

Sharing Data to Accelerate Medicine Development and Improve Neonatal Care: Data Standards and Harmonized Definitions

Kate Costeloe, MD^{1,*}, Mark A. Turner, MBChB, PhD^{2,*}, Michael A. Padula, MD, MBI³, Prakesh S. Shah, MD⁴, Neena Modi, MD⁵, Roger Soll, MD⁶, Dominique Haumont, MD⁷, Satoshi Kusuda, MD⁸, Wolfgang Göpel, MD⁹, Yun Sil Chang, MD, PhD¹⁰, P. Brian Smith, MD¹¹, Kei Lui, MD¹², Jonathan M. Davis, MD¹³, and Lynn D. Hudson, PhD¹⁴, and the International Neonatal Consortium[†]

The development and testing of the majority of medicine prescribed for newborn infants has been inadequate, with multiple barriers despite multiple legislative initiatives to encourage studies.¹ One impediment is our current limited capacity to share standardized data that would facilitate the evaluation of the effectiveness and risks of medicine given to newborn infants. Neonatal drug development requires data sharing²⁻⁴ and in this commentary, we will discuss the importance and feasibility of defining standards for definitions when sharing neonatal data in a context that meets the needs of multiple stakeholders.

There are many reasons to share data about effectiveness and risks of neonatal drug therapy. Baseline data across large and diverse populations can inform the development of clinical trials to ensure they have adequate power to detect relevant effects. The results of studies can be compared in meta-analyses. Multiple trials allow for the evaluation of generalizability across a range of settings. Placebo arms can be pooled to assess background rates of serious morbidities and other adverse events. Databases can contribute to the evaluation of a medicine across the entire pipeline of development through to postmarketing studies to monitor safety and efficacy. There is increasing recognition of the value of “real-world data,”^{5,6} but this approach needs efficient data sharing and reliable parameters for linkage.

The problems that arise when information is shared in the absence of standardized ways to define events have been highlighted in recent publications. For example, the iNeo collaborators identified 13 different definitions of bronchopulmonary dysplasia (BPD) in 628 papers⁷; see also Steinhorn et al.⁸ A comparison of outcomes for infants of very low birth weight across 8 databases in high-resource settings (Canada, United Kingdom, Sweden, Spain, Switzerland, Israel, Australia/New Zealand, and Japan) found that the incidence of a composite outcome comprising death, severe cerebral ultrasound scan abnormalities, BPD, and treated retinopathy of prematurity ranged between 26% and 42% with significant variation among sites.⁹ These differences may well be attributable to inconsistencies in

components of data and denominators (such as number of recorded births, number of admissions to neonatal units, criteria for a live birth). Variation in how clinicians interpret seemingly similar conditions as well as factors not generally recorded, such as differences in service delivery and staff numbers and skill-mix, also must be considered. Finding a way to address such variation in data capture, as has been achieved in some databases, would unlock an valuable resource.¹⁰

Thus, data sharing requires data standards, defined as “a set of rules on how a particular type of data should be structured, defined, formatted, or exchanged between computer systems.”¹¹ The definition of data standards in neonatology has been difficult because data sharing has not been prioritized sufficiently across stakeholders, there are limited drivers for consensus across multiple perspectives and continents, and the biological variation caused by the combination of ontogeny and multiple comorbidities impedes discussion.

The methodology for the development of data standards has been engineered by organizations such as the Clinical Data Interchange Standards Consortium (CDISC), which develops data standards that meet the specifications of regulatory agencies for use in clinical trials.¹² One family of CDISC standards that meet regulatory specifications relates to “foundational standards” about how data captured by all clinical trials can be recorded and presented from protocol through to reporting and

BPD	Bronchopulmonary dysplasia
CDISC	Clinical Data Interchange Standards Consortium
EVS	Enterprise Vocabulary Service
INC	International Neonatal Consortium
NCI	National Cancer Institute
NICHHD	Eunice Kennedy Shriver National Institute of Child Health and Human Development

From the ¹Paediatric Research, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London; ²Institute of Translational Medicine, University of Liverpool, Liverpool, United Kingdom; ³Division of Neonatology, Children's Hospital of Philadelphia, Philadelphia, PA; ⁴Department of Paediatrics and Institute of Health Policy, Management and Evaluation, Lunenburg Tannebaum Research Institute, Mount Sinai Hospital, Toronto, Canada; ⁵Neonatal Medicine, Imperial College London, Chelsea and Westminster Hospital Campus, London, United Kingdom; ⁶Vermont Oxford Network, Neonatology, University of Vermont College of Medicine, Burlington, VT; ⁷Department of Neonatology, Saint-Pierre University Hospital, Brussels, Belgium; ⁸Department of Neonatology, Maternal and Perinatal Center, Tokyo Women's Medical University, Tokyo, Japan; ⁹Neonatology and Paediatric Intensive Care, University of Lübeck, Department of Paediatrics, Lübeck, Germany; ¹⁰Department of Pediatrics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea; ¹¹Department of Pediatrics, Duke University School of Medicine, Durham, NC; ¹²Discipline of Paediatrics, School of Women's and Children's Health, Sydney, New South Wales, Australia; ¹³Department of Paediatrics, Floating Hospital for Children, Tufts Medical Center, Tufts Clinical and Translational Science Institute, Tufts University, Boston, MA; and ¹⁴Science, Critical Path Institute, Tucson, AZ

*Contributed equally.

[†]List of additional contributors from the International Neonatal Consortium available at www.jpeds.com (Appendix).

The International Neonatal Consortium is supported by the Food and Drug Administration (5U18 FD005320) and annual member dues. The authors declare no conflicts of interest.

0022-3476/\$ - see front matter. © 2018 Elsevier Inc. All rights reserved.
<https://doi.org/10.1016/j.jpeds.2018.07.082>

interpretation. Another family of regulatory-compliant standards relates to terminologies for morbidities in a number of therapeutic areas. We focus here on standards relating to terminologies for neonatal data because foundational standards are anticipated to be broadly applicable.

Services such as the Systematized Nomenclature of Medicine—Clinical Terms and the terms curated by the National Cancer Institute's Enterprise Vocabulary Service (NCI EVS) are likely to be used increasingly by the entire pediatric community. For example, a Pediatric Terminology Harmonization Initiative led by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) established standard coding terminology for child life stages.¹³ Subsequently, a collaborative effort between NICHD, NCI EVS, and experts from the US and other countries defined terms in Pediatric Adverse Events, Perinatology, Pediatric Rheumatology, Pediatric Endocrinology, Pediatric Infectious Diseases, and Pediatric Oncology. The NCI Thesaurus and EVS update the Pediatric Terminology sets monthly.

There are multiple initiatives and inconsistencies between initiatives that reduce the value of data standards and definitions. Accordingly, shared data standards and definitions require careful management of terminology with “control” of the terms, including synonyms and relationships to terms used in other settings. Terminologies can be integrated, as occurred when the NCI EVS terminology for pediatric adverse events was incorporated into the Medical Dictionary for Regulatory Activities.¹⁴ A collection of “common data elements” (a combination of a precisely defined question [variable] paired with a specified set of responses to the question that is common to multiple datasets or used across different studies) for clinical research is available through a National Institutes of Health repository (<http://cde.nlm.nih.gov>).¹⁵

These methodologies and services have been applied to neonatal drug development, for example, the Pediatric Terminology Harmonization Initiative integrated more than 2000 terms used by 3 neonatal research networks, but they have not been adopted widely (<https://evs.nci.nih.gov/ftp1/NICHD/About.html>). This lack of adoption may reflect content (extant vocabularies may not meet all needs) and/or implementation (that is, a lack of wide awareness or adoption).

The development of terminologies is not limited to specific case definitions for morbidities but ideally supports sharing data that can be used to ascertain multiple definitions. This is important because consensus on the best definition may not be reached, yet useful results can be obtained from analysis of data using several different definitions. In these situations, it may not be possible, or useful, to have a single definition that is used in all settings (across, eg, management, administration, coding, quality improvement, research). Seeking consensus about the best-possible definition of, for example, BPD, may not be appropriate. Instead, consensus could be reached about which definitions of BPD are worth ascertaining and which data variables are required to support the selected definitions so that the standards for capturing these variables can be defined, combined, and compared. For example, the CDISC published a Therapeutic Area guideline for asthma ([\[www.cdisc.org/standards/therapeutic-areas/asthma\]\(http://www.cdisc.org/standards/therapeutic-areas/asthma\)\) that drew on a consensus workshop about outcomes of asthma.¹⁶ Based on the consensus workshop, a number of “concept maps,” such as “what is an exacerbation of asthma,” were developed. The concept map for “exacerbation of asthma” then informed the standards for a core data set that would capture all definitions of asthma exacerbation.](https://</p></div><div data-bbox=)

It is important to note that case definitions of neonatal morbidities used in clinical trials designed for regulatory purposes may include components not included in nontrial data, eg, an outcome of a clinical trial may require a particular test or a specific assessment at a defined time point after an intervention. The rigor of such trial definitions should not be compromised or constrained by what is available “routinely” and may need to reflect de novo development of definitions and enhanced data capture. Nonetheless, there is likely to be considerable overlap in content between trial and nontrial datasets. To extend the use of nontrial datasets to support clinical trials, to promote the generalizability of results, and to contribute to postmarketing surveillance core, neonatal datasets should include the definitions used in trials and other settings.

The neonatal community should work through initiatives that already are addressing data standards, harmonized definitions, and comparability of data from heterogeneous sites. For example, the Brighton Collaboration has derived a detailed case definition for neonatal infection through a global, multistakeholder consensus group to support the evaluation of vaccines and is now working on neonatal seizures.¹⁷ The UK Neonatal Collaborative has developed a case-definition for necrotizing enterocolitis based on a whole population study.¹⁸ All neonatal units across England and Wales submit data to the National Neonatal Research Database through their data entry supplier. At present, there are more than 600 000 babies and 7 million days of care for these babies submitted into the National Neonatal Research Database (<http://www.imperial.ac.uk/neonatal-data-analysis-unit/neonatal-data/>). Other initiatives are developing core datasets that can be recorded during all studies involving a specified patient group (eg, COMET [<http://www.comet-initiative.org/>], Core Outcomes in Neonates, or COIN [<http://www.comet-initiative.org/studies/details/842>]). A global clearinghouse for data standards is provided by ELIXIR (<https://biosharing.org>). Although these initiatives are useful components of an approach to data sharing in neonates, a comprehensive approach is needed.

Creating data standards requires engagement of clinical teams, research teams, drug development teams, and other stakeholders. The harmonization of case definitions and development of data standards will require substantial work to upgrade the existing databases to allow for more extensive research use. Consensus from the broad stakeholder community needs to be reached on the case definitions for core neonatal morbidities and the specific data variables that need to be collected.

We propose the following steps toward neonatal data sharing (**Table**): (1) develop standards for data; (2) promote engagement with key stakeholders; and (3) facilitate the collection

Table. Steps to promote standards and definitions that support data sharing. The steps should proceed in parallel.

Activity	Develop standards for data	Promote engagement with stakeholders	Facilitate high quality datasets
Key Principles	Case definitions for morbidities are best drawn from a core dataset that supports the use of study-specific and other case definitions	Consensus across all relevant stakeholders is desirable about the content of each core dataset and the case definitions that can be derived from this	Terminology used in datasets needs to be specified, and used, consistently within each dataset and across datasets.
Initial steps	Define and reach broad stakeholder agreement, including the involvement of parents, on core neonatal outcomes	Identify relevant stakeholder	Define quality and completeness criteria for datasets
Main actions	Identify secondary and “nice-to-have” outcomes	Explain the rationale and purpose of this process to clinicians and others responsible for recording data	Use data standards in data collection (with other standards)
Supporting actions	For each core outcome: a. Identify the discrete data variables that are required as data “concepts” curated by CDISC/EVS b. Define the core dataset from these discrete data variables c. Identify granular components of data that support the criteria to satisfy the data definition for each outcome e.g. harmonize data elements and standards	Consider how best to incentivise complete and accurate data recording	a. Register the dataset as an information standard, when possible b. Consider standards for access to data for sharing and linkage purposes and publish these alongside each database c. Establish a mechanism for funding research infrastructure to facilitate sharing as a part of neonatal trials ¹⁰

of high-quality datasets. The involvement of parents, engagement and enthusiasm of those recording data, and quality criteria are essential.

These steps could be followed separately for clinical trials or for quality measures. Given the considerable overlap in goals, the breakthrough for neonates would be to work on clinical trials and quality measures simultaneously. Each trial, study, or clinical database could choose which data to include based on the purpose of data collection and resource constraints.

The International Neonatal Consortium (INC) has surveyed the gaps and opportunities in developing data standards for neonatology to advance drug development in this population. Key issues include creating a forum for stakeholder engagement on this topic, developing consensus regarding data standards for neonatal studies, and implementing the standards.

Opportunities for Consensus

Ideally, information on neonatal morbidities could be shared easily because definitions of diagnoses would be supported by objective clinical, anatomical, tissue, imaging, or laboratory data. This can be done through quantitative or qualitative data, as discussed for chronic pulmonary insufficiency of prematurity.⁸ Quantitative data are easier to share, validate, and combine, assuming that the quantification can be done consistently. Qualitative data or data about limited categories (such as clinician diagnosis) may be closer to clinical practice but is subject to multiple influences and biases. Many neonatal diagnoses and much quantitative data lack robust, objective definitions. These problems need solutions that meet the specific needs of neonates. Ascertainment of morbidities can be inconsistent because assessment of clinical signs is often subjective and thresholds for treatment vary. Furthermore, many definitions of neonatal

morbidities are definitions of convenience that have not been demonstrated to predict final outcomes or burdens of disease. Such inadequate definitions cannot be used as surrogate outcomes, although neonatal morbidities with robust definitions may be linked with meaningful outcomes and serve as surrogate outcomes.

INC has established a working group for data that can serve as a catalyst for developing consensus and develop data standards for neonatal studies.¹⁹ This will require collaboration with industry and regulators as well as between clinicians and databases. Funding for data infrastructure is essential. INC is evaluating CDISC data standards for use in neonatal clinical trials to develop comprehensive foundations for data sharing. Following a gap analysis between existing standards and the needs of neonatal trials, new standards will be developed as needed. This will involve a number of consensus groups that must include representation for all relevant stakeholders. Members of consensus groups need to carefully consider and represent the opinions and practice of the majority of their colleagues. Other groups and individuals will help update the concepts, definitions, and core datasets. Everybody can work toward data sharing by contributing to data recording when it will enhance care, even if the benefits are not immediately nor applicable to an individual neonate. This includes accepting the need to record data that benefits the whole community rather than a single neonate or a single unit.

Although data sharing needs an initial investment in resources, downstream benefits will be significant. These benefits are likely to include substantial cost-savings for research funders.³ The costs of developing the means to share data need explicit funding. This could include contributions from other projects (clinical research or information technology infrastructure) or dedicated project funding. Establishing a mechanism to support such infrastructure as a component of funded

trials may be an effective means to begin to bridge this gap.³

The groups that manage multi-institutional databases can promote data sharing in various ways. While acknowledging the constraints and obstacles, they can improve data recording by incorporating core datasets to existing electronic health records. Specifically, including data that can have a definitive impact globally while meeting local needs for information might be crucial. Most importantly, they can contribute to the definition of standards for data collection and their implementation.

Implementation

Standard approaches to implementation need to be considered, taking account of other demands on the time and energy of all stakeholders, particularly front-line staff who generate and enter data.²⁰ A plan that integrates the development and implementation of standardized terminologies is needed. This plan needs to share the rewards and benefits of data sharing. This includes moving beyond institutions and database organizations to promote data sharing among people who enter data and are responsible at the site level. The broader neonatal community needs to provide clear benefits at the site level to justify the effort made by staff.

Most current databases focus on infants of extremely low birth weight and those born at less than 28 weeks of gestation. Although this reflects the high burden that arises from individual cases of birth at extreme prematurity, the overall burden of preterm birth and neonatal morbidity affects babies born at all gestational ages.

In the absence of common data standards, neonatal studies will continue to lag behind the need for new and improved treatments. The status quo promotes inefficiency and wasted effort because the power and generalizability of neonatal drug development is limited by the restricted number of settings that currently contribute to data sharing. Globally, the benefits of sharing will be significantly greater than the costs of change. However, local circumstances often impede the realization of global benefits: institutional factors need to be considered in the development and implementation of standardized terminologies.

Conclusions

We have reviewed the importance of defining standards for definitions when sharing neonatal data in a context that meets the needs of multiple stakeholders. Agreeing on definitions of the most common morbidities is important but is best viewed as a way to identify the content of data standards rather than the driver of data sharing. We propose that iNeo, eNewborn, Vermont Oxford Network, UK Neonatal Collaborative, and other key stakeholders collaborate with INC to develop a consolidated approach to defining data required to document neonatal morbidities that can be shared between trial and nontrial databases on a global basis. Only then can we leverage these

important resources and optimize the outcomes for our most vulnerable population of patients. ■

Submitted for publication Nov 18, 2017; last revision received Jun 6, 2018; accepted Jul 25, 2018

Reprint requests: Mark A. Turner, MBChB, PhD, Neonatal Unit, Liverpool Women's Hospital, Crown St, Liverpool L8 7SS, United Kingdom. E-mail: mark.turner@liverpool.ac.uk

References

1. Ward RM, Benjamin D, Barrett JS, Allegaert K, Portman R, Davis JM, et al. Safety, dosing, and pharmaceutical quality for studies that evaluate medicinal products (including biological products) in neonates. *Pediatr Res* 2017;81:692-711.
2. Longo DL, Drazen JM. Data sharing. *N Engl J Med* 2016;374:276-7.
3. Warren E. Strengthening research through data sharing. *N Engl J Med* 2016;375:401-3.
4. Taichman DB, Backus J, Baethge C, Bauchner H, de Leeuw PW, Drazen JM, et al. Sharing clinical trial data: a proposal from the International Committee of Medical Journal Editors. *Rev Med Chil* 2016;188:91-2.
5. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence—what is it and what can it tell us? *N Engl J Med* 2016;375:2293-7.
6. Sherman RE, Davies KM, Robb MA, Hunter NL, Califf RM. Accelerating development of scientific evidence for medical products within the existing US regulatory framework. *Nat Rev Drug Discov* 2017;16:297-8.
7. Hines D, Modi N, Lee SK, Isayama T, Sjörs G, Gagliardi L, et al. Scoping review shows wide variation in the definitions of bronchopulmonary dysplasia in preterm infants and calls for a consensus. *Acta Paediatr* 2017;106:366-74.
8. Steinhorn R, Davis JM, Göpel W, Jobe A, Abman S, Laughon M. Chronic pulmonary insufficiency of prematurity: developing optimal endpoints for drug development. *J Pediatr* 2017;191:15-21.e1.
9. Shah PS, Lui K, Sjörs G, Mirea L, Reichman B, Adams M, et al. Neonatal outcomes of very low birth weight and very preterm neonates: an international comparison. *J Pediatr* 2016;177:144-52.e6.
10. Kush R, Goldman M. Fostering responsible data sharing through standards. *N Engl J Med* 2014;370:2163-5.
11. Center for Drug Evaluation and Research (CDER). US Food and Drug Administration. Research C for DE and electronic. Submissions to CDER—CDER Data Standards Program [Internet]. <https://www.fda.gov/drugs/developmentapprovalprocess/formssubmissionrequirements/electronicsubmissions/ucm249979.htm>. Accessed May 31, 2018.
12. Hudson LD, Kush RD, Navarro Almarino E, Seigneuret N, Jackson T, Jauregui B, et al. Global standards to expedite learning from medical research data. *Clin Transl Sci* 2018;11:342-4.
13. Kahn MG, Bailey LC, Forrest CB, Padula MA, Hirschfeld S. Building a common pediatric research terminology for accelerating child health research. *Pediatrics* 2014;133:516-25.
14. Gipson DS, Kirkendall ES, Gumbs-Petty B, Quinn T, Steen A, Hicks A, et al. Development of a pediatric adverse events terminology. *Pediatrics* 2017;139.
15. Sheehan J, Hirschfeld S, Foster E, Ghitza U, Goetz K, Karpinski J, et al. Improving the value of clinical research through the use of Common Data Elements. *Clin Trials* 2016;13:671-6.
16. Busse WW, Morgan WJ, Taggart V, Togias A. Asthma outcomes workshop: overview. *J Allergy Clin Immunol* 2012;129:S1-8.
17. Vergnano S, Buttery J, Cailles B, Chandrasekaran R, Chiappini E, Clark E, et al. Neonatal infections: case definition and guidelines for data collection, analysis, and presentation of immunisation safety data. *Vaccine* 2016;34:6038-46.

18. Battersby C, Longford N, Mandalia S, Costeloe K, Modi N, UK Neonatal Collaborative Necrotising Enterocolitis (UKNC-NEC) Study Group. Incidence and enteral feed antecedents of severe neonatal necrotising enterocolitis across neonatal networks in England, 2012-13: a whole-population surveillance study. *Lancet Gastroenterol Hepatol* 2017;2:43-51.
19. Turner MA, Davis JM, McCune S, Bax R, Portman RJ, Hudson LD. The International Neonatal Consortium: collaborating to advance regulatory science for neonates. *Pediatr Res* 2016;80:462-4.
20. Blair M. Getting evidence into practice—implementation science for paediatricians. *Arch Dis Child* 2014;99:307-9.

Appendix

The International Neonatal Consortium Data Working Group had extensive conversations that underpin this paper. Additional contributors from the International Neonatal Consortium are listed here.

Khosrow Adeli, PhD, Hospital for Sick Children, Toronto, Canada

Karel Allegaert, MD, PhD, University of Leuven, Belgium

Gerri Baer, MD, US Food and Drug Administration, Bethesda, MD, USA

Simin Baygani, MS, Eli Lilly and Company, Indianapolis, IN, USA

George Chang, DO PhD, National Cancer Institute Enterprise Vocabulary Services, Bethesda, MD, USA

Gene Dempsey, PhD, University College Cork, Ireland

Laura Fabbri, MD, Chiesi Farmaceutici S.p.A., Italy

Victoria Higgins, PhD, Hospital for Sick Children, Toronto, Canada

Shinya Hirano, MD, Osaka Medical Center & Research Institute for Maternal and Child Health, Japan

Takehiko Hiroma, MD, PhD, Nagano Children's Hospital, Japan

Steven Hirschfeld, MD, PhD, Uniformed Services University of the Health Sciences, Bethesda, MD, USA

Hiroko Iwam, MD, Osaka City General Hospital, Japan

Lauren Kelly, PhD, Mount Sinai Hospital, Canada

Thierry Lacaze, MD, CHEO Research Institute, Ottawa, Canada

Susan McCune, MD, US Food and Drug Administration, Bethesda, MD, USA

Isabella Montagna, MSc, Chiesi Farmaceutici S.p.A., Italy

Hidefumi Nakamura, MD, PhD, National Research Institute for Child Health and Development, Japan

Martin Offringa, MD, PhD, University of Toronto, Canada

Ron Portman, MD, Novartis pharmaceuticals, NJ, USA

Catherine Sherwin, PhD, University of Utah, Salt Lake City, UT, USA

Mary Short, MS, Eli Lilly and Company, Indianapolis, IN, USA

Marta Terrile, PhD, Novartis pharmaceuticals, NJ, USA

Charlie Thompson, MD, Pfizer, New York City, NY, USA

Nam Tran, MD, University of California, Davis, CA, USA

Toshimitsu Yanagisawa, MD, Shinshu University School of Medicine, Japan