Measuring Clinical Benefit in Neonatal Randomized Clinical Trials: Challenges and Opportunities

March 23, 2023

Welcome & Overview

Morgan Romine

Duke-Margolis Center for Health Policy

Statement of Independence

The Robert J. Margolis, MD, Center for Health Policy is part of Duke University, and as such it honors the tradition of academic independence on the part of its faculty and scholars. Neither Duke nor the Margolis Center take partisan positions, but the individual members are free to speak their minds and express their opinions regarding important issues.

For more details on relevant institutional policies, please refer to the Duke Faculty Handbook, including the Code of Conduct and other policies and procedures. In addition, regarding positions on legislation and advocacy, Duke University policies are available at http://publicaffairs.duke.edu/government.

Anti-Trust Compliance Policy

- Meeting participants are committed to free and open competition in the marketplace and compliance with all applicable laws, including compliance with antitrust and competition laws.
- Meetings, communications and other activities are not intended in any way to limit the individual competitive decisions of the Members or to restrict competition among them.
- It is the responsibility of all call participants to be guided by this policy of strict compliance with the antitrust laws.
- Meetings, communications and other activities shall not be proposed for, or used for the purpose of, reaching or implementing any agreement concerning the competitive activities of others.
- Any participant who has a question regarding compliance with the antitrust laws or any aspect of the meetings, communications or activities should promptly consult the participant's own legal counsel.

Antitrust policy: off-limits topics

- When participating in Collaborative activities, Members, Member representatives, and observers shall avoid discussing non-public, company-specific information relating to current or future competition in the marketplace.
- Company-specific prices, pricing methods, pricing policies, pricing plans
- Sensitive cost information, including reimbursement rates or methods, pharmacy costs, and salaries/compensation information
- Marketing and strategic plans, market or competitive evaluations
- Identity and other information about present or potential customers, healthcare providers or payers, including costs, prices, profitability, marketing plans, and product development plans
- Research & development plans
- Other confidential or proprietary activities, strategies, processes or procedures
- Refusals to deal with any company or supplier
- Strategies or plans to award business or remove business from a specific company, to participate or not participate in any particular business opportunity or type of business opportunity
- Status of negotiations with present or potential customers, suppliers, payers or healthcare providers
- Any other confidential business information that could be used to reduce competition

Audience Participation

Submit questions and comments via Slido

Join at
slido.com
#Neonatal

Meeting Agenda

9:00 a.m.	Welcome and Introduction
9:10 a.m.	Opening Remarks
9:30 a.m.	Session 1: Current Approaches to Measuring Efficacy in Neonatal Randomized Control Trials
10:30 a.m.	Session 2: Challenges in Measuring Efficacy for Neonatal Conditions with Unmet Clinical Needs
12:00 p.m.	Break for Lunch
1:00 p.m.	Session 3: Key Considerations for Endpoint Selection for Neonatal Conditions
2:30 p.m.	Session 4: Novel Approaches to Measure Clinical Benefit in Neonatal Clinical Trials
4:00 p.m.	Fireside Chat
4:25 p.m.	Closing Remarks and Meeting Adjournment
4:30 p.m.	Adjourn

Opening Remarks from FDA

Hilary Marston

U.S. Food and Drug Administration

Opening Remarks from NICHD

Diana Bianchi
National Institute of Child Health and Human Development

Session 1: Current Approaches to Measuring Efficacy in Neonatal Randomized Control Trials

Moderator: Michele Walsh, National Institute of Child Health and Human Development

Session 1: Objectives

Objectives:

- Provide an overview of currently utilized approaches to measuring clinical benefit in neonatal randomized controlled trials (RCTs)
- Highlight differences between efficacy measurement to support regulatory approval versus clinical practice change
- Discuss strategies and considerations related to endpoint selection and clinical outcome measurement in neonatal RCTs

Session 1: Presenters

- **Gerri Baer,** U.S. Food & Drug Administration
- **Barbara Schmidt**, McMaster University & University of Pennsylvania
- **Kristi Watterberg**, University of New Mexico

Gerri Baer

U.S. Food & Drug Administration

ENDPOINT SELECTION FOR NEONATAL CLINICAL EFFICACY TRIALS

Gerri Baer, MD
Lead Physician
CDER Office of New Drugs
Office of Immunology and Inflammation
Division of Hepatology and Nutrition
March 23, 2023

Conflict of Interest and Disclaimer Statement

- I have no financial disclosures regarding drugs or any medical products.
- I have no conflicts of interest.
- Views expressed in this presentation are my own and do not necessarily represent an official FDA position.

What This Talk Will Cover

- The importance of this workshop and today's discussion
- Specific regulatory considerations and options for demonstrating efficacy for medical product approval
- How you (workshop panelists and participants) can help

What's the Problem / Why are we Here?

- Short of mortality and severe morbidity (severe neurodevelopmental, pulmonary, or other significant impairments), measuring the clinical impact of therapies to treat neonates is not straightforward.

- For the purpose of developing therapies to treat or prevent neonatal conditions, agreement* on how to measure clinically meaningful change is **essential**.

* Who needs to agree? Patients and caregivers, clinical researchers, clinicians, regulators, industry partners, research funding organizations, biostatisticians...

Why Measuring Clinical Benefit is Not Straightforward (an incomplete list)

- Short term benefit may not be durable and may be accompanied by long-term tradeoffs.
- Competing endpoints can complicate efficacy assessment.
- Not everyone values the same outcomes similarly.
- Assessment of medium- and long-term endpoints (defined as anything measured after the initial hospitalization) is complicated by attrition and intercurrent experiences, including the impact of socio-economic risk factors.

Why Today's Discussion is Essential

- Some conditions that begin in the perinatal-neonatal period (chronic pulmonary disease, brain injury, congenital infections, and others) cause significant morbidity and have inadequate therapeutic options.
- Innovation (either by developing new therapies or by specifically testing re-purposed drugs) requires a road map with clear parameters for judging success.

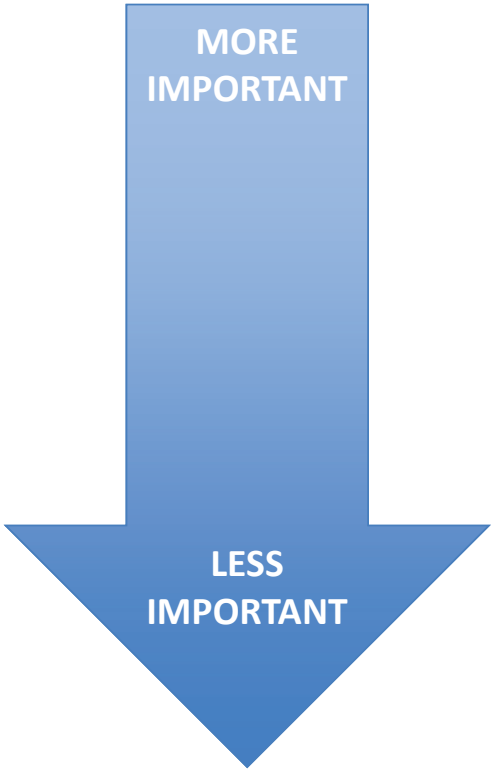
Statutory Basis for Establishing Efficacy

- Kefauver-Harris Amendments (1962) to the Food, Drug, & Cosmetic Act
  - "Substantial evidence" of effectiveness from "**adequate and well-controlled investigations**"
  - "…on the basis of which it could fairly and responsibly be concluded … the drug will have the effect it purports or is represented to have under the conditions of use prescribed…"

- For an investigation to be **adequate and well-controlled**
  - Methods of assessment of response should be well-defined and reliable
  - Can utilize clinical endpoints or, where appropriate, a surrogate endpoint
  - Endpoints should be clinically meaningful

Source: *Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products—Guidance for Industry*. U.S. Dept of HHS, Food and Drug Administration, December 2019.

Key Criteria for Selecting Endpoints



Clinical relevance and validity

What do patients/caregivers want? Can it be reliably measured?

Sensitivity to effect of treatment

Is the endpoint affected by THE expected mechanism of action?

Statistical efficiency in endpoint evaluation

How variable is measurement? What are the size and duration of trial needed?

https://www.fda.gov/drugs/development-approval-process-drugs/cder-patient-focused-drug-development

International Council for Harmonisation E9 Guideline: Statistical Principles for Clinical Trials (1998).

Clinical Endpoint Selection

- Direct measures of how a patient feels, functions, or survives
  - "Feel" for a neonate or infant is challenging to assess (consider developing observer-reported outcomes)
  - Functioning of a neonate, infant or child (consider feeding, sleep, developmental milestones, medical interventions)
  - Survive (consider relatedness to underlying condition)
- Can present challenges due to:
  - Rare events
  - Need for large studies of prolonged duration
  - Lack of precision in measurement
  - Lack of validated tools for the population

Source: *Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products—Guidance for Industry*. U.S. Dept of HHS, Food and Drug Administration, December 2019.

Clinical Outcome Assessments (COAs)

- Measures that describe or reflect how a patient feels, functions or survives
  - Patient-reported outcome (PRO) measures
  - Observer-reported outcome (ObsRO) measures
  - Clinician-reported outcome (ClinRO) measures
  - Performance outcome (PerfO) measures
- FDA reviews COAs either as a part a drug development program or via the CDER COA Qualification Program
- Qualification is a regulatory conclusion that the FDA finds the COA to be a well-defined and reliable assessment of patients' symptoms, functions, or mental state

https://www.fda.gov/about-fda/clinical-outcome-assessment-coa-frequently-asked-questions

Patient (& Caregiver) Experience Data

- Patient-focused drug development (PFDD) is a systematic approach to capture and incorporate patient experiences, needs, and priorities
- PFDD meetings are conducted with patient organizations, and can be FDA-led or externally-led
- These meetings target disease areas with
  - Identified need for patient (caregiver) input
  - Chronic conditions that affect functioning and activities of daily living
  - Aspects of disease are not formally captured in clinical trials
  - Currently few or no therapies, or available therapies do not directly affect clinical endpoints
  - Severe impact on identifiable subpopulations
- Meeting summary reports, called "Voice of the Patient," are generated

https://www.fda.gov/drugs/development-approval-process-drugs/cder-patient-focused-drug-development

Surrogate Endpoint (SE) Definitions

- A surrogate is a replacement endpoint that does not directly measure how a patient feels, functions, or survives
  - Examples: laboratory measures, imaging, physical signs
  - A drug's effect on the surrogate should **reliably predict** direct clinical benefit (requires clinical, epidemiologic, and scientific evidence)
  - SE's can be characterized by the level of clinical validation
- A biomarker – *a defined characteristic, objectively measured as an indicator of normal biological processes, pathologic processes, or response to an intervention* – can be used as an SE

BEST (Biomarkers, EndpointS, and other Tools) Resource - NCBI Bookshelf (nih.gov)

Types of Surrogate Endpoints

Candidate SE	Reasonably Likely SE	Validated SE
• Still under evaluation for its ability to predict clinical benefit	• Effect on SE expected to be correlated with clinical benefit; Supported by strong mechanistic and epidemiologic rationale; Could be used in accelerated approval	• Strong evidence that effect on SE predicts specific clinical benefit; Supported by mechanistic rationale and clinical data; Can be used to support traditional approval

BEST (Biomarkers, EndpointS, and other Tools) Resource - NCBI Bookshelf (nih.gov)
Surrogate Endpoint Resources for Drug and Biologic Development | FDA

Some Potential Surrogate Endpoint Issues

- Treatment impact on an indirect measure establishes biological activity but not necessarily direct clinical benefit
- Correlation between biomarker and clinical endpoint is necessary but not sufficient to validate the biomarker as an SE
- Biomarkers can be helpful for prognosis or diagnosis, but may not be reliable SEs
- SE may not lie on the causal pathway
- There can be uncertainty about magnitude of surrogate effect that corresponds to clinical benefit

Source: Fleming, T. R., & Powers, J. H. (2012). Biomarkers and surrogate endpoints in clinical trials. *Statistics in Medicine*, *31*(25), 2973-2984.

What are the Next Steps?

- Today's Discussions
  - Session 1: Clinical trialists past experiences and current perspectives
  - Session 2: Clinical endpoint development experiences in several neonatal conditions
  - Session 3: Panelists with diverse perspectives will discuss their important factors to consider
  - Session 4: Presentations and discussion of newer approaches to measuring clinical benefit
- **Share your experiences, perspectives, and ideas today**
- Consider engaging with public-private partnerships (like International Neonatal Consortium/C-Path) to collaborate, regulatory agencies (via PFDD or drug-development tools), and other organizations/individuals involved in the workshop

Thank you!


FDA
U.S. FOOD & DRUG
ADMINISTRATION

Barbara Schmidt

McMaster University & University of Pennsylvania

Barbara Schmidt, MD, MSc, CM

Death or Disability

A valid primary outcome for selected research questions in neonatal RCTs

Barbara Schmidt
has no financial relationships to disclose

I was the lead investigator of **TIPP** (Trial of Indomethacin Prophylaxis in Preterms), **CAP** (Caffeine for Apnea of Prematurity) trial, **COT** (Canadian Oxygen Trial), site PI in Neonatal Research Network (**NRN**) 2011-17, and member of **DSMBs** (incl. MFMU Network, MOMS, TRIGR)

Questions

1. What are the pros and cons of the primary outcome in TIPP, CAP and COT:
   Death or Disability?
2. How do composite outcomes in neonatal and adult RCTs differ?
3. Which therapies with short-term benefits require long-term follow-up?

Composite Outcome: Pros

- Avoids arbitrary choice of single outcome when several are similarly important
- Reduces problems of multiple testing
- Accounts for competing risks
- Increases trial efficiency
- Estimates “net clinical benefit” if efficacy and safety outcomes are combined

Composite Outcome: Cons

Published critiques focus on

- flaws in choice of components
- incomplete reporting of results
- misinterpretation by consumers

These problems are preventable!

N Engl J Med 2007;357:1893

Table 2. Primary Outcome of Death or Neurodevelopmental Disability.

Outcome	Caffeine Group	Placebo Group	Unadjusted Odds Ratio	Odds Ratio Adjusted for Center (95% CI)	P Value
	no./total no. (%)				
Composite					
Death or disability	377/937 (40.2)	431/932 (46.2)	0.78	0.77 (0.64–0.93)	0.008

N Engl J Med 2007;357:1893

Table 2. Primary Outcome of Death or Neurodevelopmental Disability.

Outcome	Caffeine Group	Placebo Group	Unadjusted Odds Ratio	Odds Ratio Adjusted for Center (95% CI)	P Value
	no./total no. (%)				
Composite					
Death or disability	377/937 (40.2)	431/932 (46.2)	0.78	0.77 (0.64–0.93)	0.008
Components					
Death before 18 mo†	62/974 (6.4)	63/970 (6.5)	0.98	0.97 (0.67–1.40)	0.87
Cerebral palsy‡	40/909 (4.4)	66/901 (7.3)	0.58	0.58 (0.39–0.87)	0.009
Cognitive delay‡§	293/867 (33.8)	329/858 (38.3)	0.82	0.81 (0.66–0.99)	0.04
Severe hearing loss‡¶	17/909 (1.9)	22/905 (2.4)	0.77	0.77 (0.40–1.45)‖	0.41
Bilateral blindness‡	6/911 (0.7)	8/905 (0.9)	0.74	0.74 (0.26–2.15)‖	0.58

Are these outcomes comparable?

Robert Reid

Robert Lenkiewicz

Are these outcomes comparable?

- Judgements about the importance of death vs disability are subjective and differ among clinicians as well as among parents/patients.
- Outcome selection should occur ”at the coal face” but always involves compromises.
- There are wrong approaches to outcome choice but there is not one right approach.

N Engl J Med 2014;371:140-9.

Conversation with D.A. ca 10 years ago:
"We did not foresee this problem:
IQ is the primary outcome but quite a few children cannot be tested"

ORIGINAL ARTICLE

Effects of Hypothermia for Perinatal Asphyxia on Childhood Outcomes

Denis Azzopardi, M.D., Brenda Strohm, R.N., Neil Marlow, D.M.,
Peter Brocklehurst, F.F.P.H., Aniko Deierl, M.D., Ph.D., Ova Eddama, Ph.D.,

Primary outcome at 6 yrs: Survival with IQ ≥ 85
37 of 184 survivors (20%)
"were unable to complete the (IQ) test because of physical impairment..."

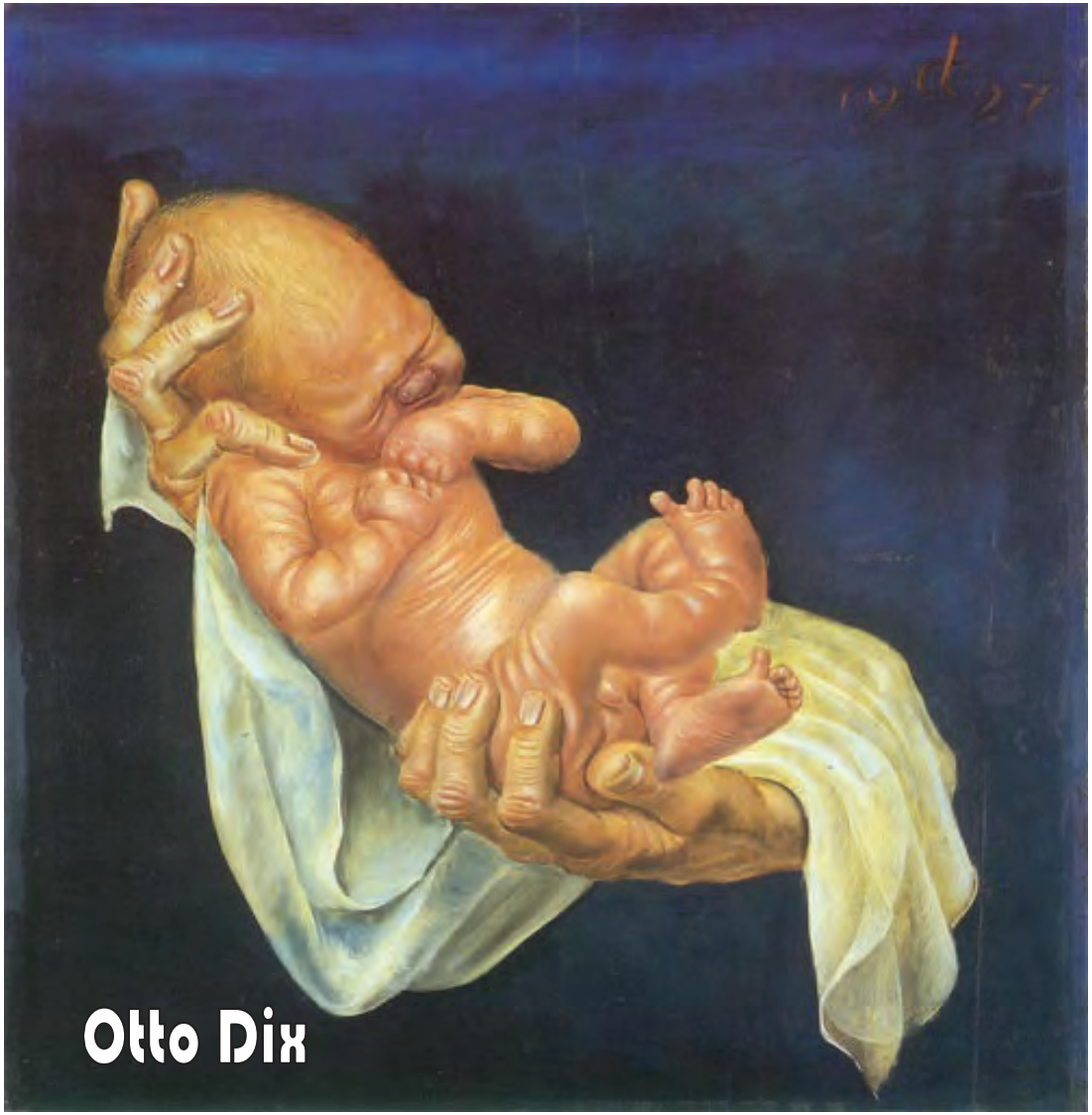
ORIGINAL ARTICLE

Effects of Hypothermia for Perinatal Asphyxia on Childhood Outcomes

Denis Azzopardi, M.D., Brenda Strohm, R.N., Neil Marlow, D.M.,
Peter Brocklehurst, F.F.P.H., Aniko Deierl, M.D., Ph.D., Ova Eddama, Ph.D.,

Questions

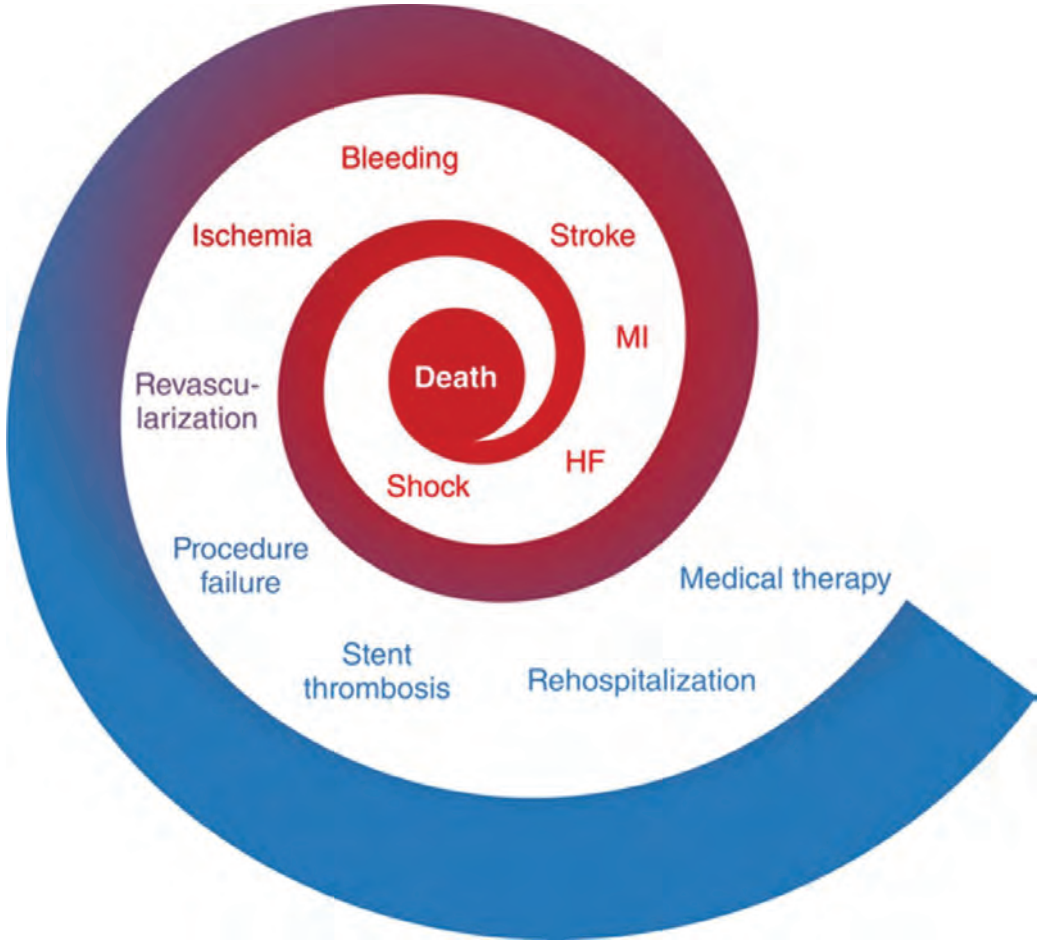
1. What are the pros and cons of the primary outcome in TIPP, CAP and COT: Death or Disability?
2. How do composite outcomes in neonatal and adult RCTs differ?
3. Which therapies with short-term benefits require long-term follow-up?


Otto Dix


Alice Neel

Components of composite outcomes in cardiovascular trials of adults.

Analysis:
Time to first event (Survival analysis)



Paul W. Armstrong. Circulation. Composite End Points in Clinical Research, Volume: 135, Issue: 23, Pages: 2299-2307, DOI: (10.1161/CIRCULATIONAHA.117.026229)

EUROPEAN SOCIETY OF CARDIOLOGY®

European Heart Journal (2012) **33**, 176–182
doi:10.1093/eurheartj/ehr352

SPECIAL ARTICLE

The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities

Stuart J. Pocock*, Cono A. Ariti, Timothy J. Collier, and Duolao Wang

Department of Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

Methods designed for survival analysis of composite outcomes in adult trials are not suitable for the outcome of “disability”

- Adult therapies are intended to prevent or delay
  PROGRESSION OF DISEASE
- Disability is a non-progressive endpoint of abnormal development

Questions

1. What are the pros and cons of the primary outcome in TIPP, CAP and COT: Death or Disability?
2. How do composite outcomes in neonatal and adult RCTs differ?
3. Which therapies with short-term benefits require long-term follow-up?

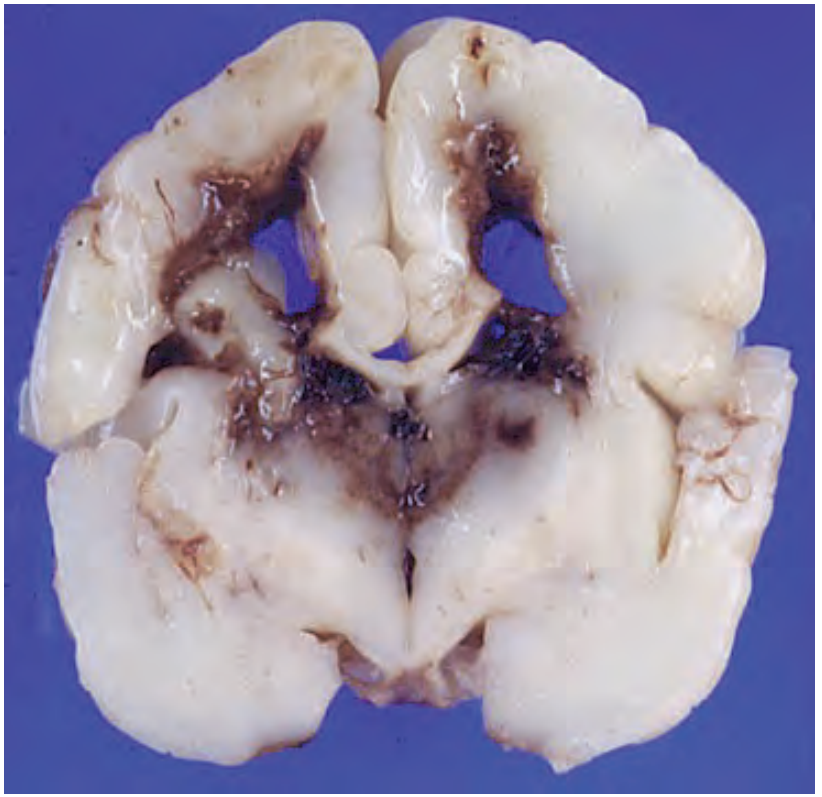
Trial of Indomethacin Prophylaxis in Preterm Infants (TIPP)

N Engl J Med 2001;344:1966-72

When we designed TIPP, it was known that:

- Indomethacin prophylaxis reduces severe IVH
- Treatment effect size is modest
- Severe IVH is quite rare

Why did we perform TIPP?



Reason for long-term primary TIPP outcome

The New England Journal of Medicine

LONG-TERM EFFECTS OF INDOMETHACIN PROPHYLAXIS IN EXTREMELY-LOW-BIRTH-WEIGHT INFANTS

BARBARA SCHMIDT, M.D., PETER DAVIS, M.D., DIANE MODDEMANN, M.D., ARNE OHLSSON, M.D., ROBIN S. ROBERTS, M.Sc., SAROJ SAIGAL, M.D., ALFONSO SOLIMANO, M.D., MICHAEL VINCER, M.D., AND LINDA L. WRIGHT, M.D., FOR THE TRIAL OF INDOMETHACIN PROPHYLAXIS IN PRETERMS INVESTIGATORS*

ABSTRACT

Background The prophylactic administration of indomethacin reduces the frequency of patent ductus arteriosus and severe intraventricular hemorrhage in very-low-birth-weight infants (those with birth weights below 1500 g). Whether prophylaxis with indomethacin confers any long-term benefits that outweigh the risks of drug-induced reductions in renal, intestinal, and cerebral blood flow is not known.

THE prophylactic administration of in methacin reduces the incidence of pat ductus arteriosus and severe intraventri lar hemorrhage in very-low-birth-weight fants (those with birth weights below 1500 g).[1] C current understanding of the mechanisms by wh indomethacin prevents intraventricular hemorrh is speculative[2] and indicates that a decrease in ce bral perfusion may be involved.[3,4] Although such a

Reason for long-term primary TIPP outcome

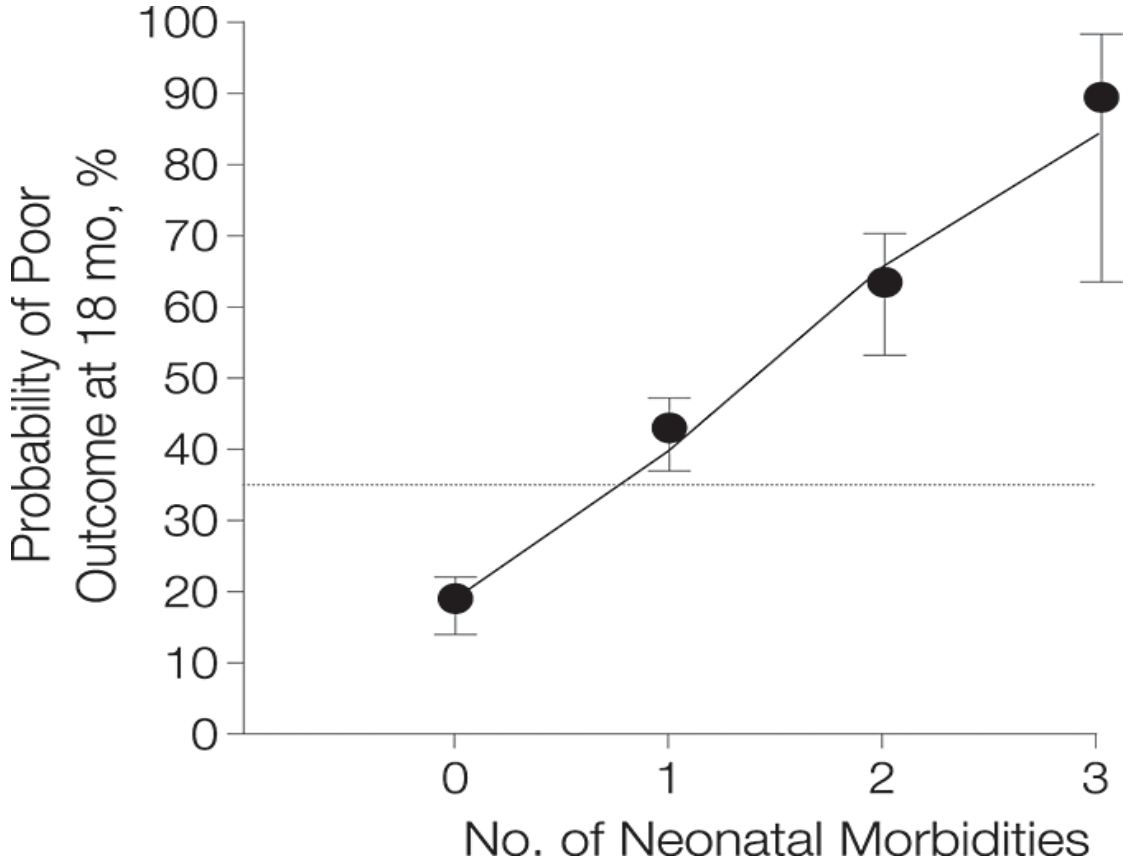
The New England Journal of Medicine

LONG-TERM EFFECTS OF INDOMETHACIN PROPHYLAXIS

SAFETY

Whether prophylaxis with indomethacin confers any long-term benefits that outweigh the risks of drug-induced reductions in renal, intestinal, and cerebral blood flow is not known.

Probability of Poor 18-Month Outcome in Study Infants (N = 910) With None, 1, 2, and All 3 Neonatal Morbidities



Schmidt, B. et al. JAMA 2003;289:1124-1129.

JAMA

Conclusions

1. Death or disability is a valid outcome for selected neonatal research questions.
2. Recommendations for design and analysis of composite outcomes in adult medicine may not apply to neonatal RCTs.
3. Common neonatal therapies with short-term benefits require long-term follow-up if long-term safety is in doubt.

Thank you!

Erik Jensen

Kristi Watterberg

University of New Mexico

Challenges and Opportunities: Measuring Benefit in Neonatal Randomized Clinical Trials

the NRN Hydrocortisone for BPD trial

Kristi Watterberg, MD

Professor Emerita of Pediatrics

University of New Mexico Health Sciences Center

Conundrum: how to evaluate both safety and efficacy in the primary outcome

- Planning for the Neonatal Research Network Hydrocortisone (HC) to decrease BPD study started around 2010
- Efficacy to be assessed short term: death/BPD at 36 weeks EGA
- But safety assessed at 2 years: NDI, and specifically cerebral palsy
- How could we put those together - and avoid repeating the dexamethasone story?
- To review a little of that history….

The dexamethasone story: A therapeutic misadventure in neonatology

- Animal studies since the 1960s showed high doses of glucocorticoids caused growth restriction in all organ systems, including brain.
- More immature animals are more susceptible.
- Dex has ~25 – 40x potency of HC.
- Yet, early studies of dexamethasone in preterm infants used high doses, commonly 0.5mg/kg/day (0.5mg =~12.5 – 20mg HC)

The dexamethasone story

- Abstracts: early anecdotes

  ('78, '80, '81)

- 'hastened weaning from ventilator' (3-day crossover, 1 month taper)

  (Mammel, 6 infants; Lancet, 1983)

- 'striking short-term improvement' (tapered over a month)

  (Avery, 16 infants; Pediatrics, 1985)

- faster weaning from IMV and O_2 (42-day tapering course)

  (Cummings, 36 infants; NEJM 1989)

- "We chose dexamethasone because of its nearly complete glucocorticoid activity and its long half-life, and because there is reasonable experience with its use in neonates and infants"

- "However, treatment cannot be recommended without further study of patient selection, dosage schedules, short and long-term side effects, and the mechanisms of its actions."

--Mammel, Lancet 1983

Studies continued, but . . .

The therapy was also adopted in clinical practice

- High dose – 0.5mg/kg/day
- Long-term – commonly a 42-day tapering course
- Starting earlier and earlier in life, until…

Babies were treated in the first postnatal week

- Rx works
  - 28 days, tapered from 0.5mg/kg/day (Yeh , Pediatrics 100:(4)E3,1997)
- Or it doesn't
  - 2 doses of 0.5mg/kg q 12 hours (Sinkin, Pediatrics 105:542, 2000)
- Rx works, but dose ↓ for GI perforation
  - 3 days tapered from 0.4mg/kg/dose (Garland, Pediatrics, 104:91, 1999)
- Study stopped for lack of efficacy and/or safety concerns
  - 12 days tapered from 0.5mg/kg/day (Vt-Ox, Pediatrics 108:741, 2001)
  - 10 days tapered from 0.15mg/kg/day (NICHD, NEJM 344:95, 2001)

School-age outcomes after dexamethasone

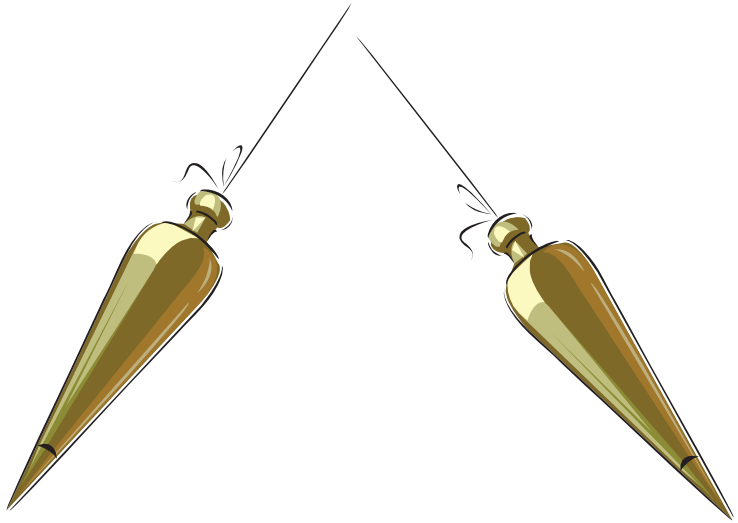
- RCT of Dex, 0.5mg/kg/day tapered over 28 d
- 146 of 159 survivors seen at age 8
- Treated children were shorter, and had:
  - Smaller head circumference
  - Lower IQ
  - More clinically significant disabilities
- “substantial adverse effects on neuromotor and cognitive function at school age”

- Yeh et al, N Engl J Med 2004;350:1304

Dexamethasone and neurologic outcome

- Dex is a risk factor for MDI <70 & abnormal neurologic exam
  - Follow up of > 1100 ELBW infants, cohort study
    - NICHD Neonatal Network, Pediatrics 105:1216, 2000
- Dex is associated with ↑CP & neurologic impairment
  - Meta-analysis of >1000 patients in RCTs
  - Follow-up assessed for 679 of those patients
  - NDI relative risk 1.34; CP relative risk of 2.02
    - Barrington KJ, BMC Pediatrics 1:1, 2001
- With that indictment of high-dose dex. . .

The pendulum swung…

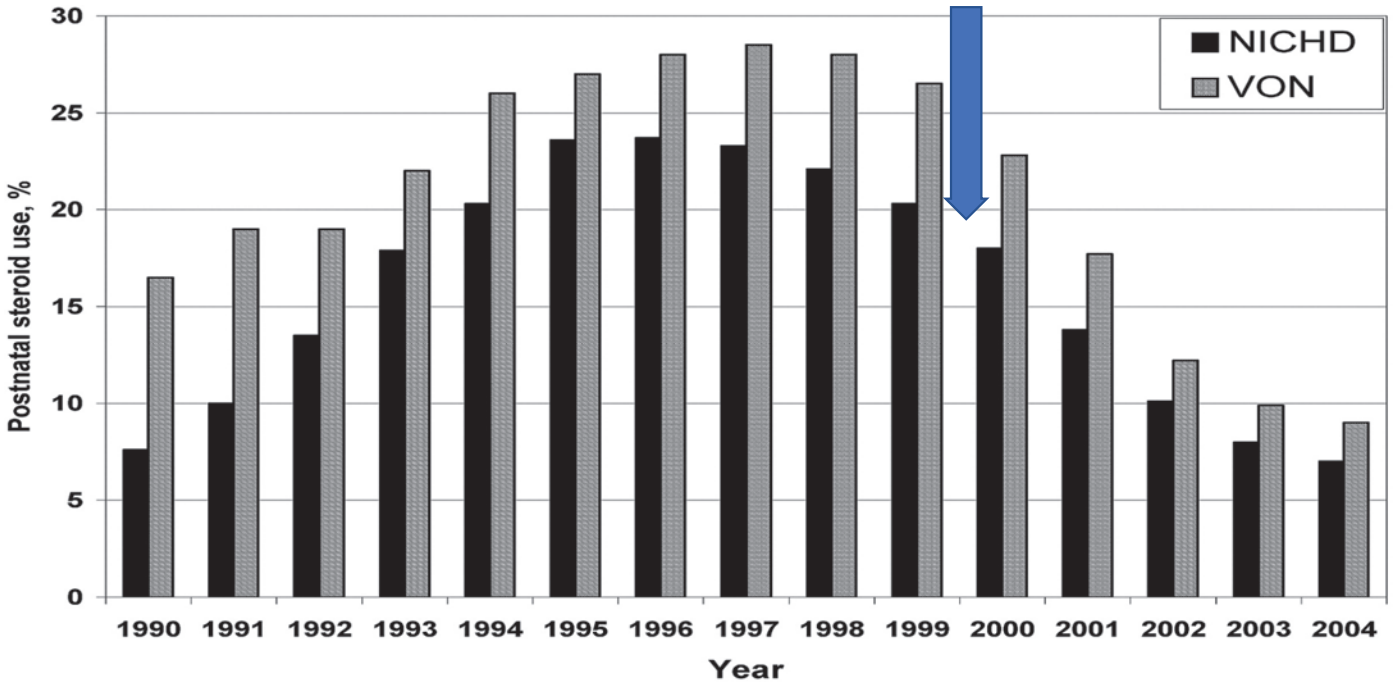


Dex is good – the more the better!

All steroids are bad – no baby should get them!

From: **Changes in the Use of Postnatal Steroids for Bronchopulmonary Dysplasia in 3 Large Neonatal Networks**

Pediatrics. 2006;118(5):e1328-e1335. doi:10.1542/peds.2006-0359



Caught in the middle: the DART trial 2000-2002

- Planned sample size 800; postnatal age >1 week
- Primary outcome: “assessment of the effects of low-dose dexamethasone (0.15mg/kg/day) on long-term rates of survival free of major neurologic disability”.
- Enrollment stopped at 70 because of other reports of adverse effects
- Right question – was the therapy safe?
- Right study – large sample size
- Wrong timing –

- Doyle et al, Pediatrics 2006;117;75
-

So, how to structure the NRN HC trial?

- Hypothesis: if hydrocortisone can improve survival without moderate or severe BPD, it will also result in improved survival without NDI
  - BPD is a risk factor for mortality and adverse neurodevelopmental outcomes.
- But a sample size of 800 not expected to show statistically significant benefit.
- Therefore, we will consider this outcome successful if either:
  - death/NDI is lower on the HC arm,
  - or there is an increase in death/NDI in the HC arm, but a one-sided 95% confidence interval for benefit (death/BPD) vs. risk (NDI) is >4; i.e., for every additional 4 infants surviving without BPD, we would have 95% confidence that no more than 1 additional infant would experience death/NDI.

So, how to structure the NRN HC trial?

- Because the primary outcome includes evaluation at 18 – 22 months, earlier study outcomes will not be reported (unless the DSMC stops enrollment for benefit or harm, or after all subjects have completed treatment there is a significant mortality benefit favoring HC ($p<0.001$).

Study outcomes

- Infants enrolled from 2011 – 2018; follow-up ended in 2020.
- Survival without BPD: 16.6% of HC group, 13.2% of placebo.
- Survival without mod/severe NDI: 36.9% of HC, 37.3% of placebo.
- Moderate/severe CP: 12% of HC-treated, 10% of placebo
- More HC-treated infants extubated during the study period
  - 44.7% vs 33.6%.
- HC-treated infants averaged 3 fewer days of mechanical ventilation.

How long is long enough?

- 2-year outcomes correlate only weakly with school-age outcomes
- What outcomes are important to parents . . . and patients?

Back to basics

- What is/are the right outcome(s) for a study intended to decrease chronic lung disease in preterm infants?
  - Does outcome at 36 weeks predict future lung function?
  - Functional outcomes at 2 years? 5 years? Adulthood?
- How do we assess what's "important"?
  - Ask the parents?
  - Ask the babies at some future point?
  - Everyone has their own point of view
    - "I was right to go on with treatment"
    - "I was right to stop"
    - "My life is worth living"
    - "My life is not worth living"

Implications for practice

- Use of dexamethasone has recently been rising
- Recent “network analysis” concluded that moderately early, moderately high dose Dex is most effective.
  - Ramaswamy et al, JAMA 2021; 75(6):e206826
- HOWEVER: no large RCTs of dex vs. placebo since early 2000s
- Most recent Cochrane statement: “This review supports . . . late systemic corticosteroids for infants who cannot be weaned from mechanical ventilation. Longer-term follow-up into late childhood is vital for assessment of important outcomes . . . such as effects . . on higher-order neurological functions, . . . and lung function. Further RCTs of late systemic corticosteroids should include longer-term survival free of neurodevelopmental disability as the primary outcome
  - Doyle et al, Cochrane Database 2021: CD001145.

Thanks for your attention!

Questions/comments

Audience Q&A

Submit questions and comments via Slido

Join at
slido.com
#Neonatal

Session 1: Current Approaches to Measuring Efficacy in Neonatal Randomized Control Trials

Moderator: Michele Walsh, National Institute of Child Health and Human Development

Session 2: Challenges in Measuring Efficacy for Neonatal Conditions with Unmet Clinical Needs

Moderator: An Massaro, U.S. Food & Drug Administration

Session 2: Objectives

Objectives:

- Highlight the challenges and considerations for developing core outcome sets for neonatal research and choosing appropriate primary endpoints for regulated trials.
- Review potential efficacy endpoints related to key neonatal conditions, such as bronchopulmonary dysplasia (BPD), neonatal seizures, neonatal opioid withdrawal syndrome (NOWS), and pain.
- Discuss best practices and key solutions for generating high-quality evidence for these conditions with high unmet clinical needs.

Session 2: Presenters

- **Kanecia Zimmerman,** Duke University School of Medicine
- **Erik Jensen,** Children's Hospital of Pennsylvania
- **Janet Soul,** Boston Children's Hospital
- **Martin Offringa,** University of Toronto

Audience Q&A

Submit questions and comments via Slido

Join at
slido.com
#Neonatal

Kanecia Zimmerman

Duke University School of Medicine

The perennial challenge of measuring the efficacy of ACUTE PAIN THERAPEUTICS in infants and young children

Kanecia Zimmerman, MD PhD MPH

March 23, 2023

COA-APTIC

Goals

I. Why is it so challenging to measure pain in this population?

II. COA-APTIC Objectives and Methods

III. COA-APTIC Findings to Date

IV. Challenges and Next Steps

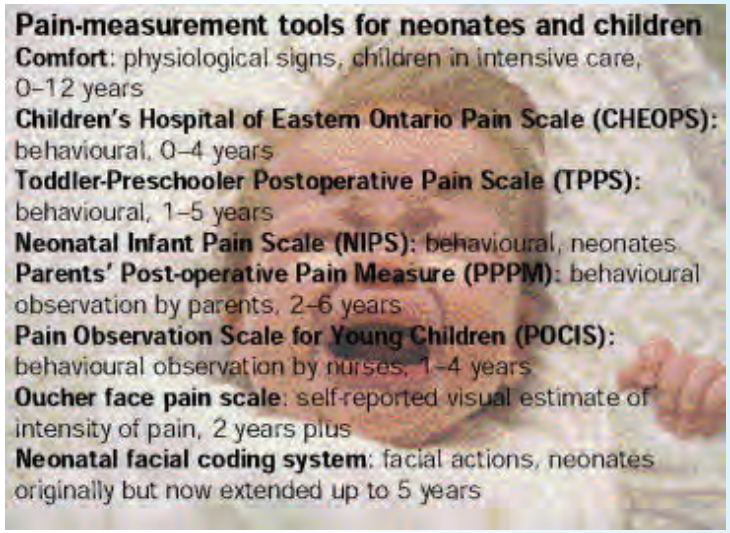
COA-APTIC

Pain…in anyone

- Many concepts
  - Pain intensity
  - Pain interference
  - Pain experience
- Subjective
- Malalignment in perceptions of pain intensity
  - Age, sex, race/ethnicity
- Pain experience intertwined with other concepts
  - Anxiety, fear, distress
- Self-report of pain intensity is GOLD STANDARD IN ADULTS

Pain in infants and young children

- Cannot rely on self report
- Focus on Clinician-Reported Outcome (ClinRO) Measure – completed by clinicians
- How do we know its “pain” versus some other “distress”?



COA-APTIC

COA-APTIC: Clinical Outcome Assessments for Acute Pain Therapeutics in Infants and young Children

Identify or develop core sets of high-quality Clinical Outcome Assessments and endpoints to assess acute pain in clinical trials of pain therapeutics in infants and young children (0 – 3 years).

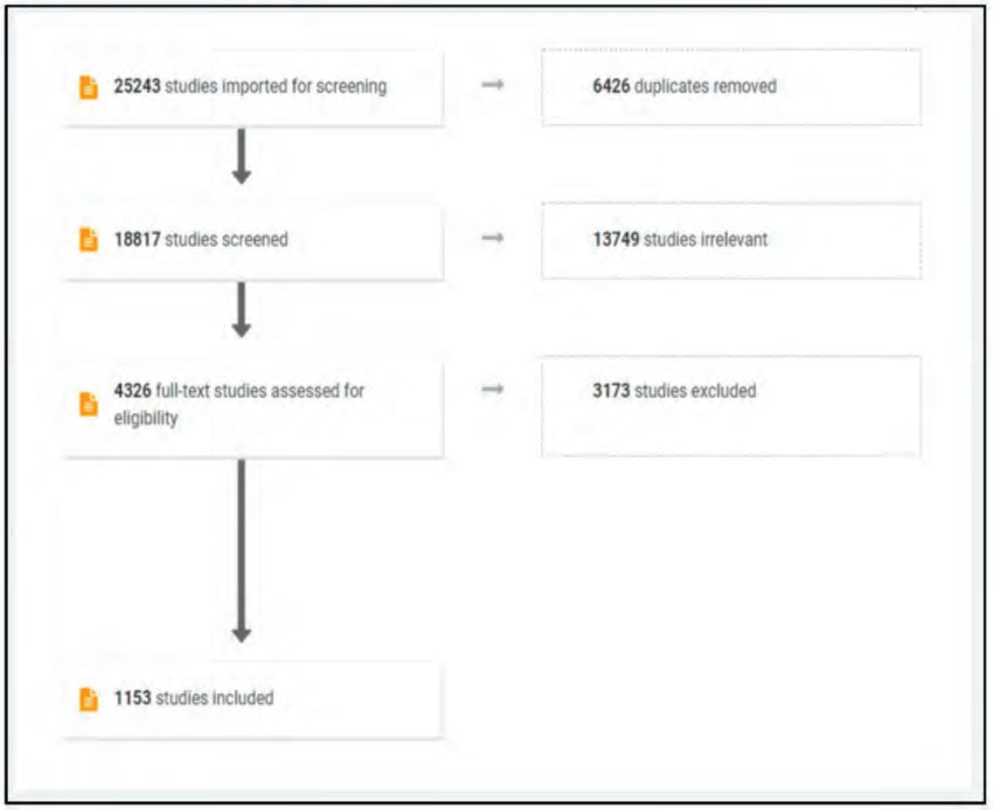
COA-APTIC

Two Phases of Project

- UG3 Phase – Planning Phase
  - Literature reviews
    - What COAs and endpoints currently exist to measure acute pain in pediatric trials?
    - What validity evidence exists for these COAs?
  - Indepth Qualitative Interviews for concept elicitation:
    - Pediatric Clinicians
    - Caregivers
  - Design Studies for UH3 Phase
- UH3 Phase – Implementation Phase
  - Carry-out both qualitative and quantitative studies to validate COAs and endpoints for acute pain

Literature reviews of RCTs

Figure 1. Flow Diagram



COA-APTIC

Literature reviews of RCTs

- Data extraction to identify the following:
  - general information for the entry citation
  - General study information (e.g. age, sample size, acute pain source, quality of study)
  - information on the pain relief intervention(s) and assessment of pain
  - specific information on the clinical outcome assessments (e.g., name of COA, frequency of collection, statistical sig)

Literature review of RCTs – key findings

- The use of specific COAs is heterogeneous and often vaguely reported in RCTs for this age group.
- There are over 83 types of ClinRO measures that have been used, and their administration methods vary
- Caregiver-reported outcome measures are used less frequently than clinician-reported outcome measures
- The quality of the eligible studies varied, with race and ethnicity of the child participants in the RCTs being reported in less than 7% of entries.

Literature review of COA validity

- What are the existing ClinRO and ObsRO measures of acute pain that have some validation evidence in infants and young children between birth and < 3 years of age?
- What pain indicators are included in the COAs?
- What is the type of validity evidence for each identified COA?
- Which COAs appear to have the strongest evidence for use in pediatric clinical trials for pain therapeutics? Of those, what are their limitations and opportunities for refinement or evaluation?

Literature review of COAs - findings

- 35 ClinRO measures reviewed; one ObsRO identified (2-7yrs)
- Most common behavioral pain indicators: facial expressions, crying, and body activity or movement
- Sparse content validity
- Lack of information on how individual items are performing; limited psychometric evidence
- Very young kids are excluded from validity data
- Prior validation studies failed to include racially and ethnically diverse populations
- Many validation studies do not blind the raters to the painful event.
- Many existing COAs are highly correlated (build on each other)

Literature review of COAs - findings

- FLACC, N-PASS, EVENDOL, and CHIPPS, relative to other ClinRO measures, capture a broad range of pain indicators and include supporting psychometric evidence for infants and young children (birth to < 3 years).
- No single COA has met the threshold of a "qualified" ClinRO or ObsRO measure for regulatory purposes; All could benefit from improvements
  - Limited resources available (money and time)
  - Prioritization of goals is necessary

How do we gather information needed to move the needle?

- **Concept elicitation interviews with Clinicians**
  - open-ended questions about the participant's professional experience recognizing and managing pain in children between ages 0 and 3 years old.
  - describe pain and non-pain distress differences and similarities across three age groups: 0 to < 2 months, 2 months to < 1 year, and 1-year to < 3 years.
- **Concept elicitation interviews with Caregivers**
  - One-hour, phone based interview using open-ended questions about acute pain assessment, treatment, and response to treatment
  - Topics of Interest: Medical history and painful experiences, Pain expression, Distress expression vs pain expression, Interventions for pain or distress

Clinician interviews- Demographics

- 18/27 (67%) white, non-Hispanic
- 16/27 (59%) female
- Age range 24-67
- 15/27 (56%) Physicians
- Representation from Anesthesiology, Critical Care, Neonatology, General Pediatrics, Heme-Onc, Surgery, Family Medicine
- Recruited through Pediatric Trials Network

COA-APTIC

Clinician interviews – key findings

- movements, facial expressions, behaviors, and vital signs are important
- Some key differences by age
- comorbidities, sedation, type of pain, illness severity, and location of pain most commonly influenced pain expression
- most commonly identified types of non-pain distress: could be grouped in general themes of separation, disruption in normal routine, and related to external conditions
- Most confident in differentiating pain from non-pain distress when there was a **discernable medical reason to suspect pain**. Eliminate other things (e.g., hunger) and seek parent's impression of pain
- Vital signs and pain scale ratings helpful; increased language skills in the older age group also increased clinician confidence in differentiating pain from non-pain distress.

Clinician interviews – key findings

- Goal of intervention: decrease pain or reduce the pain score and many emphasized the importance of reducing pain behaviors, helping the patient to be calm and comfortable, and stabilizing vital signs.
- Pain intervention success was defined as resolution of pain expression behaviors and return to normal activities as well as parental confirmation.
- Most clinicians don't use pain scales: confident that they can identify pain without a pain scale using medical context, that pain scale scores are not clinically relevant and not specific enough to differentiate pain from other non-pain distress, and that pain scales are cumbersome, time consuming, and subjective.
- Ideal pain scale: **simple and quick, able to differentiate pain from non-pain distress, and objective**

Caregiver Concept Elicitation Interviews

- Caregivers identified through sites within the Pediatric Trials Network (PTN)

Characteristics	N=44
Female	86%
Married	84%
Some college education	84%
Hispanic	16%
White	75%
Black	2%
Asian	9%
Multiracial	7%
Other	7%
Child Chronic condition causing pain	41%
Child Delayed motor	41%
Child delayed cognitive development	25%

COA-APTIC

Caregiver interviews – key findings

Behaviors	N=44
Sounds	
Crying	39
Whining/wimpering	19
Facial expression	
Scrunched face	14
Eye changes or movements	14
Face color change	13
Movements	
Indicate location	14
Posture change	13
Rocking/shaking	13
Behaviors	
Irritability and emotional dysregulation	19
Clingy/seeking comfort	10
Change in eat, sleep, play	10

Caregiver interviews- key findings

- The most commonly identified indicators for differentiating pain from non-pain distress included physical behaviors (n = 26), type of crying or changes in crying (n = 21), irritability and emotional dysregulation (n = 16), and context or time of day cues (n = 16).
- Some (n = 13) also reported using a process of elimination where they attempt to address any needs such as hunger, fatigue, or need for a diaper change in effort to first rule out non-pain distress.
- Age matters; better over time
  - parent knowing the child and the child's cues better (n = 15)
  - child being better able to communicate their pain nonverbally (n = 14) or verbally (n = 12).

UH3 Phase: Caregiver ObsRO measure?

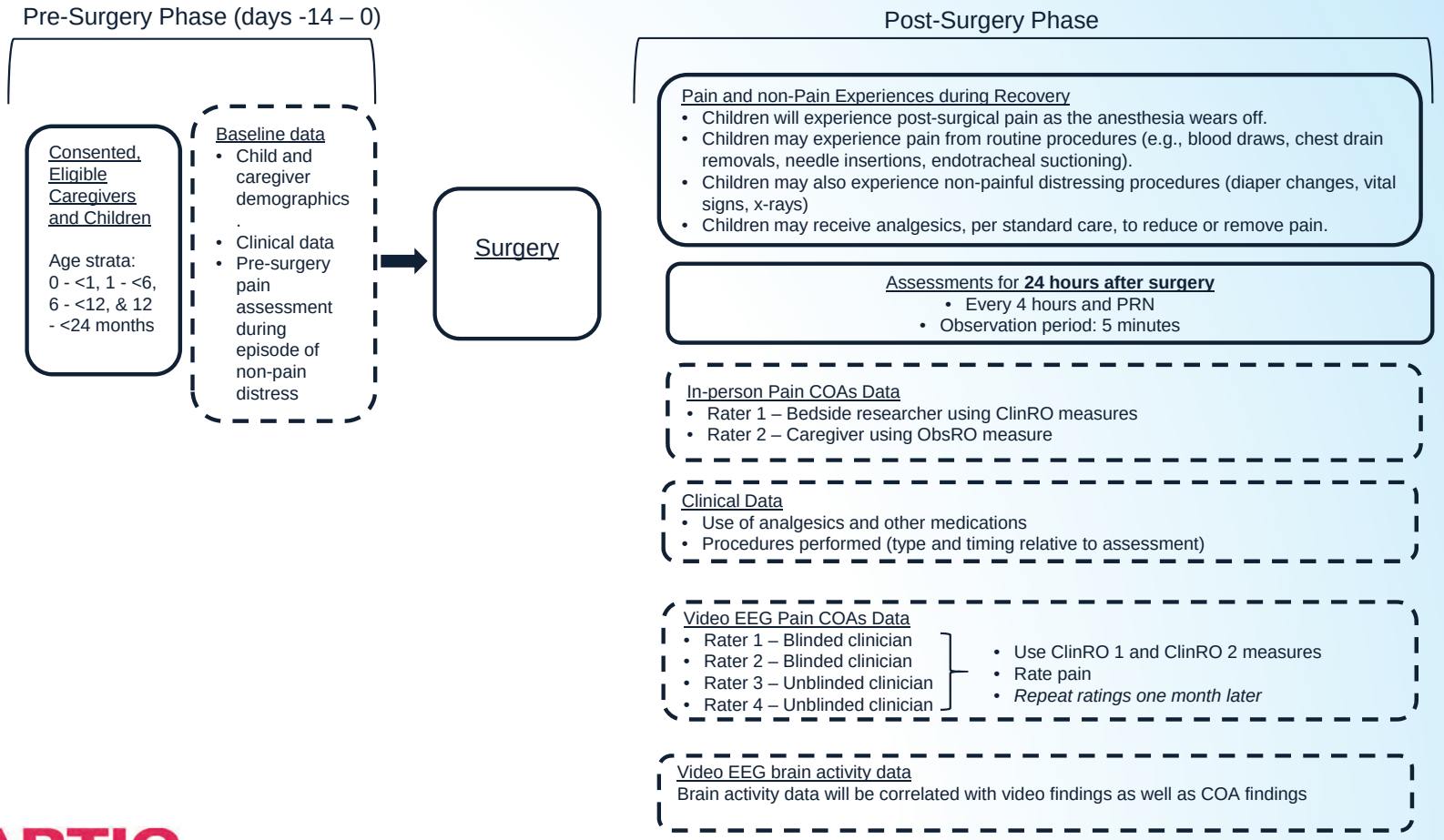
- Parents are typically the primary caregivers throughout the developmental stages of a child's life, and in particular share the responsibility with clinicians for the postoperative care for their children (Chambers et al 1996).
- Caregivers provide a unique insight into the assessment of their children's pain levels as they may be using different pain cues (Pillai Riddell et al 2008; Kappesser et al 2018). *No one knows their child better than the parents*. This can include not only the level of pain intensity, but importantly if the behaviors a child is exhibiting is related to pain or some other non-pain distress (e.g., hunger, anxiety, full diaper).
- "Parents are likewise familiar with the child's normal behavior and thus they are more able to discriminate child's pain behavior from other aberrant behavior." (Uitti et al 2018)
- Parents are also motivated to help their child and willing to complete pain questionnaires if they know it may help their child or other children with similar health issues as their child. **Not including a parent/caregiver assessment in a clinical trial feels like we are not capturing a critical source of valid pain data** (Duhn & Medves 2004).

Conclusions and Recommendations

- Lots of work to do; substantial possibility for bias in existing data
- Modify existing ClinRo measure
  - Requires permission from the developer
  - Limitation: lose ability to do historical comparison of results or rely on previous validation evidence
- Design a de novo caregiver ObsRO measure
  - !!
  - PMPP
- Methods
  - Use information gained from literature review
  - Cognitive interviews
  - Prospective psychometric evaluation (3-4)

COA-APTIC

UH3 Phase: Overview of Proposed Psychometric Study



COA-APTIC

Critical questions demonstrate remaining challenges

- Can measures differentiate pain from no pain?
- Can measures differentiate pain from non-pain distress?
- If yes, then, can they differentiate different levels of pain?

COA-APTIC

Questions?

COA-APTIC

Research Team

Kanecia Zimmerman	Principal Investigator
Bryce Reeve	Principal Investigator
Katie Gustafson	Co-Investigator
Christy Zigler	Co-Investigator
Alexy Hernandez	Clinical Research Coordinator
Kate Ehle	Clinical Research Coordinator
Laura Mkumba	Clinical Research Coordinator
Leslie Zapata Leiva	Clinical Research Specialist
Sophie Wang	Project Leader
Karylle Abella	Associate Project Leader
Christie Milleson	ANA01 Project Leader
Li Lin	Statistician
Meagan Daly	Communications
Jenny Jackman	Medical Writer

COA-APTIC

Erik Jensen

Children's Hospital of Pennsylvania

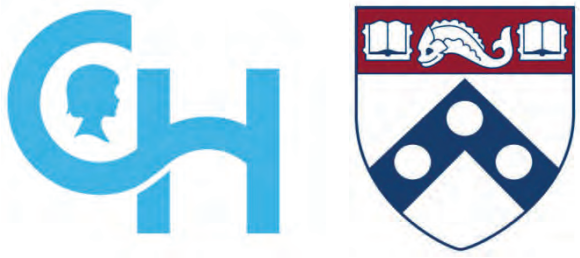
A case for why BPD is a (good) bad outcome

Measuring Clinical Benefit in Neonatal Randomized Trials

Duke-Margolis and the FDA

March 23, 2023

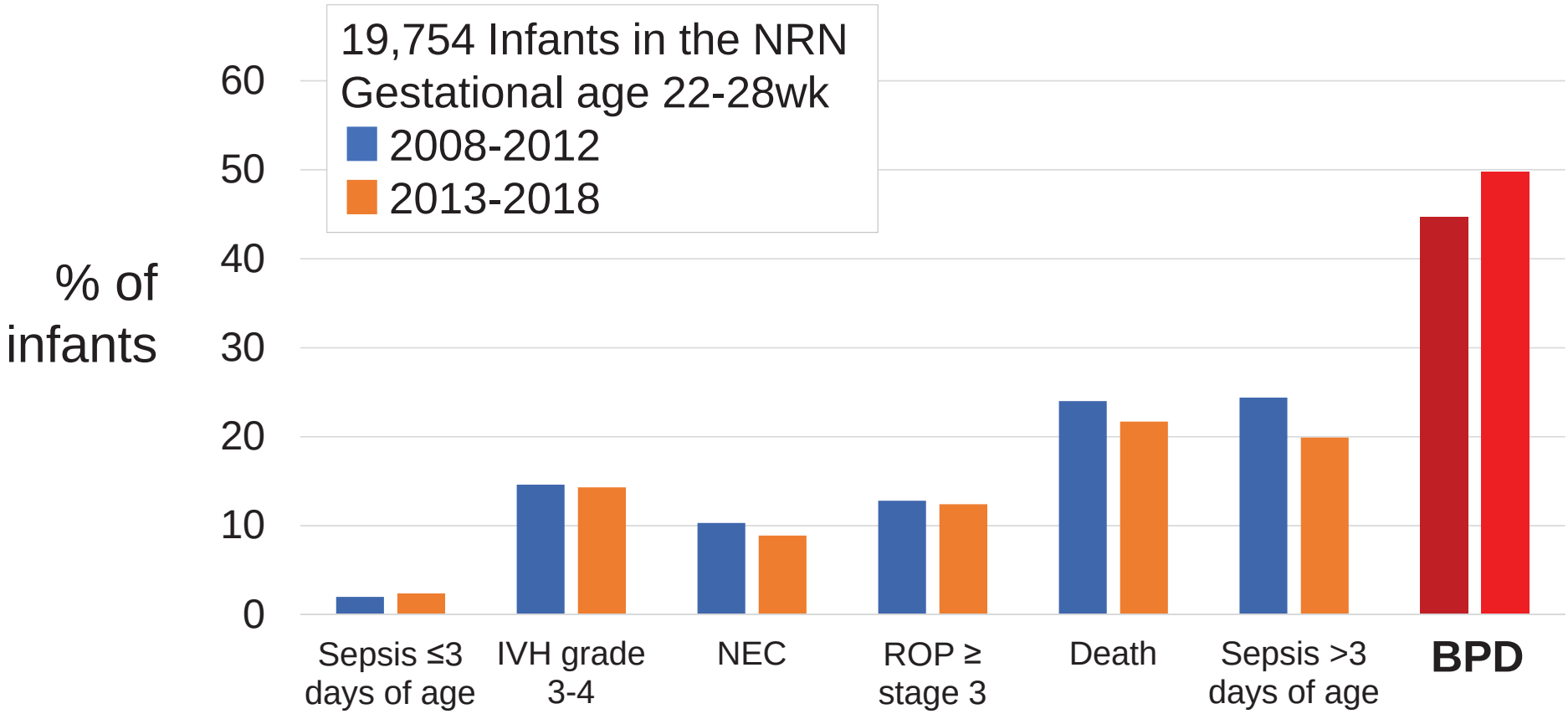
Erik A Jensen, MD, MSCE
Children's Hospital of Philadelphia
University of Pennsylvania



DISCLOSURES

I have no financial disclosures or conflicts of interest to resolve

BPD: Most common major morbidity in preterms



Bell et al. JAMA 2022

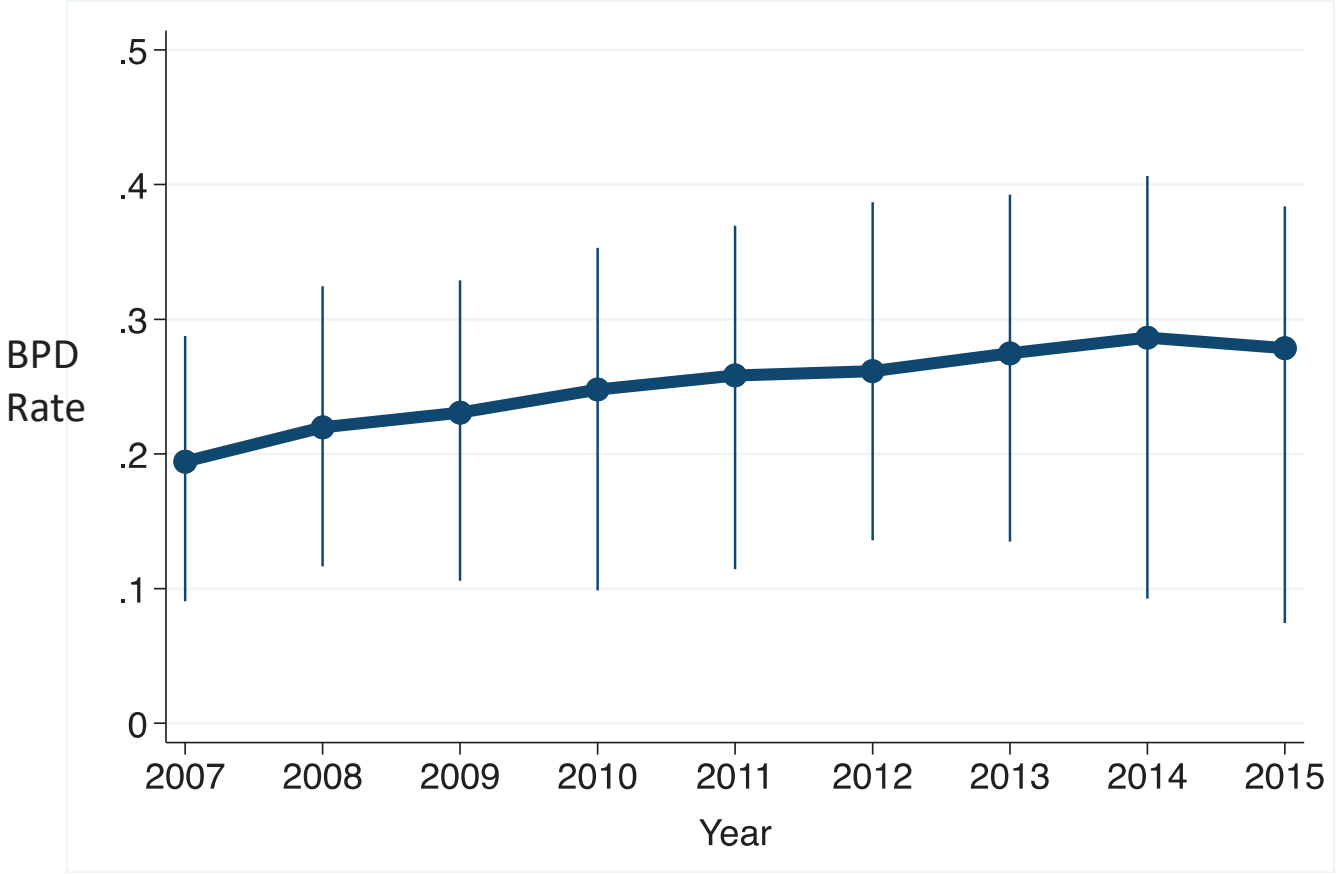
Durable benefits with BPD prevention

Caffeine for treatment of apnea of prematurity

Outcome	OR (95% CI)
BPD at 36wk PMA	0.63 (0.52-0.76)
Late death or disability at 18-21 months	0.77 (0.64-0.93)
FEV_1 z-score at 11 years	Mean improvement: 0.54 SD (0.14-0.94)

Schmidt B, et al. NEJM 2006 & 2007
Doyle LW, et al. AJRCCM 2017

BPD rates are not improving



BPD Rates

Decreased

- Canada

No change

- Finland
- Israel
- Tuscany

Increased

- Australia / New Zealand
- Japan
- Spain
- Switzerland
- Sweden
- United Kingdom

Lui K, et al. J Pediatr 2019

Diagnostic criteria for BPD remain controversial

- **Available criteria do not adequately characterize the range of lung disease severity in preterm infants**

- **Not sufficiently predictive of post-discharge outcomes**
  - Parents care more about other outcomes, not “BPD”
  - Some infants *without* BPD experience adverse post-discharge respiratory outcomes

- **Diagnosis of BPD is subjective**
  - Treatments are variable, diagnostic tests are “objective”

2001: Severity classification for BPD

NIH/ORD workshop definition

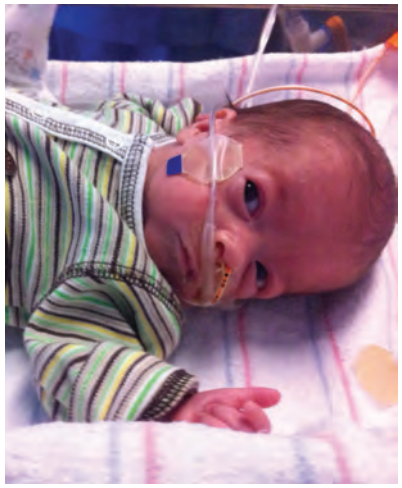
BPD Severity	Infants born <32 weeks gestation Treated with >21% O_2 for ≥28 days *plus:*
Mild	Breathing in room air at 36 weeks PMA*
Moderate	Need for <30% O_2 at 36 weeks PMA*
Severe	Need for ≥30% O_2 and/or positive airway pressure at 36 weeks PMA*

* Or discharge to home, whichever comes first

Jobe AH, et al. Am J Respir Crit Care Med. 2001

Are these cases equally “severe”?

GA = 28 weeks PMA = 36 weeks



Nasal Cannula
1L/min
30% FiO_2

?=

Nasal CPAP
6cm H_2O
30% FiO_2

?=

Invasive Ventilation
Vt 7mL/kg / PEEP 6
30% FiO_2

Pulmonary important outcomes after extremely preterm birth: Parental perspectives

NICU Outcomes

- Intubation
- Fear of death
- Steroids to prevent death
- Time on respiratory support and duration on oxygen
- Tracheostomy
- ENT problems

Post-discharge outcomes

- Home oxygen therapy
- Tracheostomy
- Hospital readmissions
- Medications
- Feeding difficulties/gastrostomy
- Exercise/school limitations

Thivierge E, et al. Acta Paediatrica 2023

Pulmonary important outcomes after extremely preterm birth: Parental perspectives

"It seems that the commonly used medical definitions of BPD are not closely associated with outcomes that are important to families.

Indeed, none of our parents spontaneously mentioned that their child had a diagnosis of BPD, nor did they ever report that their baby was on 'oxygen at 36 weeks'."

Thivierge E, et al. Acta Paediatrica 2023

NICHD
NEONATAL RESEARCH NETWORK

The Diagnosis of Bronchopulmonary Dysplasia in Very Preterm Infants: *An Evidence Based Approach*

Erik Jensen, Kevin Dysart, Marie Gantz, Scott McDonald,
Nicolas Bamat, Martin Keszler, Haresh Kirpalani, Matthew Laughon, Brenda Poindexter,
Andrea Duncan, Bradley Yoder, Eric Eichenwald, Sara DeMauro

Am J Resp Crit Care Med 2019

NIH *Eunice Kennedy Shriver* National Institute of Child Health and Human Development

Definition predictive accuracy compared

Determined which of 18 pre-specified severity-graded BPD definitions best predicts:

1. Death after 36 weeks PMA or serious respiratory morbidity at 18-26 months
2. Death after 36 weeks PMA or moderate to severe neurodevelopmental impairment

Among 2677 very preterm infants enrolled in the NICHD Neonatal Research Network

Primary Outcome: Serious Respiratory Morbidity

- Tracheostomy (any time before follow-up)
- Continued hospitalization ≥50 weeks PMA for respiratory reasons
- Use of supplemental O_2, respiratory support, or respiratory monitoring at follow-up
- ≥2 re-hospitalizations for respiratory reasons prior to 18-26 month follow-up

Definitions 1-3: Different classifications of low and high flow nasal cannula

	Treatment with >21% FiO_2 for ≥28 days and the following respiratory support at 36 weeks PMA*:							
	Room Air (no support)	NC ≤ 2L/min "low" flow		NC > 2L/min "high" flow		nCPAP NIPPV	Invasive PPV	
		FiO_2 <30%	FiO_2 ≥30%	FiO_2 <30%	FiO_2 ≥30%	Any FiO_2	FiO_2 <30%	FiO_2 ≥30%
1	Grade 1	Grade 2	Grade 3	Grade 3	Grade 3	Grade 3	Grade 3	Grade 3
2	Grade 1	Grade 2	Grade 2	Grade 3	Grade 3	Grade 3	Grade 3	Grade 3
3	Grade 1	Grade 2	Grade 3	Grade 2	Grade 3	Grade 3	Grade 3	Grade 3

* or discharge home if earlier

Definition 1 closely resembles the 2001 NIH consensus definition

Definitions 4-9: Separate severity level for invasive mechanical ventilation

	Treatment with >21% FiO_2 for ≥28 days and the following respiratory support at 36 weeks PMA*:							
	Room Air (no support)	NC ≤ 2L/min "low" flow		NC > 2L/min "high" flow		nCPAP NIPPV	Invasive PPV	
		FiO_2 <30%	FiO_2 ≥30%	FiO_2 <30%	FiO_2 ≥30%	Any FiO_2	FiO_2 <30%	FiO_2 ≥30%
4	Grade 1	Grade 2	Grade 3	Grade 3	Grade 3	Grade 3	Grade 4	Grade 4
5	Grade 1	Grade 2	Grade 3	Grade 3	Grade 3	Grade 3	Grade 3	Grade 4
6	Grade 1	Grade 2	Grade 2	Grade 3	Grade 3	Grade 3	Grade 4	Grade 4
7	Grade 1	Grade 2	Grade 2	Grade 3	Grade 3	Grade 3	Grade 3	Grade 4
8	Grade 1	Grade 2	Grade 3	Grade 2	Grade 3	Grade 3	Grade 4	Grade 4
9	Grade 1	Grade 2	Grade 3	Grade 2	Grade 3	Grade 3	Grade 3	Grade 4

* or discharge home if earlier

Definitions 10-18: No assessment of 28 days of O_2 use prior to 36 weeks PMA

	Treatment with the following respiratory support at 36 weeks PMA*:							
	Room Air (No support)	**NC ≤ 2L/min "low" flow**		**NC > 2L/min "high" flow**		**nCPAP NIPPV**	**Invasive PPV**	
		FiO_2 <30%	FiO_2 ≥30%	FiO_2 <30%	FiO_2 ≥30%	Any FiO_2	FiO_2 <30%	FiO_2 ≥30%
10	**No BPD**	**Grade 1**	**Grade 2**	**Grade 2**	**Grade 2**	**Grade 2**	**Grade 2**	**Grade 2**
11	**No BPD**	**Grade 1**	**Grade 1**	**Grade 2**	**Grade 2**	**Grade 2**	**Grade 2**	**Grade 2**
12	**No BPD**	**Grade 1**	**Grade 2**	**Grade 1**	**Grade 2**	**Grade 2**	**Grade 2**	**Grade 2**
13	**No BPD**	**Grade 1**	**Grade 2**	**Grade 2**	**Grade 2**	**Grade 2**	**Grade 3**	**Grade 3**
14	**No BPD**	**Grade 1**	**Grade 2**	**Grade 2**	**Grade 2**	**Grade 2**	**Grade 2**	**Grade 3**
15	**No BPD**	**Grade 1**	**Grade 1**	**Grade 2**	**Grade 2**	**Grade 2**	**Grade 3**	**Grade 3**
16	**No BPD**	**Grade 1**	**Grade 1**	**Grade 2**	**Grade 2**	**Grade 2**	**Grade 2**	**Grade 3**
17	**No BPD**	**Grade 1**	**Grade 2**	**Grade 1**	**Grade 2**	**Grade 2**	**Grade 3**	**Grade 3**
18	**No BPD**	**Grade 1**	**Grade 2**	**Grade 1**	**Grade 2**	**Grade 2**	**Grade 2**	**Grade 3**

* or discharge home if earlier

Optimal BPD Definition

Treatment with the following respiratory support at 36 weeks PMA*:			
Room Air No 28-day O_2 assessment	**Nasal cannula ≤ 2L/min**	**NC > 2L/min, nCPAP, or NIPPV**	**Invasive PPV**
	Any FiO_2	Any FiO_2	Any FiO_2
No BPD	**Grade 1**	**Grade 2**	**Grade 3**

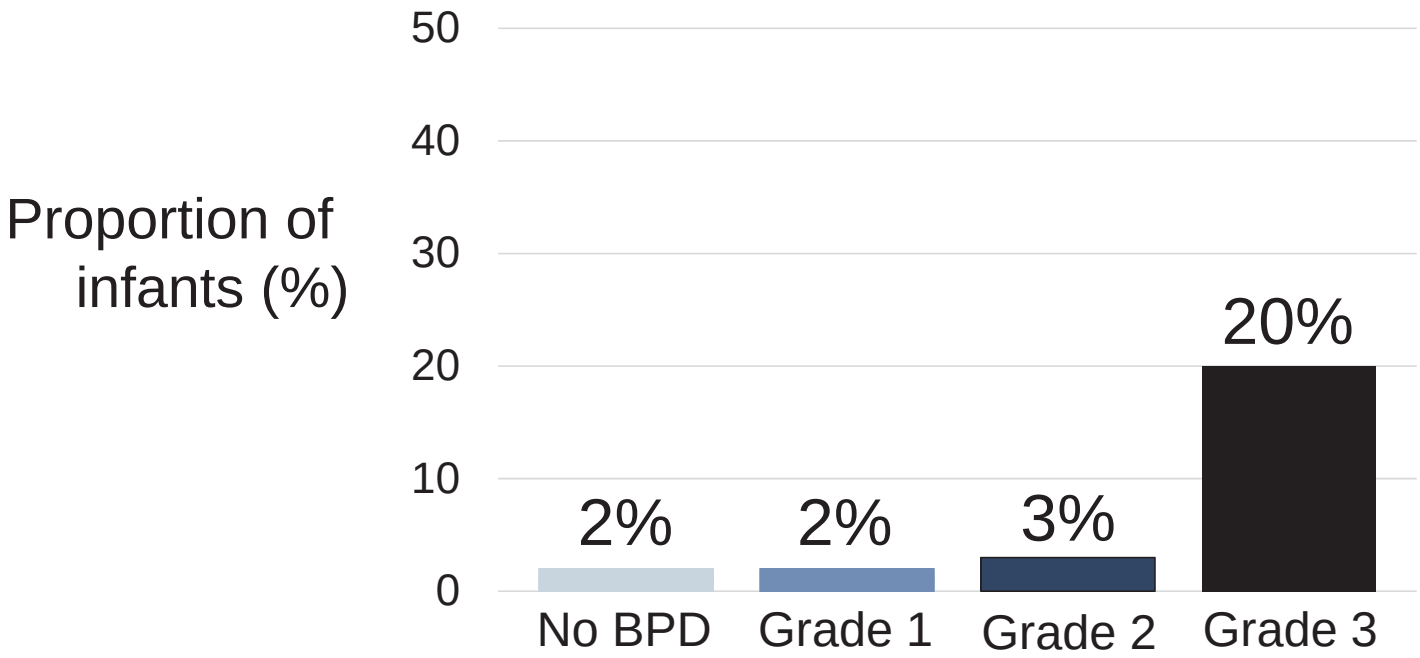
Correctly classified the presence or absence of late death or serious respiratory morbidity in **81% of study infants**

Comparison of Available BPD Definitions

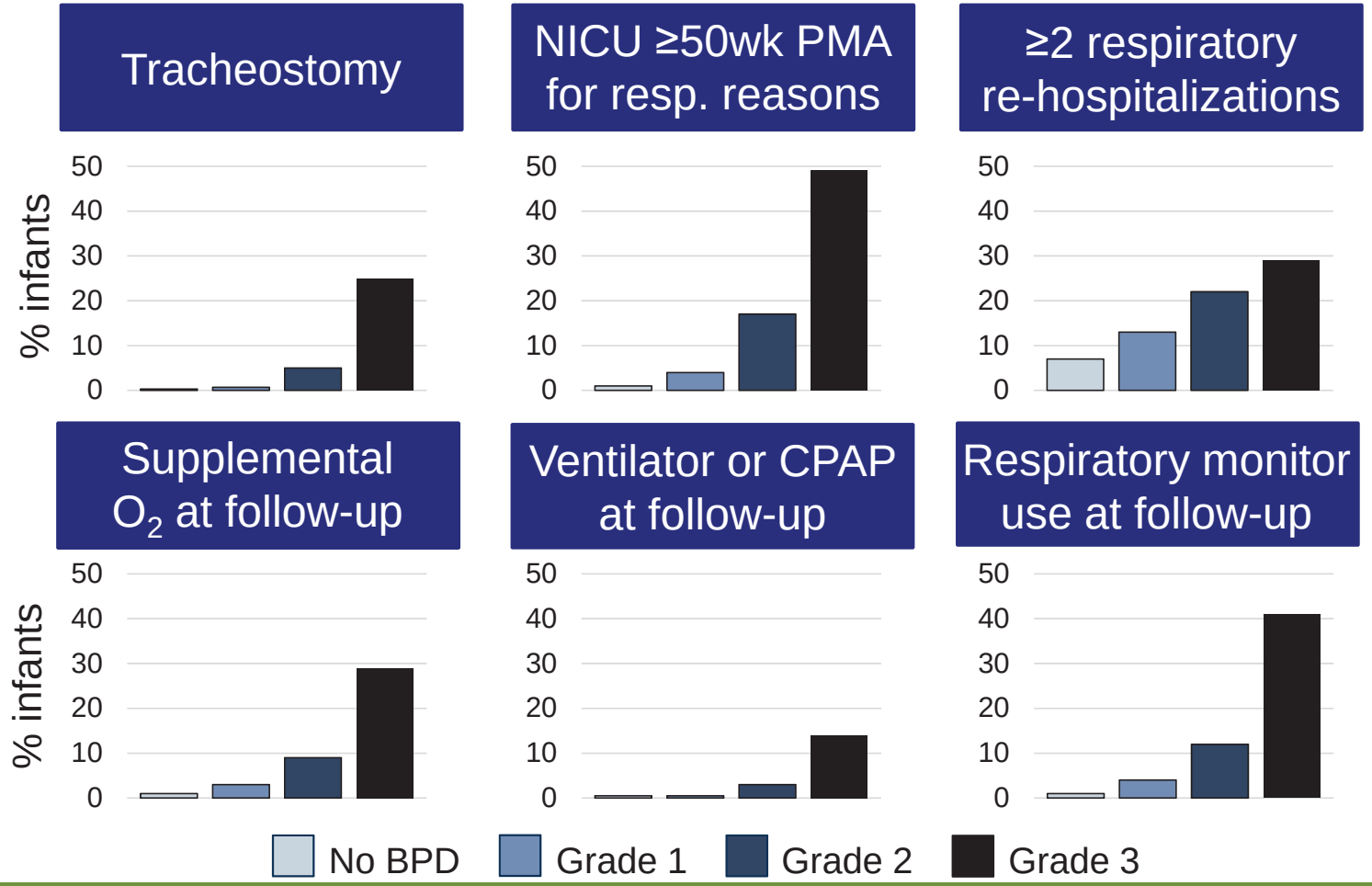
Outcome	C-statistic		
	2019 NRN	2001 NIH	P-value
Death or resp. morbidity	**0.785**	0.741	<0.001
Death or NDI	**0.747**	0.727	<0.001

Outcome	C-statistic		
	2019 NRN	2018 NICHD	P-value
Death or resp. morbidity	**0.785**	0.768	<0.001
Death or NDI	**0.747**	0.738	<0.001

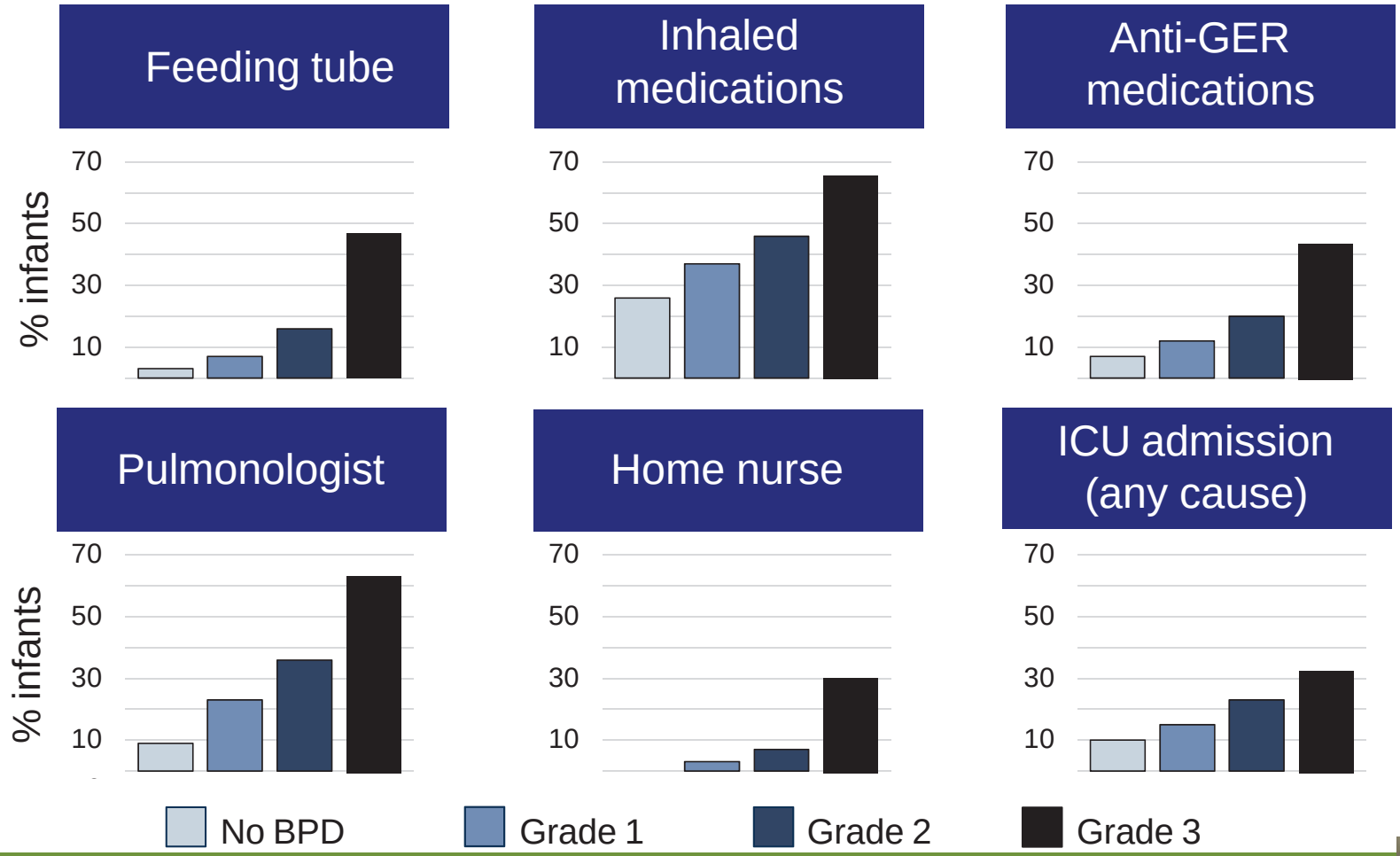
Death between 36 weeks' PMA and 18-26 month follow-up



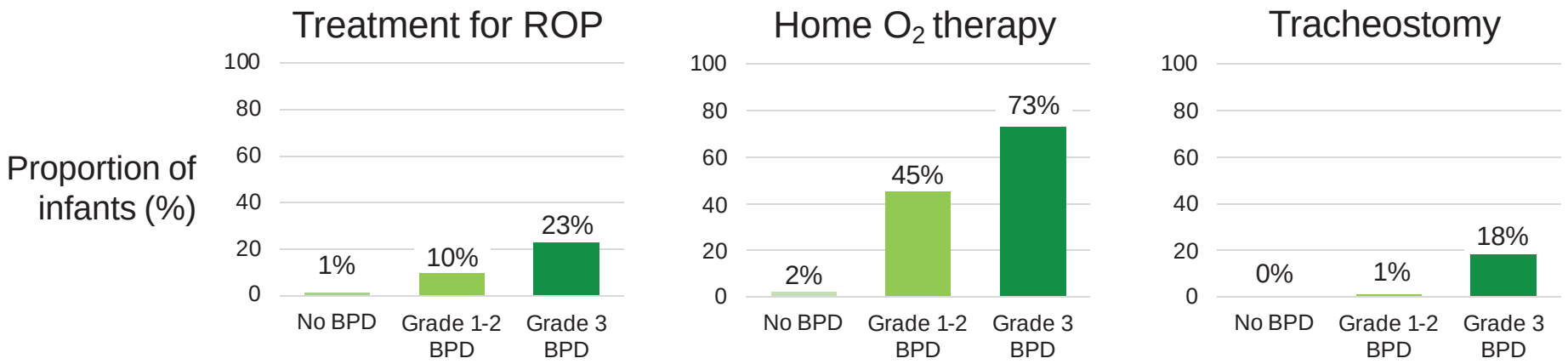
Respiratory Outcomes by BPD Severity



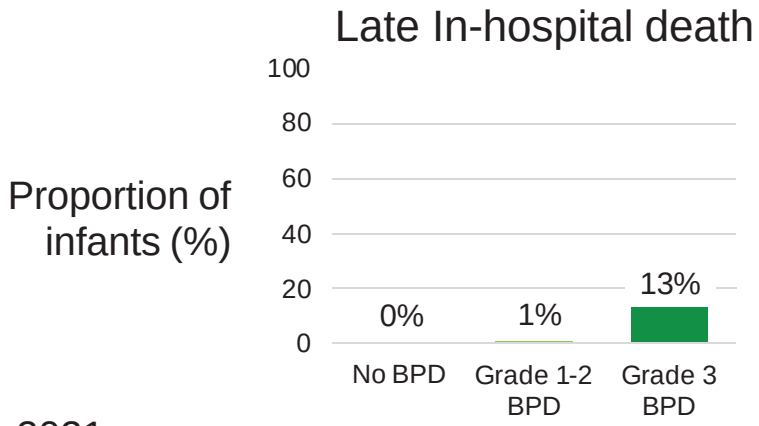
Healthcare Utilization at 2-year Follow-up

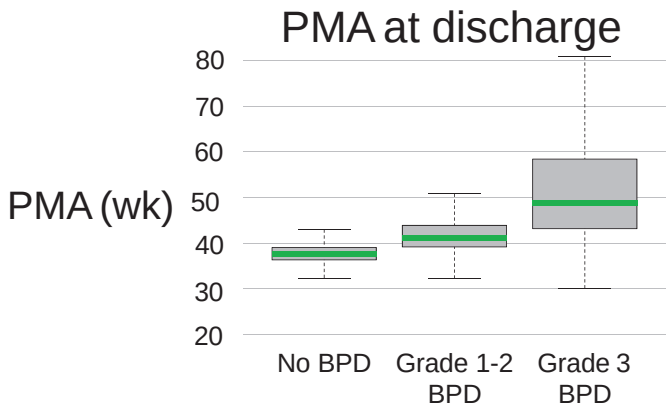


External validation: Vermont Oxford Network



24,896 infants born <30wk in 2018

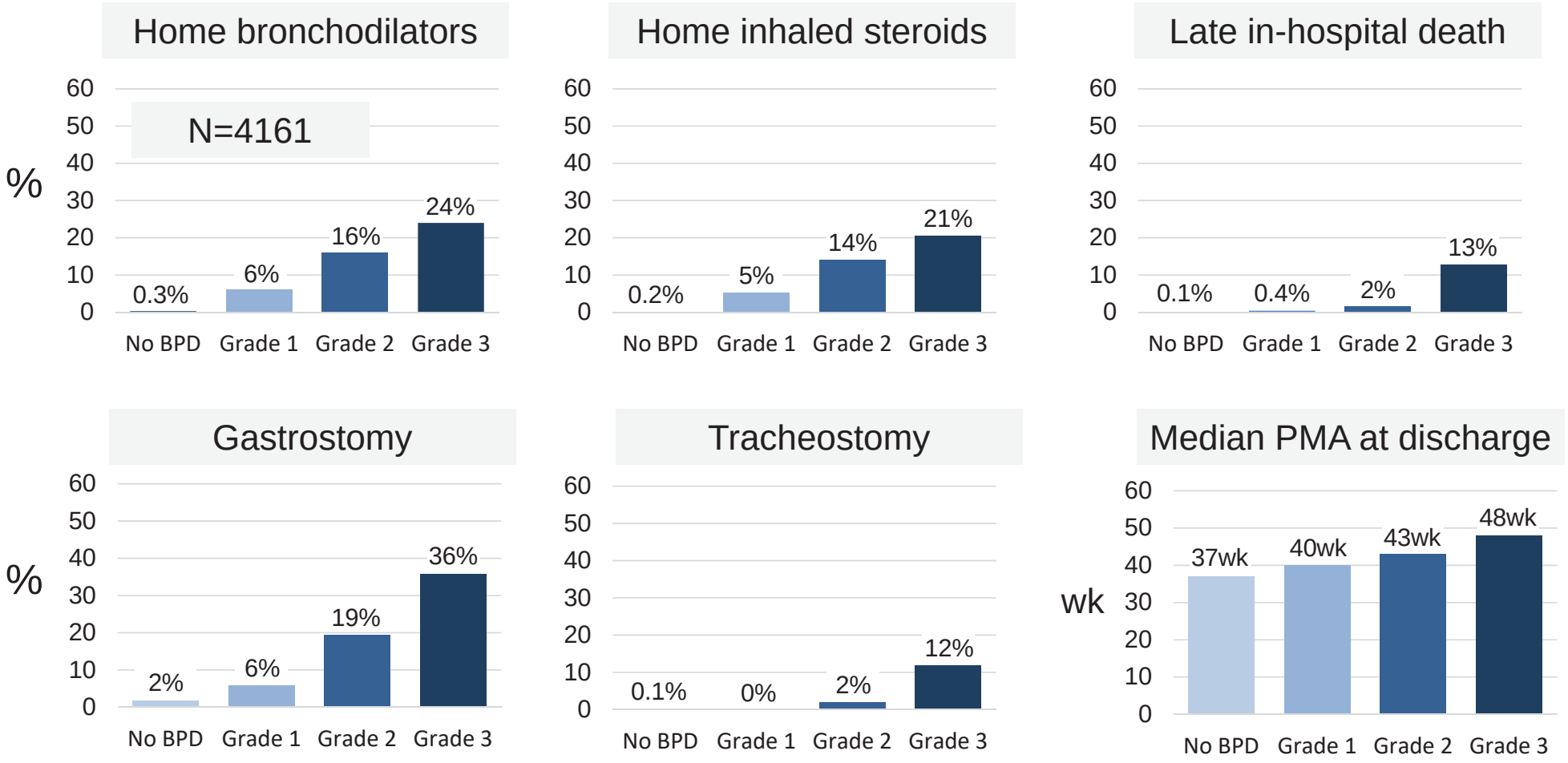




Jensen et al. Pediatrics 2021

Vermont Oxford NETWORK

External validation: Children's Hospital Consortium



Vyas-Read et al. J Perinatol 2021

Diagnostic criteria for BPD remain controversial

- **Available criteria do not adequately characterize the range of lung disease severity in preterm infants**

- **Not predictive of important post-discharge outcomes**
  - Parents care more about other outcomes, not “BPD”
  - Some infants *without* BPD experience adverse post-discharge respiratory outcomes

- **Diagnosis of BPD is subjective**
  - Treatments are variable, diagnostic tests are “objective”

Clinical Benefit

A positive clinically meaningful effect of an intervention…

A positive effect on how an individual:

- Feels
- Functions, *or*
- Survives

FDA U.S. FOOD & DRUG ADMINISTRATION

High blood pressure is a serious illness.

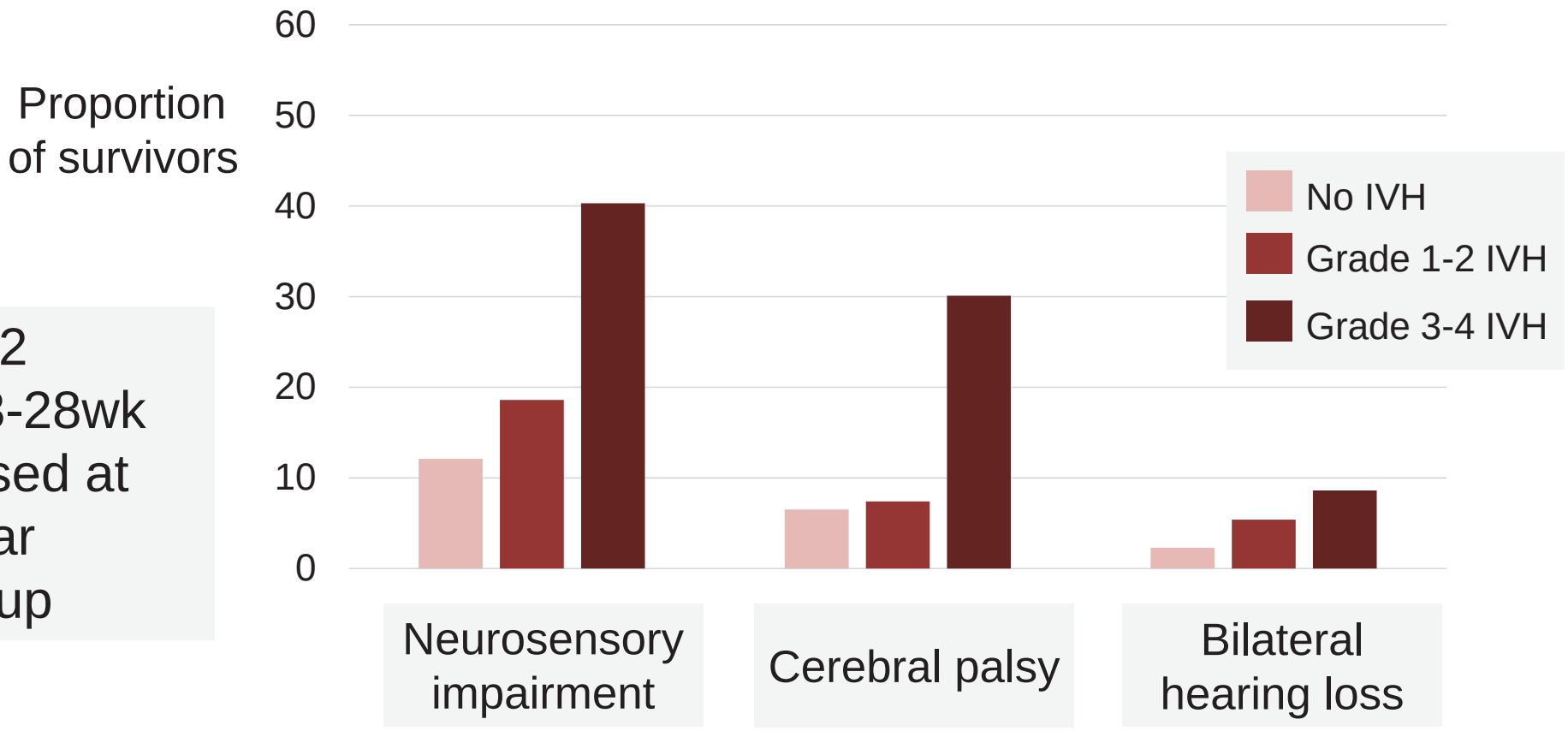
High blood pressure is often called a "silent killer" because you can have it and not know it. You may not feel sick at first. **Over time, if you do not get treated for high blood pressure, you can get very sick or even die.**

There are various FDA-approved products that are currently available to treat high blood pressure. Learn more about your options and use this information to help you talk to your healthcare provider about your blood pressure medicine. You will also find some general information to help you use your medicine wisely.

https://www.fda.gov/consumers/free-publications-women/high-blood-pressure

IVH: An important but *imperfect* predictor of disability

N=1472
GA: 23-28wk
Assessed at 2-3 year follow-up



Bolisetty S, et al. Pediatrics 2014

Is it objective? Diagnosis of IVH

Local vs. central reader interpretation of head ultrasound studies

	Central reader #1		Central reader #2	
IVH/PVL	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
Grade 1	53	95	28	97
Grade 2	47	95	50	93
Grade 3	62	94	42	95
Grade 4	82	96	86	92
PVL	44	98	20	97

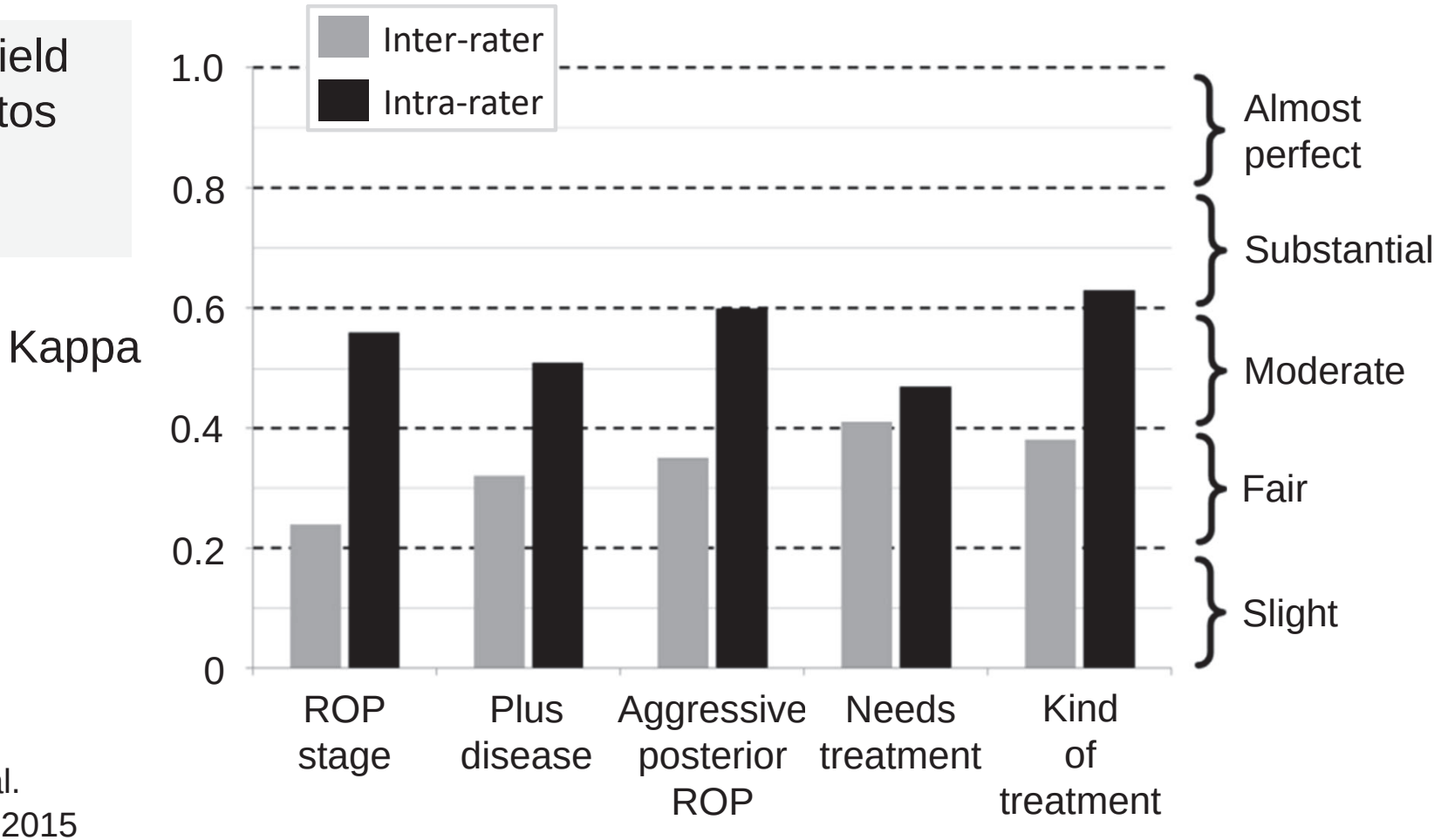
Sensitivity: Proportion "positive" by local reader among all "positive" by the central reader

Specificity: Proportion "negative" by local reader among all "negative" by the central reader

Hintz SR, et al. J Pediatr 2007

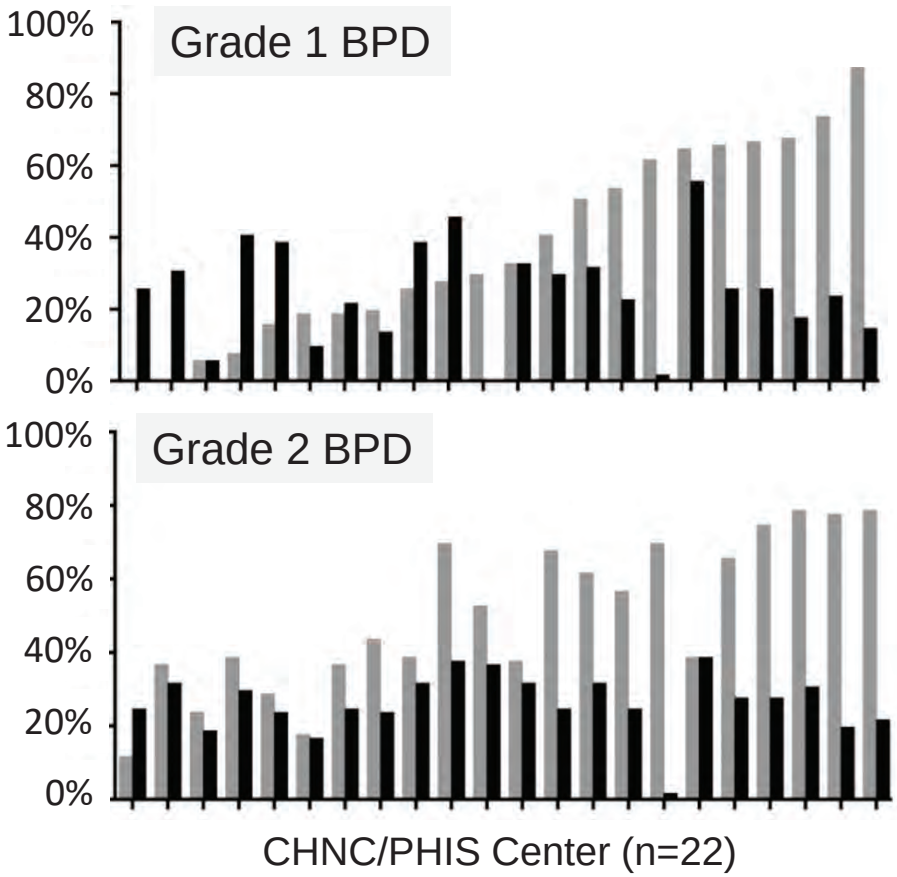
Is it objective? Diagnosis and treatment of ROP

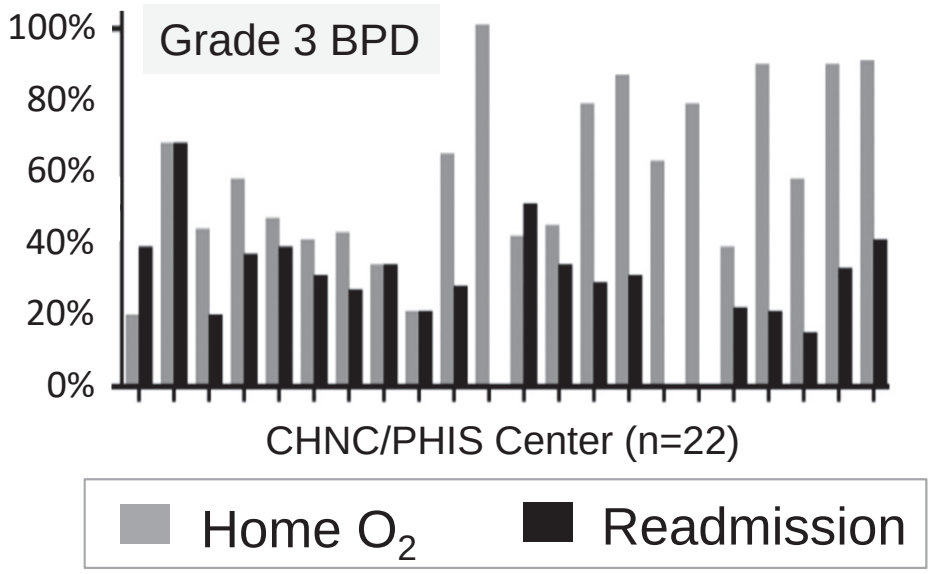
- 260 wide-field digital photos
- 52 infants
- 7 experts



Gschliesser A, et al.
Am J Ophthalmol. 2015

Home oxygen use and **readmissions** through 1 year vary widely by centers





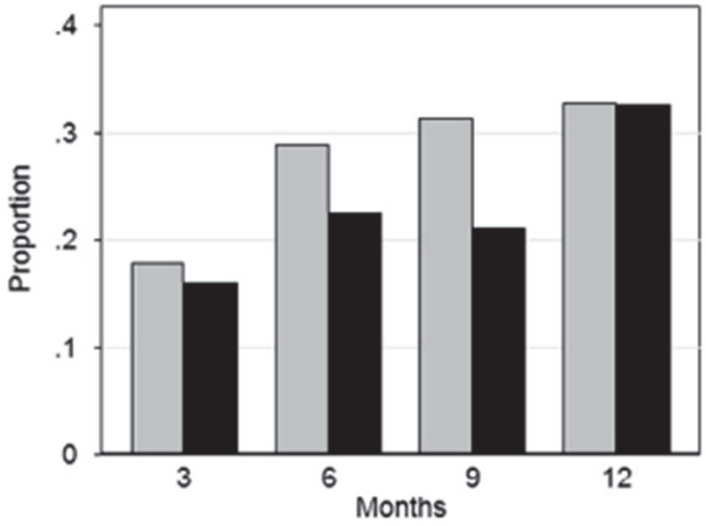
Clinical risk factors explained 72% of readmission risk

Lagatta J, et al. J Pediatr 2020

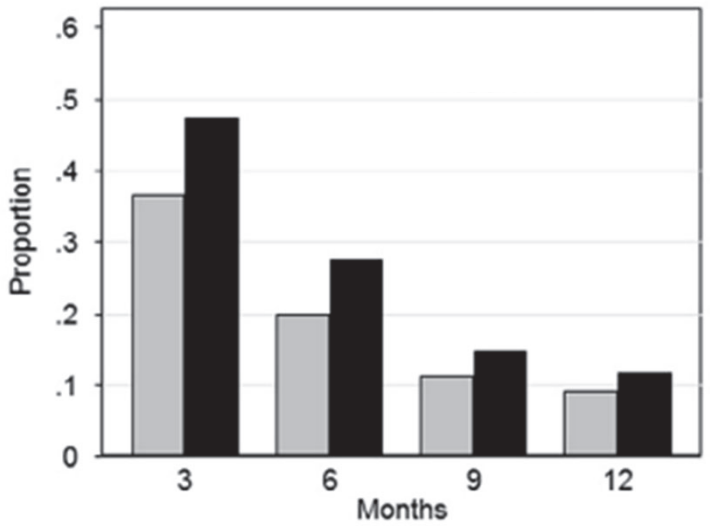
Tradeoffs in post-discharge respiratory outcomes?

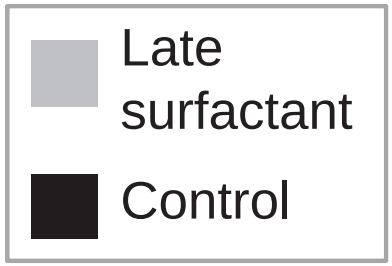
The Randomized, Controlled Trial of Late Surfactant: Effects on Respiratory Outcomes at 1-Year Corrected Age

Inhaled bronchodilators



Home respiratory support





Keller RL, et al. J Pediatr 2017

Waiting for the perfect bad respiratory outcome at the cost of preventing good bad outcomes?

"*Le mieux est le mortel ennemi du bien*"

“The best is the mortal enemy of the good”

Essays on Montesquieu and on the Enlightenment, Voltaire Foundation at the Taylor Institution

Voltaire 1694-1778

Summary

- We have proposed a data-driven definition of BPD that grades disease severity and improves prognostication of childhood outcomes
- Most outcome definitions/diagnostic criteria have limitations
- Preventing BPD should be one (but not the only) focus on improving respiratory outcomes in preterm infants

Janet Soul

Boston Children's Hospital

Session 2:
Challenges in Measuring Efficacy for Neonatal Conditions with Unmet Clinical Needs

Neonatal Seizures

Janet Soul, MD, Boston Children's Hospital

Objectives:

1)Highlight the challenges and considerations for developing core outcome sets for neonatal seizures and choosing appropriate primary endpoints for regulated trials

2)Review potential efficacy endpoints related to neonatal seizures

3)Discuss best practices and key solutions for generating high-quality evidence for these conditions with high unmet clinical needs

Importance of neonatal seizures

- **Highest lifetime seizure incidence is in newborns**
  - **Incidence of 2-4/1000 live births**

- **Associated with long-term neurologic disability**
  - **Epilepsy**
  - **Intellectual, motor and sensory disability**

Higher neonatal seizure burden associated with worse outcome

Short-term (neonatal)

- Greater mortality
- Longer length of hospital stay
  - Glass, Soul, J Peds 2016 Neonatal Seizure Registry

Long-term outcome

- Abnormal neurologic outcome
  - McBride Neurol 2000, Kharoshankaya DMCN 2016
- Epilepsy

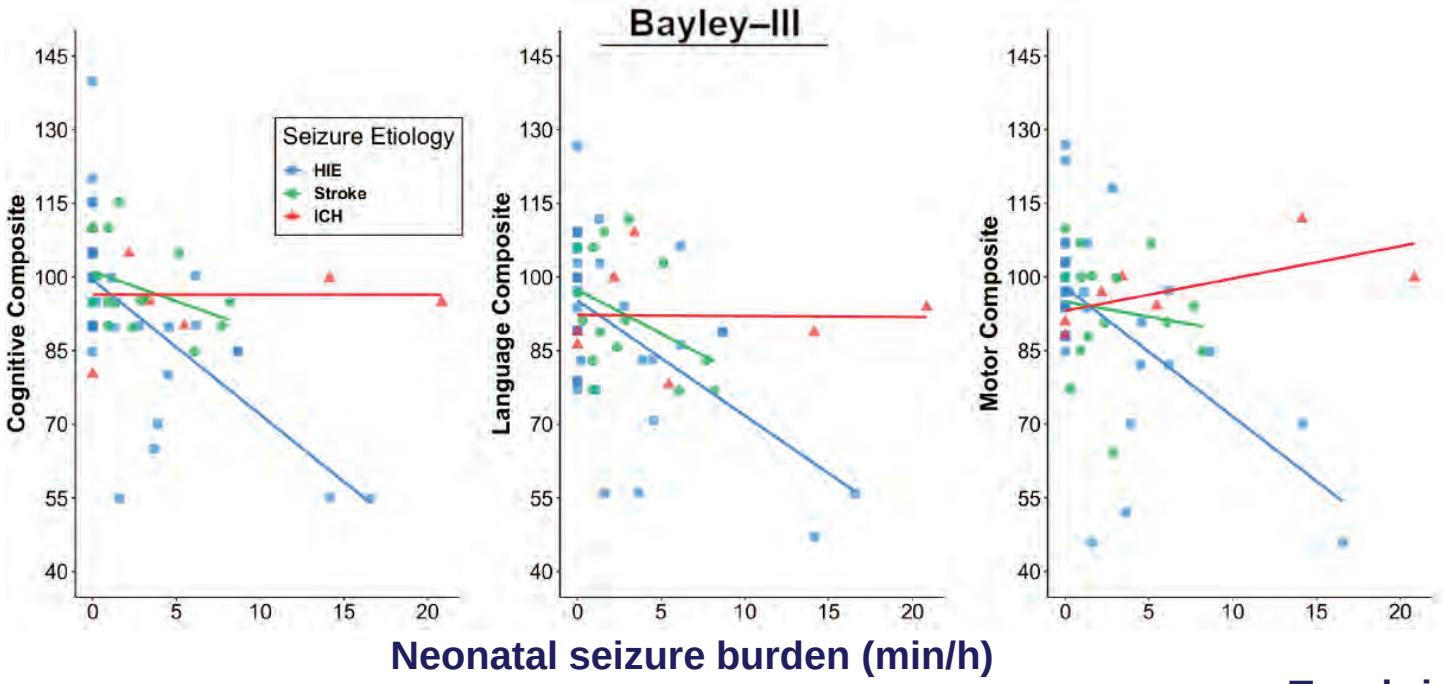
Chicken or egg?

- **Worse brain injury causes more seizures?**

OR

- **More seizures cause worse brain injury?**

Association of neonatal seizure burden with long-term outcome depends on seizure etiology



Blue = HIE
Green = Stroke
Red = Intracranial hemorrhage

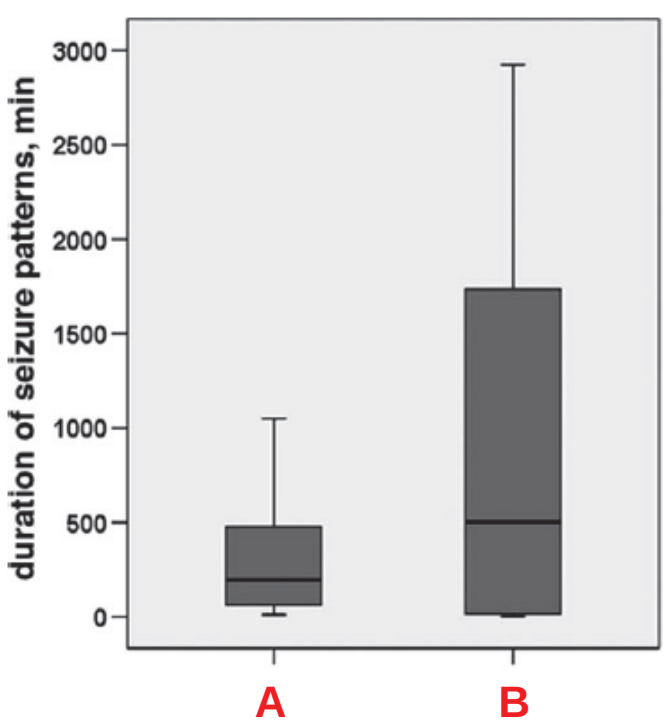
Trowbridge Annals of the CNS 2023
(Outcome data from Boston Bumetanide Trial)

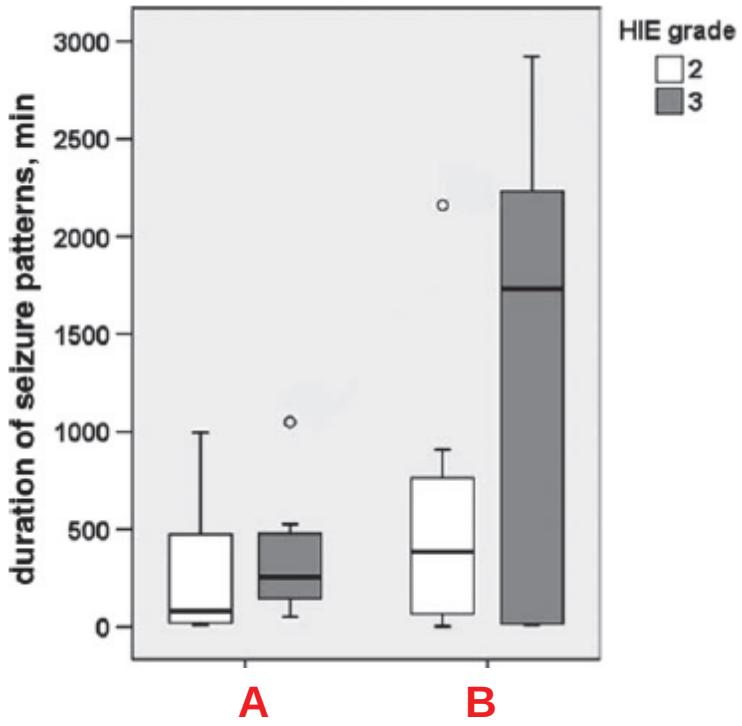
Two Trials of Seizure Treatment

- Newborns with HIE
- Treated clinical vs. EEG seizures
  - One trial with aEEG (Europe)
    - N=33
  - One trial with cvEEG (Washington Univ.)
    - N=35

Van Rooij Peds 2010
Srinivasakumar Peds 2015

Treatment of aEEG seizures reduced seizure duration

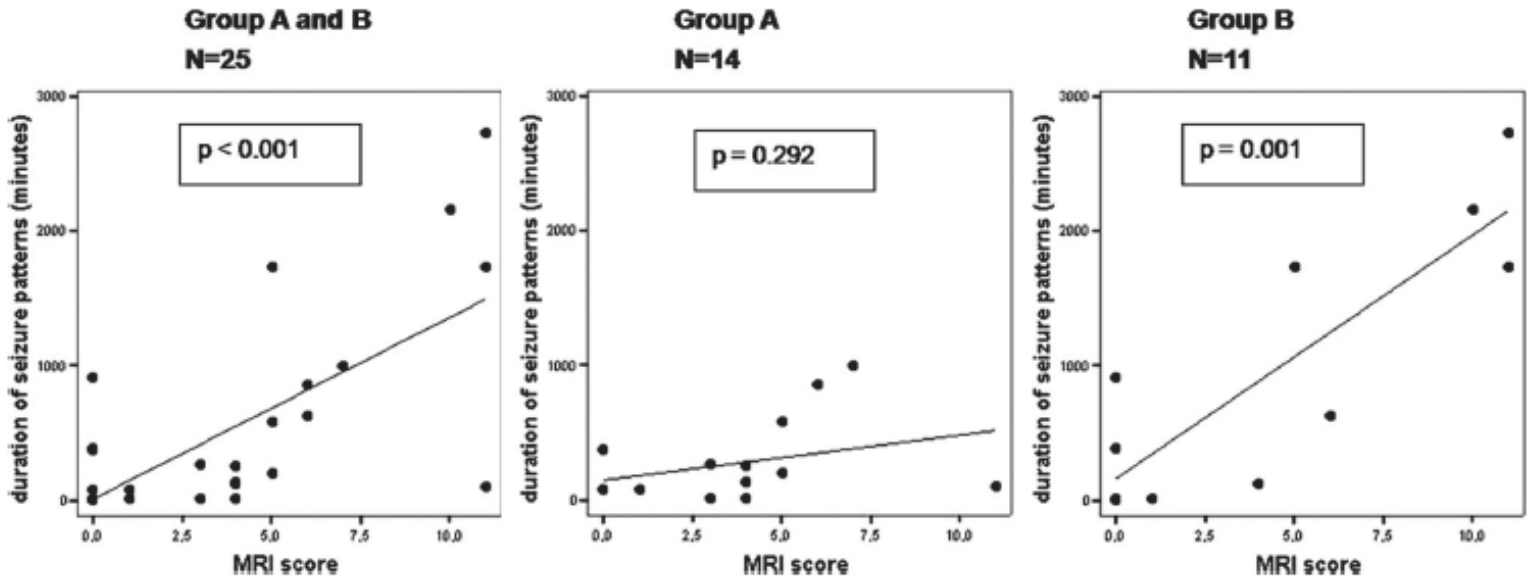




A = Treat aEEG seizures
B = Treat clinical seizures

Van Rooij Peds 2010

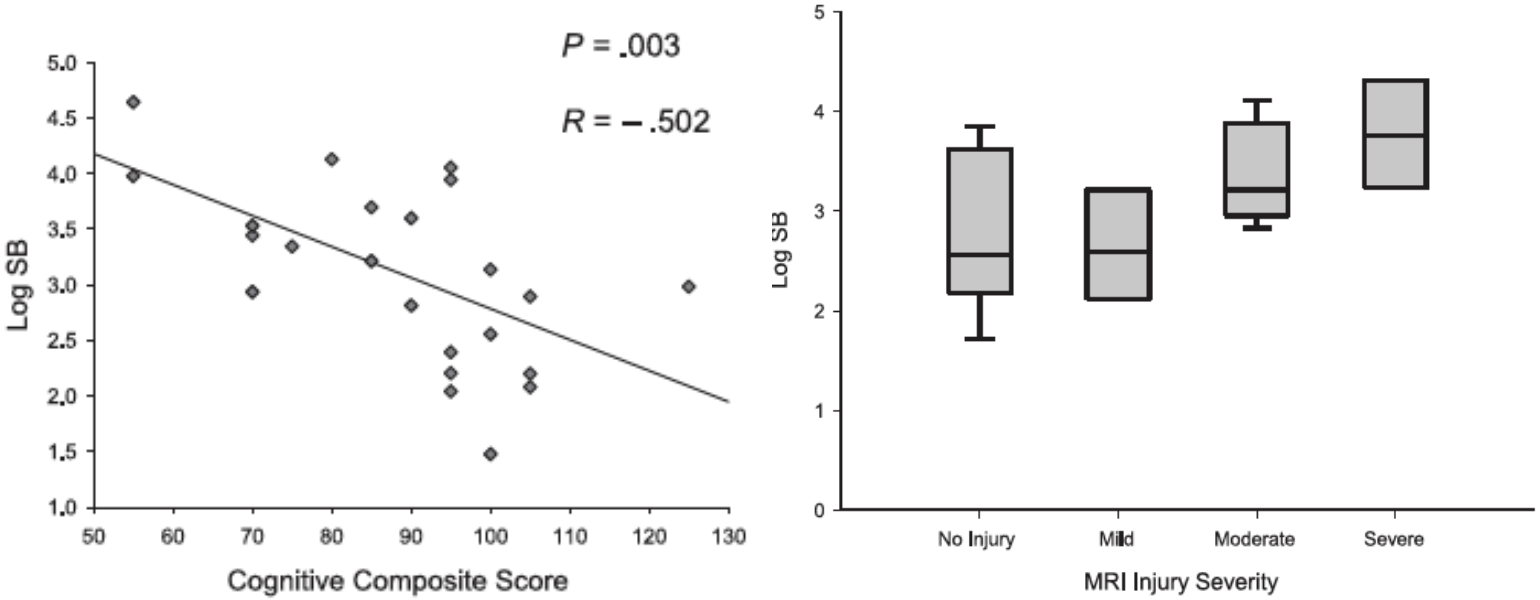
Treatment of aEEG seizures improved MRI score of injury



A = Treat aEEG seizures
B = Treat clinical seizures

Van Rooij Peds 2010

EEG trial: Higher seizure burden associated with worse outcome



Srinivasakumar Peds 2015

Trials of EEG vs. Clinical Seizure Treatment

- **Conclusions:**
  - **ASMs reduce seizure burden**
  - **Treatment of neonatal EEG seizures might actually improve neurologic outcome**

Van Rooij Peds 2010
Srinivasakumar Peds 2015

Which anti-seizure medications (ASMs) should we test to improve outcome?

Anti-seizure Medications (ASMs) studied to date

- Phenobarbital (use for >100 years!!!)
  - Approved by FDA Nov 2022
- Phenytoin / fosphenytoin (>80 years!)
- Levetiracetam (~20 years)
- Lidocaine
- Midazolam
- Others: bumetanide, carbamazepine, topiramate, other ASMs used in older children

➢ Phenobarbital first ASM for >90-95%

Trials for Neonatal Seizures: Real World Data

- Randomized trials: 1st and 2nd line ASMs
  - 3 EEG / aEEG monitoring trials
  - 2 cross-over trials of 1st line therapy
  - 1 small trial of 2nd line therapy
  - 1 early phase, controlled trial
- Several small open label trials
- Retrospective studies

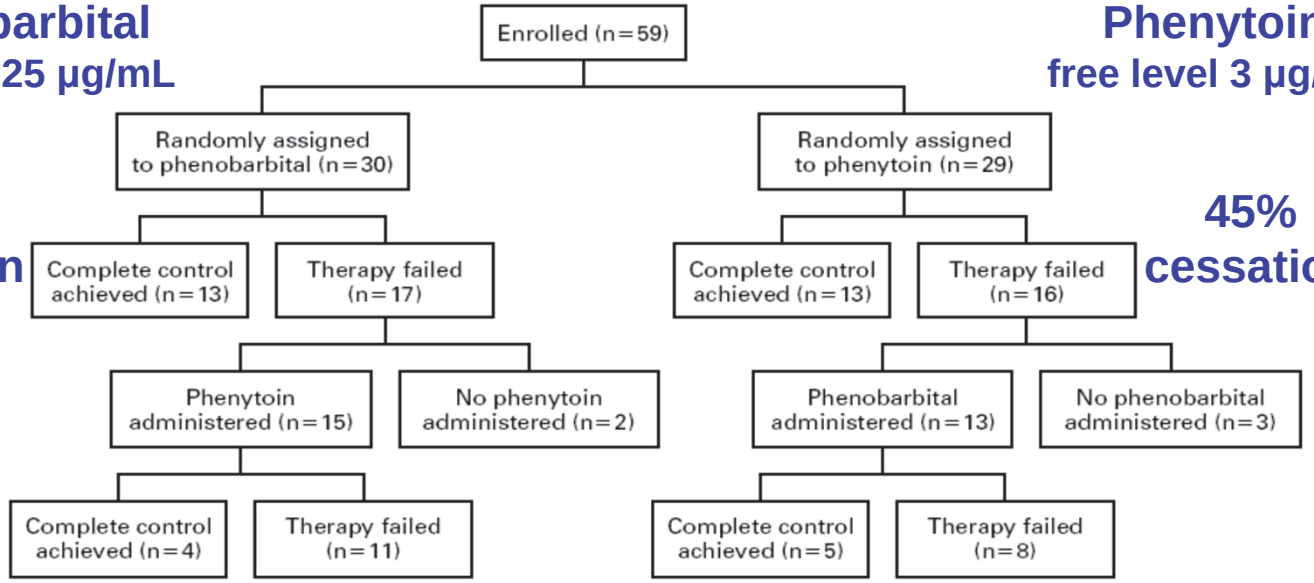
Phenobarbital vs. phenytoin trial
Primary outcome: seizure cessation

Painter NEJM 1999 (n = 59)

1-hour EEG:
Enrolled if seizures

Phenobarbital
free level 25 µg/mL

Phenytoin
free level 3 µg/mL



43% cessation

45% cessation

Figure 1. Treatment Assignment and Outcomes of 59 Neonates with Seizures Treated with Phenobarbital or Phenytoin.

57% cessation with both ASMs

Phenobarbital vs. phenytoin trial
Efficacy affected by:

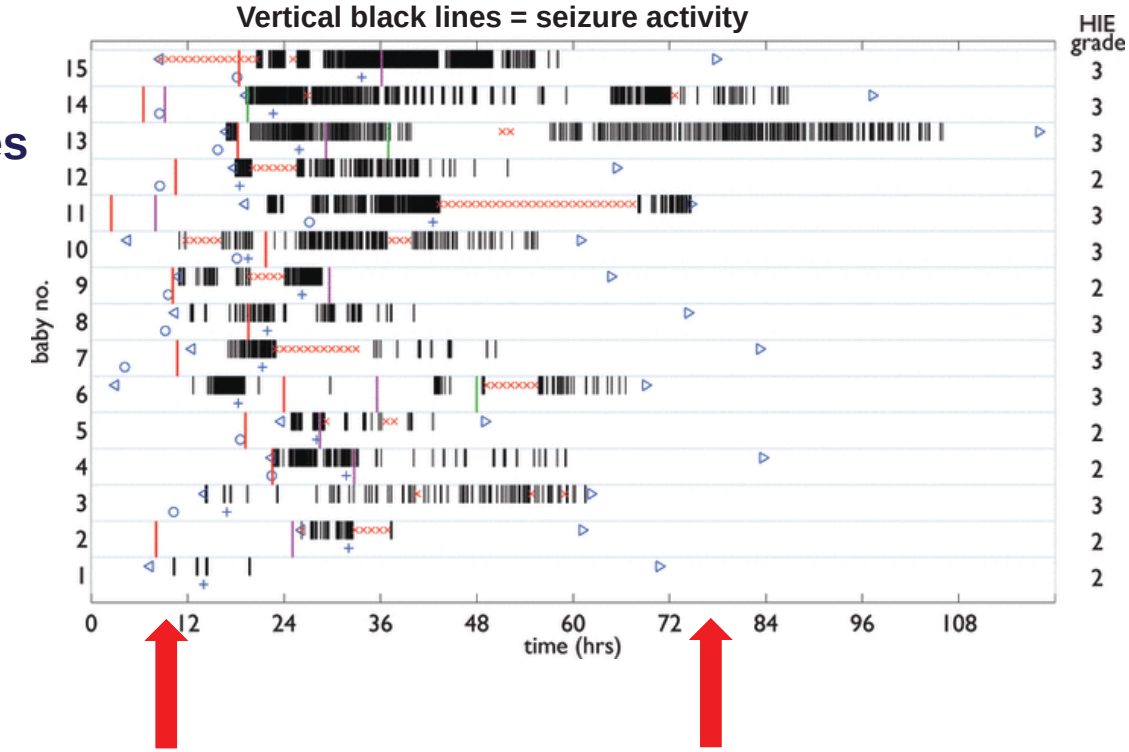
1. Seizure severity / burden
   - 88% mild vs. 10% severe with cessation

2. Seizure course, i.e., timing of randomization
   - 81% decreasing vs. 30% increasing

Painter NEJM 1999

Acute Provoked Neonatal Seizures: Resolve in hours to days

Example:
15 neonates with HIE



Lynch Epilepsia 2012

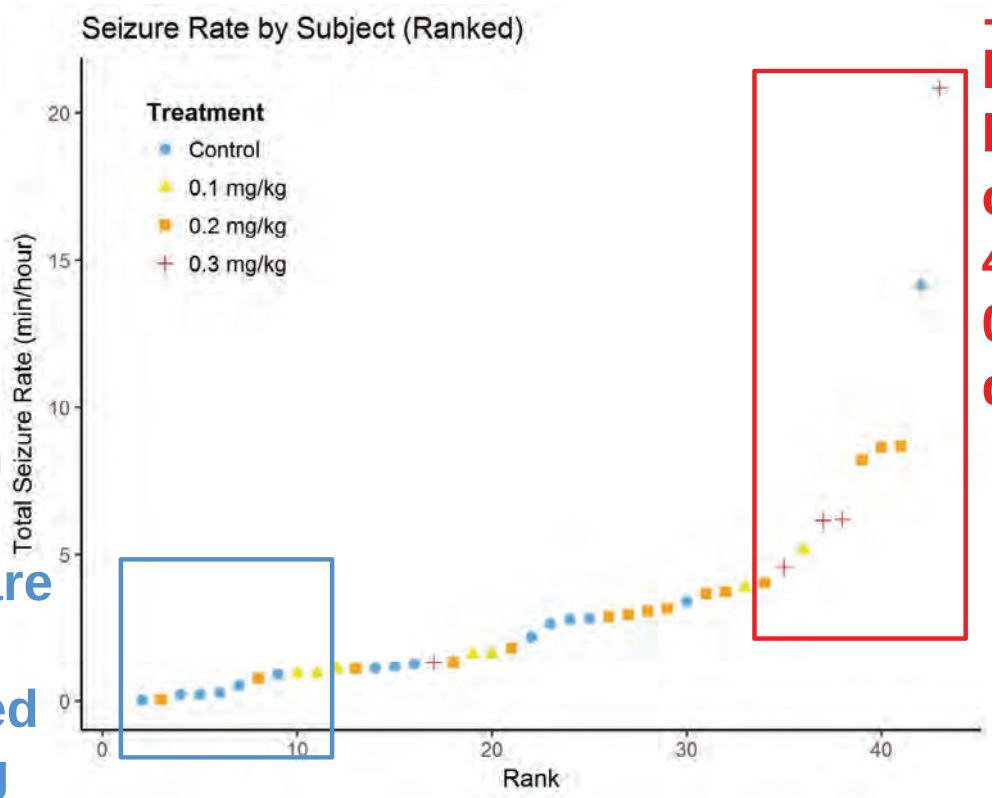
Real-world challenges for neonatal seizure treatment trials

- Seizure severity/burden highly variable
  - Major determinant of ASM efficacy
  - Cannot currently be measured rapidly at time of enrollment or randomization
- Timing of randomization is key
  - Need to randomize early in the course of seizures

Boston Bumetanide Trial Design

- Newborns with acute seizures refractory to phenobarbital load
- Add-on bumetanide vs. standard therapy control group
- Randomized, double blind
- Dose escalation design to test consecutively higher doses of bumetanide

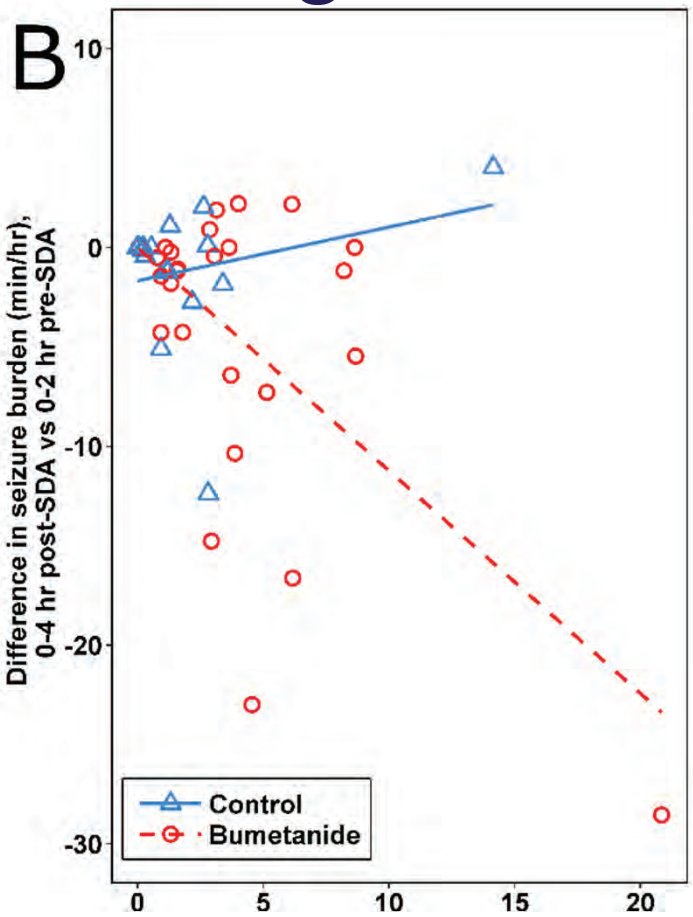
Total seizure burden in min/h
(ranked by subject)



1/10 with highest burden were controls, but 4/5 of the 0.3mg/kg dose group

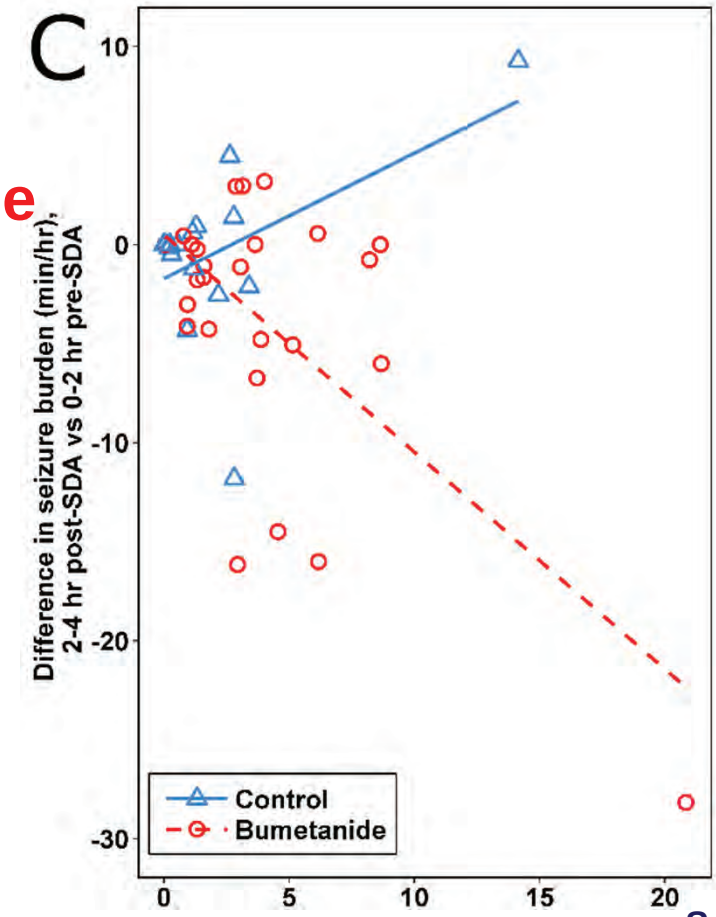
6/10 with lowest burden are controls, 2 received 0.1mg/kg

Quantitative seizure reduction greater in bumetanide than control



Control
Bumetanide

p=0.008

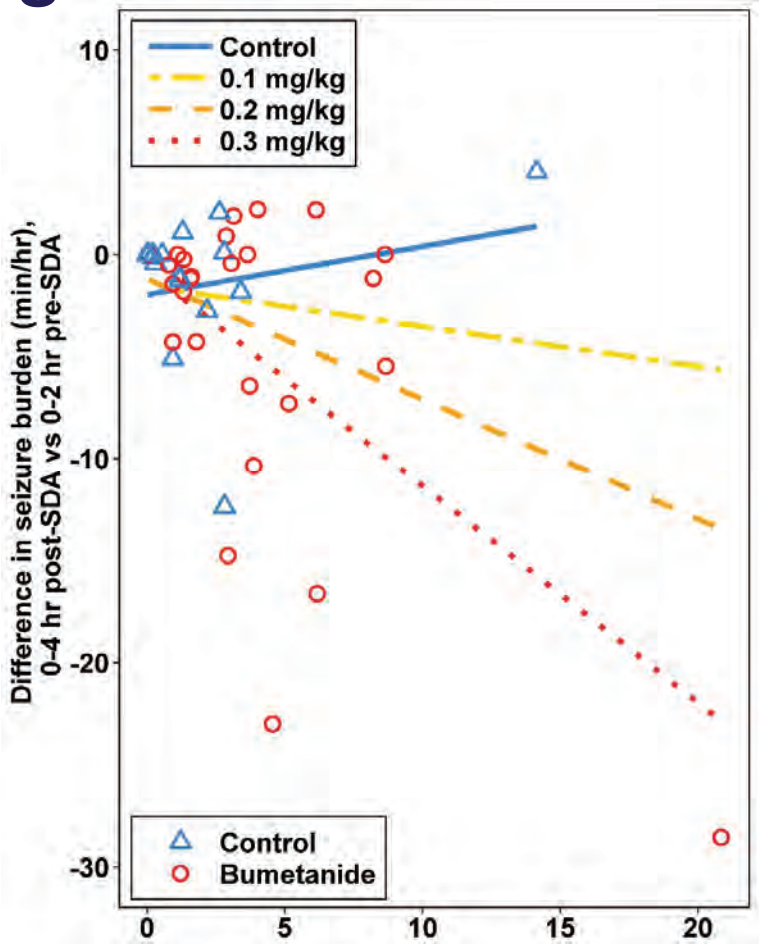


Control
Bumetanide

p=0.0004

Soul Ann Neurol *2021*

Greater seizure reduction with higher bumetanide dose exposure



- Control
- BTN 0.1
- BTN 0.2
- BTN 0.3

p=0.008

Soul Ann Neurol *2021*

Levetiracetam (LEV) Trial

NCT01720667 Sharpe *Pediatrics* 2020

- Trial compared LEV 40-60 mg/kg to Phenobarbital 20-40 mg/kg as first line therapy
  - Randomized, double-blind
  - Continuous cvEEG monitoring
- If seizures persisted, subjects crossed over to receive other ASM
- Phase II trial Primary Outcome: Compared rate of seizure cessation between treatment groups

Levetiracetam Trial: Efficacy Data

NCT01720667: Sharpe *Pediatrics* 2020

% of newborns with seizure cessation

Time (hours)	Phenobarbital (%)	Levetiracetam (%)
1 h	93	49
24 h	80	28
48 h	64	17

Trial endpoints/outcome:

1. **Short-term outcome**
   - reduce seizure burden, morbidity, hospital stay

2. **Reduce long-term neurologic disability**
   - probably mediated by reduced seizure burden

3. **Reduce rate of later epilepsy**
   - probably mediated by reduced seizure burden

Long-term Outcomes of Trials

- No long-term outcome data
  - PB vs. PHT trial - NEJM 1999
  - NEMO: stopped early, enrolled 14 subjects, 3 died – Lancet Neurol 2015
  - Levetiracetam vs. Phenobarbital trial - Pediatrics 2020
  - ANSeR (monitoring) trial - Lancet Child Adolesc Health 2020

- Long-term outcome data available:
  - EEG/aEEG vs. clinical seizure treatment trials – Pediatrics 2010, 2015
  - Boston Bumetanide Trial – Annals of Neurology 2021
    - Outcome data – Annals of Child Neurology Society 2023

➢ Need adequate funding, requirement to obtain long-term outcome data required to assess both ASM efficacy and safety

Trial Design to Achieve Primary Outcome

- **Short term outcomes:**
  - **Quantitative ASM response – early phase**
  - **Compare seizure burden among groups**
  - **Safety data – adverse events**
    - **Duration of stay, other short-term morbidities**

- **Long-term outcome**
  - **Rates of neurologic impairments – many measures**
  - **Rates of epilepsy**
  - **Measures ASM efficacy and safety**

Soul Pediatr Res *2019*

Trial Design to Achieve Primary Outcome

Trial Design needs:

- Careful choice of seizure etiologies
  - affects long-term outcome, depends on trial phase
- Prolonged conventional EEG monitoring
- Control group to assess ASM efficacy & safety
- Rapid randomization early in seizure course
- Balance/Analyze effect of seizure severity
  - Requires sufficient sample size
- Measure ASM levels to analyze dose exposure
- Long-term follow-up data!

Soul Pediatr Res *2019*

Martin Offringa

University of Toronto

Session 2: Challenges in Measuring Efficacy for Neonatal Conditions with Unmet Clinical Needs

NOWS & Core Outcomes for regulated trials

Martin Offringa, MD PhD
Division of Neonatology
Child Health Evaluative Sciences,
The Hospital for Sick Children Research Institute
University of Toronto
Martin.offringa@sickkids.ca

SickKids RESEARCH INSTITUTE

TORCH Toronto Outcomes Research In Child Health

Psychiatry UNIVERSITY OF TORONTO

Outline

1. The NOWS COS
2. Challenges and Considerations - developing COS
3. Justifying Proposed Efficacy Endpoints
4. Best Practices - choosing appropriate primary endpoints for regulated trials

Therapeutic Innovation & Regulatory Science
1-9

DOI: 10.1177/21684790
tirs.sagepub.com

Applying Regulatory Science to Develop Safe and Effective Medicines for Neonates: Report of the US Food and Drug Administration First Annual Neonatal Scientific Workshop, October 28-29, 2014

Martin Offringa, MD, PhD[1], Jonathan M. Davis, MD[2], Mark A. Turner, PhD, MBChB, MRCP(UK), MRCPCH, DRCOG[3], Robert Ward, MD[4], Ralph Bax, MD, MA, PhD[5], Sam Maldonado, MD, MPH, FAAP[6], Vikram Sinha, PhD[7], Susan K. McCune, MD[8], Anne Zajicek, MD, PharmD, FAAP[9], Daniel K. Benjamin Jr, MD, PhD[10], Christina Bucci-Rechtweg, MD[11], and Robert M. Nelson, MD, PhD[12]

Abstract

The First Annual Neonatal Scientific Workshop focused on the needs of the neonate by addressing the basic quest information is required to inform decision making both at the regulatory level and at the bedside? Priority therape include neonatal lung, brain, and gastrointestinal injury, retinopathy of prematurity, sepsis, and neonatal abstinence s Scientific progress in these therapeutic areas, regulatory standards, and the acceptable design and conduct of clinical tria aligned. This report will review potential approaches to enhancing neonatal drug development.

This paper provides an overview of the discussions at the Workshop, which concentrated on better defining regulatory standards, scientific knowledge, and clinical feasibility. Workshop participants addressed both challenges and potential approaches for developing safe and efficacious treatments for neonates, including the following:

1. What therapeutic areas are in highest need of clinical drug development;
2. How best to study new and existing drugs in neonates, with an emphasis on appropriate clinical outcome measures;
3. What study designs to apply;
4. How extrapolation and modern pharmacometrics can be used in designing trials and analyzing data;
5. Which innovative trial designs and approaches are regulatory ready for use allowing "smart designs" to make decisions earlier/faster;
6. Criteria for initiating trials in neonates;
7. The need for age appropriate formulations with particular attention to potential toxicities of excipients;
8. A consistent definition of the neonate needs to be established for clinical trials purposes and regulatory considerations.

NOWS

preterm and term infants will develop serious bacterial and viral infections resulting in death or NDI. More than 25% of preterm infants will develop early-onset or nosocomial sepsis, with an increasing number of infections resistant to traditional antibiotic and antiviral agents.

e. **Retinopathy of prematurity (ROP)**. Many preterm infants develop ROP, with more severe forms requiring treatment with laser ablation or anti-angiogenic agents (off label use). The risk of visual impairment or blindness in this high-risk population remains unacceptably high. New approaches to the prevention and treatment are urgently needed.

f. **Neonatal abstinence syndrome (NAS)**. In the US, there has been a dramatic increase in the incidence of NAS resulting from in utero exposure to opiates, with rates tripling in the last 10 years. There is significant uncertainty on who to treat, when to treat, and how to treat affected infants. In-depth studies with better assessment techniques and short- and long-term outcome measures are urgently needed.

Biomarkers and Endpoints

Biomarkers

The careful application and assessment of the right biomarker(s) in the right population are pivotal in drug development. Biomarkers can support the diagnosis, prognosis, initiation of treatment, and the response to treatment (see Table 1). The development of a biomarker may include a comprehensive evaluation, ensuring it is fit for purpose and then qualifying it for a particular purpose within a tightly defined context of use. Biomarkers can be used in an individual drug development program or qualified for a particular context of use across multiple drug development programs.

A large number of biomarkers and clinically important outcomes have been described in several systematic reviews.[5-10] Overall, many biomarkers may have some utility in clinical practice, but insufficient high-quality data are available to support their use in neonatal drug development. In neonates, clinically meaningful surrogate outcomes or biochemical biomarkers borrowed from other age groups or therapeutic contexts

1) Challenge: to evaluate new (pharmacological) interventions
2) Problem: Heterogeneity in existing outcomes
3) Solution: a Core Outcome Set

Shan *et al. Trials* (2020) 21:262
https://doi.org/10.1186/s13063-020-4183-9

Trials

REVIEW Open Access

Outcome reporting in neonates experiencing withdrawal following opioid exposure in pregnancy: a systematic review

Flora Shan[1], Sonya MacVicar[2], Karel Allegaert[3,4], Martin Offringa[5], Lauren M. Jansson[6], Sarah Simpson[7], Wendy Moulsdale[8] and Lauren E. Kelly[1,9*]

Abstract

Background: Neonatal withdrawal secondary to in utero opioid exposure is a growing global concern stressing the psychosocial well-being of affected families and scarce hospital resources. In the ongoing search for the most effective treatment, randomized controlled trials are indispensable. Consistent outcome selection and measurement across randomized controlled trials enables synthesis of results, fostering the translation of research into practice. Currently, there is no core outcome set to standardize outcome selection, definition and reporting. This study identifies the outcomes currently reported in the literature for neonates experiencing withdrawal following opioid exposure during pregnancy.

Methods: A comprehensive literature search of MEDLINE, EMBASE and Cochrane Central was conducted to identify all primary research studies (randomized controlled trials, clinical trials, case-controlled studies, uncontrolled trials, observational cohort studies, clinical practice guidelines and case reports) reporting outcomes for interventions used to manage neonatal abstinence syndrome between July 2007 and July 2017. All "primary" and "secondary" neonatal outcomes were extracted by two independent reviewers and were assigned to one of OMERACT's core areas of "pathophysiological manifestation", "life impact", "resource use", "adverse events", or "death".

Results: Forty-seven primary research articles reporting 107 "primary" and 127 "secondary" outcomes were included. The most frequently reported outcomes were "duration of pharmacotherapy" (68% of studies, $N = 32$), "duration of hospital stay" (66% of studies, $N = 31$) and "withdrawal symptoms" (51% of studies, $N = 24$). The discrepancy between the number of times an outcome was reported and the number of articles was secondary to the use of composite outcomes. Frequently reported outcomes had heterogeneous definitions or were not defined by the study and were measured at different times. Outcomes reported in the literature to date were mainly assigned to the core areas "pathophysiologic manifestations" or "resource use". No articles reported included parent or former patient involvement in outcome selections.

(Continued on next page)

Conclusions:

Inconsistent selection and definition of primary and secondary outcomes exists in the present literature of pharmacologic and nonpharmacologic interventions for managing opioid withdrawal in neonates.

No studies involved parents in the process of outcome selection. These findings hinder evidence synthesis to generate clinically meaningful practice guidelines. The development of a specific core outcome set is imperative.

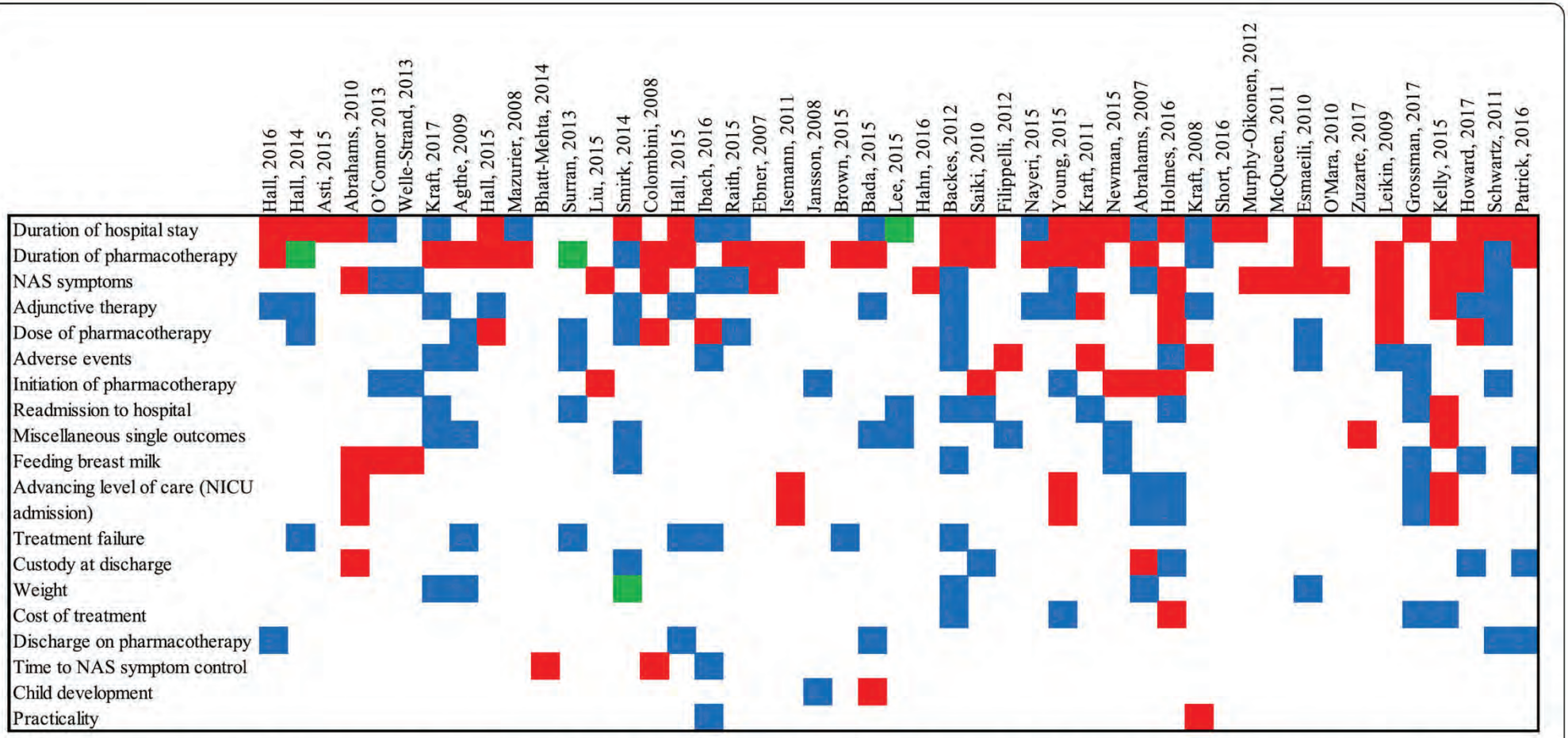


Fig. 2 Outcome matrix of 234 outcome terms (rows) for 47 primary research studies (columns). Red signifies a primary outcome. Blue signifies a secondary outcome. Green signifies a primary and secondary outcome reported in the same article that fit within the same outcome category. NAS neonatal abstinence syndrome, NICU neonatal intensive care unit

Table 2 Variation in definition of outcomes reported

Outcome	Reported outcome definition
Duration of pharmacotherapy ($n = 32$)	Days of opioid treatment
	Duration of treatment for NAS symptoms
	Duration of pharmacotherapy for NAS/NOWS
	Days of treatment
	Duration of treatment required for NAS resolution
	Number of days on morphine
	Duration of opioid treatment
	Duration of oral morphine treatment
	Length of opioid treatment
	Length of methadone therapy
	Duration of phenobarbital treatment
	Total phenobarbital treatment days
	Total treatment duration (hospital + home)
	Length of treatment
	Days on DTO
	Days on any form of pharmacologic support
Duration of hospital stay ($n = 31$)	Days of inpatient hospitalization
	Neonatal length of stay
	Length of inpatient stay
	Duration of hospital stay
	Mean infant hospital stay
	Average length of stay between hospitals

Most common

2nd Most common

Problem:
outcome definitions

PEDIATRICS®

Content ⌄ Authors/Reviewers ⌄ Collections ⌄ Multimedia ⌄ Blogs Si

Volume 146, Issue 1
July 2020

ARTICLES | JULY 01 2020

A Core Outcome Set for Neonatal Opioid Withdrawal Syndrome

abstract

BACKGROUND: As rates of neonatal opioid withdrawal are increasing, the need for research to evaluate new treatments is growing. Large heterogeneity exists in health outcomes reported in current literature. Our objective is to develop an evidence-informed and consensus-based core outcome set in neonatal opioid withdrawal syndrome (NOWS-COS) for use in studies and clinical practice.

METHODS: An international multidisciplinary steering committee was established. A systematic review and a 3-round Delphi was performed with open-ended and score-based assessments of the importance of each outcome to inform clinical management of neonatal opioid withdrawal. Interviews were conducted with parents and/or caregivers on outcome importance. Finally, a consensus meeting with diverse stakeholders was held to review all data from all sources and establish a core set of outcomes with definitions.

RESULTS: The NOWS-COS was informed by 47 published studies, 41 Delphi participants, and 6 parent interviews. There were 63 outcomes evaluated. Final core outcomes include (1) pharmacologic treatment, (2) total dose of opioid treatment, (3) duration of treatment, (4) adjuvant therapy, (5) feeding difficulties, (6) consolability, (7) time to adequate symptom control, (8) parent-infant bonding, (9) duration of time the neonate spent in the hospital, (10) breastfeeding, (11) weight gain at hospital discharge, (12) readmission to hospital for withdrawal, and (13) neurodevelopment.

CONCLUSIONS: We developed an evidence-informed and consensus-based core outcome set. Implementation of this core outcome set will reduce heterogeneity between studies and facilitate evidence-based decision-making. Future research will disseminate all the findings and pilot test the validity of the NOWS-COS in additional countries and populations to increase generalizability and impact.

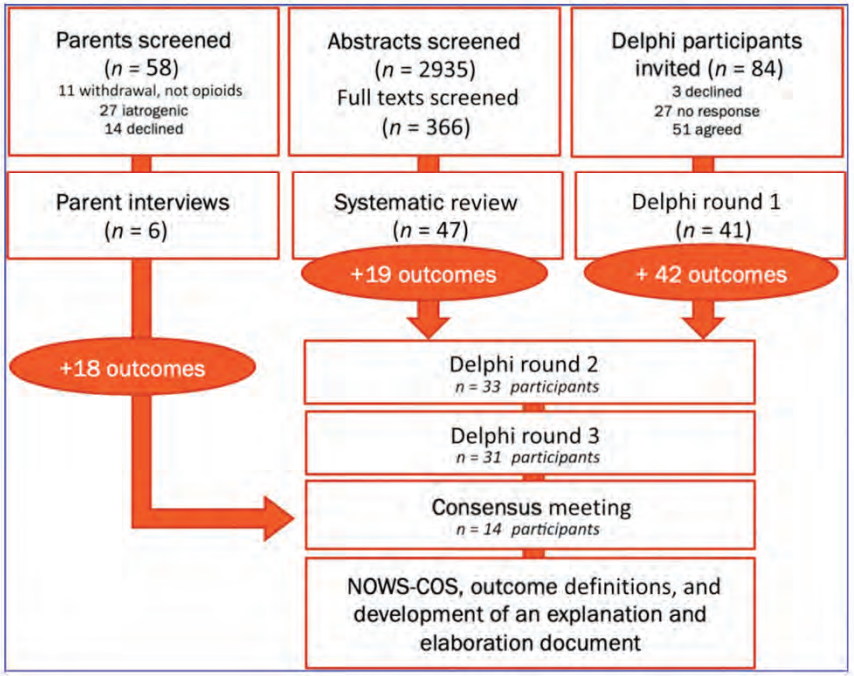


FIGURE 1
NOWS-COS overall methods and results.

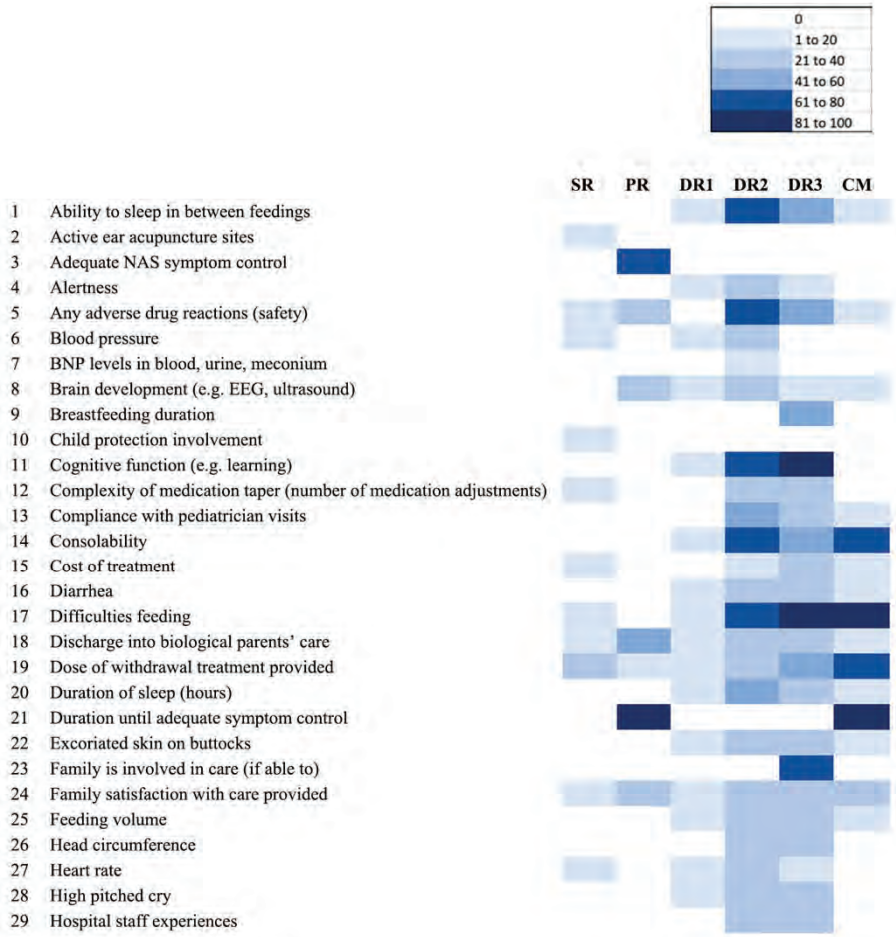


FIGURE 2
Scoring for COS candidate items. BNP, brain natriuretic peptide; CM, consensus meeting (percentage who voted in); DR1, percentage of participants who reported this outcome in response to open-ended questions; DR2, percentage of participants who reported as critical in the second round of the Delphi; DR3, percentage of participants who reported as critical in the third round of the Delphi; ED, emergency department; NEC, necrotizing enterocolitis; PR, percentage of parent who reported this outcome; SR, percent of included studies in the systematic review that reported.

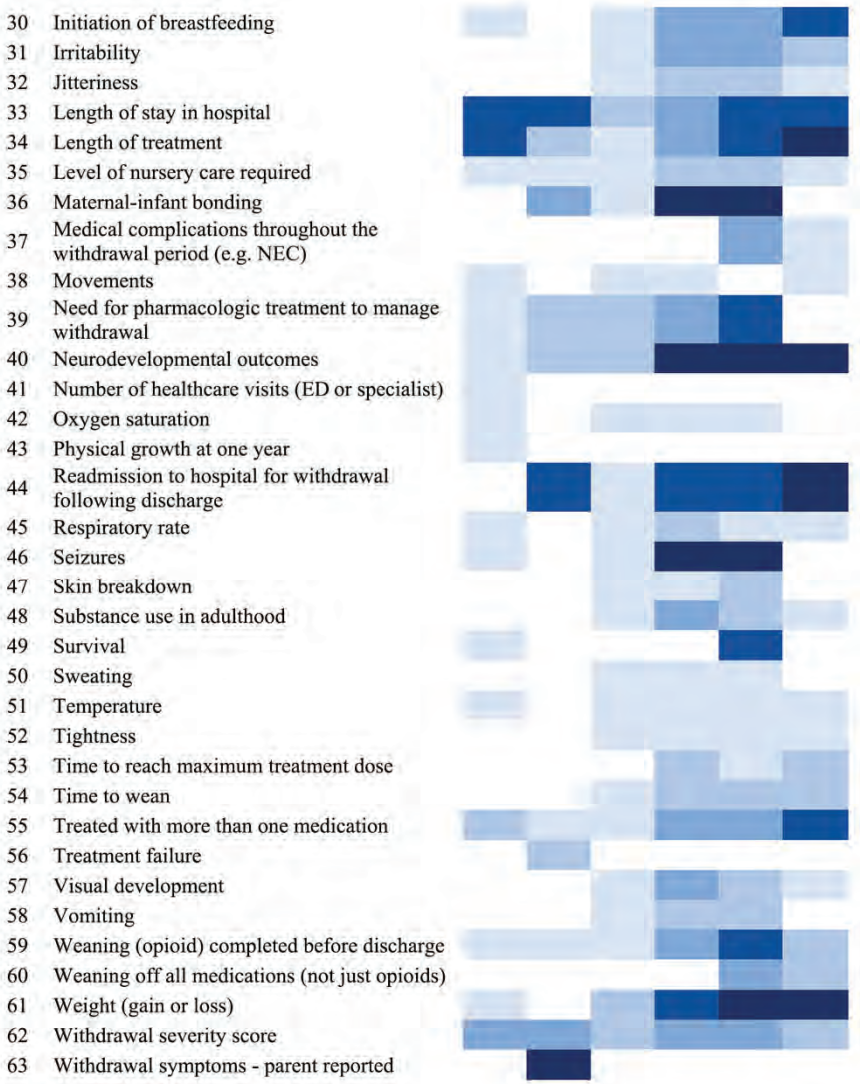


FIGURE 2
Continued.

Parent interviews

- 58 total women screened
  - 11 non-opioid withdrawal (9 SSRI, 4 BDZ)
  - 27 iatrogenic withdrawal
  - 14 declined to participate
  - 6 women recruited

- ✓ REB approval
- ✓ Consent and recorded
- ✓ Transcription
- ✓ Thematic analysis in duplicate

What do Parents think?

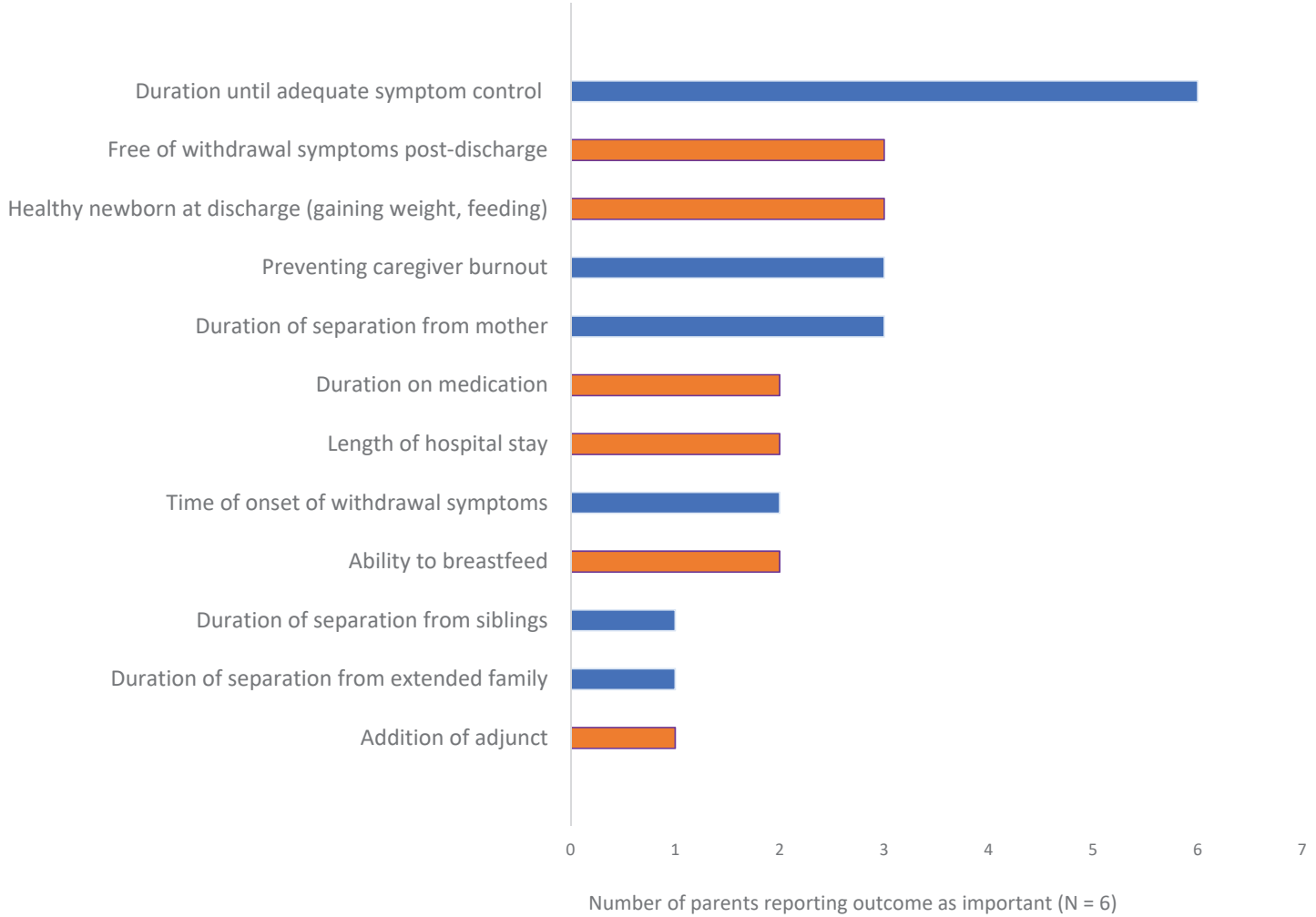


TABLE 1 COS for NOWS

Outcome	Definition
Pharmacologic treatment required to manage neonatal withdrawal, "yes" or "no"	Any oral or IV drugs given for withdrawal symptoms
Total dose of opioid treatment received to manage neonatal withdrawal, mg	Including before and after discharge if weaning at home
Duration of treatment received, d	From initiation to discontinuation, both in and out hospital
Adjuvant therapy, "yes" or "no"	More than one medication received to manage neonatal opioid withdrawal
Difficulties feeding, "yes" or "no"	Inability to feed in adequate quantities to support survival without tube or intravenous feeding intervention
Consolability, "yes" or "no"	Ability to settle in care provider's arms
Time to adequate symptom control, h	From initiation of both drug and nondrug interventions
Parent-infant bonding, "yes" or "no"	Dyadic communication: the ability of the infant to transmit meaningful cues and the capacity of the parent (birth or foster) to understand and act on them
Duration of time the neonate spent in the hospital as an inpatient (in all levels of care), No. calendar d	From admission to discharge
Receiving any breast milk at discharge, "yes" or "no"	Breastfeeding, expressed maternal or donor milk, whether orally or by nasogastric tube
Wt gain (from birth) at hospital discharge, g/kg per d	From birth to discharge
Readmission to hospital for withdrawal concerns, "yes" or "no"	Return to hospital once discharged for symptoms related with neonatal withdrawal
Neurodevelopment	Definition and measurement appropriate for time of follow-up and the child's age

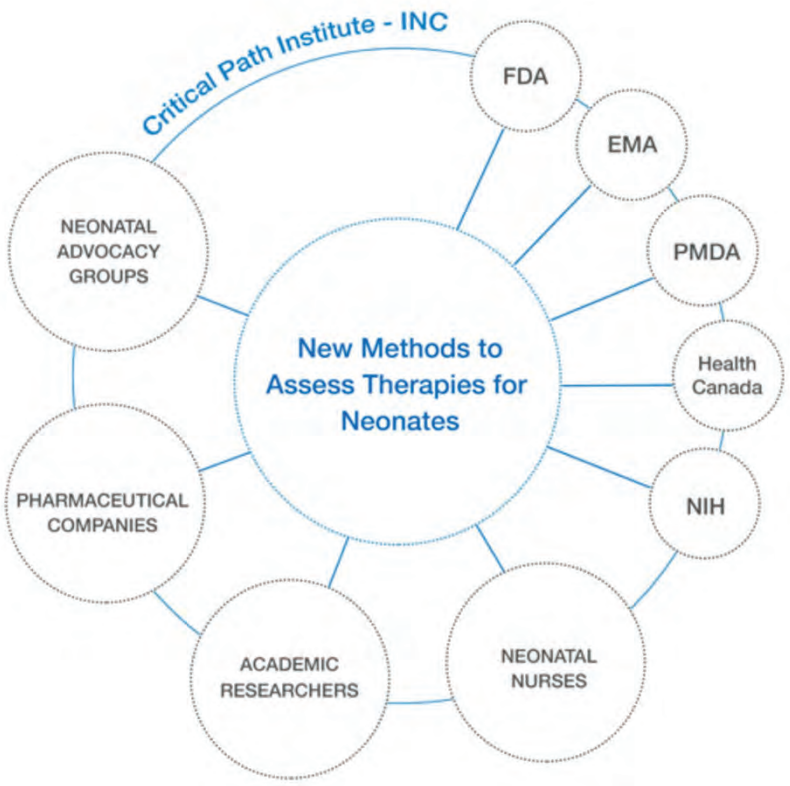
IV, intravenous.

NOWS COS

13 Outcomes With definitions

NOWS COS Knowledge Translation

- International neonatal consortium



2. COS Challenges and Considerations

1) Content Validity of the Core Set
    1) Rigor in Development
    2) Consensus on Definitions
    3) Consensus on Measurement Instruments
2) Minimal Reporting Standards

ELSEVIER

Journal of Clinical Epidemiology 155 (2023) 13–21

Journal of Clinical Epidemiology

REVIEW

Pediatric core outcome sets had deficiencies and lacked child and family input: A methodological review

Katherine Goren[a], Andrea Monsour[a], Emma Stallwood[a], Martin Offringa[a,b,c], Nancy J. Butcher[a,d,*]

[a]*Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, Ontario, Canada*
[b]*Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada*
[c]*Division of Neonatology, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada*
[d]*Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada*

Accepted 12 December 2022; Published online 15 December 2022

Abstract

Objectives: The Core Outcome Set-STAndards for Development (COS-STAD), published in 2017, contains 11 standards (12 criteria) describing minimum design criteria for core outcome set (COS) development. We aimed to identify and appraise all pediatric COS published prior to COS-STAD, and assess methods of child and family involvement in their development.

Study Design and Setting: This methodological review included documents that described the development of pediatric COS up to and including 2017. Reviewers independently assessed each COS against COS-STAD criteria, and methods of involvement were synthesized.

Results: A total of 56 pediatric COS were identified, meeting a median of five COS-STAD criteria. Nearly all met criteria on COS scope specification for setting, health condition, and population; 41% met criteria for intervention. Standards were more often met for the involvement of researchers/health professionals (64%) than for patients or their representatives (29%). Few met standards for achieving COS consensus (4–23%). Methods of child and family engagement varied and were limited.

Conclusion: A large proportion of pediatric COS developed prior to COS-STAD recommendations show gaps in design methodology. Updated and newly developed pediatric COS would benefit from the inclusion of the child and family voice, implementing a priori criteria for COS consensus, and clear reporting.

Table 2. Frequency of COS-STAD criteria met, unclear, or not met in 56 pediatric core outcome sets

Domain	Standard number	Standard	Standard met, *n* (%)	Standard unclear, *n* (%)	Standard not met, *n* (%)
Scope specification	1	The research or practice setting(s) in which the COS is to be applied	56 (100)	0 (0)	0 (0)
	2	The health condition(s) covered by the COS should be considered	55 (98)	1 (2)[a]	0 (0)
	3	The population(s) covered by the COS should be considered	54 (96)	2 (4)	0 (0)
	4	The intervention(s) covered by the COS should be considered	23 (41)	33 (59)	0 (0)
Stakeholders involved	5	Those who will use the COS in research	36 (64)	19 (34)	1 (2)
	6	Health care professionals with experience of patients with the condition	36 (64)	18 (32)	2 (4)
	7	Patients with the condition or their representatives	16 (29)	2 (4)	38 (68)
Consensus process	8	The initial list of outcomes considered both health care professionals' and patients' views	13 (23)	5 (9)	38 (68)
	9a	A scoring process was described a priori	2 (4)	48 (86)	6 (11)
	9b	A consensus definition was described a priori	6 (11)	45 (80)	5 (9)
	10	Criteria for including/dropping/adding outcomes were described a priori	2 (4)	54 (96)	0 (0)
	11	Care was taken to avoid ambiguity of language used in the list of outcomes	6 (11)	50 (89)	0 (0)

Abbreviations: COS-STAD, Core Outcome Set-STAndards development; COS, core outcome set

See Supplemental Table 2 for COS-STAD assessment criteria.

[a] One COS was scored "unclear whether standard is met" for standard 2 as it was for procedural sedation and did not specify which health condition(s) the COS was intended.

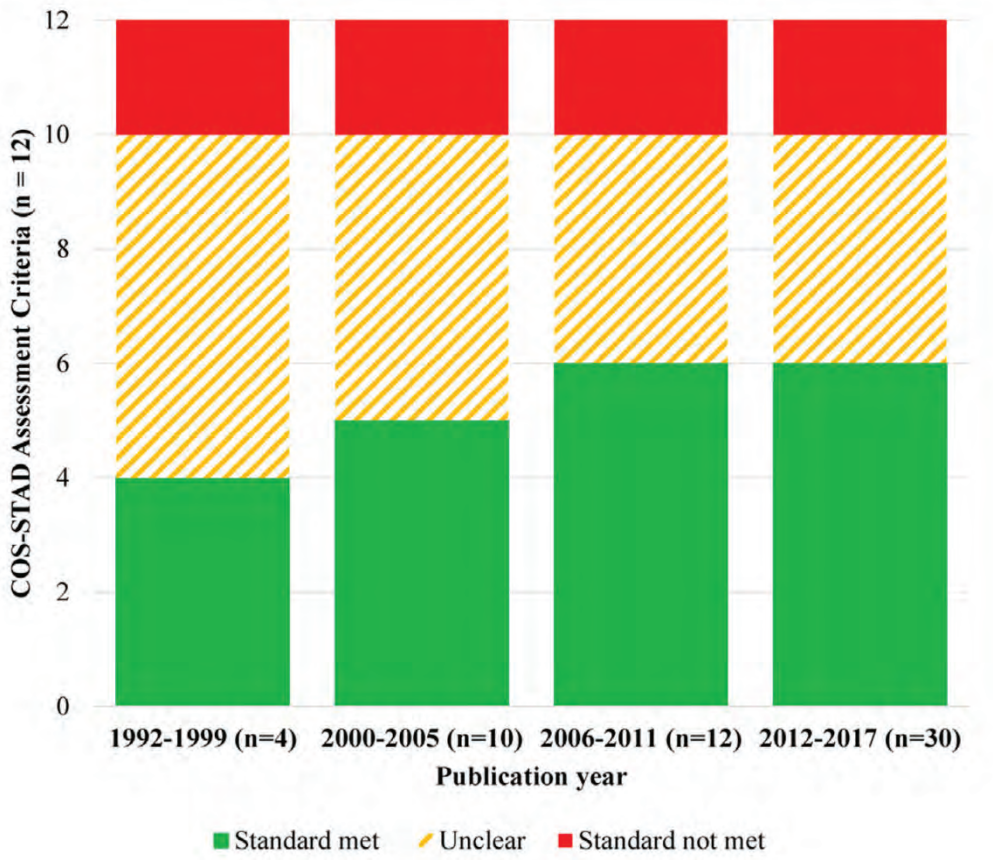


Fig. 1. Median number of Core Outcome Set STAndards for Assessment met, unclear, or not met in 56 pediatric core outcome sets published between 1992 and 2017.

Conclusions:

A large proportion of pediatric COS show important gaps in design methodology.

Pediatric COS development methods would benefit from the inclusion of the patient and family voices in the COS development process, defining and applying a priori criteria on how to reach COS consensus, and clear methods reporting.

Clinical researchers should consider using pediatric COS that have been developed according to current best practices in designing their studies.

PLOS | MEDICINE

GUIDELINES AND GUIDANCE

Core Outcome Set–STAndards for Reporting: The COS-STAR Statement

Jamie J. Kirkham[1], Sarah Gorst[1], Douglas G. Altman[2], Jane M. Blazeby[3], Mike Clarke[4], Declan Devane[5], Elizabeth Gargon[1], David Moher[6], Jochen Schmitt[7], Peter Tugwell[8], Sean Tunis[9], Paula R. Williamson[1]*

1 MRC North West Hub for Trials Methodology Research, Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom, **2** Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, United Kingdom, **3** MRC ConDuCT II Hub for Trials Methodology Research, School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom, **4** Northern Ireland Hub for Trials Methodology Research, Centre for Public Health, Queen's University Belfast, Belfast, United Kingdom, **5** National University of Ireland Galway and HRB Trials Methodology Research Network, Ireland, **6** Ottawa Methods Centre, Ottawa Hospital Research Institute, Ottawa, Canada, **7** Center for Evidence-Based Healthcare, Medizinische Fakultät, Technische Universität Dresden, Dresden, Germany, **8** Institute of Population Health, University of Ottawa, Ottawa, Canada, **9** Center for Medical Technology Policy, Baltimore, Maryland, United States of America

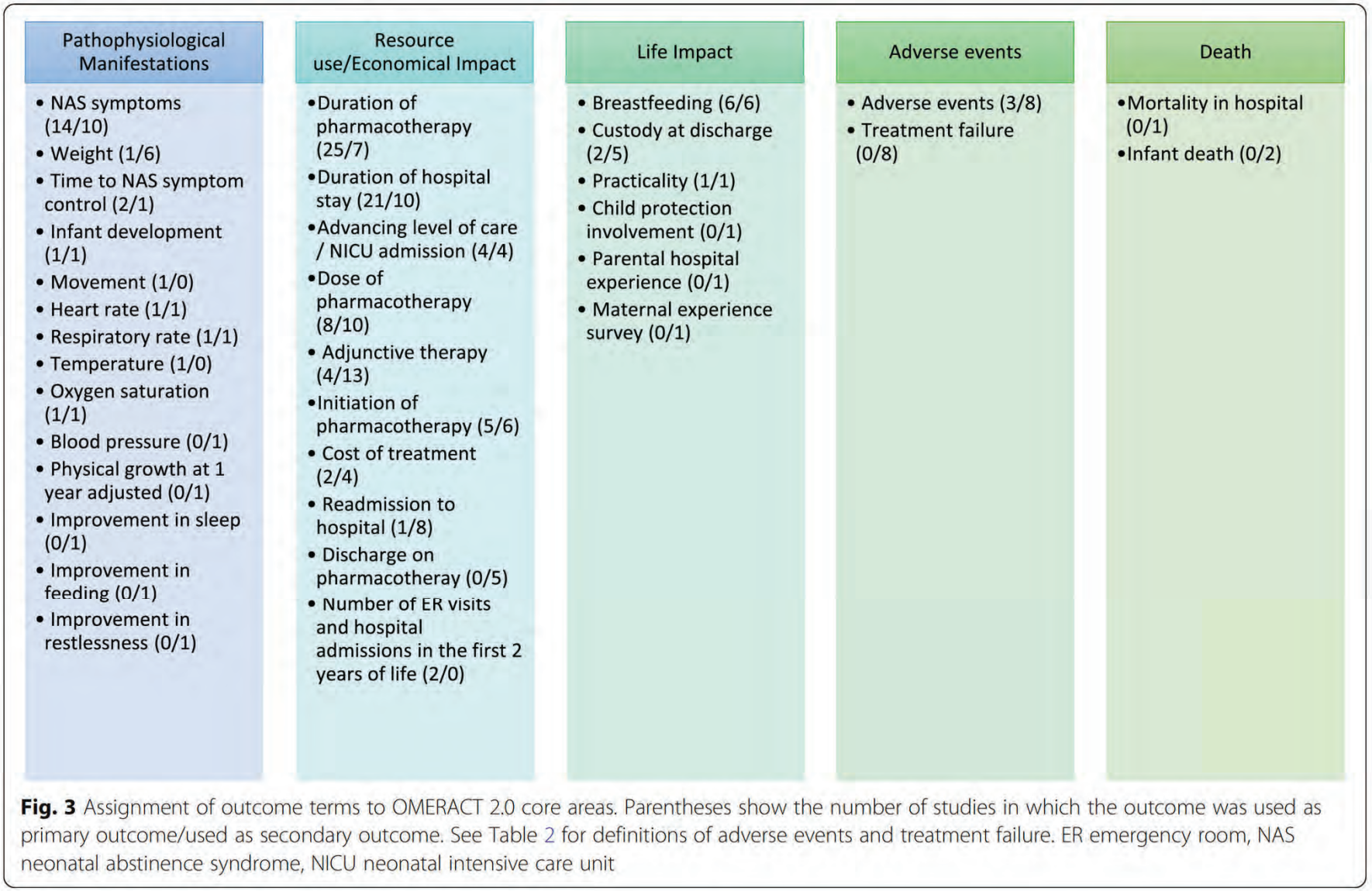


Fig. 3 Assignment of outcome terms to OMERACT 2.0 core areas. Parentheses show the number of studies in which the outcome was used as primary outcome/used as secondary outcome. See Table 2 for definitions of adverse events and treatment failure. ER emergency room, NAS neonatal abstinence syndrome, NICU neonatal intensive care unit

Original article

OPEN ACCESS

Core outcomes in neonatology: development of a core outcome set for neonatal research

James William Harrison Webbe [1], James M N Duffy,[2] Elsa Afonso,[3] Iyad Al-Muzaffar,[4] Ginny Brunton,[5] Anne Greenough [6], Nigel J Hall [7], Marian Knight [8], Jos M Latour,[9,10] Caroline Lee-Davey,[11] Neil Marlow [12], Laura Noakes,[13] Julie Nycyk,[14] Angela Richard-Löndt,[13] Ben Wills-Eve,[15] Neena Modi [16], Chris Gale [1]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/archdischild-2019-317501).

For numbered affiliations see end of article.

Correspondence to
Dr James William Harrison Webbe, Academic Neonatal Medicine, Imperial College London, London, UK;
j.webbe@imperial.ac.uk

Received 1 May 2019
Revised 1 October 2019
Accepted 30 October 2019

ABSTRACT

Background Neonatal research evaluates many different outcomes using multiple measures. This can prevent synthesis of trial results in meta-analyses, and selected outcomes may not be relevant to former patients, parents and health professionals.

Objective To define a core outcome set (COS) for research involving infants receiving neonatal care in a high-income setting.

Design Outcomes reported in neonatal trials and qualitative studies were systematically reviewed. Stakeholders were recruited for a three-round international Delphi survey. A consensus meeting was held to confirm the final COS, based on the survey results.

Participants Four hundred and fourteen former patients, parents, healthcare professionals and researchers took part in the eDelphi survey; 173 completed all three rounds. Sixteen stakeholders participated in the consensus meeting.

Results The literature reviews identified 104 outcomes; these were included in round 1. Participants proposed 10 additional outcomes; 114 outcomes were scored in rounds 2 and 3. Round 1 scores showed different stakeholder groups prioritised contrasting outcomes. Twelve outcomes were included in the final COS: survival, sepsis, necrotising enterocolitis, brain injury on imaging, general gross motor ability, general cognitive

What is already known on this topic?

- Inconsistent reporting of outcomes of limited relevance to former patients, parents and healthcare professionals is an important cause of research waste.
- There is a lack of evidence to guide many neonatal practices, leading to variation in both the care provided and outcomes for patients.
- Core outcome sets (agreed, standardised outcomes to be reported by all trials) have been developed in other fields to improve outcome selection and facilitate meta-analysis.

What this study adds?

- Former patients, parents, doctors, nurses and researchers show differences in how they prioritise neonatal care outcomes.
- We have identified 12 outcomes that are important to these stakeholders.
- If these outcomes are reported in a standardised manner by all neonatal research, this will enhance future evidence synthesis.

COIN

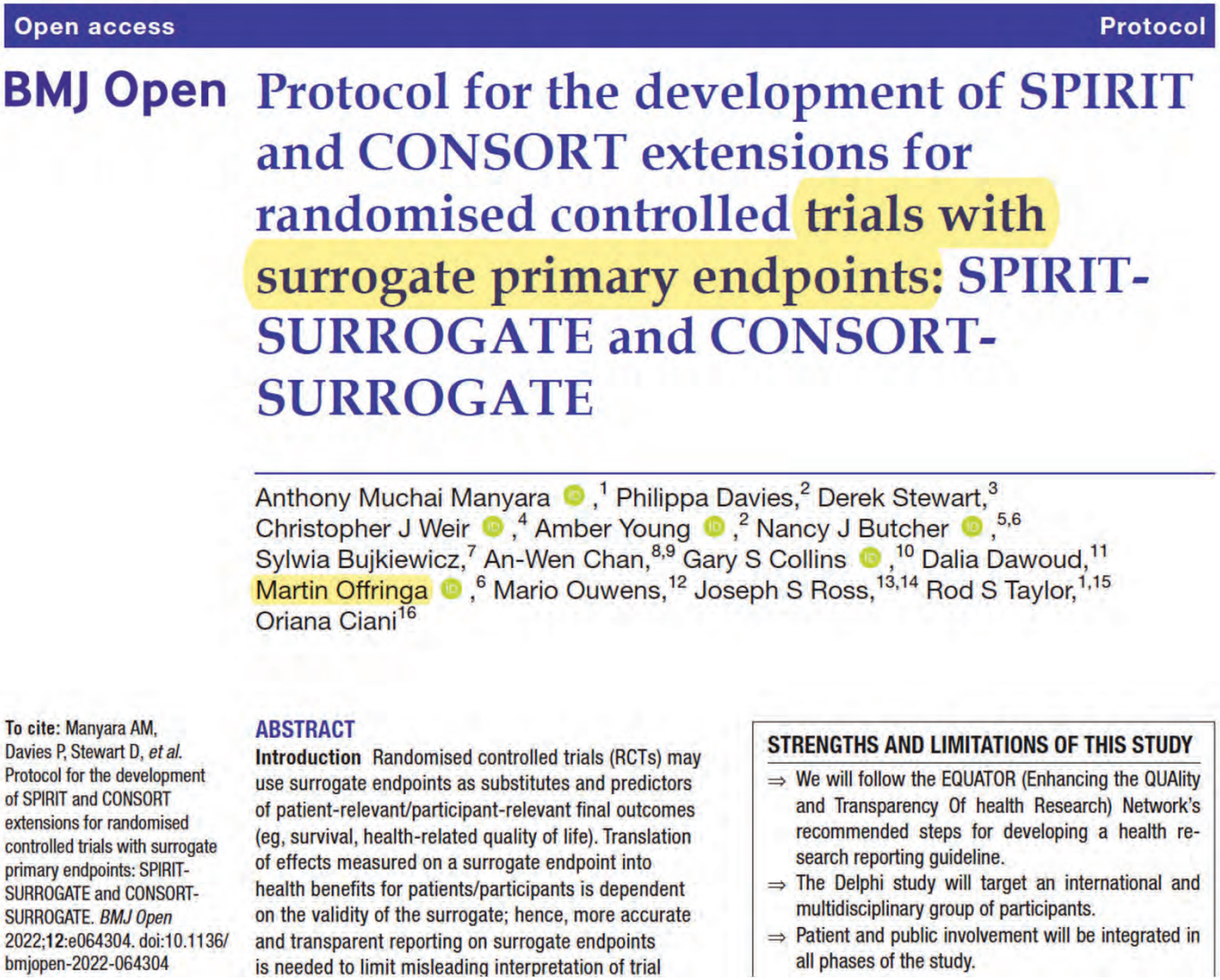
Conclusions and relevance

A COS for clinical trials and other research studies involving infants receiving neonatal care in a high-income setting has been identified.

This COS for neonatology will help standardise outcome selection in clinical trials and ensure these are relevant to those most affected by neonatal care.

Final COIN Core Outcome Set

1. Survival
2. Sepsis
3. Necrotising enterocolitis
4. Brain injury on imaging
5. General gross motor ability
6. General cognitive ability
7. Quality of life
8. Adverse events
9. Visual impairment or blindness
10. Hearing impairment or deafness
11. Retinopathy of prematurity *(preterm only)*
12. Chronic lung disease / bronchopulmonary dysplasia *(preterm only)*



Open access Protocol

BMJ Open

Protocol for the development of SPIRIT and CONSORT extensions for randomised controlled trials with surrogate primary endpoints: SPIRIT-SURROGATE and CONSORT-SURROGATE

Anthony Muchai Manyara ,[1] Philippa Davies,[2] Derek Stewart,[3] Christopher J Weir ,[4] Amber Young ,[2] Nancy J Butcher ,[5,6] Sylwia Bujkiewicz,[7] An-Wen Chan,[8,9] Gary S Collins ,[10] Dalia Dawoud,[11] Martin Offringa ,[6] Mario Ouwens,[12] Joseph S Ross,[13,14] Rod S Taylor,[1,15] Oriana Ciani[16]

To cite: Manyara AM, Davies P, Stewart D, *et al*. Protocol for the development of SPIRIT and CONSORT extensions for randomised controlled trials with surrogate primary endpoints: SPIRIT-SURROGATE and CONSORT-SURROGATE. *BMJ Open* 2022;**12**:e064304. doi:10.1136/bmjopen-2022-064304

ABSTRACT

Introduction Randomised controlled trials (RCTs) may use surrogate endpoints as substitutes and predictors of patient-relevant/participant-relevant final outcomes (eg, survival, health-related quality of life). Translation of effects measured on a surrogate endpoint into health benefits for patients/participants is dependent on the validity of the surrogate; hence, more accurate and transparent reporting on surrogate endpoints is needed to limit misleading interpretation of trial

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ We will follow the EQUATOR (Enhancing the QUAlity and Transparency Of health Research) Network's recommended steps for developing a health research reporting guideline.

⇒ The Delphi study will target an international and multidisciplinary group of participants.

⇒ Patient and public involvement will be integrated in all phases of the study.

Evolving definition of a Surrogate Endpoint:

"… a substitute for a direct measure of how a patient feels, functions, or survives. A surrogate endpoint does not measure the clinical benefit of primary interest in and of itself, but rather is expected to reliably predict that clinical benefit or harm based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence."

Glasgow, 13 March 2023

JAMA | Special Communication

Guidelines for Reporting Outcomes in Trial Protocols
The SPIRIT-Outcomes 2022 Extension

Nancy J. Butcher, PhD; Andrea Monsour, MPH; Emma J. Mew, MPH, MPhil; An-Wen Chan, MD, DPhil; David Moher, PhD; Evan Mayo-Wilson, DPhil; Caroline B. Terwee, PhD; Alyssandra Chee-A-Tow, MPH; Ami Baba, MRes; Frank Gavin, MA; Jeremy M. Grimshaw, MBCHB, PhD; Lauren E. Kelly, PhD; Leena Saeed, BSc, BEd; Lehana Thabane, PhD; Lisa Askie, PhD; Maureen Smith, MEd; Mufiza Farid-Kapadia, MD, PhD; Paula R. Williamson, PhD; Peter Szatmari, MD; Peter Tugwell, MD; Robert M. Golub, MD; Suneeta Monga, MD; Sunita Vohra, MD; Susan Marlin, MSc; Wendy J. Ungar, PhD; Martin Offringa, MD, PhD

IMPORTANCE Complete information in a trial protocol regarding study outcomes is crucial for obtaining regulatory approvals, ensuring standardized trial conduct, reducing research waste, and providing transparency of methods to facilitate trial replication, critical appraisal, accurate reporting and interpretation of trial results, and knowledge synthesis. However, recommendations on what outcome-specific information should be included are diverse and inconsistent. To improve reporting practices promoting transparent and reproducible outcome selection, assessment, and analysis, a need for specific and harmonized guidance as to what outcome-specific information should be addressed in clinical trial protocols exists.

OBJECTIVE To develop harmonized, evidence- and consensus-based standards for describing outcomes in clinical trial protocols through integration with the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) 2013 statement.

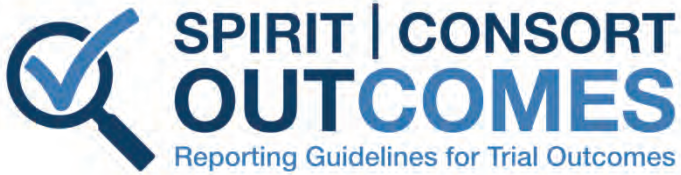
EVIDENCE REVIEW Using the Enhancing the Quality and Transparency of Health Research (EQUATOR) methodological framework, the SPIRIT-Outcomes 2022 extension of the SPIRIT 2013 statement was developed by (1) generation and evaluation of candidate outcome reporting items via consultation with experts and a scoping review of existing guidance for reporting trial outcomes (published within the 10 years prior to March 19, 2018) identified through expert solicitation, electronic database searches of MEDLINE and the Cochrane Methodology Register, gray literature searches, and reference list searches; (2) a 3-round international Delphi voting process (November 2018-February 2019) completed by 124 panelists from 22 countries to rate and identify additional items; and (3) an in-person consensus meeting (April 9-10, 2019) attended by 25 panelists to identify essential items for outcome-specific reporting to be addressed in clinical trial protocols.

FINDINGS The scoping review and consultation with experts identified 108 recommendations relevant to outcome-specific reporting to be addressed in trial protocols, the majority (72%) of which were not included in the SPIRIT 2013 statement. All recommendations were consolidated into 56 items for Delphi voting; after the Delphi survey process, 19 items met criteria for further evaluation at the consensus meeting and possible inclusion in the SPIRIT-Outcomes 2022 extension. The discussions during and after the consensus meeting yielded 9 items that elaborate on the SPIRIT 2013 statement checklist items and are related to completely defining and justifying the choice of primary, secondary, and other outcomes (SPIRIT 2013 statement checklist item 12) prospectively in the trial protocol, defining and justifying the target difference between treatment groups for the primary outcome used in the sample size calculations (SPIRIT 2013 statement checklist item 14), describing the responsiveness of the study instruments used to assess the outcome and providing details on the outcome assessors (SPIRIT 2013 statement checklist item 18a), and describing any planned methods to account for multiplicity relating to the analyses or interpretation of the results (SPIRIT 2013 statement checklist item 20a).

CONCLUSIONS AND RELEVANCE This SPIRIT-Outcomes 2022 extension of the SPIRIT 2013 statement provides 9 outcome-specific items that should be addressed in all trial protocols and may help increase trial utility, replicability, and transparency and may minimize the risk of selective nonreporting of trial results.

JAMA. 2022;328(23):2345-2356. doi:10.1001/jama.2022.21243
Published online December 13, 2022.

Supplemental content



Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Nancy J. Butcher, PhD, Peter Gilgan Centre for Research and Learning, The Hospital for Sick Children, 686 Bay St, Toronto, ON M5G 0A4, Canada (nancy.butcher@sickkids.ca).

JAMA | Special Communication

Guidelines for Reporting Outcomes in Trial Reports
The CONSORT-Outcomes 2022 Extension

Nancy J. Butcher, PhD; Andrea Monsour, MPH; Emma J. Mew, MPH, MPhil; An-Wen Chan, MD, DPhil; David Moher, PhD; Evan Mayo-Wilson, DPhil; Caroline B. Terwee, PhD; Alyssandra Chee-A-Tow, MPH; Ami Baba, MRes; Frank Gavin, MA; Jeremy M. Grimshaw, MBCHB, PhD; Lauren E. Kelly, PhD; Leena Saeed, BSc, BEd; Lehana Thabane, PhD; Lisa Askie, PhD; Maureen Smith, MEd; Mufiza Farid-Kapadia, MD, PhD; Paula R. Williamson, PhD; Peter Szatmari, MD; Peter Tugwell, MD; Robert M. Golub, MD; Suneeta Monga, MD; Sunita Vohra, MD; Susan Marlin, MSc; Wendy J. Ungar, MSc, PhD; Martin Offringa, MD, PhD

Supplemental content

IMPORTANCE Clinicians, patients, and policy makers rely on published results from clinical trials to help make evidence-informed decisions. To critically evaluate and use trial results, readers require complete and transparent information regarding what was planned, done, and found. Specific and harmonized guidance as to what outcome-specific information should be reported in publications of clinical trials is needed to reduce deficient reporting practices that obscure issues with outcome selection, assessment, and analysis.

OBJECTIVE To develop harmonized, evidence- and consensus-based standards for reporting outcomes in clinical trial reports through integration with the Consolidated Standards of Reporting Trials (CONSORT) 2010 statement.

EVIDENCE REVIEW Using the Enhancing the Quality and Transparency of Health Research (EQUATOR) methodological framework, the CONSORT-Outcomes 2022 extension of the CONSORT 2010 statement was developed by (1) generation and evaluation of candidate outcome reporting items via consultation with experts and a scoping review of existing guidance for reporting trial outcomes (published within the 10 years prior to March 19, 2018) identified through expert solicitation, electronic database searches of MEDLINE and the Cochrane Methodology Register, gray literature searches, and reference list searches; (2) a 3-round international Delphi voting process (November 2018-February 2019) completed by 124 panelists from 22 countries to rate and identify additional items; and (3) an in-person consensus meeting (April 9-10, 2019) attended by 25 panelists to identify essential items for the reporting of outcomes in clinical trial reports.

FINDINGS The scoping review and consultation with experts identified 128 recommendations relevant to reporting outcomes in trial reports, the majority (83%) of which were not included in the CONSORT 2010 statement. All recommendations were consolidated into 64 items for Delphi voting; after the Delphi survey process, 30 items met criteria for further evaluation at the consensus meeting and possible inclusion in the CONSORT-Outcomes 2022 extension. The discussions during and after the consensus meeting yielded 17 items that elaborate on the CONSORT 2010 statement checklist items and are related to completely defining and justifying the trial outcomes, including how and when they were assessed (CONSORT 2010 statement checklist item 6a), defining and justifying the target difference between treatment groups during sample size calculations (CONSORT 2010 statement checklist item 7a), describing the statistical methods used to compare groups for the primary and secondary outcomes (CONSORT 2010 statement checklist item 12a), and describing the prespecified analyses and any outcome analyses not prespecified (CONSORT 2010 statement checklist item 18).

CONCLUSIONS AND RELEVANCE This CONSORT-Outcomes 2022 extension of the CONSORT 2010 statement provides 17 outcome-specific items that should be addressed in all published clinical trial reports and may help increase trial utility, replicability, and transparency and may minimize the risk of selective nonreporting of trial results.

JAMA. 2022;328(22):2252-2264. doi:10.1001/jama.2022.21022

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Nancy J. Butcher, PhD, Peter Gilgan Centre for Research and Learning, The Hospital for Sick Children, 686 Bay St, Toronto, ON M5G 0A4, Canada (nancy.butcher@sickkids.ca).

Table 2. The 5 Core Elements of a Defined Outcome[a]

Element No.	Element term	Definition used	Example 1	Example 2	Example 3
1	Domain[b]	Title or concept to describe ≥1 outcomes	Blood pressure	Depression	Death
2	Measurement variable or specific measurement	Corresponds to the data collected directly from the trial participants; description includes the instrument used to assess the outcome domain			
		• Descriptive name	Systolic blood pressure measured with Omron upper arm blood pressure monitor	MADRS	All-cause mortality per the hospital database
		• If applicable, the total score or the subscales that will be analyzed	Not applicable	MADRS total score	Not applicable
3	Specific metric	Participant-level unit of measurement (eg, change from baseline, final value or a value at a time point, time to event) for the analysis	Value at a time point	Change from baseline	Time to event
4	Method of aggregation	The procedure for estimating the treatment effect			
		• If the outcome will be treated as a continuous, categorical, or time-to-event variable	Continuous variable	Binary variable	Time to event
		• For continuous variables, a measure of central tendency (eg, mean value); for categorical and time-to-event data variables, proportion with an event and, if relevant, the specific cutoff values or categories compared	Mean value	Proportion of participants with ≥50% decrease	Incidence density and between-group incidence density rate
5	Time point	The timing of follow-up measurements			
		• When outcome measurements will be obtained	2, 4, and 12 wk after randomization	2, 4, 6, and 8 wk after randomization	Daily
		• Which of the outcome measurements will be analyzed	12 wk after randomization	8 wk after randomization	End of follow-up

Abbreviation: MADRS, Montgomery-Åsberg Depression Rating Scale.

[a] Content adapted from Zarin et al,[39] Mayo-Wilson et al,[10] and Chan et al.[35]

[b] An explicit and specific description of the outcome domain should be provided in the trial protocol, as appropriate, when defining the trial outcome. If an outcome domain is broad, such as pain, a specific protocolized domain definition might be the daily average of the intensity of the sensation of pain expressed on a range from no pain to worst pain imaginable over a 24-hour window during an average day.[40]

COS Best Practice

Develop COS using uniform methods

Generate transparent reports (use COS-STAR)

Invest* in Definitions, Measurement Properties, and the 5 Core Elements of a Defined Outcome

Make harmonized efforts to validate

“surrogate endpoints”

Session 2 Discussion Questions

- What challenges exist in measuring efficacy and selecting endpoints for neonatal RCTs? How have these challenges impacted meaningful evidence generation?
- What are the best approaches for developing core outcome sets for key neonatal conditions and how can core outcome sets be used in demonstrating efficacy?
- What are the best approaches for justifying proposed efficacy endpoints for a neonatal trial?

Session 2: Challenges in Measuring Efficacy for Neonatal Conditions with Unmet Clinical Needs

Moderator: An Massaro, U.S. Food & Drug Administration

Lunch Break – 1 Hour

Session 3 will begin promptly at 1:00 pm