charts help us understand variation; and plan-do-study-act cycles can test our predictions and shape our theory of knowledge.^{1 2} However, similar tools are lacking for the fourth domain. Although psychology is a key driver of system performance, we lack an effective and simple means to measure and describe human factors behind improvement.³

Psychology of change is defined by the Institute for Healthcare Improvement as ‘the science and art of human behavior as

it relates to transformation,' and is known to impact the adoption and sustainability of an improvement initiative and the function of the improvement team.⁴ Although there are numerous success stories in the *QI* literature, a substantial gap remains between what we know and how we practice. Evidence-based practices are often slow to be implemented, or fail to be adopted altogether.^{4–6} Even in successfully implemented projects, barriers may prevent innovations from becoming sustainable, when a 'critical mass' of individuals accepts and adopts an innovation as routine practice.^{5–7} Standardisation of practice and reducing variation among providers, even in the absence of defined best practices, can be an important driver of quality.⁸ Given the complex and multidisciplinary nature of healthcare, understanding individual provider practice preferences, and how combined preferences inform culture within a clinical unit, may help us target psychology as an improvement domain and address quality gaps.

Vignettes are brief, written cases based on realistic clinical situations that can be used to identify practice preferences. They have shown effectiveness for exploring provider decision-making and identifying variations in practice, as compared with standardised patient scenarios and chart abstractions.^{9–11} Vignettes have recently been used to study variation in providers' preferences in neonatology.¹²

In the neonatal intensive care unit (NICU), very preterm infants such as those born at less than 29 weeks gestational age frequently need prolonged mechanical ventilation and are at risk for developing bronchopulmonary dysplasia (BPD), the most common long-term morbidity of prematurity.^{13 14} The management of respiratory support in the NICU contributes substantially to outcomes, and is therefore a common focus of local and collaborative *QI* efforts targeting BPD. However, improvement efforts in this area have shown limited success.¹⁵ Several factors may drive this limited improvement. Despite a large body of research, equipoise between multiple approaches still exists in many specific aspects of neonatal respiratory care, making standardisation of practice more difficult.¹⁶ In addition, neonatal respiratory care is particularly multidisciplinary, involving at a minimum neonatal nurses, respiratory therapists, advanced practice providers, and neonatologists; achieving team consensus and a unified approach to care may be more challenging than in other areas.¹⁵

With the goal of informing the psychology of change within *QI* efforts, we sought to create a novel metric based on clinical vignettes that could be used to measure individual practice preferences and unit culture within NICU regarding the respiratory care of preterm infants. Our team developed a vignette-based survey called the Neonatology Survey of Interdisciplinary Groups in Healthcare Tool (NSIGHT) to elicit preferences regarding specific respiratory care scenarios from the multidisciplinary team in our unit. We hypothesised that preference agreement and variation between individual providers and

between disciplines could be measured, and could be used as a *QI* metric to help identify barriers to change and to help inform improvement efforts by informing unit culture and facilitating the development of shared mental models of practice.

METHODS

Survey instrument development

We developed a 14-item clinical vignette-based survey targeting themes identified by our NICU's respiratory care *QI* team. The vignettes were piloted by a multidisciplinary group of clinicians and unit leaders and iteratively revised. Survey items are shown in online supplemental table 1. Each vignette briefly described a common clinical scenario involving a preterm infant requiring respiratory support, and two options for further treatment. The scenarios were designed such that both options for treatment would be considered appropriate by current standards of practice.

Clinical setting and participants

Our hospital is a large tertiary academic centre with a level III NICU.¹⁷ We care for approximately 5200 births and 950 NICU admissions per year. Approximately 90% of NICU admissions are inborn, with the remainder transferred to our NICU after birth.

All active NICU clinical staff were surveyed, including attending neonatologists (MDs), advanced practice providers (APPs), clinical nurses (RNs) and respiratory therapists (RTs). Surveys were sent via email using a REDCap survey tool.^{18 19} All responses were anonymous, identified only by discipline. Staff were asked to indicate their personal practice preference for a clinical scenario between two choices (options A and B). Preferences were indicated using a modified 5-point Likert scale: strongly favour A, favour A, neutral, favour B, and strongly favour B. Staff could also select 'unable to answer' if they felt they did not have sufficient knowledge or information to answer the question. At least five responses per discipline were needed to be included in data analysis.

The initial survey was conducted February to March 2017 and the same survey was repeated August to October 2019. Between the two survey periods, local improvement efforts focused on increasing the use of non-invasive positive pressure support, including the development of a unit guideline for use of continuous-positive-airway-pressure (CPAP), and staff education about BPD and ventilator-induced lung injury. Of note, when the survey was conducted at our hospital in 2019, a multisite cohort of additional NICUs simultaneously conducted the NSIGHT, using the same 14 vignettes.

Our institutional review board determined that this project was *QI* and did not constitute human subject research.

Statistical analyses

Net preference

A net preference for a particular vignette was defined as a clearly favoured practice among respondents. To measure net preference, survey responses were assigned a numerical value from 1 to 5 (strongly favour A=1; favour A=2; neutral=3; favour B=4, strongly favour B=5). The net preference was calculated as the median value of all responses. The Wilcoxon Signed Rank Test was applied to determine whether there was a significant preference for choice A or B. A p value less than 0.05 indicated a net preference that significantly differed from 3.0, or neutral. A p value at or above 0.05 indicated that there was no clearly favoured practice among respondents. Net preference was determined for each vignette for the unit as a whole and for individual disciplines.

Agreement

Standard deviation (SD) around the mean was used to measure agreement among respondents.²⁰ A smaller SD reflected a narrow distribution of responses and thus higher agreement, while a larger SD reflected a wide distribution of responses and thus lower agreement. SDs from the coinciding multicentre cohort were used to generate a scale for benchmarking. All SDs from all vignettes from all centres were ordered and divided into terciles. High agreement was defined as SD within the lowest tercile (33.3 percentile and below), low agreement was defined as SD within the highest tercile (66.7 percentile and above), and medium agreement was defined as SD within the middle tercile.

Between year

Change in agreement and net preference for each clinical vignette was examined from 2017 to 2019. Levene's test was used to assess for a significant change in agreement (SD) among all staff between the two periods.^{21 22}

Between discipline

Net preference was compared between the four discipline groups and between each pairing of two disciplines using the pairwise Kruskal-Wallis test. A p value less than 0.05 indicated a significant difference of at least one discipline's preference from the rest. In addition, to allow for further granularity, degree of difference of preference between discipline pairs was measured using additional typically reported p value thresholds, ranging from p value greater than 0.1 indicating most similar preference to a p value less than 0.0001 indicating the most different preference.

Examined patterns of response

Based on the above analyses for preference and agreement, we categorised responses for each vignette into six possible response patterns: (1) high agreement and net preference; (2) high agreement and no net preference; (3) medium agreement and net preference; (4) medium agreement and no net preference; (5) low agreement

and net preference; and (6) low agreement and no net preference.

Data analysis was performed using SAS software, V.9.4 (SAS Institute).

RESULTS

Response rates are shown in online supplemental table 2. In 2017, 103 of 202 staff (51%) completed the survey. In 2019, 130 of 230 staff (57%) completed the survey. The response rate among neonatologists, advanced practice providers and respiratory therapists was between 92% and 100% in both years. The response rate among nurses was 40% in 2017 and 47% in 2019. As the largest group in the NICU, nurses comprised 62% of all responses in 2017 and 68% in 2019.

Staff-wide preferences and agreement

Figure 1 shows multiple analyses on one graph, including net preference and agreement level for each vignette, and between year comparisons.

Net preference

In 2017, 5 of the 14 vignettes (1, 4, 5, 6 and 7) showed no significant net preference among all respondents. In 2019, 4 of the 14 vignettes (5, 6, 7 and 10) showed no practice preference. The remainder showed a significant unit-wide practice preference for either option A or option B.

Agreement level

In 2017, 5 vignettes (1, 3, 7, 11 and 14) had low agreement levels and 2 (2 and 13) had high agreement. In 2019, 2 (7 and 11) had low agreement and 8 (1, 2, 4, 5, 6, 8, 9 and 13) had high agreement.

Between years

Between 2017 and 2019, 2 vignettes (1 and 4) showed staff developed a practice preference, with both also seeing an increase in agreement level. One vignette (10) showed staff had a preference in 2017 but did not have a significant preference in 2019, with agreement level remaining medium. Eight vignettes showed staff maintained the same preference from 2017 to 2019, and 3 showed staff maintained a lack of significant preference. Nine vignettes showed an increase in agreement level, with 3 of these (1, 8, 14) being a significant change. None of the scenarios showed a decrease in agreement level.

Practice preference by discipline

Figure 2 shows staff practice preferences across all disciplines for three example vignettes. In vignette 1, while a statistical difference in preferences among the disciplines was seen in both 2017 and 2019, a shift is visually evident towards the preference for choice B among all disciplines. Vignette 2 illustrates a clear practice preference for choice A, which is consistent across all the disciplines and between years. Vignette 7 demonstrates notable differences in preferences among the disciplines,

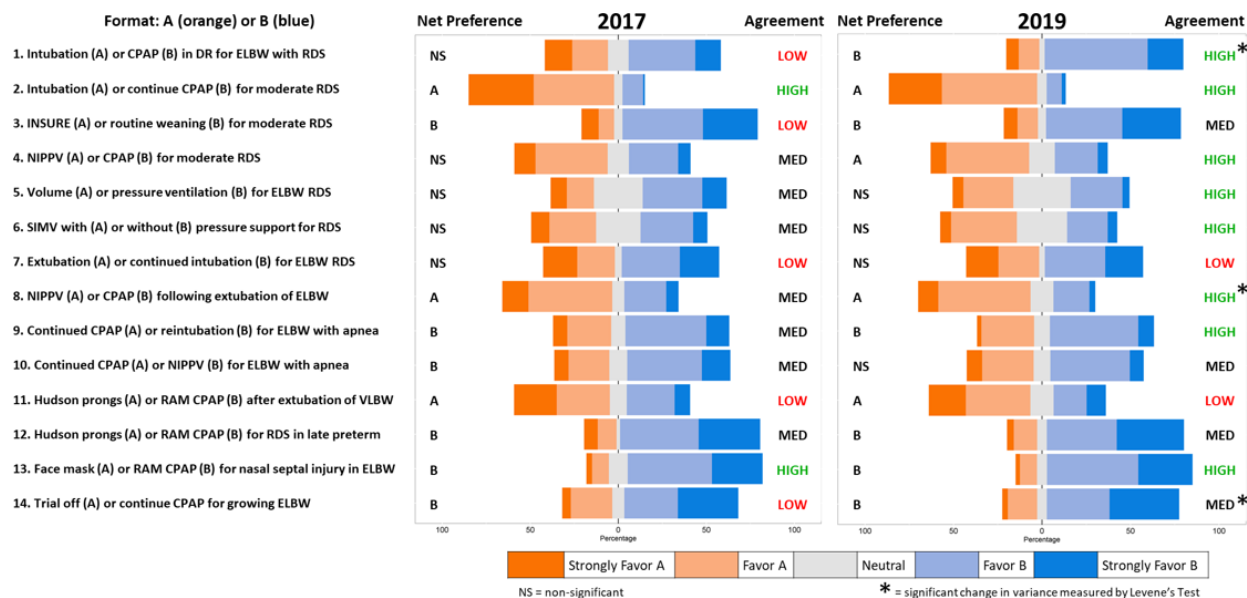


Figure 1 NICU staff preference and agreement level in 2017 and 2019. Preference distribution of all staff, for all 14 vignettes, in 2017 and 2019. The net preference is displayed on the left hand column of each figure; agreement level, as measured by SD tercile score, is displayed on the right hand column. An asterisk (*) indicates a significant change in agreement between 2017 and 2019, as measured by the Levene's test. CPAP, continuous positive airway pressure; NICU, neonatal intensive care unit; ELBW, extremely low birth weight; RDS, respiratory distress syndrome; INSURE, Intubation-SURfactant-Extubation; NIPPV, nasal intermittent positive pressure ventilation; SIMV, synchronized intermittent mandatory ventilation; VLBW, very low birth weight.

with a strong preference towards choice A for the MDs and APPs and choice B for RTs and RNs, and no change or improvement between the 2 years. All of these results are consistent with the overall patterns for these vignettes

seen in [figure 1](#). Online supplemental figure 1 shows staff preferences by discipline for the remainder of the vignettes.

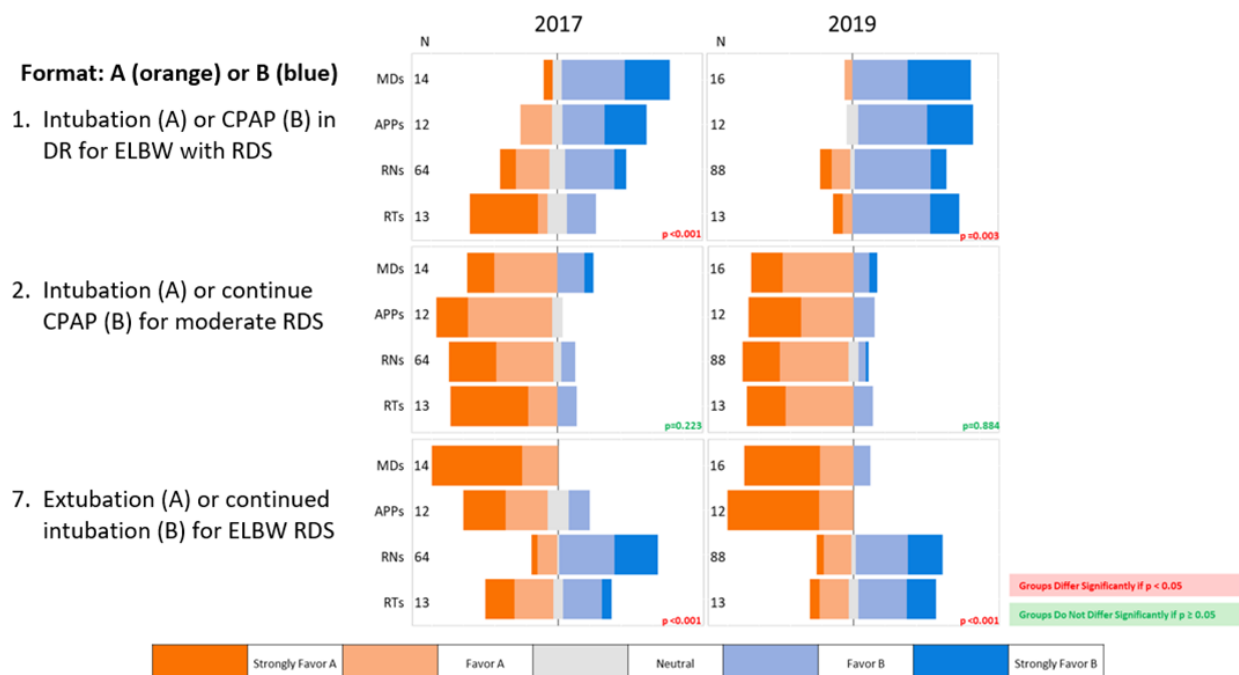


Figure 2 NICU staff preferences for non-invasive or invasive respiratory support, by discipline in 2017 and 2019. Staff practice preferences by discipline, with vignettes 1, 2 and 7 highlighted as examples. These vignettes explore the decision of invasive or non-invasive respiratory support. Each discipline's preference is shown by the colours orange or blue, in 2017 and 2019. A p value of < 0.05 , as measured by the Kruskal-Wallis test, signifies that at least one discipline differed significantly. CPAP, continuous positive airway pressure; DR, delivery room; ELBW, extremely low birth weight; RDS, respiratory distress syndrome.

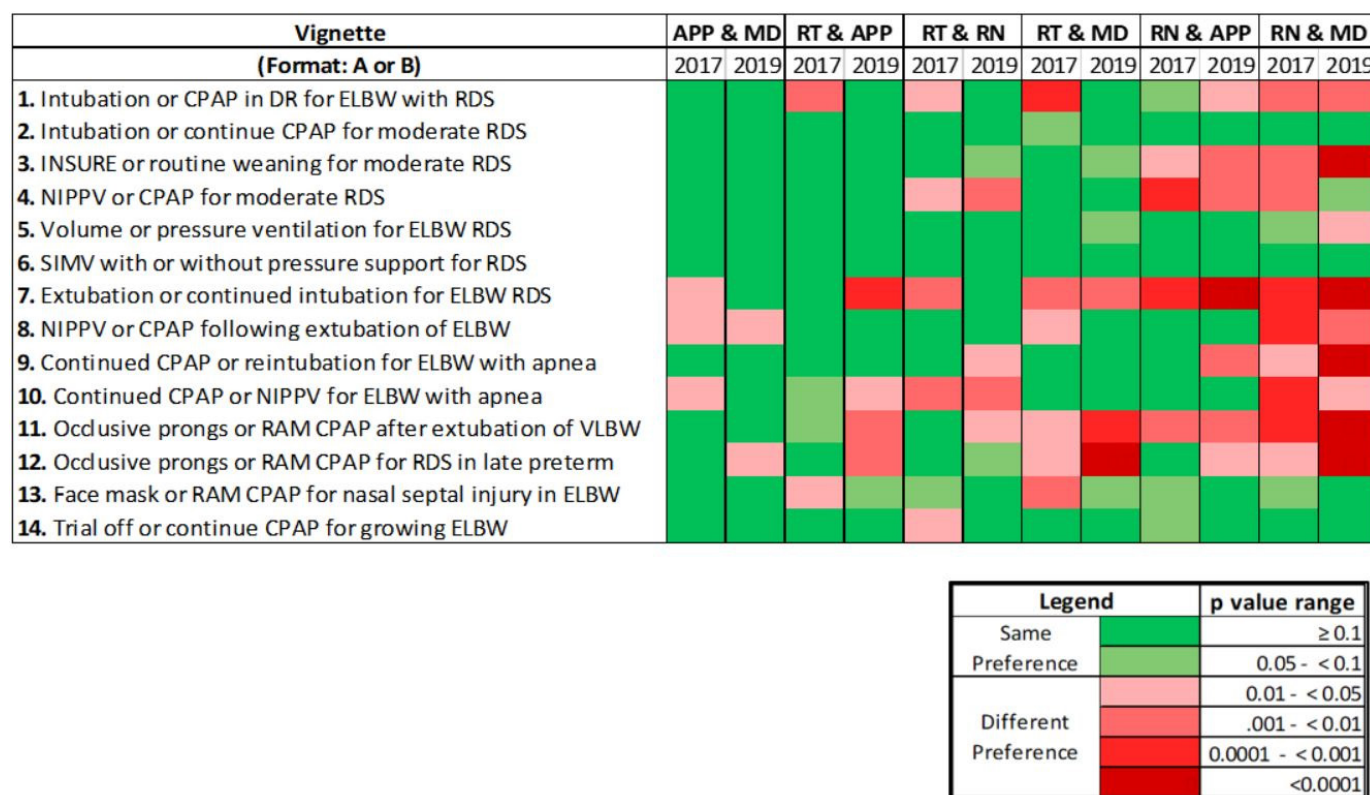


Figure 3 Preference comparison between discipline pairs in 2017 and 2019. The comparisons of each discipline pair in both 2017 and 2019, with a p value threshold used to determine a difference of preference indicated by shade of red or green. P values were calculated using the pairwise Kruskal-Wallis test. CPAP, continuous positive airway pressure; DR, delivery room; ELBW, extremely low birth weight; RDS, respiratory distress syndrome; NIPPV, nasal intermittent positive pressure ventilation; INSURE, Intubation-SURfactant-Extubation; VLBW, very low birth weight.

Figure 3 shows the comparisons of practice preference by discipline pairs in both 2017 and 2019. Degree of difference of preference is indicated by colour, based on p value thresholds. Pairs with the highest degree of similar preference are those with a p value of greater than 0.1, and are indicated with dark green. Pairs with the highest difference in preference are those with a p value of less than 0.0001, and are indicated in dark red. Overall, APPs and MDs had the same preference most often, followed by the pairing of RTs and APPs. In contrast, RNs and MDs had the most differences in preference.

Patterns of response

Based on the analyses described above, six patterns of response were possible, ranging from high agreement with a net preference (pattern 1) to low agreement without a net preference (pattern 6). Patterns of responses are named and defined in table 1. The response patterns demonstrated in our NICU in 2017 and 2019 are categorised in table 2. In 2017, two vignettes were patterns 1 or 2, seven were pattern 3 or 4, and five were pattern 5 or 6. In 2019, eight vignettes were pattern 1 or 2, four were pattern 3 or 4, and two were pattern 5 or 6. Nine vignettes showed a change in pattern from 2017 to 2019.

DISCUSSION

Unit culture undoubtedly impacts quality, including in neonatal intensive care.^{23 24} Current QI methods have limited tools for formally assessing the human psychology that underlies culture. We propose that the NSIGHT tool described above can help measure individual provider practice preferences and inform improvement efforts by targeting the ‘human side of ‘change’. Three findings give us confidence regarding the validity of NSIGHT.

First, the results in our NICU were generally consistent in two surveys completed 2 years apart. The visual patterns displayed in figure 1 for 2017 and 2019 are remarkably similar, and practice preference remained the same for 11 out of 14 clinical scenarios. This consistency among responses for over 100 NICU staff members suggests the survey tool had high reliability in eliciting practice preferences.

Second, areas in which the most change was seen between the two surveys were areas of focus for QI efforts in the NICU during the interval time period. The vignette with the largest differences between 2017 and 2019 was vignette 1, which moved from no preference to a net preference for option B and low agreement to high agreement (figure 1). This vignette asked about preferred mode of initial respiratory support for a preterm infant with respiratory distress syndrome (RDS) in the delivery room. In

Table 1 NSIGHT response patterns

Pattern	Preference	Agreement	Comments
Shared preference	Yes	High	This pattern suggests a routine practice that is part of unit culture.
Shared neutrality	No	High	This pattern may arise if there are insufficient data to support one practice; however, there may still be benefit to developing consensus on a practice choice.
Partially shared preference	Yes	Medium	This pattern suggests unit consensus with some degree of disagreement, and could reflect early stages of adoption of a new practice and/or a minority group with a different preference.
Indeterminate neutrality	No	Medium	This pattern suggests no unit consensus with some degree of disagreement.
Unshared preference	Yes	Low	This pattern suggests unit consensus but a substantial portion of staff not in agreement; apparent consensus could be driven by one group, and not reflect a shared mental model throughout the unit.
Disagreement	No	Low	This pattern suggests no unit consensus with widely varying individual preferences; this likely will impede buy-in and joint decision-making.

NSIGHT, Neonatology Survey of Interdisciplinary Groups in Healthcare Tool.

2017, the staff was fairly evenly distributed across preference for invasive support with intubation and mechanical ventilation and non-invasive support with CPAP. By 2019, there was significant preference and agreement for non-invasive support. This likely reflects substantial improvement efforts that occurred over that period to promote non-invasive respiratory support as an initial mode of support, including release of an updated guideline, extensive staff education and improvements in devices

and interfaces used for CPAP delivery. The improvement efforts also emphasised continuing CPAP for a longer time in growing preterm infants; this may be reflected in the higher agreement seen on vignette 14 in 2019.

Third, the findings overall and the differences measured between disciplines are consistent with the authors' experiences leading the clinical team in our NICU. While we were pleased to see increased agreement on use of CPAP for preterm infants with RDS (vignette 1), we know that

Table 2 NSIGHT response patterns of agreement and practice preference in the NICU

2017				2019		
Vignette	Agreement	Preference	Pattern	Agreement	Preference	Pattern
1	Low	No	6	High	Yes	1
2	High	Yes	1	High	Yes	1
3	Low	Yes	5	Medium	Yes	3
4	Medium	No	4	High	Yes	1
5	Medium	No	4	High	No	2
6	Medium	No	4	High	No	2
7	Low	No	6	Low	No	6
8	Medium	Yes	3	High	Yes	1
9	Medium	Yes	3	High	Yes	1
10	Medium	Yes	3	Medium	No	4
11	Low	Yes	5	Low	Yes	5
12	Medium	Yes	3	Medium	Yes	3
13	High	Yes	1	High	Yes	1
14	Low	Yes	5	Medium	Yes	3

Bolded vignettes are those that demonstrated a change in pattern.

NICU, neonatal intensive care unit; NSIGHT, Neonatology Survey of Interdisciplinary Groups in Healthcare Tool.

significant variation in preferences still exists in other important areas. Vignettes 5 and 6 asked about modes of mechanical ventilation, and showed evenly distributed preferences among staff between volume and pressure ventilation as well as use or non-use of pressure support (figure 1). This matches growing variation in ventilator modes being used in our NICU, and indicates a need for developing consensus. Similarly, vignette 11 asked about preference between two interfaces for CPAP for smaller preterm infants, and showed low agreement between use of the more occlusive but more cumbersome Hudson prongs or the less occlusive but more convenient RAM cannula (figure 1); this also matches regular bedside discussions in the NICU.

As a tool, NSIGHT can not only evaluate the impact of improvement efforts, it can also drive the approach to future interventions. The response patterns may be particularly useful indicators to guide improvement efforts. In general, clinical leaders strive for standardisation of practice to drive quality.^{8 25} When standardisation is the goal, Pattern 1 (shared preference) is likely the most favourable, suggesting widespread consensus among providers with a shared mental model. In 2019, we saw this pattern in several areas, including non-invasive support in the delivery room, intubation for moderate RDS, NIPPV over CPAP for moderate RDS, NIPPV over CPAP following extubation, and reintubation for a small preterm infant with apnoea (table 2). These results are informing our efforts around the use of NIPPV and CPAP, two approaches to non-invasive respiratory support.

Low agreement and lack of consensus would likely be a barrier to change, with patterns 5 (unshared preference) and 6 (disagreement) being unfavourable. In our unit, this was most evident for vignette 7, which showed pattern 6 in both 2017 and 2019 (table 2). This vignette asks about early extubation versus continued intubation for extremely preterm infants; strong evidence of best practice for this population is lacking, and we have not focused local improvement work in this area. Not surprisingly, this vignette showed evenly distributed preferences across all staff with low agreement (figure 1). When examined by discipline, however, this vignette showed remarkable and consistent discipline-specific differences, with MDs and APPs preferring early extubation and RNs and RTs preferring continued intubation (figure 2). This granularity provides important insight into the drivers of low agreement; the differences between disciplines are an important reflection of unit culture in this area, and without efforts to achieve buy-in, the lack of consensus will likely impair efforts to standardise practice.

Other response patterns require cautious interpretation. Pattern 2 (shared neutrality) occurs when responses generally agree on being neutral; this may be appropriate when there is true equipoise in the evidence, and may then reflect an opportunity for developing consensus and standardisation. Pattern 4 (indeterminate neutrality) reflects an overall lack of unit preference, but disagreement among staff with

some preferences for each option. This may require targeted discussion among those staff with identified preferences in order to reach consensus and achieve buy-in.

Overall discipline comparisons can also be informative. The 'heat map' in figure 3 readily demonstrates that the most disagreement between disciplines in our NICU occurs between RNs and MDs. The perspectives of these groups in neonatal intensive care are undeniably different, and it is likely other NICUs would see similar patterns. While it would not be difficult to postulate drivers of these differences between RNs and MDs, the dramatic demonstration of this divide by NSIGHT reinforces the importance of addressing the human side of change in clinical leadership and QI, and the need to work collaboratively across all disciplines.

Important limitations exist for the use of a vignette-based tool to measure practice preferences. First, a written scenario cannot account for all the potential clinical factors that drive practice decisions, and the formal validity and reliability of this survey tool has not been tested. However, the concordance of results with our knowledge of the unit support the validity of the questions, and the consistency of responses suggests adequate reliability. Second, the utility of any survey depends on response rate. We were pleased that virtually all of our MDs, NPs and RTs completed both surveys, but recognise that less than half of RNs did. While the RN response was still robust, future uses of NSIGHT in our unit will include specific efforts to encourage RN participation. Third, our vignette-based approach requires selection of two practice choices for each scenario that are both reasonable based on current evidence. There may be some subjectivity to this assessment, and evidence may change over time; vignettes should be created with input from a multidisciplinary team, and reviewed regularly to ensure they remain topical and accurate.

CONCLUSIONS

NSIGHT is a vignette-based survey tool that measures provider practice preferences, and can help to quantify the human factors behind change. Response patterns allow for interpretation of net preference and agreement levels, and can assess culture around particular clinical care topics. Used in concert with other traditional QI methods, NSIGHT may help target potential barriers to change and inform interventions, and thereby impact the success and sustainment of improvement efforts.

Acknowledgements We acknowledge and thank Kanekal Suresh Gautham for the initial stimulus to explore this aspect of quality, and for continued intellectual support and collaboration.

Contributors All authors of this manuscript are responsible for the reported research, and all authors have participated in study design, data collection,

interpretation of results, and drafting and revisions of the manuscript. All authors have approved the final manuscript as provided. Acting guarantor E.W.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants but an Institutional Board exempted this study. Review completed by: Committee on Clinical Investigations, Beth Israel Deaconess Medical Center. Protocol #: 2021D000721

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Emily Whitesel <http://orcid.org/0009-0001-9119-1329>

REFERENCES

- Deming WE. The new economics for industry, government, education. In: *The New Economics for Industry, Government, Education*. 3rd ed. Cambridge, Massachusetts: The Massachusetts Institute of Technology Press, 30 October 2018.
- Langley GJ, Moen RD, Nolan KM, et al. n.d. The improvement guide. *San Francisco Jossey-Bass*
- Baker GR, King H, MacDonald JL, et al. Using organizational assessment surveys for improvement in neonatal intensive care. *Pediatrics* 2003;111(4 Pt 2):e419–25.
- Hilton K, Anderson A. IHI psychology of change framework to advance and sustain improvement. In: *IHI White Paper*. Boston, Massachusetts: Institute for Healthcare Improvement, 2018.
- Institute of Medicine (U.S.). *Committee on Quality of Health Care in America. Crossing the quality chasm: a new health system for the 21st century*. Washington, D.C: National Academy Press, 2001.
- Balas EA, Boren SA. Managing clinical knowledge for health care improvement. *Yearb Med Inform* 2000;65–70.
- Rogers EM. *Diffusion of Innovations* 5th ed. Simon and Schuster, 2003.
- Balakrishnan M, Raghavan A, Suresh GK. Eliminating undesirable variation in neonatal practice: balancing standardization and Customization. *Clin Perinatol* 2017;44:529–40.
- Veloski J, Tai S, Evans AS, et al. Clinical vignette-based surveys: a tool for assessing physician practice variation. *Am J Med Qual* 2005;20:151–7.
- Peabody JW, Luck J, Glassman P, et al. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA* 2000;283:1715–22.
- Peabody JW, Luck J, Glassman P, et al. Measuring the quality of physician practice by using clinical vignettes: a prospective validation study. *Ann Intern Med* 2004;141:771–80.
- Payton KSE, Wirtschaffer D, Bennett MV, et al. Vignettes identify variation in antibiotic use for suspected early onset sepsis. *Hosp Pediatr* 2021;11:770–4.
- Horbar JD, Edwards EM, Greenberg LT, et al. Variation in performance of neonatal intensive care units in the United States. *JAMA Pediatr* 2017;171:e164396.
- Stoll BJ, Hansen NI, Bell EF, et al. Trends in care practices, morbidity, and mortality of extremely Preterm neonates, 1993–2012. *JAMA* 2015;314:1039–51.
- Healy H, Croonen LEE, Onland W, et al. A systematic review of reports of quality improvement for Bronchopulmonary dysplasia. *Semin Fetal Neonatal Med* 2021;26.
- Pfister RH, Goldsmith JP. Quality improvement in respiratory care: decreasing Bronchopulmonary dysplasia. *Clin Perinatol* 2010;37:273–93.
- American Academy of Pediatrics Committee on Fetus And Newborn. Levels of neonatal care. *Pediatrics* 2012;130:587–97.
- Harris PA, Taylor R, Minor BL, et al. The Redcap consortium: building an international community of software platform partners. *J Biomed Inform* 2019;95.
- Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (Redcap)--A Metadata-driven methodology and Workflow process for providing Translational research Informatics support. *J Biomed Inform* 2009;42:377–81.
- Norman G. "Likert scales, levels of measurement and the "laws" of Statistics". *Adv Health Sci Educ Theory Pract* 2010;15:625–32.
- Derrick B, Ruck A, Toher D, et al. Tests for equality of Variances between two samples which contain both paired observations and independent observations. *Journal of Applied Quantitative Methods* 2018;13.
- Gastwirth JL, Gel YR, Miao W. The impact of Levene's test of equality of Variances on statistical theory and practice. *Statist Sci* 2009;24:343–60.
- Kaempf JW, Wang L, Dunn M. Using a composite morbidity score and cultural survey to explore characteristics of high proficiency neonatal intensive care units. *Arch Dis Child Fetal Neonatal Ed* 2019;104:F13–7.
- Profit J, Sharek PJ, Cui X, et al. The correlation between neonatal intensive care unit safety culture and quality of care. *J Patient Saf* 2020;16:e310–6.
- Wennberg JE. Time to tackle unwarranted variations in practice. *BMJ* 2011;342.