Published in final edited form as:

Arch Pediatr Adolesc Med. 2012 February ; 166(2): 191–194. doi:10.1001/archpediatrics.2011.810.

Perils and Opportunities of Comparative Performance Measurement

Jochen Profit, MD, MPH^{a,b,c} and LeChauncy D. Woodard, MD, MPH^{b,c}

^a Department of Pediatrics, Baylor College of Medicine; Texas Children's Hospital; Houston, TX, USA

^b Section of Health Services Research, Department of Medicine, Baylor College of Medicine, Houston, TX, USA

^c Houston Veterans Affairs (VA) Health Services Research and Development Center of Excellence, Health Policy and Quality Program, Michael E. DeBakey VA Medical Center; Houston, TX, USA

Guillén and colleagues conducted a systematic review of studies performed in high-income countries on the effect of loss to follow-up at on neurodevelopmental outcomes for extremely preterm infants (1). This study showed substantial variation in neurodevelopmental impairment (NDI) rates between cohorts, with the worst rates reported mostly from studies in the United States (US). The authors demonstrated that this can be partly explained by higher loss to follow-up of healthy patients among US cohorts.

These findings raise several important questions:

1. What are the reasons for the large variation in NDI?
2. Why are follow-up rates in the US lower than those in other high-income countries?
3. What are the implications for comparative performance measurement?

What are the reasons for the large variation in NDI?

For the purpose of policy-setting or quality improvement, comparative measurement of outcomes often serves to spotlight actionable differences in quality of healthcare delivery. In their systematic review, Guillén et al. found NDI rates varying from 12.4 to 57.5%. Since most US cohorts fall within the upper end of this range, a superficial assessment might imply that the US healthcare system does a poor job at caring for these infants. While US cohorts experienced significantly greater loss to follow-up of healthier infants, this component accounted for only 38% of the variation in NDI rates. Therefore, additional factors contribute to higher NDI rates among US cohorts. Higher NDI rates may result from differences in structures of care delivery at the hospital or healthcare system level (e.g., care

Correspondence to: Jochen Profit, MD, MPH Assistant Professor Baylor College of Medicine Section of Neonatology Houston Center for Quality of Care & Utilization Studies VA HSR&D (152) 2002 Holcombe Boulevard Houston, TX 77030 Tel: 713.794.8609 Fax: 713.748.7359 profit@bcm.edu.

The authors have no financial relationships relevant to this article to disclose.

integration, regionalization, or patient-to-nurse ratios); care processes (e.g., use of potentially harmful soy-based intravenous lipid emulsions in the US; non-adherence to care guidelines); or culture of care delivery (e.g., safety culture) (2-7).

Beyond these factors, NDI rates reflect the complex interactions between the patient's (and his parents') biology, as well as the socioeconomic context in which healthcare is delivered (8). Since biology and socioeconomic context are not usually under the control of healthcare providers, they need to be adequately accounted for in statistical analyses before variation between countries can be attributed to differences in quality of care.

The four major sources of variation in outcomes are risk, chance, true differences in care, and bias (see Table 1) (9). Differences in quality of care originate from differences in personnel, practices, resources, technology, organizational structure, and culture. In order to attribute differences in outcomes to differences in quality of care, comparative analysis of performance must account for risk and chance.

In performance measurement, chance is addressed by allowing absolute or relative results to vary within a predefined range, such as a 95 or 99.9% control limit. Chance is influenced by sample size, with smaller samples having a higher probability for an outlier result. Indeed, some of the single center and smaller studies have among the lowest (10-13) and highest (14) rates of NDI. Funnel plots could highlight whether individual studies truly have outlier NDI rates or whether these fall within random variation among the group.(15)

Risk is a characteristic that elevates the probability of an adverse outcome but is beyond the control of the agent responsible for the outcome (9). Comparative performance assessments should account for such risks. In neonatal intensive care, low gestational age or birth weight, male gender, and plurality have been shown to be associated with higher mortality rates. Additional factors commonly predicting outcome include birth at a perinatal center (16-18); cesarean section, which appears to confer a survival benefit to very low birth weight (VLBW) infants who are also small for gestational age (19;20); and the existence of congenital anomalies (21;22). Guillén and colleagues' study is an important addition to this literature and highlights the importance of accounting for attrition rates in comparative outcome assessments.

The relative contributions of socioeconomic and biological influences (e.g., race, ethnicity, poverty, education, social support, and health system functionality) to NDI can be difficult to measure and are often excluded from international comparisons. However, these factors may be highly influential and help to identify high-risk groups to target through health policy. We think that in future studies every effort should be made to collect this data in a standardized fashion.

The methods used to define outcomes may also account for variation in NDI rates, even beyond those demonstrated by Guillén et al. Despite the authors' attempts at defining a homogeneous sample for comparison and statistical adjustments for clinical risk, residual heterogeneity across studies may bias results. For example, NDI rates at high-risk perinatal centers tend to be higher than population-based rates. Cohorts with more inclusive definitions of NDI (e.g., those including seizures) may have increased rates. Selection bias

may exist with regard to continuing care in the face of predictable adverse neurologic outcome. Thus, it is unclear whether NDI rates reported in US studies are spuriously higher than in other high-income countries, or whether they result from systematic differences in quality of care or socioeconomic context.

Why are follow-up rates in the US lower than those in other high-income countries?

Multiple factors, including patient and family characteristics, healthcare system structure, and social environment, affect whether families keep their follow-up appointments more than one year post hospital discharge. Families must weigh whether the value of the visit exceeds the burdens associated with it. American families face several barriers, which may negatively influence this value equation. For example, social support systems, such as paid job-protected parental leave, are much less generous in the US than in most European countries (zero weeks compared to sixty eight weeks in Sweden) (23). US employment conditions may also be less forgiving, such that keeping appointments may result in wage losses, and endanger employment or chances for promotion.

Parents in the US may also be more worried about expenses associated with follow-up visits, such as, transportation, or co-payments. US health insurance coverage tends to be less comprehensive for speech and occupational therapy. Even when covered, families face higher out-of-pocket costs. These barriers limit healthcare utilization and may result in greater loss to follow-up.

In addition, labor mobility is higher in the US, compared to many European countries, and families may relocate to different geographic areas (24). Depending on how follow-up is ascertained in a given cohort, patients may be lost to follow-up even if they access care at a new location. Again, financial considerations may influence families' decision to move as they seek social support, higher wages, or more comprehensive health insurance. The impact of health expenditures on household financial health was illustrated in a study by Himmelstein and colleagues, who showed that these accounted for nearly half of personal bankruptcies.(25)

Guillén and colleagues show that infants lost to follow-up have less NDI. Is it possible, that lower US follow-up rates merely reflect more efficient healthcare resource use? Do US families underuse resources, or do other countries, because of more comprehensive social support and insurance coverage, overuse resources? While we cannot definitively answer these questions, the literature suggests a benefit of early education programs for high-risk infants of the kind initiated and tracked through follow-up programs (26). In this regard, loss to follow-up remains concerning.

What are the implications for comparative performance measurement?

Comparative performance measurement has become ubiquitous. We use it to decide which movie to watch, which restaurant to visit, and which schools to choose for our children. In recent years, comparative performance measurement has become increasingly common in

healthcare (27-30). Frequently, comparisons attempt to promote improvements in the quality of healthcare delivery through the use of monetary or reputational incentives (31-33). While users of this information (e.g. consumers, health plans, and policy-makers) may appreciate information on provider, hospital, and healthcare system performance, they tend to be less sensitive to the methodological weaknesses of approaches to measurement. Further, their desire for information may outpace measurement science. However, for those being evaluated, the stakes are high, and they demand that measurement be valid and reliable (34). The tension between the desire for information by users and the demand for accuracy by providers is not easily reconciled.

NDI rates are attractive for judging performance of neonatal intensive care. They represent the bottom-line of what the neonatal community, families, and policy-makers care about. However, for the reasons mentioned above, international comparisons of NDI rates should be interpreted with caution. Much of the differences in NDI rates between the US and other high-income countries may be explained by systematic bias and confounding.

Yet, it is possible that there are real differences in the care that US infants receive, which promote higher NDI rates. Given the current imperfect state of knowledge regarding potential socioeconomic confounding, we recommend that comparisons of NDI rates between countries be viewed as a starting point for further inquiry and testable hypotheses. These hypotheses should be studied using a data collection framework that includes clinical and contextual measures.

We recognize that such a research agenda will take many years to bear fruit. In the meantime, the following approaches may improve the value and policy relevance of international comparisons (Table 1). First, as done in this systematic review, international comparisons of NDI rates should always use statistical methods that account for risk and chance (35;36). Attrition rates could be handled in a manner similar to a missing data analysis, with the potential to impute values based on the characteristics of those lost to follow-up (8). Analogous to methods used in decision science, the effect of potential biases on NDI rates could be investigated using sensitivity analysis. For example, in simulation models follow-up rates in laggard cohorts could be set to 100%, or be allowed to vary within a defined range.

Another way to improve comparative performance measurement is to focus less on one outcome measure and instead, develop a dashboard of measures relevant to families, clinicians, and policy-makers. In related work, we are working to develop a composite indicator of neonatal intensive care quality for VLBW infants, the Baby-MONITOR, based on data items collected by members of the California Perinatal Quality Care Collaborative and the Vermont Oxford Network (37). A mix of process and outcome measures, such as those used in the Baby-MONITOR, might provide deeper insights into reasons for performance variation (38). Given the influence that multiple disciplines have on neonatal outcomes, such a scorecard should be expanded longitudinally to include perinatal and post-neonatal care. Measurement should reflect the multidisciplinary nature of neonatal care by assessing integration of care and coordination among disciplines. A multi-dimensional

evaluation of the perinatal medical home may provide users with more actionable information on how to best serve our patients.

In summary, international comparisons can result in oversimplification of policy messages. Nevertheless, such comparisons, if done well, offer the opportunity to learn from existing differences and generate hypotheses for improvement that would be otherwise remain hidden.

Acknowledgments

Grant support: Jochen Profit's contribution is supported, in part, by the Eunice Kennedy Shriver National Institute of Child Health and Human Development #1 K23 HD056298-01 (PI: Profit). Dr. Woodard's contribution is supported in part by a VA Health Services Research and Development Pilot Award (PPO 09-316). Drs. Profit and Woodard also receive support from a Veterans Administration Center Grant (VA HSR&D CoE HFP90-20).

Abbreviations

NDI	neurodevelopmental impairment
NICU	neonatal intensive care unit
VLBW	very low birth weight

Reference List

- Guillen U, DeMauro S, Ma L, Zupancic J, Roberts R, Schmidt B, et al. Relationship between attrition and neurodevelopmental impairment rates in extremely preterm infants at 18-24 months: a systematic review. *Arch Pediatr Adolesc Med*. 2011
- Profit J, Zupancic JA, McCormick MC, Richardson DK, Escobar GJ, Tucker J, et al. Moderately premature infants at Kaiser Permanente Medical Care Program in California are discharged home earlier than their peers in Massachusetts and the United Kingdom. *Arch Dis Child Fetal Neonatal Ed*. 2006; 91:245–50.
- Lorch SA, Myers S, Carr B. The regionalization of pediatric health care. *Pediatrics*. 2010; 126:1182–90. [PubMed: 21041285]
- Profit J, McCormick MC, Escobar GJ, Richardson DK, Zheng Z, Coleman-Phox K, et al. Neonatal intensive care unit census influences discharge of moderately preterm infants. *Pediatrics*. 2007; 119:314–9. [PubMed: 17272621]
- Deshpande G, Simmer K. Lipids for parenteral nutrition in neonates. *Curr Opin Clin Nutr Metab Care*. 2011; 14:145–50. [PubMed: 21192257]
- Profit J, Cambric-Hargrove AJ, Tittle KO, Pietz K, Stark AR. Delayed pediatric office follow-up of newborns after birth hospitalization. *Pediatrics*. 2009; 124:548–54. [PubMed: 19651578]
- Profit J, Etchegaray J, Petersen LA, Sexton BJ, Mei M, Thomas EJ. Safety attitudes among neonatal intensive care units vary widely. *E-PAS*. 2010:4425, 595.
- Profit J, Typo KV, Hysong SJ, Woodard LD, Kallen MA, Petersen LA. Improving benchmarking by using an explicit framework for the development of composite indicators: an example using pediatric quality of care. *Implement Sci*. 2010; 5:13. [PubMed: 20181129]
- Gould JB. Vital records for quality improvement. *Pediatrics*. 1999; 103:278–90. [PubMed: 9917471]
- Kutz P, Horsch S, Kuhn L, Roll C. Single-centre vs. population-based outcome data of extremely preterm infants at the limits of viability. *Acta Paediatr*. 2009; 98:1451–5. [PubMed: 19575767]
- Tommiska V, Heinonen K, Lehtonen L, Renlund M, Saarela T, Tammela O, et al. No improvement in outcome of nationwide extremely low birth weight infant populations between 1996-1997 and 1999-2000. *Pediatrics*. 2007; 119:29–36. [PubMed: 17200268]

12. Jacobs SE, O'Brien K, Inwood S, Kelly EN, Whyte HE. Outcome of infants 23-26 weeks' gestation pre and post surfactant. *Acta Paediatr.* 2000; 89:959–65. [PubMed: 10976839]
13. Wilson-Costello D, Friedman H, Minich N, Siner B, Taylor G, Schluchter M, et al. Improved neurodevelopmental outcomes for extremely low birth weight infants in 2000-2002. *Pediatrics.* 2007; 119:37–45. [PubMed: 17200269]
14. Rijken M, Stoelhorst GM, Martens SE, van Zwieten PH, Brand R, Wit JM, et al. Mortality and neurologic, mental, and psychomotor development at 2 years in infants born less than 27 weeks' gestation: the Leiden follow-up project on prematurity. *Pediatrics.* 2003; 112:351–8. [PubMed: 12897286]
15. Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Stat Med.* 2005; 24:1185–202. [PubMed: 15568194]
16. Chien LY, Whyte R, Aziz K, Thiessen P, Matthew D, Lee SK. Improved outcome of preterm infants when delivered in tertiary care centers. *Obstetrics & Gynecology.* 2001; 98:247–52. [PubMed: 11506840]
17. Lee SK, McMillan DD, Ohlsson A, Boulton J, Lee DS, Ting S, et al. The benefit of preterm birth at tertiary care centers is related to gestational age. *Am J Obstet Gynecol.* 2003; 188:617–22. [PubMed: 12634630]
18. Pietz K, Gould JB, Kowalkowski MA, Petersen LA, Profit J. Outborn birth of very low birth weight infants increases morbidity and mortality. *E-PAS.* 2011:2914, 199.
19. Lee HC, Gould JB. Survival Rates and Mode of Delivery for Vertex Preterm Neonates According to Small- or Appropriate-for-Gestational-Age Status. *Pediatrics.* 2006; 118:e1836–e1844. [PubMed: 17142505]
20. Lee HC, Gould JB. Survival Advantage Associated With Cesarean Delivery in Very Low Birth Weight Vertex Neonates. *Obstet Gynecol.* 2006; 107:97–105. [PubMed: 16394046]
21. Synnes A, Berry M, Jones H, Pendray M, Stewart SD, Lee SK, et al. Infants with congenital anomalies admitted to neonatal intensive care units. *Am J Perinatology.* 2004; 21:199–208.
22. Suresh GK, Horbar JD, Kenny M, Carpenter JH. Major birth defects in very low birth weight infants in the Vermont Oxford Network. *The Journal of Pediatrics.* 2001; 139:366–73. [PubMed: 11562615]
23. Tanaka S. Parental leave and child health across OECD countries. *Economic Journal.* 2005; 115:F7–F28.
24. Layard, R.; Nickell, SJ. Labour Market Institutions and Economic Performance.. In: Ashenfelter, O.; Card, D.; Layard, R.; Nickell, SJ., editors. *Handbook of Labour Economics.* North-Holland; Amsterdam: 1999. 1999
25. Himmelstein DU, Warren E, Thorne D, Woolhandler S. Illness and injury as contributors to bankruptcy. *Health Aff (Millwood).* 2005; (Suppl):W5. Web Exclusives. [PubMed: 15689369]
26. McCormick MC, Brooks-Gunn J, Buka SL, Goldman J, Yu J, Salganik M, et al. Early intervention in low birth weight premature infants: results at 18 years of age for the Infant Health and Development Program. *Pediatrics.* 2006; 117:771–80. [PubMed: 16510657]
27. Premier Hospital Quality Incentive Project. Premier Inc.; Available at: <http://www.premierinc.com/all/quality/hqi/resources/top-performer-summary.pdf>. [5-19-2006]
28. Profit J, Zupancic JA, Gould JB, Petersen LA. Implementing pay-for-performance in the neonatal intensive care unit. *Pediatrics.* 2007; 119:975–82. [PubMed: 17473099]
29. World Health Organization. [4-3-2006] World health report 2000.. Health systems: improving performance. Available at: <http://www.who.int/whr/2000/en/index.html>.
30. Davis, K.; Schoen, C.; Stremikis, K. [6-8-2011] Mirror, Mirror on the Wall: How the Performance of the U.S. Health Care System Compares Internationally 2010 Update.. The Commonwealth Fund. 2010. Available at: http://www.commonwealthfund.org/~media/Files/Publications/Fund%20Report/2010/Jun/1400_Davis_Mirror_Mirror_on_the_wall_2010.pdf.
31. Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? *Ann Intern Med.* 2006; 145:265–72. [PubMed: 16908917]
32. Mandel KE, Kotagal UR. Pay for performance alone cannot drive quality. *Arch Pediatr Adolesc Med.* 2007; 161:650–5. [PubMed: 17606827]

33. Profit J, Petersen LA. Pay for performance is growing up. *Arch Pediatr Adolesc Med.* 2007; 161:713–4. [PubMed: 17606837]
34. Woodard LD, Petersen LA. Improving the performance of performance measurement. *J Gen Intern Med.* 2010; 25:100–1. [PubMed: 19953336]
35. Draper D, Gittoes M. Statistical analysis of performance indicators in UK higher education. *Journal of the Royal Statistical Society.* 2004; 167:449–74.
36. Schulman J, Spiegelhalter DJ, Parry G. How to interpret your dot: decoding the message of clinical performance indicators. *J Perinatol.* 2008; 28:588–96. [PubMed: 18633418]
37. Profit J, Gould JB, Zupancic JA, Stark AR, Wall KM, Kowalkowski MA, et al. Formal selection of measures for a composite index of NICU quality of care: Baby-MONITOR. *J Perinatol.* Feb 24.2011 [Epub ahead of print].
38. Jha AK. Measuring hospital quality: what physicians do? How patients fare? Or both? *JAMA.* 2006; 296:95–7. [PubMed: 16820553]

Table 1

Strategies to enhance usefulness of comparative performance assessments

Factors	Strategies for Mitigation
Risk	Statistical adjustment for -Illness severity -Socioeconomic factors
Chance	Control limits around parameter estimate to separate random from non-random variation Increase sample size
Quality of care	Focus on processes of care Composite indicators
Bias	Definition of homogeneous sample Standardized measurement of process or outcome Repeated sampling