Cost-effectiveness of Alternative NICU Designs

by

Timothy Disher

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
August 2020

Table of Contents

List of Tables

List of Figures

Abstract

The design of the physical space of the neonatal intensive care unit (NICU) has been proposed as a target for intervention to improve neonatal outcomes. Most NICUs are currently designed so that infants are cared for in a shared space which is commonly referred to as an open bay design. The purpose of this study was to identify the cost-effective NICU design from a Canadian public payer perspective using a lifetime time-horizon. The study used a decision model approach informed by individual participant data for estimating the baseline history of disease and the impact of morbidities on length of stay. Efficacy parameters were informed through analysis from an ongoing randomized controlled trial (RCT) of single-family room design and a systematic and targeted literature search. Meta-analysis of efficacy parameters was conducted using a multivariate network meta-analysis. A stochastic multi-criteria acceptability analysis (SMAA) was conducted to provide an additional perspective limited to clinical outcomes of alternative room designs in addition to an ordinal ranking of their degree of family centeredness. The network meta-analysis included a total of 25 studies (2 RCTs, 23 observational studies) evaluating old open bay, new open bay, half-wall, private room, combined single-family room and open bay, and single-family room only designs. When using a life-time horizon all designs had a higher expected value of net monetary benefit than old open bay designs although there was substantial uncertainty in relative ranking and the probability that any given design was cost-effective. Half-wall designs had the highest mean expected value of net monetary benefit over a wide range of values of willingness to pay. Conclusions were robust to sensitivity analyses, including the results of the SMAA. Results were limited by the strong assumptions required to create a connected network that assessed all outcomes of interest. The decision to undertake new construction and the decision on which design aspects to integrate should reflect the complex and multi-factorial nature of the decision problem.

List of Abbreviations Used

Abbreviation	Definition
BPD	Bronchopulmonary Dysplasia
BSID	Bayley Scales of Infant and Toddler development
CAD	Canadian Dollar
CADTH	Canadian Agency for Drugs and Technology in Health
CI	Confidence Interval
CIHI	Canadian Institute for Health Information
CINAHL	Cumulative Index to Nursing and Allied Health Literature
CN-FUN	Canadian Neonatal Follow-up Network
CNN	Canadian Neonatal Network
CP	Cerebral palsy
EQ-5D	EuroQoL 5-D
GA	Gestational Age
GMFCS	Gross Motor Function Classification System
ICD	International Classification of Disease
ICER	Incremental Cost-Effectiveness Ratio
ICU	Intensive Care Unit
ICUR	Incremental Cost-Utility Ratio
IQR	Inter-Quartile Range
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
IVH	Intraventricular Hemorrhage
IWK	Izaak Walton Killam
JAGS	Just Another Gibbs Sampler
MD	Mean Difference
MEDLINE	Medical Literature Analysis and Retrieval System Online
NA	Not Available
NDI	Neurodevelopmental Impairment
sNDI	Severe Neurodevelopmental Impairment
NEC	Necrotising Enterocolitis
NICE	National Institute for Health and Care Excellence
NICU	Neonatal Intensive Care Unit
NIDCAP	Newborn Individualized Developmental Care and Assessment Program
NMA	Network Meta-Analysis
OECD	Organisation for Economic Co-operation and Development
OR	Odds Ratio
PDA	Patent Ductus Arteriosus
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSA	Probabilistic Sensitivity Analysis
QALY	Quality-adjusted Life Year
RCT	Randomized Controlled Trial
ROP	Retinopathy of Prematurity
SD	Standard Deviation
SFR	Single Family Room
SMAA	Stochastic Multi-criteria Acceptability Analysis
SMD	Standardized Mean Difference
SNAPPE-II	Score for Neonatal Acute Physiology with Perinatal Extension-II
SSC	Skin-to-skin contact

Abbreviation	Definition
UDC	Universe of Developmental Care
UK	United Kingdom
USD	United States Dollar

Chapter 1: Introduction

Intensive Care Unit Admissions in Canada

In Canada, approximately 60% of neonatal intensive care unit (NICU) admissions are for infants delivered preterm (Canadian Neonatal Network, 2019). Preterm birth is defined as an infant delivered before the completion of 37 weeks post-conceptual age (World Health Organization, 2018). The World Health Organization classifies sub-categories of preterm birth based on post conception gestational age measured in weeks. Extremely preterm infants are those born at less than 28 weeks, very preterm from 28 to less than 32 weeks, and moderate to late preterm from 32 to less than 37 weeks (World Health Organization, 2018). The rates of preterm live births in Canada have generally remained stable at approximately 8% between 2000 and 2013 (Statistics Canada, 2016). Based on the most recent Canadian birth estimates, this translates to roughly 31,193 preterm infants born across Canada. The Canadian Neonatal Network (CNN) was founded in 1995 and maintains a database of admissions and outcomes of all infants admitted to one of its participating partner sites across Canada. According to the 2019 CNN Annual Report (Canadian Neonatal Network, 2019) approximately 16% of preterm births admitted to participating NICUs were extremely preterm, 19% very preterm, and 65% moderate to late preterm.

Characteristics of term born infants requiring intensive care treatment are less well documented in Canada. While the CNN does report common serious diagnoses for this cohort, they do not document the primary reason for admission. In the 2018 report, term infants accounted for 41.3% of all included NICU admissions. Of note, approximately 3% of term infants admitted to Canadian NICUs are diagnosed with respiratory distress syndrome, 6.3% are diagnosed with a pneumothorax, 4.7% require surgery, and 12.1% are diagnosed with a major

congenital anomaly (e.g. spina bifada, hypoplastic left heart syndrome). In a retrospective cohort of 133,691 term babies across 163 neonatal units in England the primary reasons for admission were respiratory disease (24%), infection (18%), hypoglycaemia (10%) and jaundice (5%). The median (IQR) total length of stay was 5 (3-7) days (Battersby et al., 2017). As a result of the general low acuity of these admissions, most literature concerning NICU design has focused on impacts on preterm infants.

The Burden of Prematurity in Canada

According to the 2019 CNN Annual Report (Canadian Neonatal Network, 2019) survival to discharge reaches 90% by 26 completed weeks, and 98% by 29 completed weeks (Canadian Neonatal Network, 2019). These survival rates have been stable or improving across most gestational age groups for the last eight years, with the most preterm babies seeing the greatest substantial gains in survival. For example, those born at 23 weeks and receiving intensive care have seen an improvement from ~32% survival in 2010 to nearly 49% in 2018 (Canadian Neonatal Network, 2019). Despite these improvements in overall survival, there has been less consistent progress in the reduction of major morbidities.

The CNN defines major morbidities as chronic lung disease (also known as bronchopulmonary dyspasia [BPD]), severe retinopathy of prematurity (ROP, stage 3-5 (Jefferies & Canadian Paediatric Society: Fetus and Newborn Committee, 2016) and/or those who receive treatment), severe neurological injury (Intraventricular Hemorrhage (IVH) grade 3 or grade 4 or periventricular leukomalacia), necrotizing enterocolitis (NEC, Bell stage two or greater (Walsh & Kliegman, 1986)), and late onset sepsis (positive culture after two days of age). While some gestational age groups have seen significant improvements in certain outcomes, these gains have been partially offset by matching decline elsewhere. As a result, the percentage of infants with

mortality or major morbidity has remained nearly unchanged since 2010 (Canadian Neonatal Network, 2016). This trend of improved overall survival accompanied by heterogeneous change in rates of major morbidities across preterm subgroups and modest or no change in rates of survival without major morbidity is consistent with findings from other countries. The most recent evaluation of trends in morbidity and mortality of extremely preterm infants in the United States saw an increase in survival from 70-79% between 1993 and 2012, with those born at 25-28 reaching comparable levels to those in Canada (81-94%) (Stoll et al., 2015). Only those born 27-28 weeks, however, have seen considerable gains in survival to discharge without major morbidity (29% in 1993 to 47% in 2012, and 38% in 1993 to 56% in 2012 respectively). Infants less than 27 weeks have seen only small improvements (1-4%) (2015). Of these morbidities, only late-onset sepsis has seen a large reduction (38% in 1993 to 20% in 2011), with IVH and periventricular leukomalacia seeing very small downward trends, ROP and NEC remaining virtually unchanged, and chronic lung disease seeing a significant increase (32% vs 42%).

This trend towards increased survival with limited improvements in rates of morbidity is important for two main reasons (a) major morbidities are associated with later disability and developmental delays, and (b) their treatment complicates and extends hospital treatment leading to increased treatment-related costs (Synnes et al., 2017). The Canadian Neonatal Follow-up Network (CN-FUN) monitors preterm infants at elevated risk for developmental delays. Two important outcomes that they monitor are the rates of neurodevelopmental impairment (NDI) and serious NDI (sNDI). Categorization is based primarily on the Bayley Scales of Infant and Toddler Development-Third Edition (BSID-III) (Bayley, 2006). This standardized scale evaluates motor, cognitive, and language development with CN-FUN using NDI cut offs for scores less than 85 and sNDI for scores less than 70. Children with mixed hearing loss or unilateral visual

impairment are considered to have NDI whereas hearing aid/cochlear implant or bilateral vision impairment are deemed sNDI. In their 2019 annual report following 5842 NICU survivors born less than 29 weeks between 2009 and 2016 with follow-up data at 21 months corrected age, 28% met the cut off for mild-moderate NDI and 17% were classified as having significant NDI (Canadian Neonatal Follow-Up Network, 2019). A separate analysis of a subset of this cohort found chronic lung disease, NEC, and brain injury were all retained in a stepwise logistic regression model predicting NDI, and ROP, sepsis, and brain injury were all retained the model for sNDI (Synnes et al., 2017). The domain most commonly associated with a score indicating NDI in this cohort was language development, a finding consistent with an independent cohort of very preterm infants followed from two to 13 years of age (Nguyen et al., 2018). Two considerations of the results of findings based on a mix of Bayley III scores and clinical findings are relevant: The first is the need to recognize the heterogeneity of disability definitions across networks (Haslam et al., 2018), and the second is that disability in later life is difficult to predict based on assessments in early childhood. Haslam and colleagues (Haslam et al., 2018) assessed the incidence of neonatal disability by applying criteria used by a range of follow-up programs and found that incidence of severe NDI ranged from 3.5% using the most stringent definition to 14.9% using the least stringent. The ability of the Bayley III scales to predict school age cognitive delay was evaluated in a convenience sample of children born preterm assessed at 18 and 30 months, with a cut off of 2 SDs (<80) used to classify a screen as positive. Results were compared to those from the Wechsler Intelligence Scale for Children administered at 6-8 years of age. While the Bayley scales had modest sensitivity (53%) and adequate specificity (88%), the low prevalence of cognitive delay (0.07%) meant the scale had low positive predictive value (24%), but high negative predictive value (96%) (Schonhaut et al., 2020).

The financial costs associated with preterm birth in Canada are significant, and primarily driven by the cost of care in hospital (Johnston et al., 2014). The only report that quantifies this burden in a Canadian cohort estimated that the total national cost for early preterm infants is $123.3 million per year, with moderate preterm and late preterm estimates of $255.6, and $208.2 million respectively (2014 CAD) (Johnston et al., 2014). In this economic model, authors estimated costs associated with care from a Quebec database, with disability and resource use in later life modeled using figures adapted from an analysis of the United Kingdom and Wales (Mangham et al., 2009). The model followed children in this cohort until an age of 10 years. Contributors to hospital costs are multifactorial, but the total direct cost is roughly proportional to the intensity of resources required to care for a given infant multiplied by the length of their hospital stay. Infants who experience major morbidity require more resource intensive care (e.g. increased use of nursing time), and models of incremental costs associated with morbidities conducted in the United States reflect this (Johnson et al., 2012). Using a case-costing approach, Johnson et al. (2012) found that the incremental increase in costs associated with major morbidities were $31,565, $15,440, $12,048, and $10,055 for chronic lung disease, NEC, brain injury, and sepsis respectively (2011 USD). Beyond costs, a 2011 Finnish cross-sectional study four-year-old children born very preterm found small (0.4 to 0.3) four-year quality adjusted life-year (QALY) decrements associated with major morbidity experienced during hospitalization (Korvenranta et al., 2010).

Role of environment

The womb provides the developing fetus with the appropriate stimulation required for optimal development. The fetus is exposed primarily to low-frequency sounds from the mother and her environment, and visual stimuli are attenuated (Haumont, 2013). Developing organs

including the skin, respiratory, and gastrointestinal systems are protected from noxious stimuli until they reach maturity. At birth, parents and babies are kept together, promoting breastfeeding and interactions that lay the early foundations for attachment and language development (Pineda, Guth, et al., 2017). Preterm birth interrupts this process, exposing the infant to loud, high-frequency sound, bright light absent the regular rhythm of day and night, frequent skin-breaking procedures, separation from parents, and crowded environments. Since the 1980s, considerable theory generating work has been conducted to relate these environmental insults to short and long-term effects on infants and generate protective interventions (Als, 1982; Gibbins et al., 2008).

Of 466 reviews maintained by the Cochrane neonatal group six are directly related to the potential effects of NICU environment on outcomes of interest in neonatal populations. This includes one review of the effect of light on ROP (Jorge et al., 2013), one concerned with developmental care while in the NICU (an approach to care that attempts to mitigate negative effects of the environment) (Symington & Pinelli, 2006), and four focused on the physical environment of the NICU (Morag & Ohlsson, 2017). Thus, less than two percent of Cochrane neonatal group reviews are concerned with the potential benefits associated with the physical environment of the NICU. Most of the reviews contain a small number of studies, with small numbers of infants in each. A review of kangaroo mother care – defined as vertical, ventral skin-to-skin contact (preferably continuous) between the infant and their mother/caregiver's bare chest (skin to skin contact [SSC]), frequent/exclusive breastfeeding, and attempts at early discharge (World Health Organization, 2016) – focused primarily on mortality and morbidity was the sole review evaluating an intervention that is implemented by parents (Conde-Aguedelo & Díaz-rossello, 2016). Only cycled lighting (length of stay) (Morag & Ohlsson, 2017), and

kangaroo mother care (breastfeeding in high-income countries) (Conde-Aguedelo & Díaz-rossello, 2016) has statistically significant evidence of benefit as sole interventions in high-income countries. Recent trials of interventions targeted at improving parental presence and involvement in care (O'Brien et al., 2018; Welch et al., 2013) suggest improvement in both neonatal and parent outcomes.

An emerging line of research that combines efforts to improve the care environment for infants and increase parent involvement has been the physical design of the NICU itself (Aija et al., 2019; White, 2010). Of the designs described in the literature, private room care, where each infant is provided their own room with or without additional living space for parents (the latter referred to as single family room design), has received considerable attention (Shahheidari & Homer, 2012; van Veenendaal et al., 2019). Research suggests private rooms can reduce the length of stay, improve breastfeeding rates, reduce infections, and decrease the occurrence of re-admissions while limiting noxious environmental stimuli and allowing greater individualized care (Shahheidari & Homer, 2012; van Veenendaal et al., 2019). These findings are the basis of a paradigm shift towards NICU designs that are intended to support infant development and encourage parent presence and involvement in care.

The problem

Based on these findings, many of the largest NICUs in Canada have committed to or are considering new construction including private rooms. A broader movement towards new room design is a pressing issue with calls to build new private room NICUs from some academic circles becoming increasingly urgent, and direct (Stevens et al., 2015). The current environment, therefore, has the potential to create the impression that the science appears settled, despite lingering questions regarding which design represents the best use of scarce resources. The

potential impact of inappropriate design is significant, with the Canadian Association of Neonatal Nurses' most recent list of Canadian NICUs sitting at 177. Thus, while it seems clear that a change from crowded traditional NICU design is needed, there is also a need to consider the available options carefully. For example, private rooms can further be categorized into private rooms without a sleep space for parents (simply private rooms) or designs that include a parent sleep space and thus focusing on the entire family (i.e. single-family rooms).

Despite substantial research effort, there are several major and minor limitations remaining in the literature intended to help Canadian decision makers choose the optimal NICU design. The most serious issue is the lack of a robust clinical and economic evaluation of the expected benefits and costs of alternative NICU designs in the Canadian setting. The Canadian NICU context is unique, and the use of costs from existing American sources may over or underestimate the potential value of NICU construction. Beyond this Canadian specific concern, existing economic analyses have focused entirely on costs and have not appropriately explored potential adverse effects over a sufficient time horizon using the cost-utility analysis approach recommended by the Canadian Agency for Drugs and Technology in Health (CADTH, 2016). Pineda et al. (2014) identified decreased BSID-III scores in infants cared for in private rooms, and analyses that focus purely on costs ignore these impacts as well as the potential benefits related to decreased morbidity. Furthermore, trial designs in which modern units were compared against crowded open bays with low square footage per baby have dominated published assessments of private room care, creating an artificial dichotomy of choices when there are several design layouts available.

Other important data for decision makers which have yet to be included in analysis are (a) the potential burdens and health effects experienced by families, (b) concerns for health

equity, and (c) consideration of potential differences in benefit for subgroups of infants. The proposed dissertation will seek to address these issues by combining data from published literature, national healthcare cost databases, and individual participant data from ongoing and planned research to take place during the IWK Health Centre's (Halifax, NS, Canada) move from an open bay to a single family room care design. The IWK Health Centre is a tertiary care hospital that includes a level III/IV NICU that serves as the regional referral hospital for Nova Scotia, New Brunswick, and Prince Edward Island.

Objectives

The primary objective of this research dissertation is to create a decision model to compare the cost-effectiveness of alternative types of NICU design. The secondary objective was to explore of the potential for some groups of families to experience a more significant share of the risk of benefit or harm.

Chapter 2: Literature Review

Chapter Two expands on the concepts outlined in Chapter One and is intended to provide the necessary clinical and methodological background for the dissertation. The first section of the chapter consists of a review of the clinical evidence covering (a) the history and rationale for consideration of NICU design; (b) the theoretical underpinnings of the clinical benefits of private room care, (c) empirical evidence of benefits associated with alternative designs, (d) potential mediators and moderations of effects, and (e) issues of health equity. The second section provides a review of economic evidence including (a) key concepts vital to the understanding of economic evaluations of health technologies, (b) review of existing economic evidence for NICU designs.

Clinical Review

NICU design. The consideration of the potential effects of NICU design on patient outcomes has been simultaneously influenced by the adult and pediatric literature as well as a shift in care to consider developmentally appropriate care that first became influential in the early 1980s (Als, 1982). In Canada and most countries around the world, NICUs take the form of large ward-style rooms, with minimal square footage dedicated to each care space. Open bay style designs are made primarily with the care-staff in mind, as they allow for easy monitoring of and access to all infants in an area (Brown & Taquino, 2001). As early as 1981, investigators recognized the potential effect of the physical environment on neonatal outcomes, with reports from Goldmann, Durbin, and Freeman (1981) and Larson, Hargiss, and Dyk (1985) linking increased square footage per baby to reduction in nosocomial sepsis. In response to growing calls for greater consideration of how the physical space can support developmentally supportive and family-centred care, most contemporary research is concerned with evaluation of new private,

semi-private, and single family rooms (Stevens, Thompson, Helseth, & Pottala, 2015; White & Whitman, 1992).

History and rationale for private room care. While the first suggestion that NICUs use private rooms appeared in the academic literature in the early 1990s, there is anecdotal evidence that they were first implemented almost a decade earlier in hopes of reducing infections (Walsh et al., 2006). Initial publications documented how the transition team supported the move to private rooms, provided theoretical justification for how the new unit will benefit infants and parents, and highlighted potential issues for staff (Brown & Taquino, 2001). Staff in this initial unit were concerned with difficulty monitoring patients and other team members and easy access to required equipment (2001). The perceived potential benefits for infants and families paired with manageable staff drawbacks led to construction of new units despite the lack of empirical evidence (White, 2003). It was not until almost ten years after the first publications of private room experiences that the first and only published randomized controlled trial of private room care was conducted in a level II unit in Sweden (Ortenstrand et al., 2010). Parents were required to be present for 24 hours a day in order to be eligible to be included and were there on average 22 hours a day. Findings included a reduction in the mean length of stay by 5.3 days for the entire cohort and 10.1 days for infants born at less than 30 weeks gestational age (2010). While no difference in infections were observed, there was a reduction in severe bronchopulmonary dysplasia (2010). Since then, the positive effects of NICU design in general, and private room care specifically, have been established in numerous publications in units from Europe and the United States (van Veenendaal et al., 2019). When compared against older open-bay designs, private rooms have been associated with far reaching benefits observed in neonatal (breastfeeding after discharge, reduced apnea, sepsis, total parenteral nutrition days, time to full

enteral feeds, reduced re-hospitalization), family (improved satisfaction, more presence, improved involvement in care), and staff (subjective assessment of improved environment and quality of care) outcomes (Shahheidari & Homer, 2012; van Veenendaal et al., 2019). The only costing study published to date has found that the daily direct costs of private room care are likely at least no more expensive, and potentially cost savings for the healthcare system (Stevens et al., 2014). These initial publications captured many of the hopes for private room designs but also drew attention to what would become some of the largest challenges: staff and parents could feel isolated, and the high price of rooms would need to be justified with objective improvements in health outcomes (White, 2003). Importantly, these early publications highlighted how entirely private designs might not be necessary to achieve some of the health benefits of private rooms (White, 2003). While some argue that the current evidence base provides enough justification for new construction, alternative designs that may share many of the same underlying benefits have received less attention and should be considered as potential options.

Developmental care. Some of the initial arguments for a change in NICU design were based on the ethical imperative implied by developmental care (White & Whitman, 1992). A concept analysis of developmental care by Macho (2017) traces its roots back to the early 1960s with the initial work by Brazelton to understand how interactions between the baby and their environment supported positive development. The term developmental care, however, is credited to Heidelise Als whose synactive theory of development underpins the Newborn Individualized Care and Assessment Program (NIDCAP) and continues to influence contemporary operationalizations of developmental care (Als, 1982). Most relevant to the concern of NICU design is how developmental care calls for individualized manipulation of the environment and close involvement of family which was not possible within a large open bay (White, 2003).

Matching light and sound exposure to the needs of each infant, for example, is difficult or impossible within large open bays units.

Since its formation, developmental care has moved beyond its close connection with the synactive theory and has taken on broader, more operational definitions, such as interventions intended to minimize stress and individualize care for infants in order to "maximize neurological development and reduce long-term cognitive and behavioural problems" (Macho, 2017, p.169). This results in broad interventions falling under the umbrella of developmental care including positioning devices, incubator covers, ear muffs, multidisciplinary care, family-centered or family-integrated care, providing comfort/reducing pain, and parent education (Macho, 2017). Consideration of the evidence base is complicated by this use of developmental care as a theoretical or conceptual framework in addition to interventions or combinations of interventions being presented as implementations of developmental care as an intervention (Burke, 2018). From 2006 to 2017 at least 19 articles (N [total infants] = 1370) were published that assessed the impact of interventions described as developmental care (Burke, 2018). Of these, eight were randomized controlled trials, with the remaining studies using a variety of non-randomized designs. Components of interventions varied, with NIDCAP including parent support, child support, and multiple approaches to parent education while others (e.g. Family Nurture Intervention) did not provide parent support and used two approaches to education. Greater heterogeneity was present in dosing and level of expertise of those implementing the intervention. A modified Mother-Infant Transaction Program was delivered in eight 1-hour hospital sessions, and four home visits, while NIDCAP was implemented in either daily or weekly formats, and a kangaroo care-based intervention took place over two interfeeding periods (2018). Professionals responsible for implementing interventions ranged from nurses with

specialty training to developmental psychologists, pediatric physical therapists, and research staff (2018). Intervention effects were mixed, with outcomes from NIDCAP broadly failing to reach statistical significance, the Infant Behavioural Assessment and Intervention Program showed varying results across time points without a statistically significant linear trend. The family nurture intervention was associated with improvements in sleep and electroencephalography coherence at term age as well as higher cognitive and language scores on the BSID-III at 18 months corrected age. Results from three variations of Mother-Infant Transaction Program were mixed and difficult to interpret without a coherent statistical model as effects vary across time and outcome (2018). Massage and multisensory intervention were associated with mental scale BSID-III improvements at 12 months, and language and motor domains at two to three years of age while kangaroo care was associated with significant difference in electroencephalography complexity at term age (2018).

When results from Burke are read within the context of existing Cochrane and non-Cochrane reviews of developmental care (Ohlsson & Jacobs, 2013; Symington & Pinelli, 2006) and aspects of developmental care (e.g. lighting (Morag & Ohlsson, 2011) and sound (Almadhoob & Ohlsson, 2015), it appears that the most stable and convincing effects are related to those interventions that are primarily executed by parents (e.g. kangaroo care, family nurture intervention). This aligns well with interpretations of results from several evaluations of private room care, which point towards a pattern wherein increased parental involvement is related to better neurodevelopmental outcomes (Lester et al., 2016; Pineda et al., 2013, 2014; Vohr et al., 2017). The added privacy of a private room is hypothesized to increase the likelihood that parents will be physically present and involved in the care of their infant (e.g. provide skin to

skin contact and other beneficial aspects of care), and to provide the ability for parents to create a homelike space at the hospital (White, 2003).

Types of designs.

Open-bay-NICU. The open-bay design represents the current standard of care in Canada and around the world (White, 2014). Open-bay designs care for all infants in a common shared space assigning the smallest square footage to individual infants and typically only separate bed spaces with a cloth curtain. Most units do not have dedicated parent spaces, but when they do, they typically take the form of communal areas. Parents who wish to stay overnight would generally sit at the bedside or, in some units/hospitals, can apply for one of a limited number of parent rooms that are assigned on a day to day basis.

Half-walls or semi-private NICU. These NICU designs are intended to maximize privacy without committing to a single room for each infant (White, 2003). Semi-private NICUs use an intentionally spread out design and separate spaces with half-walls.

Private room. Infants are housed in a private room with a door and full walls separating them from neighbors. Typically, a bench or small seating area is provided for parents and visitors, but no permanent sleeping or bathroom facilities. These designs require greater square footage per infant than the private or semi-private room (White, 2014).

Single-family room. This design is similar to the private room but with additional square footage allocated to provide a small, fully furnished parent suite (White, 2014). In some variations, the parent area can be closed off from the care area, allowing privacy from staff. For the purpose of this dissertation, the defining feature of a single-family room is a parent sleep space. A version of this design which includes a private shower in each room is being

implemented at the IWK Health Center and requires the largest square footage per baby site because of the extra parent space.

Theoretical underpinnings. Any consideration of the influence of physical space on the health of inpatients would be incomplete without considering Florence Nightingale's environmental theory which was first described in written publication through her notes on nursing (Nightingale, 1974). Nightingale is often credited as the founder of modern nursing and in many ways she is the archetype of what a nurse should be (Karimi & Alavi, 2015; Strickler, 2017). Her approach to care was holistic and emphasized the importance of trusting relationships, her involvement in politics allowed her to advocate for change, and her comfort and dedication to collecting and analyzing data allowed her to identify problems and assess whether her interventions were addressing them. She was the first female member of the Statistical Society of London and was an honorary member of the American Statistical Association (Strickler, 2017). In 1860, she opened the first nursing school and dedicated the remainder of her life to developing standards for nursing. She is perhaps most famous for her work during the Crimean war, where her focus on the importance of the environment in promoting healing was credited with a reduction in mortality in hospitalized British soldiers from 60% to 2% (Karimi & Alavi, 2015). In her textbook, Nightingale describes her beliefs regarding the important influence of fresh air, clean water, rapid and appropriate waste disposal, cleanliness, light, and sound. Each of these factors is considered vital to place patients in an environment that supported healing (Nightingale, 1974). Her work on the importance of the environment for nurses is so central to modern nursing, that it is part of the metaparadigm of

nursing (Fawcett, 1984). To her credit, these factors capture much of what has been adapted and specialized to form the theoretical basis for NICU design.

Synactive theory of development. The most influential theoretical framework for understanding the impact of NICU design has been the synactive theory of development first described by Als (1982). This framework forms the basis of developmental care and has been influential in helping practitioners to conceptualize the important relationship between neonatal environment, development, and the role of caregivers. The synactive theory of development imagines four subsystems: (a) autonomic, (b) motor, (c) state, and (d) attention/interaction. The name synactive means "Together in action" and is intended to draw focus to the interdependence of subsystems for appropriate neurodevelopment (Als, 1982). Assessments based on this theory instruct practitioners to consider the underlying neurodevelopment of an individual infant and how they can provide care to support that development. A classic example is the provision of a soother to encourage non-nutritive sucking and support motor development, or swaddling to encourage infants to self-sooth (Gibbins et al., 2008).

Universe of developmental care. The synactive theory remained the linchpin of developmental care from its development until the recently published universe of developmental care (UDC) (Gibbins et al., 2008). The UDC positions itself as an extension to synactive theory and makes developmental care more amenable to empirical investigation. Gibbins et al describe the central contribution of this model as being the concept of the "shared surface", which forms the boundary between the infant's developing central nervous system and the environment (2008). Naturally, the point of reference is the infant's skin which authors argue is central to caregiving activities regardless of the specific disease being treated. It is argued that while the synactive theory draws attention to the central nervous system, the UDC's focus on the shared

surface of the skin promotes a more natural way to appreciate the interaction of multiple subsystems (Gibbins et al., 2008). The UDC probably represents less of a major development in the actual theory of developmental care and more of an operationalization of how it might look within terms that are easily understood by clinicians. The focus on four subsystems of neurodevelopment have been replaced with nine physiological systems (nervous, integumentary, metabolic, etc.) that can each be influenced through care practices associated with nine care domains (feeding, positioning, skin care, etc.…). The family's centrality in care is made explicit through their placement closest to the infant, and authors are careful to emphasize that care on one "planet" (e.g. skin care) can affect and be affected by influence on another "planet" (e.g. infection control). Authors have since described how the UDC can be used to identify key performance indicators, which has been re-packaged in a simplified but conceptually identical mother and child integrative developmental care model (Altimier, 2011; Coughlin et al., 2009).

Model of NICU design and infant outcome. Theories of developmental care can capture much of the actual and potential influence of private room care, but a more focused short-range theory would be able to help draw closer attention to the specific interactions of interest for those focused on design. Lester et al's (2011) model of NICU design and infant outcome is the only published conceptual model developed specifically to explain the connections between design and outcomes (Figure 1). It proposes that the improvement in neonatal outcomes observed in the switch to private room care is mediated through family centered care, developmental care, parent/family factors, staff behavior/attitudes, and medical practices. Private room designs may support a family centered care approach because of its commitment to recognizing that creating a parent family space encourages their presence and participation in care (Lester et al., 2011). Similarly, private rooms typically allow for more careful control of light and noise and facilitate

parent involvement of care all of which are components of developmental care. Parent and family factors include their presence and involvement in care as well as the impact that private rooms potentially have on stress and quality of life of parents. Finally, the model recognizes the potential impact that private room designs can have on staff stress/burnout as well as the practices of nurses, physicians, and other health care providers. When taken together, the authors suggest that these factors are key to explaining the potential for private and single family rooms to reduce infections (e.g. less crowded spaces, increased care by parents, increased sink to bed ratio), necrotizing enterocolitis (NEC) (reduced stress, improved feeding tolerance, increased breastfeeding), and other neonatal morbidities (Lester et al., 2011).

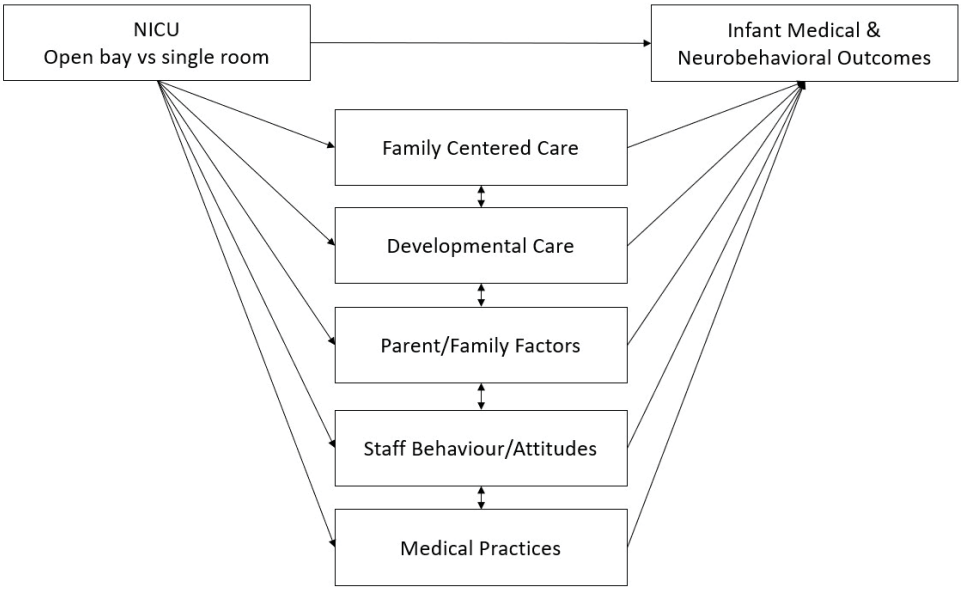


Figure 1. Model of NICU design and infant outcome, adapted from Lester et. al (2011). Arrow between NICU open bay vs private room to infant medical and neurobehavioural outcomes represents the observed relationships in the literature. Arrows that travel through hypothesized mediators provide the hypothesized explanation for the main effect. Bi-directional arrows between hypothesized mediators are intended to communicate shared influence of all factors on each other. For example, even though there is no direct arrow between family centred care and medical practices it is still believed that they interact.

Clinical benefits of alternative room designs. To address the problem of the burden of prematurity in Canada, NICU design would be required to provide some combination of improvement (or maintenance) of clinical outcomes, with reduced cost of care (Johnston et al., 2014). In this section, the potential benefits and harms of alternative NICU designs will be reviewed. The focus will be limited to parent, family, and staff effects that have been highlighted in the literature. Economic evidence will be reviewed in a subsequent section.

Infant effects.

Benefits. When compared against traditional crowded open bays, several alternative designs have shown promise of improving neonatal outcomes. Four studies (Chen et al., 2017; Goldmann et al., 1981; Jones et al., 2012; Larson et al., 1985) have compared a traditional open bay unit to modern, more spacious (Increase of 10-110 square feet for infant) open bay units. All four studies used a prospective pre-post study design but did not include a control group or adjust for baseline imbalances in infants. Only Goldmann et al. (1981) matched patients. All four studies saw reductions in the occurrence of late-onset sepsis of varying magnitudes (range of absolute risk reduction 2.1%-7.1%). Interpretation of these findings is complicated by at least two studies confounding their move to a new unit with the implementation of quality improvement initiatives targeted at reducing catheter associated sepsis (Jones et al., 2012).

Only one study (Altimier et al., 2005) has compared a semi-private pin-wheel style design to a traditional open bay using an un-controlled pre-post prospective design. In this unit, individual sites were separated by half-walls and infants were provided with 110-125 square feet of space, approximately double the space in the older open bay. The move was associated with decreased light and noise, as well as statistically significant reductions in the rates of severe intraventricular hemorrhage (Absolute risk reduction = 8%, ventilator days (5.6 vs 4.3) and

length of stay (mean difference of 21 days for 24-27 weeks gestational age; 13 for 38-30 weeks; and 11 for 31-34 weeks). Move to the new unit was accompanied by a developmental care bundle and no adjustments for baseline differences were made, although groups were similar on the limited characteristics reported (gestational age and birthweight).

Four reports (Julian et al., 2015; Milford et al., 2008; Pineda et al., 2012, 2014) compared private rooms to traditional open bays. One (Julian et al., 2015) used a retrospective cohort design, one (Milford et al., 2008) a prospective pre-post, and the remaining two studies were conducted by the same unit whose mixed-design allowed for a quasi-experimental design. Only Pineda et al. (2014) controlled for baseline differences in infants, and when assessing developmental outcomes. Across all studies, authors reported data on mortality (Julian et al., 2015; Pineda et al., 2012), NEC, ROP (Pineda et al., 2014), BPD (Pineda et al., 2014), sepsis (Julian et al., 2015; Pineda et al., 2012, 2014), cerebral impairment (Pineda et al., 2012), any breastfeeding at discharge (Pineda et al., 2012), and length of stay (Milford et al., 2008; Pineda et al., 2012, 2014). Small, statistically insignificant differences favoring private room designs were reported for sepsis, cerebral impairment, and mortality. Heterogeneous findings were reported for the length of stay (mean difference of 1.49 days (Milford et al., 2008) to 8 days (Pineda et al., 2014).

Single family rooms have seen the most extensive research effort with 16 reports (Alessio, 2011; Domanico et al., 2011; Erdeve et al., 2008; Erickson et al., 2011; Feeley et al., 2020; Lester et al., 2014, 2016; Ortenstrand et al., 2010; Puumala et al., 2020; Smithgall, 2011; Stevens et al., 2014; Stevens, Helseth, et al., 2011; Tandberg, Frøslie, et al., 2019; Vohr et al., 2017; W. Walsh et al., 2006). Of these, seven were prospective un-controlled pre-post (Domanico et al., 2011; Erdeve et al., 2008; Erickson et al., 2011; Feeley et al., 2020; Lester et al., 2014,

2016; Puumala et al., 2020; Stevens et al., 2014; Tandberg et al., 2014; Vohr et al., 2017; Walsh et al., 2006), three were retrospective un-controlled pre-post (Alessio, 2011; Smithgall, 2011; Stevens, Helseth, et al., 2011) and one was a randomized controlled trial (RCT) (Ortenstrand et al., 2010). Changes in room designs were accompanied with an additional 100-150 square feet of space per baby. Five studies (Alessio, 2011; Domanico et al., 2011; Erdeve et al., 2008; Lester et al., 2016; W. Walsh et al., 2006) did not adjust for potential confounders in any analysis, although Lester et al (2016) stratified by degree of parent involvement (above or below median).

Considering the non-randomized studies first, statistically significant improvements in outcomes favoring single family room were reported for sepsis (Stevens, Thompson, et al., 2011; W. Walsh et al., 2006), time to achieve full-enteral feeds (Alessio, 2011; Erickson et al., 2011), time to any enteral feed (Domanico et al., 2011), breastfeeding at follow-up (Erdeve et al., 2008), discharge weight and weight gain per day (Erickson et al., 2011; Lester et al., 2016; Stevens, Helseth, et al., 2011; Vohr et al., 2017), weight gain in grams/kg/day (Vohr et al., 2017), risk of readmission (Erdeve et al., 2008), and BSID III developmental scores at 18 (Lester et al., 2016) and 24 months (Vohr et al., 2017). Lester (2016) found that BSID-III scores were higher in infants whose mothers were more highly involved regardless of room type. Most recently, Puumala et al. (2020) conducted a large prospective study that found a statistically significant decrease in length of stay and sepsis for extremely preterm and very preterm infants with an offsetting increase in the risk for term infants. Interpretation of the causal effect of the switch is not clear however, as the interrupted time series analysis suggested that the decreased length of stay was the continuation of a trend over time.

The lone randomized control trial (RCT) of single family room design was conducted in Sweden by Ortenstrand et al. (2010). Results indicated that infants cared for in single family

rooms spent 5.2 fewer days in the hospital and were at a slightly decreased risk of developing severe bronchopulmonary dysplasia. Parents included in this study were required to spend 24 hours a day in the hospital, and authors were careful to emphasize the important role that this presence likely played in reducing the length of stay. They cautioned that results might not be similar in locations where parents are not offered the same supports as in Sweden, where national programs provide parents with approximately 80% of their regular income while on leave.

Harms. Amidst this largely positive publication history, however, has been a spattering of publications that have found poorer neonatal and family outcomes associated with private room designs in particular (Pineda et al., 2012, 2014; Stevens et al., 2015). The most influential of these have been a pair of publications from Pineda et al. (2012, 2014). An advantage of the design in the Pineda unit has been that they are the only unit publishing that has elected to use a mixed design, which means they have been able to report results comparing outcomes of infants that were cared for in alternate designs at the same time. In their neonatal outcome focused publication, they reported on findings from 107 infants of gestational age range 23-30 weeks (mean = 26.6). They found no differences in clinical outcomes including NEC, ROP, cerebral injury, confirmed sepsis, BPD, or length of stay. When controlling for family functioning, infant acuity, cerebral injury, and social risk score they did, however, find a statistically significant six-point decrease in BSID-III language scores, and a trend to lower motor scores ($p = 0.02$) and increased externalizing behaviors ($p = 0.04$). These findings were found in the context of functional neuroimaging results indicative of delayed development. Authors hypothesized that results might have been caused by the relative sensory deprivation that may occur in private rooms within the context of their low parental presence. Importantly, these are the first and only empirical data that explicitly suggest that some babies may be more appropriate for private room

care than others. The only other study investigating developmental outcomes to date found support for the hypothesis that the degree of parent involvement may be an important mediator in developmental outcomes, although families in their single family rooms performed better on developmental outcomes than those in open bay (Lester et al., 2016). Interestingly, authors did not adjust for the degree of developmental care provided despite being identified as an important mediator of weight gain and early developmental assessments in the original study (Lester et al., 2014, 2016). The incidence of bronchopulmonary dysplasia reported by Stevens et al. (2011), Lester et al. (2014), and Vohr et al. (2017) indicate increased risk for BPD in a single family room, although results do not reach statistical significance. This analysis is limited by its reliance on unadjusted published data but raises additional questions regarding which infants would benefit most from alternative NICU designs.

Outside of Ortenstrand's (2010) randomized trial, all other identified single-family room studies saw an increased length of stay in the new unit despite what appear to be equivalent groups at baseline. This is surprising when considering that private (i.e., no sleeping space) and semi-private designs were generally associated with decreased length of stay. It is expected that a developmentally appropriate environment that creates additional space for parents to be involved in care should encourage increased weight gain and parent caregiving confidence, therefore it is difficult to explain why lengths of stay would increase in those units. This may be explained by unobserved confounders or could suggest that the observed statistically insignificant increases in neonatal morbidities had an outsized impact on length of stay.

Parent effects.

Benefits. Generally speaking, discussion of benefits to families has been limited to measures of satisfaction, presence, and involvement in care although measures of stress and

depressive symptoms are reported more often in recent publications (Feeley et al., 2020; Tandberg, Flacking, et al., 2019).

In their systematic review of the impact of design on neonatal intensive care unit staff, families, and infants, Shahheidari and Homer (2012) conclude that parents expressed private room units made them feel less like visitors and provided them with the privacy required to express joy and sorrow. The synthesis was based on three studies and noted that greater physical space and policy changes allowing unlimited parental presence might offer some explanation of parent experience beyond the design itself.

In Smithgall's (2010) comparison of single family room to traditional open bay NICU, the Parental Stress Scale: Neonatal Intensive Care Unit survey was administered to 26 parents from an open bay unit and 20 parents of gestational age-matched infants from a single family room. She found that the environment made no difference in reported stress scales and that high-stress scores were equally likely from both groups. Importantly, she found that parents of infants cared for in single-family rooms were present more days on average (mean difference 109.58 days, $p < 0.001$) despite only a small, statistically insignificant increase in the length of stay in the single family room (mean difference = 4 days, $p = 1.06$). This increase in parent presence in single family room and private room designs appears to be a robust finding across centres, with similar findings reported by Pineda et al. (2012), Lester et al. (2014;2016), and in a large 11 site, six country study published by Raiskila et al. (Raiskila et al., 2017). Most recently, Feeley et al. (Feeley et al., 2020) assessed differences in presence and participation in a Canadian unit that switched from an open bay design to a combined single family room/pod design. Consistent with earlier evidence, results showed a trend towards increased parent presence, with the most robust difference appearing at discharge assessment where parents were reported as present for nearly

twice as long as those in open bay (Feeley et al., 2020). An even more marked difference in presence between single family room and open bay units was observed in a Norwegian study in which parents stayed on average three times longer per day (21 vs 7 hours per day) (Tandberg, Flacking, et al., 2019).

Harms. Pineda et al. (2012) were the first to publish results indicating that parents in their private rooms reported increased stress in comparison to those in their open bay. They found that while parents were present more often in the private rooms, there was a slight increase in stress and a small but statistically insignificant decrease in the number of infants discharged receiving breastmilk (Pineda et al.). Feeley et al. (2020), however, found a similarly small trend in the opposite direction, suggesting that effects of single family room designs on parent stress may be small and highly variable. Similarly, while Pineda found improved parent presence, there was no statistically significant increase in SSC or holding, despite policies that encouraged these practices even for mechanically ventilated children (Pineda et al., 2013). This is consistent with a recently published survey of parent presence and involvement in care in 11 NICUs in 6 European Countries (Raiskila et al., 2017). Despite large, consistent differences in total presence times between private room and open bay designed NICUs there was much more variability in the associated parent involvement regarding SSC and holding. Thus there appears to be some inconsistency in the magnitude and direction of effects associated with changes in NICU design, which needs to be explored (Raiskila et al., 2017).

Mediators and moderators. There have only been limited efforts to measure potential mediators of improved outcomes, but one such analysis was undertaken by Lester et al. in their before and after evaluation (Lester et al., 2014). Using structural equation modeling, they found that statistically significant relationships between NICU environment and weight at discharge

and rate of weight gain were mediated through increased developmental support (Lester et al., 2014). Developmental support was defined as the number of occupational therapy visits. A separate structural equation model explained a small but statistically significant portion of the reduction in neonatal stress and mean pain score through the NICU environment's effect on increased parent involvement. These findings are important because they emphasize that private room designs often bring a commitment to developmental care with them. They also need to be interpreted with abundant caution, as mediation analyses are at high risk for collider stratification bias, which may result in changes in the magnitude and even direction of effect (Hernán & Robins, 2020). This raises the question of to what degree do observed outcomes depend on the design itself as opposed to a commitment to developmental care, and what happens if the design itself does not improve those mediators?

Health inequity. There has been surprisingly little consideration of the potential health inequities that can be introduced through construction in the neonatal intensive care environment. No studies of NICU design have examined the potential moderating influence of race, marital status, or socio-economic factors on neonatal outcomes in a prospective fashion. Most reports of neonatal outcomes only report characteristics of mothers, and we are aware of only five that collected data regarding socio-economic status at all (Lester et al., 2014, 2016, Pineda et al., 2012, 2014; Vohr et al., 2017). Inferences based on these data are limited by the observational nature of the studies and the potential for ecological bias present when aggregate characteristics are compared between studies (Borenstein et al., 2009). With those limitations in mind, there is a noted difference in the proportion of parents on Medicaid in studies by Lester et al. and Pineda et al. (30% vs. 67%). The potential for health inequity is likely to occur through two avenues: a potential imbalance in infants and parents who bear the weight of detrimental

effects, and the expectation of increased presence is increasing the financial burden on those same families. These effects may be small when observed from an aggregate level but may represent substantial concerns for equity if consistently concentrated amongst a group of families with shared characteristics. Who these groups may be is currently unclear, but the first efforts to address this question can be addressed through the incorporation of existing data in a model that can appropriately capture these effects.

Economic Analysis

The observed and proposed benefits of private room care provide sufficient justification to hypothesize that the optimal room design could reasonably expect to reduce costs of care and/or improve outcomes sufficiently to justify more costly care. Reduced length of stay is theoretically feasible and has been shown empirically across most designs compared against open bays. If it is as large (5.2 days) as suggested by some, the reduction in per-baby costs reimbursed by the public payer could be significant. Similarly, reductions in rates of major neonatal morbidities (e.g., sepsis) can be expected to potentially improve long-term infant outcomes, which may improve quality of life for children and parents. The evidence also suggests that less favorable outcomes may be plausible, and thus an investment in new construction could result in increased harm. In this section, we provide a brief review of the fundamental concepts in the economic evaluation of health technologies, and discuss the existing empirical evidence surrounding NICU design.

Key terms and concepts in economic analyses.

Health-technology. Despite the seemingly narrow scope that the term conveys, health technology refers broadly to equipment, drugs, or interventions intended to improve health outcomes.

Opportunity cost. A key concept in all evaluations of health technology is that budgets are assumed to be fixed, and thus there is a finite amount of money that can be invested in any given portfolio of health technologies. Opportunity cost describes the concept that since money spent on one technology cannot be spent on another, the true cost of introducing a new technology is "not the number of dollars appearing on the program budget, but rather the value of the benefits achievable in some other program that has been forgone by committing those resources to the first program" (Drummond et al., 2015, p. 3).

Decision problem. Economic evaluations are intended to inform decisions and are not designed to test hypotheses. CADTH (2016) guidelines recommend that the decision problem capture the perspective to be taken, the interventions being compared, the measures used to compare them, and the time-horizon over which the decision will be made.

Reference case. The reference case is a term used by CADTH to describe the recommended set of methods for analysis that allows for apples to apples comparisons of technologies across sectors (Canadian Agency for Drugs and Technology in Health (CADTH), 2017). In addition to the reference case, economic assessments can include non-reference cases which may be of relevance to a decision problem. Non-reference cases can be thought of as analogous to a sensitivity analysis.

Cost-effectiveness analysis/cost-utility analysis. Cost-effectiveness analyses allow for the simultaneous consideration of costs, benefits, and harms for a given set of interventions

relevant to a decision problem (CADTH), 2016, p.24). Technologies are considered cost-effective when their adoption provides more value than the opportunity cost associated with forgoing the investment in an alternate technology (Drummond et al., 2015). When the unit of interest is expressed in quality-adjusted life years, it is sometimes referred to as a cost-utility analysis, although the terms are often used interchangeably. CADTH states that the reference case ought to be a cost-utility analysis, as this allows for the use of a single unit to compare value across all technologies (Canadian Agency for Drugs and Technology in Health (CADTH), 2017).

Incremental cost-effectiveness ratio or incremental cost-utility ratio (ICER/ICUR). Terms are used interchangeably, with the ICUR being a special case where the outcome of effect is the quality adjusted life year (QALY). The ICER/ICUR is the cost associated with one additional unit of effect (e.g., price per additional QALY) of new technology compared to an alternative technology (Briggs, Claxton, & Sculpher, 2006, p.24). If the ICER is less than the decision maker's willingness to pay threshold (the maximum price they will pay for one additional unit of effect), then the new intervention is considered cost-effective in comparison to the old intervention (CADTH, p.24). In the UK, National Institute for Health and Care Excellence (NICE) sets the willingness to pay for one additional QALY at £20,000-30,000. No public threshold exists in Canada, although it is often considered to be $50,000/QALY. These signposts are intended to represent the point at which a new technology displaces more health than it generates, although the empirical justification for such a statement is weak (Drummond et al., 2015).

Perspective. The perspective taken for an economic evaluation outlines the boundaries for what costs and health effects are considered relevant. Since Canada has publicly funded health insurance, CADTH (2016) guidelines require the reference case to be from the perspective

of the public payer. This perspective includes direct and indirect costs to the healthcare system but would exclude costs to private insurers, social services, and non-health effects to informal care givers (Canadian Agency for Drugs and Technology in Health (CADTH), 2017). These additional costs can be explored through broader non-reference case analyses such as the private payer, broader government payer, or societal perspectives.

Time horizon. The time horizon is the period over which health effects and costs are calculated to address the decision problem. CADTH (2016) guidelines suggest that time-horizons capture all possible outcomes and costs associated with an intervention.

Discounting. When time-horizons beyond a year are used, it is important to capture the value that society typically places on costs and benefits that occur immediately compared to those that happen in future years (CADTH, 2016). The term used to capture this concept is discounting and is a type of compounding interest that reduces the value of future costs, benefits, or harms, typically at a fixed rate. The most recent version of the CADTH (2016) guidelines suggests that the reference case use a discount rate of 1.5%, with a sensitivity analyses using a 3% rate for costs and outcomes. This rate would imply that 5 QALYs gained in 10 years would have equivalent value to 4.31 QALYS earned today. In other words, we should be just as happy with an intervention that returns 4.31 QALYs immediately, as an identically priced intervention that returns 5 QALYs ten years from now. These rates were derived under the guidelines' assumption of a societal decision-maker perspective in which representatives of duly elected officials administer budgets with the goal of maximizing health. From this perspective, it is argued that the most relevant discount rate is the typical return of a provincial bond since this is a

routinely expressed preference for money today versus a larger amount in the future (CADTH, 2016).

Dealing with uncertainty. It is important to highlight a significant way in which the health economic approach to decision making differs from typically used hypothesis testing approaches. Under typical use, a p-value greater than 0.05 is interpreted as an inability to reject the null hypothesis, and routinely referred to as indicating "no difference." This approach is problematic for decision makers for a number of reasons, not the least of which is that it can routinely create situations in which they would be guided to implement an intervention with a lower expected value (e.g., mean score) (Claxton, 1999). This approach also ignores the role of future research to increase precision in an estimate and does not provide a method to prioritize future studies based on a combination of the uncertainty in their estimate and their effect on the decision (Claxton, 1999). As an example, using a strict "no difference" interpretation of the risk of an adverse event would result in either the exclusion of that event from a decision model or the entry of that value as identical to that of the one it was compared against. If this estimate was based on a small study and was only marginally non-significant, then there may actually be a high probability of a large difference that could significantly alter which intervention is optimal (Briggs et al., 2006). For this reason, it is often argued that inference is irrelevant for decision models (Claxton, 1999).

In the absence of p-value driven inference, decision models use the expected value as a point estimate and then specify a probabilistic distribution around it (Briggs et al., 2006). The model is then simulated, for e.g., 10,000 times, with a new random draw from each of the probability distributions at each simulation. This allows for simultaneous consideration of all

model uncertainty, and can provide credible intervals for the net monetary benefit (Briggs et al., 2006).

Existing economic analyses of NICU design.

Trial based. Stevens et al. (Stevens et al., 2014) conducted the only primary economic evaluation of single family room to date. In their uncontrolled prospective pre-post design, which included 269 open bay and 306 single family room infants in the final analysis, study authors collected information on all direct costs (labor and benefits, supplies, and depreciation) and compared designs using a generalized linear model with log-transformed costs as the dependent variable. Explanatory models were built in a hierarchical manner beginning with an unadjusted model, then adding admission variables, "severity measures," duration of respiratory support, and duration of hospitalization in a stepwise fashion. The final model had an R^2 of 0.799, $p = 0.0095$, and an estimated cost ratio of 1.11 indicating that open bay care was 11% more expensive than care in the single unit.

While these findings are interesting, there are a number of questions remaining that create doubt. The most important of these is the lack of clarity regarding which "severity measures" were included in the model. The inclusion of severity measures leads to a sign change in the cost ratio (i.e., changes the estimate to suggest single family room care is less expensive than care in open bay), but it is unclear which of the 24 severity measures described in their table were included in the analysis. This may be of importance because several acquired complications that could be attached to room design (BPD, ROP, late-onset sepsis, pneumonia, days of continuous positive airway pressure, days of oxygen, days of mechanical ventilation) were higher in the single-family room group. Of these, days of continuous positive airway pressure, and suspected late sepsis were statistically significant and thus more likely to be selected for inclusion,

although most of the above outcomes reached borderline p-values (e.g., 0.06). These measures would likely be inappropriate to add to an explanatory model since they may be caused by the room design and thus separating out their explanatory power from unit design could artificially lower the associated costs or flip them entirely as a result of collider bias. Since this step is the one associated with a flip in a sign of the design coefficient, conclusions should be drawn cautiously.

An additional point of concern with analysis of the data by Stevens et al (2014) is their decision to use the duration of hospitalization as an explanatory variable. The inclusion of this step resulted in an overall p-value of 0.0095 for the model and is the primary support for the author's argument that single-family room care may be less expensive than care in an open bay unit. Given that this is an outcome that is affected by room design itself, the coefficient for design in the final model including length of stay is interpreted as the indirect effect of single-family room on total costs after controlling for clinical outcomes. This is a research question of little interest to the dissertation and can be misleading since it could be that a newly required unit requires less maintenance costs compared to a 20 year old unit at the end of its life span but could actually lead to increased costs if it had a negative effect on neonatal outcomes. The authors argue, however, that length of stay differences was not actually an outcome but a proxy for unmeasured clinical factors that differed between the two timepoints. This is a considerable assumption given the extensive admission and (acquired) severity measures already included in the model. An alternative hypothesis would be that there is a chance that after controlling for other factors babies remained in single family rooms for a longer period. This would encourage the use of the previous step (despite its lack of statistical significance), which only included clinical outcomes and provided a more modest cost ratio of 1.05.

Decision models. To date, there have been two attempts to capture the relevant economic factors associated with NICU construction using a decision model. The first was developed by Shepley, Smith, Sadler, and White using a very simple return on investment approach based on a narrative review of the literature (Shepley et al., 2014). The authors included costs of construction, direct costs of care, and estimates of the reduced length of stay in a deterministic approach. They concluded that the cost-savings would provide a return on investment within the first year, with a $1.2 million 2014 USD cost savings in each of the following years (Shepley et al., 2014). More recently, Sadatsafavi, Niknejad, Shepley, and Sadatsafavi (2017) conducted a probabilistic return-on-investment analysis comparing open bay to an adaptable single family room design. They improved on the methods of the earlier study by including a systematic review that captured a total of five studies that could provide information for the analysis. Because of the focus on costs only, they conducted three separate analyses based on the observed reductions in nosocomial infections, reduced length of stay (multiplied against a per diem rate), and reduced direct costs of care (Sadatsafavi et al., 2017). Their results supported a positive benefit/cost ratio when either the length of stay or direct cost figures were used. The strengths of this study included its use of a probabilistic approach which allows for visualization of uncertainty in the estimate.

A major limitation of both models is that they have framed their decision problem in such a way that they only considered two designs at a time, which does not accurately reflect the decision problem. Our clinical review suggests that other designs may seem similar, and potentially more favorable, outcomes at a lower cost. Excluding them from consideration ignores the possibility that other designs may represent a more optimal use of funds. Of equal concern is that heterogeneity in the evidence base as described is ignored in decision models. At a

minimum, models should include a sensitivity analysis assessing the economic consequences of the less optimistic length of stay reductions. Similarly, if effectiveness outcomes (e.g., sepsis) are to be used from observational trials, then other major morbidities which show less favorable results in alternative designs (e.g., BPD in single family rooms) should also be included. Lastly, as the previously discussed costing study by Stevens et al. is the only one of its kind, any decision model relying on its findings suffers from the same limitations. Any model using this study as their direct cost foundation must explore the implications of adjustment for acquired severity measures and length of hospitalization.

Summary of the Evidence

When the evidence base is taken as a whole, there appears to be justification for some change in infant and parent effects related to design, although a more rigorous data-driven approach is required before confident recommendations for practice or future research can be recommended.

Strengths. There is enough theoretical basis to support the empirical findings of improvements in neonatal outcomes associated with NICU design. The evidence appears to be the most consistent for weight gain outcomes, and increased parent presence, which aligns with theoretical expectations. It should be noted that the informal "vote counting" (i.e., identifying statistically significant findings) approach used here, however, does not take the place of formal evidence synthesis (Borenstein et al., 2009). The economic literature is less well-developed but, with re-analysis of costing data, could likely offer meaningful insight at least in the American context.

Limitations. The current clinical evidence base suffers from a lack of well-controlled trials, with some exceptions. This is particularly troublesome outside of the domain of single-

family room studies, which will represent a challenge for decision makers hoping to consider all available options. Synthesis of observational and randomized trials is a growing area of interest (Zhang et al., 2015). However a robust and valid combination of these trials within a single statistical framework (e.g., network meta-analysis) will be challenging for the following reasons:

1. Lack of adjustment for, or assessment of, baseline differences in groups for most trials.
2. Most changes in construction have been accompanied by additional co-interventions (e.g., developmental care bundles, increased square footage, quality improvement initiatives), confounding the effect of the design itself.
3. Lack of control groups in pre-post studies undermines the validity of findings.
4. Apparent contradictions between theoretical expectations, and findings from newly constructed open bays (large decrease in the incidence of sepsis), private rooms (no difference in sepsis), and single-family rooms (moderate decrease in sepsis rates).
5. Unstable findings related to developmental outcomes and BPD.
6. Disagreement between randomized and non-randomized studies (e.g., length of stay).
7. Exclusion of term born infants.

Taken together, these limitations represent a significant challenge and should encourage clinicians and decision makers to assess the vigor with which new construction is pursued. There are, however, several promising avenues that may allow for the risks associated with these limitations to be mitigated.

Despite the apparent limitations of the current evidence base, the reality of NICU construction plans in Canada necessitates a careful synthesis of the available evidence to inform further decision making. New developments in Bayesian hierarchical synthesis that allow for bias adjustment combined with the increased accessibility of methods for combining individual participant and aggregate data may allow a synthesis that appropriately adjusts for and considers uncertainty in introduced bias while offering recommendations for future research (Zhang et al., 2015). A planned randomized controlled trial at the IWK Health Center comparing single family room to open bay design will be a welcome addition to the evidence base and may help reconcile some of the observed differences between randomized and non-randomized trials (Dr. Mike Vincer, personal communication, May 28, 2018).

Limitations of the economic evidence are potentially more straightforward to address. The first issue is the lack of relevant costing data for the Canadian context. The Canadian Institute for Health Information (CIHI) collects costing data based on standardized methods that are specific to the Canadian healthcare system (Canadian Institute for Health Information, 2011). These data are available for access and can be linked to relevant clinical data, allowing for robust estimation of incremental costs associated with alternative clinical courses. An additional limitation of the economic evidence is directly limited to questions of equity, which is whether the increased parent presence required to maximize infant outcomes is associated with significant additional out of pocket expenses.

Canadian decision makers will be best served by an economic analysis based on a decision model that allows consideration of short and long-term outcomes across all viable design options. This should take the form of a cost-utility analysis, include parent and infant health outcomes, and allow for consideration of wider societal consequences as well as an

exploration of possible equity implications. For example: if lower parent presence is associated with poorer outcome, how can decision makers ensure all parents are provided equitable support to maximize time at the bedside while minimizing burden? Consideration of the implications of those supports on optimal NICU design should be included. Most of the costs associated with preterm birth are incurred during the initial hospitalization, but continuing costs are nonignorable (Johnston et al., 2014) and developmental consequences of design should be considered in terms of quality of life.

Dissertation Objectives

The primary objective of this research dissertation was to create a decision model to compare the cost-effectiveness of alternative types of NICU design. The secondary objective was to exploration the potential for some groups of families to experience a more significant share of the risk of harm.

Research questions

Primary Research Questions.

1. What is the most cost-effective NICU design from the perspective of the Canadian public payer over a lifetime time horizon?
2. Does the preferred design change if a broader, societal perspective is adopted?

Secondary Research Questions.

1. Are there groups of babies or families that consistently experience undue burden or harm in any design?
    a. If these families are compensated (e.g., parking or meals), does the cost-effective design change?

Chapter 3: Methods

This chapter contains a detailed outline of the methods and approaches used in this dissertation. It includes a description of economic model specification, parameter sources, and methods for analysis and knowledge translation.

Decision problem

Reference case. The problem facing decision makers is whether to construct a new NICU, and which design to use if so. Since the decision to construct or renovate an intensive care unit is made by hospital administrators in collaboration with provincial health officials, the perspective to be taken in the base-case will be that of the public payer over life-time horizon. Informal consultation with local decision makers has also emphasized the interest in two additional non-reference cases.

Non-reference case 1. Takes the societal perspective and represents local leadership's willingness to consider costs and effects outside those covered in the reference case.

Non-reference case 2. The second non-reference case was a clinician decision maker perspective developed in response to conversation with NICU staff and neonatologists and takes the form of a multi-criteria acceptability assessment focused only on neonatal outcomes and the perceived family centeredness of different designs.

Participants

Infants requiring admission to the neonatal intensive care unit. As infants are under the care of at least one caregiver, spillover health effects were considered in the base case. Additional out-of-pocket expenses for caregivers were considered in the non-reference case.

Intervention

Interventions being compared were new open bay construction, private room designs including single-family rooms, mixed designs, and designs implementing half walls or other semi-private designs.

Comparator

Traditional open bay NICU.

Outcomes

Health effects were measured in quality-adjusted life years, and the primary outcome is the net monetary benefit under a range of willingness to pay thresholds. The influence of parameter uncertainty was examined through probabilistic sensitivity analysis.

Type of Evaluation

As per CADTH (2016) recommendations, this evaluation took the form of a cost-utility analysis with outcomes expressed as quality-adjusted life years (QALYs). This decision simplifies comparisons across interventions and allows for simultaneous consideration of how an intervention affects the number of years lived, and the quality of the years lived (CADTH, 2016).

Creation of Model

Consideration of modeling approaches. Guidelines from the International Society for Pharmacoeconomics and Outcome Research guidelines and National Institute for Health and Care Excellence state that model choice should reflect a balance between realism and tractability (Davis et al., 2014; Siebert, 2012). Candidate models include a simple decision tree approach or combination of decision tree and state-transition model modeled at either the cohort or individual level using a Markov or discrete event simulation approach. Existing models of the burden of prematurity in Canada (Johnston et al., 2014) and the UK (Mangham et al., 2009) use a decision

tree like approach to simulate the hospital period with a cohort level Markov state-transition model to capture movement between disability states after discharge.

Decision trees are appropriate when states are short lived or the exact timing of events are unimportant, and thus modeling the hospital course in this way is sound (Briggs et al., 2006). State transition models specify a number of states (e.g. no disability, mild disability, moderate disability, severe disability, death) with transition probabilities that are evaluated at a constant time period (e.g. one year) (2006). This approach allows for travel between states over time which is appropriate when the underlying disease process is subject to change over time, which is consistent with the expected pattern of neonatal disability across the lifespan (Mangham et al., 2009). These models can either be modelled at the cohort or individual level.

The biggest advantages of modeling a decision process at the level of the cohort are that it is comparatively easier to build and debug, and that it has significant advantages for computation time for consideration of uncertainty in parameters (Davis et al., 2014). Typical probabilistic sensitivity analysis involves evaluating the model with 10,000 draws of parameters from their probability distributions. Individual level models often use 10,000 simulated individuals to provide stable estimates of outputs. Running a traditional probabilistic sensitivity analysis on an individual level model thus requires a total of 100,000,000 simulations (2014).

Individual level models are appropriate when individuals interact (e.g. to spread infection), when event risks change as a function of time spent in state, when previous history changes the probability of a new event (for e.g. the risk of having a heart attack increases with a previous heart attack), or (most importantly for this project) there are characteristics that differ between patients (e.g. gestational age) and those characteristics have a non-linear relationship with outcomes (costs or QALYs) (Davis et al., 2014). In the latter case, use of a cohort model

with average values entered into equations will result in a biased estimate (2014). The need to use microsimulation models in the last case can be avoided if variables of interest can be broken into crude groups, although this becomes unfeasible and increases the difficulty in interpreting code when the number of variables of interest is high or contains continuous variables (Davis et al., 2014).

To properly address the secondary research question of whether there are subgroups of patients who may experience an inequitable share of harm in alternative room designs, heterogeneity in infant gestational age, sex, and characteristics anticipated to predict parent presence (e.g. distance from hospital, children at home, level of education, availability of support person, and income) were required to be simulated. This is owing to the current hypothesis that parent presence is an effect modifying variable for neurodevelopmental outcomes (Lester et al., 2016; Pineda et al., 2014). Heterogeneity in gestational age is important because it is associated with both risk for morbidity and later disability (Canadian Neonatal Network, 2016; Synnes et al., 2017). In the reference and non-reference cases, it was also important to be able to model effects on infants and their caregivers as a family unit.

Model structure. To accomplish these goals, the analysis used a microsimulation approach which allowed for families to be sent through the complete model one at a time (Briggs et al., 2006). As described in Figure 2, this takes place in four main steps:

1. The model starts by using a bootstrapped draw of an infant and their parents from patient data simulated using characteristics from the IWK perinatal follow-up database, performance analytics, and the CHEZ-NICU Home study. CHEZ-NICU Home is an Atlantic Canada Opportunities Agency (ACOA), IWK Health Centre and Cisco Canada partnered funded study that is assessing parent presence and

involvement in care at the IWK before and after the move the a single-family room design, and designing and implementing an e-Health intervention to provide families with education, resources, and support needed to maximize their involvement in the care of their infant. Using a bootstrapped sample ensures that correlations between patient and family characteristics are maintained (Degeling et al., 2017).

2. Simulated families travel through a hospitalization module that estimates costs, length of stay, death and morbidity.
3. A Markov model with yearly cycles simulates the remainder of life for infants and parents from discharge to death (Figure 3).
4. Once all families have been simulated, costs and QALYs are calculated for each NICU design. Additional outcomes (e.g. rates or morbidities) are captured for scenario and sensitivity analyses.

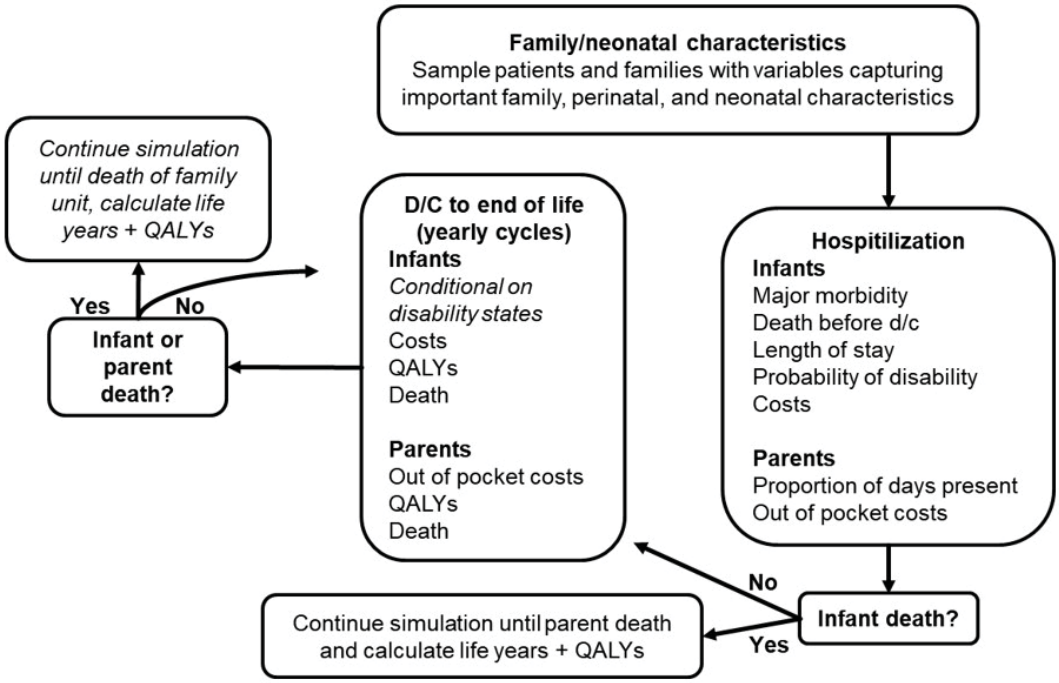


Figure 2. Cost effectiveness model structure. Arrows indicate flow of patients through model. D/C: Discharge; QALY: Quality adjusted life year

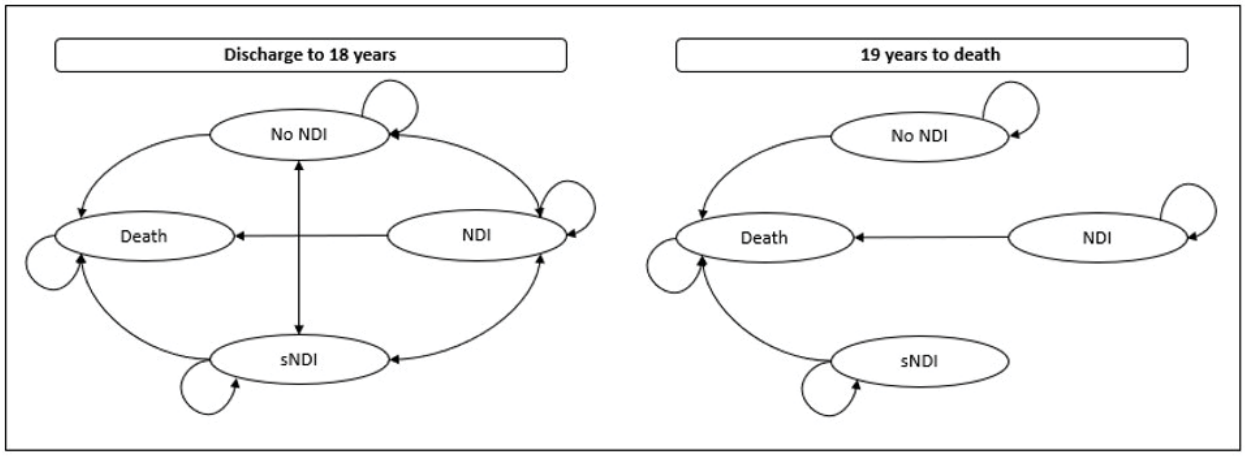


Figure 3 Markov model diagram for discharge to death. Arrows indicate possible direction of travel at each cycle of the Markov chain. From discharge to 18 years it is possible to transition from any disability state to any other state. From 19 years to death disability is considered stable and all simulated patients remain in that state until death. NDI: Neurodevelopmental impairment; sNDI: severe neurodevelopmental impairment.

Model structure was informed by previously published burden of prematurity models and adapted to include family outcomes in addition to those of infants (Johnston et al., 2014;

Mangham et al., 2009). All analyses and model construction were done using the statistical program R (R Core Team, 2015).

Perspective

In the reference case, the health care payer perspective was used (Canadian Agency for Drugs and Technology in Health (CADTH), 2017). Two non-reference cases were developed: the first taking a societal perspective and the second taking a clinician-decision maker perspective.

Time horizon

In the reference case, a life-time horizon for children and parents was used. The same horizon was used for non-reference cases from a local decision maker's perspective. No additional births were simulated over time. A hospital only horizon based on a stochastic multi-criteria acceptability approach (Tervonen et al., 2015) was used for the clinician decision-maker perspective.

Half-cycle correction

Half-cycle corrections were applied to costs and health effects at the first and last cycle to be consistent with the assumption that transitions occur on average in the middle of a cycle. This prevents over- and underestimation of expected values that have been identified by assigning payoffs at the beginning or end of a cycle (Siebert, 2012).

Number of individual simulations

Good modeling practices suggest that the number of individuals simulated in a microsimulation model be sufficient to achieve stability of estimates, where stability is defined qualitatively via trace plots of convergence (Krijkamp et al., 2018; Siebert, 2012).

Parameter sources

A complete list of the parameters to be estimated, and how they were included in the model for this dissertation is available in Table 1. In order to include the term population in the model, it was assumed that gestational age is not an effect modifying variable. This means that while baseline rates are expected to differ across gestational ages, the relative effect (e.g. odds ratio) of room design will not be anticipated to change across different gestational ages. Reasonable efforts were made to access data from the preferred data source. While it was expected that these data would be available in all cases, alternate data sources were used when access was either not granted or not be granted within a reasonable timeframe. Data for this dissertation were included from the following sources:

1. Chez-NICU Home: A study led Dr. Marsha Campbell-Yeo that recruited families that have been randomized to either the single-family room unit or open bay unit at the IWK (Marsha Campbell-Yeo, personal communication, May 28, 2018).
2. IWK Single Family Room RCT: From April 2018 until April 2019 all infants admitted to the IWK NICU were randomized to either the single-family room or open bay unit as part of a unit-led initiative.
3. IWK Follow-up Database: The IWK perinatal follow-up service follows all infants born less than 31 weeks of age and collects detailed information on maternal and neonatal characteristics including disability. This database was also used to develop the baseline history model and replaced the Canadian Neonatal Network database in that role because of data access issues.

4. IWK Performance Analytics: Contains patient costed data for all interactions at the IWK health centre. This data source replaced the Canadian Institute for Health Information (CIHI) database because data from CIHI were not available within a reasonable time frame.
5. Network meta-analysis: The primary efficacy parameters (length of stay and risk of morbidities and disability) were estimated through a systematic review and network meta-analysis.
6. Supplemental searches: In keeping with National Institute for Health and Care Excellence (NICE) guidelines (Kaltenthaler et al., 2011), a full systematic review was only conducted for treatment effects. Less vital parameters were informed through supplemental searches of the literature.

Table 1. Parameters and their primary/alternate sources

Parameter	Preferred data source	Data Source Used in Final Model	Relevant perspectives
	Hospitalization module		
Infant and parent demographics	Chez-NICU Home and CNN annual report	Chez-NICU home augmented with data from perinatal follow-up and IWK performance analytics	All
Major morbidity base rates	CNN database	IWK Follow-up Data	All
Relative risk of morbidity given room design	IWK follow-up database	IWK follow-up database	All
Death before discharge	Supplemental search	IWK follow-up database	All
Baseline length of stay	IWK follow-up database	IWK follow-up database	All
Effect of room design on length of stay	IWK Single Family Room RCT + Network meta-analysis	IWK Single Family Room RCT + Network meta-analysis	Health care payer, societal
Probability of disability given hospital course	IWK Follow-up database	IWK Follow-up database	All
Hospital costs given length of stay and morbidities	IWK Performance Analytics	IWK Performance Analytics	Health care payer, societal
Parent presence	Chez-NICU Home	Chez-NICU Home	Health care payer, societal
Out of pocket costs	Chez-NICU Home	Chez-NICU Home	Societal
Ordinal ranking of family centeredness of NICU designs	Assumption	Assumption	Clinician decision maker
	Discharge to end of life		
Transition between disability states	Supplemental search	Supplemental search	Health care payer, societal

Parameter	Preferred data source	Data Source Used in Final Model	Relevant perspectives
Costs associated with disability	Supplemental search	Supplemental search	Health care payer, societal
Neonatal QALYs associated with disability	Supplemental search	Supplemental search	Health care payer, societal
Background probability of death	Statistics Canada life tables	Statistics Canada life tables	All
Relative risk of death given disability state	Supplemental search	Supplemental search	All
Parent QALYs associated with infant disability	Supplemental search	Supplemental search	Health care payer, societal
Parent out of pocket costs	Chez-NICU Home	Chez-NICU Home	Societal

NICU: Neonatal Intensive Care Unit; CNN: Canadian Neonatal Network; QALY: Quality adjusted life year

Convergence of Bayesian Models.

Analyses were primarily conducted within a Bayesian framework because the methods are a good theoretical pairing to decision models which are intended to integrate over all sources of uncertainty to provide recommendations for action. Bayesian methods also provide a number of pragmatic advantages over traditional analyses: Draws from the posterior automatically account for correlations between parameters, simplifying sampling of parameter uncertainty; Bayesian models are flexible and modular, allowing analyses to be flexible to data needs; and the use of genuine prior information can improve model stability and protect against implausibly large effects (Gabry et al., 2017). All of the Bayesian models used to support this decision model were fit using Markov Chain Monte Carlo (MCMC) methods, which allows sampling from posterior distributions when there is no closed form solution as is common in most applied examples (Gelman et al., 2015). While the software used differs across analyses (Gibbs sampling via JAGS (Plummer, 2003) and OpenBUGS (Spiegelhalter et al., 2012) for meta-analysis and Hamiltonian Monte-Carlo (Carpenter et al., 2017) via Bayesian regression models using Stan [BRMS] (Bürkner, 2018) for all individual patient level analyses), all require assessment of convergence, effective sample size, and Monte-Carlo standard error (Gelman et al., 2008). All MCMC methods are based on sampling techniques which all converge to the same posterior given sufficient sampling (Gelman et al., 2015). All samples prior to convergence to this posterior should be discarded, with parameter estimates derived from additional samples taken after this point (Gelman et al., 2015). Convergence is monitored quantitatively using the latest implementation Gelman-Rubin diagnostic (called Rhat) based on four chains (Vehtari et al., 2019). This new implementation captures non-convergence from stationary but non-overlapping chains, over-lapping non-stationary chains, chains with heavy tails, and chains with different

variance. Samples were considered to have converged if Rhat was equal to or less than 1.05. After convergence has been reached, concerns turn to whether there are sufficient independent samples for stable estimates. The newest version of effective sample size and Monte-Carlo squared error estimation were used to ensure sufficient post-convergence samples were taken to support inference (Vehtari et al., 2019). If the rank-normalized effective sample size was greater than 400 (i.e. 100 per chain) then samples were taken to ensure that Monte Carlo squared error was small enough to allow for stable estimates to at least one decimal place (i.e. Monte Carlo squared error = 0.002 or less) (Vehtari et al., 2019). All assessments of effective sample size and Monte Carlo squared error were made for each quantile that is reported. If all Rhats and MCSEs were acceptable, convergence was confirmed visually using rank plots of posterior draws (Vehtari et al., 2019). Final models were thinned to provide a total of 10000 posterior samples that meet the above requirements. This allowed for standardization across other samples required for assessment of parameter uncertainty in the health economic model.

Baseline History Model

Cost of a new unit. Cost of construction was based on methods described in an earlier probabilistic return on investment study (Sadatsafavi et al., 2017), which assumed US $550 per gross square foot in a medium sized NICU (40 beds total) with square footage requirements of 165 square feet for single family room and 120 square feet in the open bay unit. For single family room, private room, and combined single family room and pod designs, the analysis also used estimated increases in operating costs of $8, $10, and $7 per gross square foot respectively and 1 additional full-time equivalent nurse at $50,000/year, consistent with previous models (Sadatsafavi et al., 2017). It was assumed that private rooms without a parent sleep space would use 150 square feet and half wall designs 130 square feet. Based on approaches used in previous

health economic models, costs of construction were divided by effective lifespan of the new unit, which was assumed to be 30 years (Johnston et al., 2014).

Hospital module.

Major morbidity and length of stay base rates. Data sources. The baseline prevalence of neonatal mortality and major morbidities were derived from the IWK Perinatal Follow-up Database. This database captures all very preterm infants born in Nova Scotia, Canada since 1993. The dataset used for analysis included all live born very preterm infants from 2004-2018. Patients were excluded from the analysis if they had a major congenital anomaly, were not offered intensive care, or withdrew intensive care. Infants included in the database were cared for in one of three tertiary care hospitals: IWK Health Centre, The Cape Breton Regional Hospital, and the Moncton City hospital. Patients were also included in the follow-up database if they had been transferred to one of the above hospitals from a regional centre. All data at the IWK were entered and reviewed for accuracy by Dr. Mike Vincer (IWK Neonatologist and perinatal follow-up director) and verified by on-site visits at the other two hospitals by the same. Outcomes of interest from this database included mortality, and major morbidities potentially affected by single room care (NEC, sepsis, BPD, ROP, IVH) and length of stay.

Since reliable data on morbidities and mortality including relevant confounders were not available for infants born greater than 31 weeks gestational age, the economic model assumed that events rates for all morbidities and mortality was 0.1% for all infants older than 32 weeks. This allowed for minimal extrapolation of these data, which could result in unrealistic predictions given the rapid decline in complications as gestational age increases (Canadian Neonatal Follow-Up Network, 2019). This assumption was further verified against available data in the IWK Performance Analytics Database. Estimates for length of stay were extrapolated to

the full range of gestational ages under consideration, with extrapolations checked for logic visually using the results from the IWK single family room RCT as a comparison.

Causal Assumptions. Health economic models assume an underlying causal structure that varies in complexity depending on the question at hand. This is particularly true of decision models, where multiple sources of information are required to patch together a model that can be used to predict costs and health outcomes under various treatments. Analysis of the hospital module assumes the causal structure described in Figure 4, which was developed based on review of the literature and discussion with clinical experts. The described structure shows that a properly identified causal model requires either series of piecewise regressions that condition on mediating morbidities thus introducing the potential for collider bias, or a joint model that captures correlations between morbidities without explicitly conditioning on them.

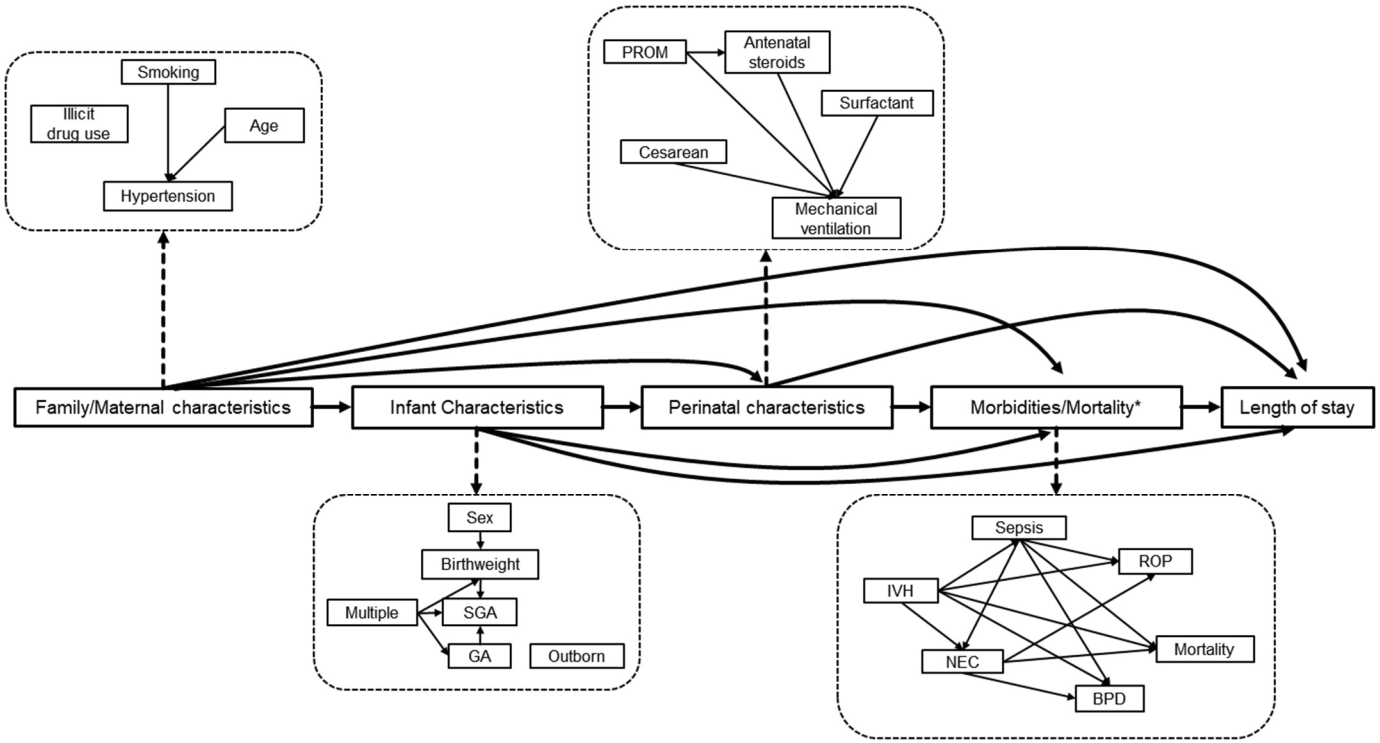


Figure 4. Directed acyclic graph for prediction of morbidities and length of stay, accounting for correlation between morbidities. Detailed assumptions regarding morbidities in box below morbidity term. SGA = small for gestational age; GA = gestational age; PROM = premature rupture of membranes; IVH = intraventricular hemorrhage; NEC = necrotizing enterocolitis; BPD = bronchopulmonary dysplasia; ROP = retinopathy of prematurity

Analysis. Prior to analysis, patient characteristics were explored using summary tables and plots to identify missing data or illogical entries. Continuous variables were summarized using mean and SD for and the number of patients and percentages for binary and categorical variables. Mortality and morbidity outcomes were analyzed as binary variables, with sepsis coded as one or zero indicating whether an infant ever had sepsis. A gaussian copula was used to capture correlation between outcomes (Aas & Berg, 2009) . Copulas are flexible tools for specifying dependence between variables with arbitrary marginals and allow for marginals to be modeled separately from their dependence structure. A two-step procedure was used to first predict each outcome of interest conditioning on all confounding variables, and then fitting a copula to pseudo observations by holding the marginals constant (Andersen, 2005). This allowed for modeling of the residual correlation, i.e. the dependence that remains after removing correlation induced by confounding. This approach was selected because it allows for models to estimate morbidities which can then have a treatment effect applied to them from the NMA without creating issues of double counting of treatment effects.

Owing to the large number of candidate variables required to predict, at times, events with small counts (e.g. IVH), a two-step pragmatic approach was used. In the first step the number of effective variables that could be realistically fit was determined based on methods described by Riley et al (2019), assuming that models would have a modest R^2 of 0.2 and using the observed prevalence or standard deviation as inputs. For binary outcomes, the criteria for determining the required sample size for a predictive model are: 1) A small degree of overfitting defined as an expected shrinkage of predictors by 10% or less; 2) A small difference between the models estimated and adjusted R^2; and 3) Precise estimation (within 5%) of the average risk in the population. If the required sample size was greater than what was available in the dataset, the

proposed number of variables was decreased by one. This process was then repeated until the estimated required sample size was equal to or less than the available observations. This number of parameters was then used as the "prior guess" of the number of true parameters in a regularized horseshoe prior (Piironen & Vehtari, 2017b). As the focus of the morbidity and mortality analysis was primarily concerned with the estimation of an unbiased effect of gestational age on morbidity for the purpose of stratification within the economic model, the proposed regression was further simplified by focusing on maternal age, smoking, illicit drug and illicit drug use as the strongest common causes of gestational age and morbidities. This ignores the potential confounding effect of twin birth, but this residual confounding is anticipated to be small. These types of priors can be thought of as a continuous version of stochastic search variable selection (Yi et al., 2003), which chooses a subset of variables that optimize the likelihood. This allows variables to be both partially in and out of the model and has favourable properties compared to other variable selection approaches such as p-value screening (Piironen & Vehtari, 2017b). While no method can offer optimal performance in exceptionally sparse datasets, this approach was chosen in favour of variable reduction and selection approaches since small datasets are not well suited for those approaches (Piironen & Vehtari, 2017a).

Results are displayed by each outcome of interest, using calibration plots and area under the receiver operating curve plots to visualize calibration and discrimination respectively. Calibration plots were fit by predicting the probability of outcomes from the model and then fitting a flexible local regression (LOESS) smoother (Cleveland & Devlin, 1988) predicting actual event status by predicted probability. Plots are then made to show actual probability (the output from the LOESS smoother) against predicted probability, with perfect calibration represented by a diagonal line with slope of one. The area under the receiver operating curve

shows an approximation of how well an algorithm does at separating risk histograms in the population. The higher the area under the curve the better job the model does at splitting individuals into high and low risk categories. The area under the curve is bound between 0 and 1, with 0.5 indicating random chance and values above 0.7 generally being considered good discrimination. Since the focus of the models was on estimation and splines were used the model non-linear effects of continuous variables, models are summarized using partial dependence plots. Since the models included are known to condition along a number of causal paths, these plots were primarily used to assist logic checks and cannot be meaningfully interpreted as total or even direct effects of any particular variable on any particular outcome.

Infant cost.

Data sources. All data for infant cost were provided by the IWK performance analytics department. The data consisted of all infants admitted to the IWK NICU from September 2013 to March 2016. Infant diagnoses are coded by trained abstractors using codes from the International Classification of Diseases, Tenth Revision (ICD-10) (World Health Organisation, 2015). Patient costs are assigned based on Canadian Institute for Health Information MIS Patient Costing Methodology and represent true costs based on IWK ledger and patient activity (Canadian Institute for Health Information, 2019). Costs include all hospital costs including direct and indirect costs fixed and variable costs, traceable supplies, operating room costs, etcetera and were adjusted to the 2020 Canadian dollars using the health and personal component of the Statistics Canada Consumer Price Index. At the time that costing data was developed nursing workload was not captured with enough detail to allow true costing of nursing time. Based on clinician feedback, infants are assigned a per-diem cost based on birthweight categories. In this approach, infants born >= 2500 grams were treated as the baseline per-diem,

those 2000-2499 grams given an amount equal to twice that, and those < 2000 grams given three times the weight. This limits the interpretation of the incremental costs of morbidities since these will partially be absorbed by the increased per diem in different birthweight risk categories.

Occurrence of morbidity and mortality were assigned based on presence of ICD 10 codes as described in Table 2. Infants were not excluded if they had a major anomaly. The total infant length of stay was computed as the number of hospital days from birth until discharge from the hospital.

Table 2. ICD 10 codes for neonatal cost analysis

Condition	ICD 10 code
Bronchopulmonary Dysplasia	P271
Necrotizing enterocolitis	P77
Intraventricular Hemorrhage grade III-IV	P5220, P5221
Sepsis	P36, A40, A41
Retinopathy of Prematurity	H351

Causal Assumptions. The assumed causal structure for cost data was simpler than that used for analyses of morbidities and length of stay, primarily for two reasons: The common causes of morbidities outlined in Figure 4 were assumed to exert their effects on costs through

effects on morbidities and length of stay and because the IWK costing data only contained limited information for adjustment. Assumptions of the analyses are outlined in Figure 5.

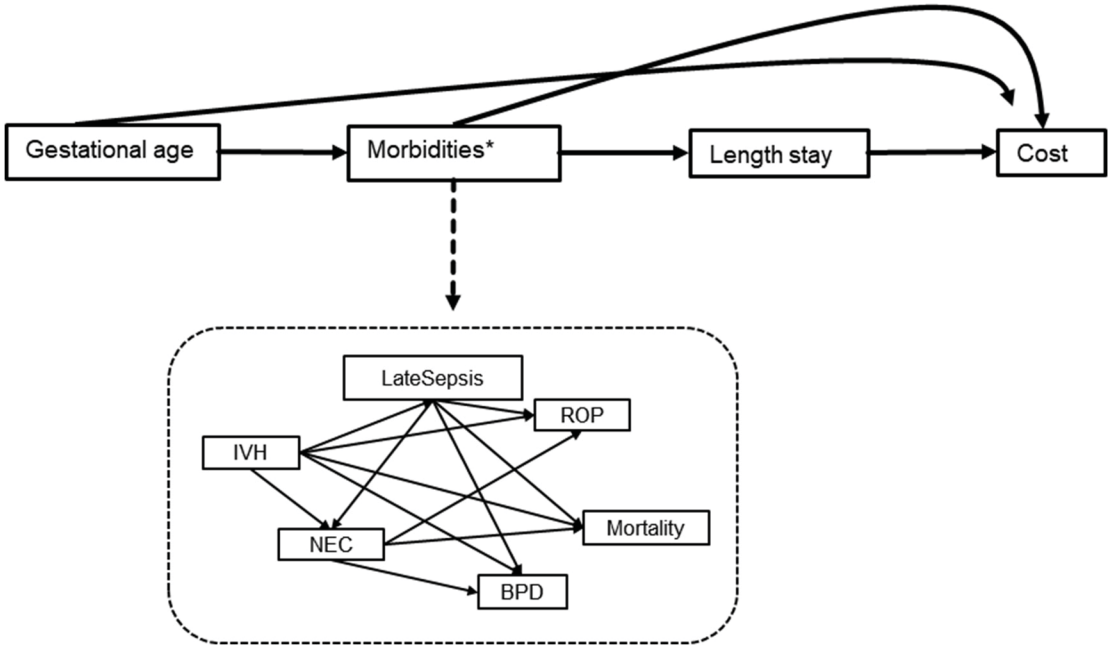


Figure 5. Directed acyclic graph for neonatal hospital costs. Detailed assumptions regarding morbidities in box below morbidity term. ROP = retinopathy of prematurity; BPD = bronchopulmonary dysplasia; NEC = necrotizing enterocolitis; IVH = intraventricular hemorrhage.

Analysis. Costs were visualized and summary statistics calculated prior to analysis in order to identify potential errors in data entry. Costs were visualized as a function of gestational age and birthweight using a LOESS smoother. The probability of ever receiving one of the morbidities of interest or of dying before discharge were visualized using the same methods. Since costs are a zero bounded variable with a potentially long tail (i.e. a small number of patients can have large costs) a log-normal distribution was considered appropriate. A log link was used which results in coefficients being interpreted as cost multipliers. Variables included in the model included binary indicators for BPD, NEC, IVH grade 3-4, sepsis, ROP, and death as well as a smooth cubic spline on gestational age. It was also assumed that the standard deviation of the model could change in response to gestational age (Umlauf & Kneib, 2018). Gestational

age was selected instead of birthweight because the two variables were highly correlated and gestational age was more relevant for subgroup analysis in the economic model.

Parent presence, out of pocket costs, and quality of life.

Data sources. Data for parent presence and out of pocket costs were gathered as part of the CHEZ-NICU baseline assessment. The first phase of the study took place from October 2016 to March 2018 at the IWK health centre. All infants during this period were cared for in open-bay units as construction was under way for the single-family room units. Participants included infants admitted to the NICU and their parents, with the only exclusion criteria being that they were anticipated to stay greater than five days, were anticipated to survive, were recruited within 10 days post-natal age, and that caregivers were able to read and write English. Parents who were included were asked to complete daily parent participation and financial cost diaries for each day they were present in the NICU. At baseline, two-weeks, and discharge mothers were administered the EQ-5D-5L quality of life questionnaire. The EQ-5D-5L is a standardized instrument for use as a measure of health outcomes and is applicable to a wide range of health conditions and treatments. It provides a simple descriptive profile as well as a single index value for health status. It has been extensively validated and is the preferred quality of life questionnaire for the National Institute for Health and Care Excellence (NICE). Data from a second phase of this study were also included. The second phase was identical in all aspects

except it occurred during a period where parents were randomized into open bay or single-family rooms.

Causal assumptions. Assumptions regarding the causal structure of the analysis of parent out of pocket costs and parent presence are summarized in Figure 6. The directed acyclic graph used for building the model for maternal utility at discharge is described in Figure 7.

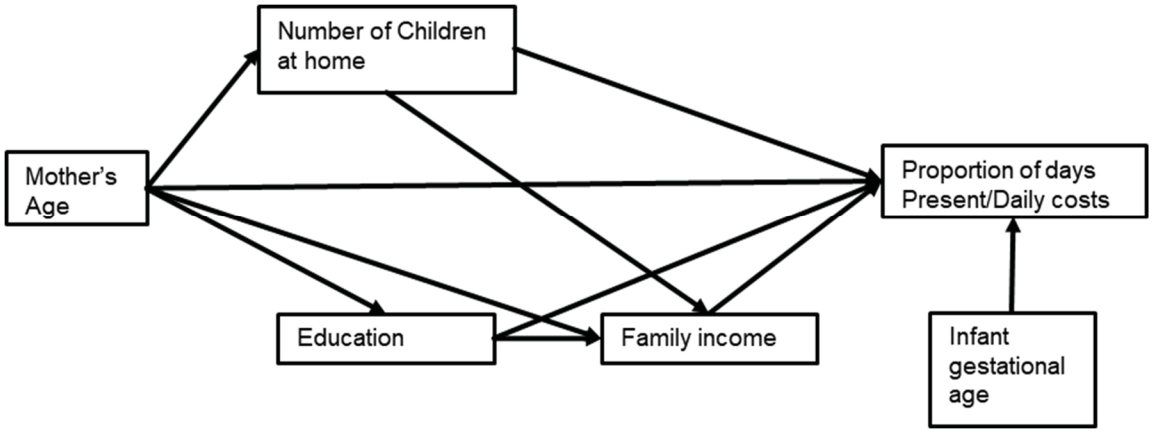


Figure 6. Directed acyclic graph for analysis of presence and out of pocket costs.

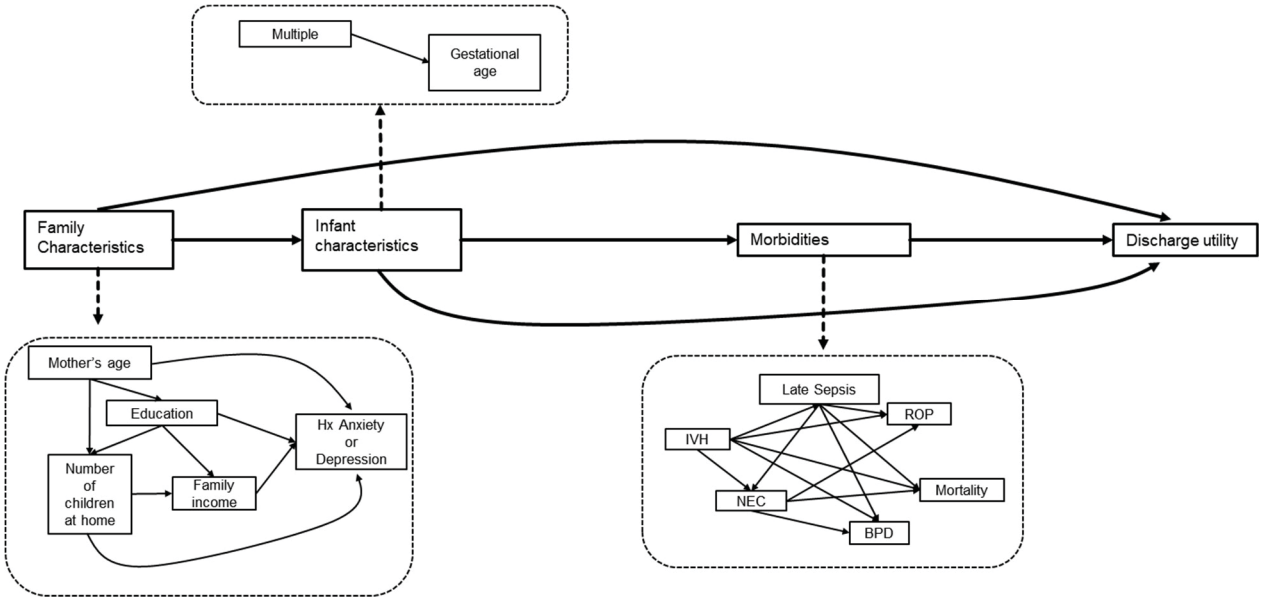


Figure 7. Directed acyclic graph for analysis of discharge utility. Detailed assumptions regarding morbidities in box below morbidity term. Hx = history; ROP = retinopathy of prematurity; BPD = bronchopulmonary dysplasia; NEC = necrotizing enterocolitis; IVH = intraventricular hemorrhage.

Analysis. Following data cleaning and inspection of summary statistics and descriptive plots, parent presence was estimated using a mixed effect model that clustered on family, allowing estimation for both mothers and fathers. Parent presence was estimated as the rate of days present using a Poisson model with a log link and log time offset for infant length of stay clustering on family units. Predictions from these model's respect boundaries, thus eliminating the chance of impossible predictions. The model included all variables in Figure 7. Family income was an ordinal income ranging from under $29,000 per year to > $150,000, patients also had the option to state they prefer to not answer. The ordinal nature of income was captured by specifying it as a monotonic effect (i.e. modeling it as an ordinal variable), which constrains increased family income to increasing the proportion of time present (Bürkner & Charpentier, 2020).

Parent out of pocket costs were assessed using a mixed model clustered on family, using daily observations. This approach was chosen instead of collapsing to average expenditure since it allows for increased power if observations within families are modestly correlated. A Bayesian hurdle gamma model was fit using brms which bounds predictions at zero but allows for exact zero expenditures without the need for continuity corrections (Hu et al., 2011). Predictors in this model included those in Figure 6.

Maternal EQ-5D-5L scores were converted the Canadian utilities using the formula published in Xie et al (2016). Utilities were converted to disutility and modeled using a log normal likelihood. Variables included in the model are described in Figure 7. Because this represents a large number of variables relative to observations, a regularized horseshoe prior was placed on all variables, with the expected number of true parameters as three (Piironen & Vehtari, 2017b). These types of priors can be thought of as a continuous version of stochastic

search variable selection (Yi et al., 2003), which chooses a subset of variables that optimize the likelihood. This allows variables to be both partially in and out of the model and has favourable properties compared to other variable selection approaches such as p-value screening (Piironen & Vehtari, 2017b).

Discharge to end of life. For the reference case and societal non-reference case, a life-time horizon was used. Horizons of this length are appropriate when studying chronic conditions, or those that may result in benefits or harms that occur in the distant future (CADTH, 2016).

Probability of disability given hospital course.

Data sources. The baseline prevalence of neonatal mortality and major morbidities and developmental outcomes were derived from IWK Perinatal Follow-up Database. Details of this database are as described earlier, in the Major morbidity and length of stay base rates. section. The outcomes of interest were Bayley III language, cognitive, and motor scores corrected at 18-21 months corrected gestational age and cerebral palsy severity as measured by the Gross Motor Function Classification System (GMFCS) (Palisano et al., 1997). As less than 2% of the population had GMFCS greater than or equal to three, any cerebral palsy was defined as GMFCS of one of greater. Developmental assessments were conducted by a qualified multidisciplinary team including a neonatologist and other allied health professionals.

Causal assumptions. Causal assumptions for the disability model are outlined in Figure 8. These are broadly similar to the hospital course module but also include parent socio-economic status as measured by Hollingshead four factor index of social status (Hollingsead, 1975), and specifies the relationship between disabilities.

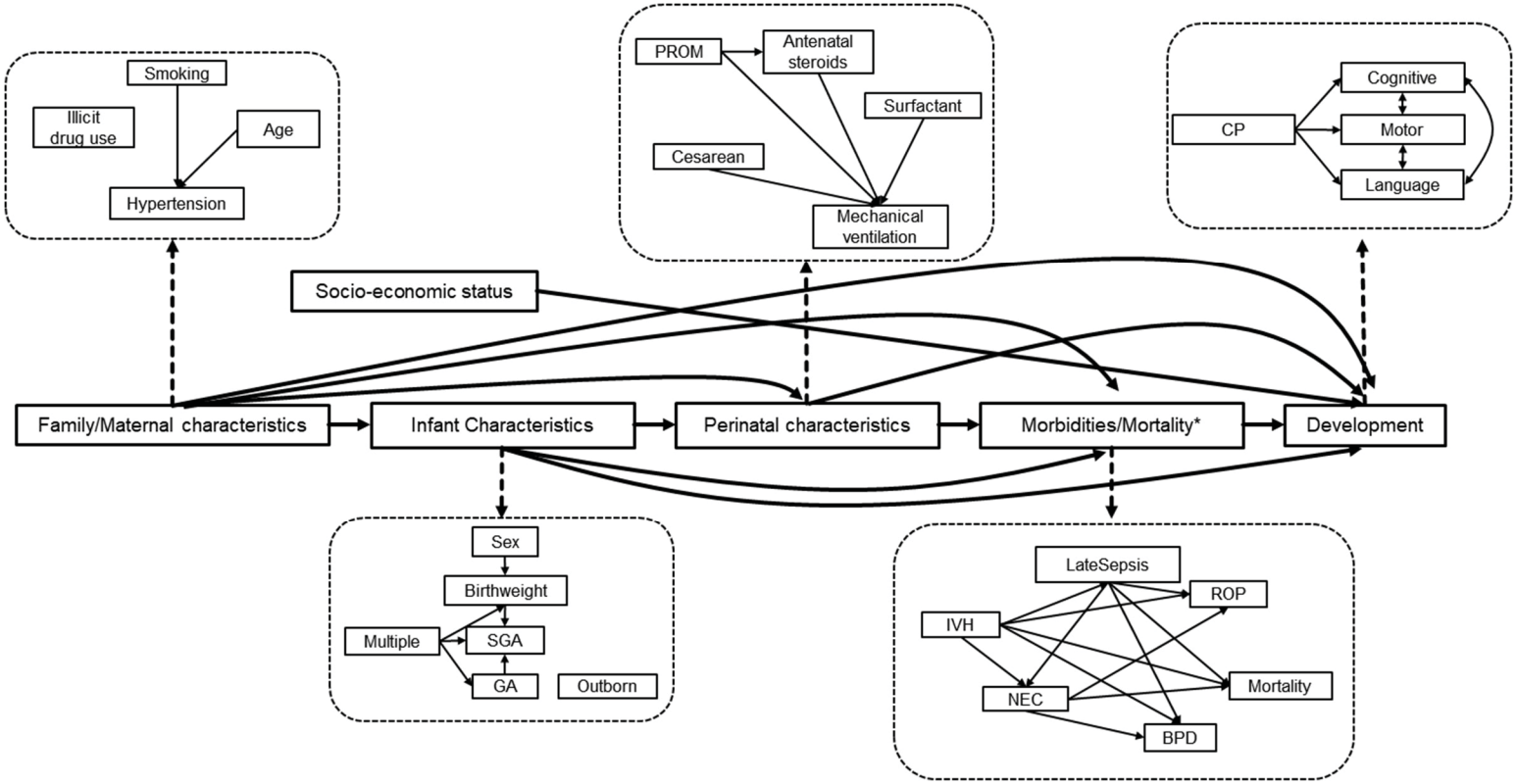


Figure 8. Directed acyclic graph for disability. Detailed assumptions regarding morbidities in box below morbidity term. SGA = small for gestational age; GA = gestational age; CP = cerebral palsy; ROP = retinopathy of prematurity; BPD = bronchopulmonary dysplasia; NEC = necrotizing enterocolitis; IVH = intraventricular hemorrhage.

Analysis. Prior to analysis patient characteristics were explored using summary tables and plots to identify missing data or illogical entries. Continuous variables were summarized using mean and SD for the number of patients and percentages for binary and categorical variables. Bayley corrected cognitive, language, and motor sub-scores were estimated using a Bayesian model with a gaussian likelihood and identity link. Confounding factors included in these models included all variables preceding morbidities/mortality in Figure 8. Socio-economic status was modeled as an ordinal variable using methods outlined by Burkner and Charpentiere (2020) which allows for more efficient estimation of predictors with ordered categories. A model was also placed on the standard deviation of the normal likelihood, using a smooth spline on gestational age. This captures the belief that infants of different gestational ages may not only have different predicted mean scores but different variability in scores. Probability of having a GMFCS score of one or greater was modeled using binomial likelihood with monotonic effects for socio-economic status and smooth splines for maternal age, number of cigarettes smoked per day, and infant gestational age.

Incorporation of disability predictions in the economic model. Incorporation of Bayley III scores in the economic model was achieved by predicting a set of scores for an infant directly, then classifying them as having NDI if any score was less than 85 but greater than 70, and sNDI if any score was 70 or less. These cut-points were combined with predicted cerebral palsy status score wherein any Bayley < 85 or positive cerebral palsy was considered NDI. Based on only 2% of the follow-up dataset showing GMFCS > 3, this was not modeled directly. Based on the observation that approximately 10% of infants with any cerebral palsy had severe cerebral palsy, this simplifying assumption was used in the model. Classification of disability states has also relied on the presence or absence of levels of hearing and visual deficits (Synnes et al., 2017),

however this was not incorporated in the present analysis. The decision not to use these outcomes for definition of NDI and sNDI was based on there being too few events to estimate them.

Transition between disability states. To remain consistent with the most recent publications from CN-FUN, disability states included no disability, NDI, and sNDI (Figure 3). Since the environment a child is raised in can be supportive or harmful, and because certainty in degrees of disability changes over time, the model allows for transition between disability states from discharge to 18 years of age (Mangham et al., 2009).

Data sources. Transition probabilities were adapted from a previously published burden of premature model for England and Wales (Mangham et al., 2009). Data for estimation of disability transitions in this model were calculated from individual level data of a cohort of UK children from 1991-1992 and followed-up at two, five, and eight years. In order to align definitions of disability, the moderate and severe categories from this study were combined, since the definition of moderate disability was similar to that used as sNDI in CN-FUN analyses. Disability in this study was classified using the Wechsler Intelligence Scale for Children, fourth edition (Wechsler, 1949) or Bayley II for developmental delay, the GMFCS for cerebral palsy, and hearing and vision tests. Mild disability was defined as more than a one SD decrease in developmental quotient/IQ score or mild cerebral palsy, and moderate/severe was defined as greater two SD decrease in developmental quotient/IQ score and moderate-severe cerebral palsy.

Analysis. In order to align definitions of impairment with the economic model, moderate and severe disability were merged together. This was deemed appropriate based on the similarity of the moderate impairment category to the Synnes sNDI definition (Synnes et al., 2017). From 19 years until end of life, disability was assumed constant. The transition matrix provided was to

parameterize a Dirichlet distribution, which is appropriate for specifying uncertainty in a multinomial outcome (Briggs et al., 2006). This model relies on the assumption that the transition hazard (i.e. the instantaneous probability of any given transition state) is constant over the given time period.

Costs associated with disability. In the reference case, costs included all direct and indirect costs of construction and infant care during the inpatient course, and any additional interactions with the health care system including developmental assessments, early intervention, and hospital or physician visits. The probability of need for additional resources was estimated from summary data available in a Canadian health resource use study conducted by Nassel et al. (2019) using a logistic regression to predict medical complexity status based on level of disability. In the non-reference case, a societal perspective was taken capturing additional costs related to need for special education (Johnston et al., 2014). No productivity or out of pocket costs were considered for infants/children.Costs for these portions of the model are summarized in Appendix Table A 7. All costs were appropriately inflated using the Bank of Canada consumer price index healthcare basket of goods if estimated from a Canadian source and the Organisation for Economic Co-operation and Development price index otherwise.

QALYs associated with disability. Health state utilities were sourced from a systematic review of studies assessing preference-based health related quality of life outcomes in preterm infants (Petrou et al., 2019). While the overall quantitative estimates from this review were not relevant to the decision problem since they were not restricted to disability states, the review included several studies that estimated utility decrements based on disability categories assessed across the lifespan. Since studies that investigated utility amongst different groupings based on disability generally combined no and mild disability, the same approach was used for the present

model. Saigal et al. (Saigal et al., 2016) was used to provide estimates of mean utility decrement and its standard error, which were then used to parameterize a gamma model for probabilistic uncertainty analyses. Scores for adolescent utility were used from discharge to age 16.

Background probability of death. For parents and infants, Canadian life tables from Stats Canada based on estimation from 2016-2018 were used to provide the background probability of death.

Probability of death given disability. No studies were identified that provided an estimate of the incremental increase in the hazard of death given disability states, and thus to be consistent with previous studies (Mangham et al., 2009), estimates from Strauss et al. (2008) were used for the incremental risk of death given severe disability (2009). Probability for NDI and no disability were both only based on life tables.

Parent QALY associated with infant disability. The approach to inclusion of spillover burden for parents was based on a recently conducted National Institute for Health and Care Excellence Decision Support Unit review of incorporation of spillover in NICE single technology assessment submissions (Pennington & Wong, 2019). Parent QALYs were thus measured using a two-step procedure. First, based on findings from Song et al. (Song et al., 2011) a utility decrement was applied if the infant died and the parent was in bereavement. As duration of bereavement varies by individual (Rogers et al., 2008) the number of years in bereavement was assumed to be at least one, with the number of additional years sampled from a negative binomial distribution with size of 2 and mean of 2 which creates a distribution with a long right tail where 50% of the probability is concentrated on a total duration of grieving from one to two years. The decrement applied to parents that are grieving death was -0.04 (standard error = 0.02) for both mothers and fathers (Song et al., 2011). If children were alive, then utility

was based on the disability state of the infant. While no direct estimate of this was identified in the literature, a systematic review by Wittenberg and Prosser identified a small number of studies evaluating disutility of illness for caregivers and families of children with chronic and acute illness. Based on this review, the spillover for parents of children with mild disability was taken from parents of children with activity limitations assessed by Kuhlthau et al. (2010) (-0.07, 95% CI: -0.10 to 0.03), and for serious disability the value for serious congenital anomalies from Poley et al. (2012) (-0.1) were used (Appendix Table A 7).

Parent out of pocket costs. Parent out of pocket and productivity costs were estimated using resource use estimates as identified by (Nassel et al., 2019) using the same strategy of using a logistic regression to predict medical complexity given disability status. The impact of hospitalizations and outpatient visits were estimated using the same approach employed by Johnston et al. (2014). Outpatient visits were associated with two hours taken off work and inpatient days were associated with eight hours off work, with an assumed hourly wage of $23.18 (Johnston et al., 2014).

Treatment Effect Parameters

Morbidity, mortality, and length of stay.

Data sources. Data for length of stay and mortality came from admissions during the randomization phase of the IWK's transition to single-family room care. Data consisted of all infants admitted between April 2018 and April 2019. Allocation to room was based on a minimized randomization schedule based on unit acuity at the time of admission. The dataset

provided included the probability of randomization to each team at the time of admission, total length of stay in days (counting admission as day 1) and the gestational age of each infant.

During the period that infants were randomized into single-family room (SFR) or open bay rooms randomization was based on a minimization algorithm intended to approximately balance rooms in terms of acuity. The algorithm was based on the acuity of the child to be admitted as well as unit specific Winnipeg Assessment of Neonatal Nurse Needs Tool (Winnipeg Regional Health Authority, 2013) scores that were completed twice daily. Based on these factors the probability that an infant was randomized into one unit or another was altered to maintain comparable staffing requirements.

Infants were randomized to one of two teams made up of nurses and allied health professionals as well as neonatologists, fellows, and residents. Staffing was organized specifically to ensure that nurses and neonatologists were assigned to both teams equally, although this was not formally randomized. Randomization was determined by a computer algorithm at the time of admission. Owing to the nature of the intervention, it was not possible to blind staff or parents to room allocation. All quality improvement initiatives were initiated in both room designs and support staff were available equally. During the period that both units were open, the total square footage was greater than in the original open bay alone and thus modifications were made to allow for beds alongside each patient.

In addition to the full set of data on mortality and length of stay, a smaller subset of patients was also recruited into a second phase of the CHEZ-NICU study, which had the same inclusion/exclusion criteria as the phase that was restricted to open-bay only. This overlapping cohort provided information on neonatal morbidities not captured in the larger cohort used for

length of stay and mortality. This dataset was also used to calculate a treatment effect single family room on the rate of days present.

Analysis. Prior to analysis, the distributions of length of stay and summary statistics for mortality were assessed visually. Since data collection began on the first day that the SFRs were open, a scatter plot of length of stay by admission date stratified by room design was also created and a locally estimated scatterplot smoother was used to assess for signs of a trend over time. The dataset provided to support these analyses also contained the exact probability of randomization for each infant, which were used to create propensity scores weights to allow for the calculation of the average treatment effect. For length of stay a weighted gaussian likelihood with a log link was used in order to provide the same type of estimate as that used for the subsequent meta-analysis. Use of the log link ensures that predictions of mean lengths of stay less than zero are impossible and leads to interpretation of the treatment effect as being a length of stay multiplier. The effect of treatment of the probability of death was estimated using a weighted logistic regression. Standard errors were adjusted using robust sandwich estimators. In order to be compatible with the format required for the network meta-analysis models, the model for death was estimated by arm and confirmed to provide the same point estimates and standard errors.

Parent presence and out of pocket costs.

Data sources. Data consisted of the CHEZ-NICU phase 2 data which gathered information regarding parent presence and out of pocket costs a manner similar to that used for the baseline history mode.

Analysis. All analyses were conducted using the same approaches used for the baseline natural history model analysis of parent presence and out of pocket costs, with the addition of a

term for treatment effect. For presence, the treatment effect is interpreted as a multiplier of the underlying rate of days present and for costs as a cost multiplier.

Meta-analysis. For each room design, a set of relative risks were derived for major morbidities, length of stay, and Bayley III scores. In keeping with NICE guidelines, these estimates were derived from a systematic review and network meta-analysis of the existing literature (Kaltenthaler et al., 2011). Reporting of the results of the review is consistent with Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) statement and the corresponding extension.

Identification and selection of relevant studies. An electronic search of the literature was developed in collaboration with an experienced trained systematic review librarian. The original search was conducted in Feb 2017 and updated Feb 2020. A full search strategy is available in Appendix Table A 1. Databases searched included MEDLINE, Embase, CINAHL, Cochrane CENTRAL, and Web of Science.

Search strategies included both key words and controlled vocabulary (e.g. MeSH headings), with no limitations on dates. The search was not limited by trial design type or language, and no additional filters were used. Duplicate citations were removed prior to screening. In addition to the electronic search the reference list of included articles was searched, and Google Scholar was used to identify references to included articles.

In the initial search in Feb 2017 and 2020 update, two reviewers screened abstracts at the title and abstract and full-text phase, with disagreements settled by a third reviewer if consensus could not be reached. Reviewers included Tim Disher, Justine Dol, and Brianna Richardson who are all doctoral trainees in the lab of Dr. Marsha Campbell-Yeo and have received formal training

in the conduct of systematic reviews. Titles were screened based on eligibility criteria consistent with the scope of the decision problem (Table 3).

Table 3. Eligibility Criteria for Systematic Literature Review

Item	Inclusion Criteria
Population	Infants cared for in a neonatal intensive care unit or special care nursery
Intervention/Comparators	Any move to a new unit, compared against the prior unit
Outcomes	Length of stay, late-onset sepsis, bronchopulmonary dysplasia (as defined by study), intraventricular hemorrhage grade III or greater, retinopathy of prematurity requiring treatment, mortality, Bayley III cognitive, motor, and language sub scales
Study Design	Any comparative study
Publication type	Peer-reviewed publications

Data collection. In the initial search, data extraction was conducted independently by two reviewers (TD, BR, JD) and in the update was conducted by a single reviewer (TD). Data were extracted using a standardized form designed in Microsoft Excel (Microsoft Corportation, 2019). Characteristics of interest included neonatal and maternal characteristics (e.g. gestational age, socio-economic status), outcome definitions, and characteristics of room designs (e.g. sleeping space for parents). Access to a parent sleep space was considered the defining feature of a single-family room, with bedside chairs or benches in an individual patient room defined as being a private room. Dichotomous outcomes were extracted as counts of events and patients analyzed, and continuous outcomes were extracted as means and standard deviations.

Risk of Bias Assessment. As the PICOs for this review would include studies with multiple designs including both randomized and non-randomized studies, quality appraisal was conducted using the quality appraisal checklists of the Joanna Briggs Institute. These checklists include a review of study domains relevant to the quality of randomized (Joanna Briggs Institute, 2017b), non-randomized cohort (Joanna Briggs Institute, 2017a), and quasi experimental studies

(The Joanna Briggs Institute, 2017). Risk of bias assessment was conducted independently by two reviewers (TD, JD, BR) in the initial search, and by a single reviewer for the update (TD).

Assessment of suitability for synthesis. Prior to conducting any meta-synthesis, characteristics of the included trials anticipated to modify the effectiveness of treatments were compared qualitatively across treatments, studies, and outcomes using tables of summary characteristics. This included patient level (e.g. gestational age) as well as study level characteristics (e.g. design).

Pairwise meta-Analysis. Prior to the conduct of any network meta-analysis, pairwise meta-analyses of each outcome for each treatment were conducted. The results of the analyses were summarized using forest plots and were intended to assist the identification of potentially influential studies and explore sources of heterogeneity. Heterogeneity was summarized using both the I^2 statistic, which describes the amount of variability between studies not accounted for by sampling error as well as qualitative inspection of plots (Julian P T Higgins & Thompson, 2002). This was done since when analyses include trials with high precision, they can lead to very large I^2 values even if the treatment effects are similar across included studies in terms of magnitude and direction of effect. Qualitative assessment of statistical heterogeneity is also beneficial in cases where there are very few studies, since estimates of heterogeneity and I^2 can be unstable and uninformative. All pairwise meta-analyses were conducted within a random effects framework, which assumes estimates of study effect can vary across included trials (Dias et al., 2011). This was considered appropriate from a theoretical perspective since

implementations of new room designs can vary in terms of quality, population of patients included, and residual bias from confounding or other sources.

Since the final model code for the network meta-analysis used requires that data for dichotomous outcomes be entered as log-odds (instead of raw counts), pairwise meta-analyses were conducted on this scale using inverse variance weights. Similarly, since the network-meta analysis for length of stay included data entered both in arm format and contrast format all data for this analysis were entered as treatment differences. Room designs were assumed to exert their effect on the log ratio of means scale, as this reflects that length of stay has a minimum of zero and a long maximum tail (Friedrich et al., 2008). For the Bayley scales, no synthesis was possible because of overlap in trial populations across room designs, but results were still summarized using forest plots.

Network meta-analysis. Analyses were conducted in a Bayesian framework, with original open bay chosen as the reference treatment. The structure of the network of evidence was examined using network plots, where larger nodes indicate more patients receiving a given exposure and thicker lines connecting nodes indicating more studies making that comparison (Dias et al., 2011).

Models for dichotomous and continuous variables. Outcomes evaluated in the NMA included a mix of binomial and continuous outcomes. A typical strength of the use of Bayesian methods is the ability to use exact likelihoods as opposed to relying on synthesis of odds ratios or mean differences directly (Dias et al., 2011). However, given that outputs of the economic model are a non-linear function of treatment effects and studies included in the analysis rarely reported all outcomes of interest, multivariate models were used (Achana et al., 2014; Bujkiewicz et al., 2019). The primary disadvantages of these models are that they rely on the normal

approximation the log odds to hold which may not be the case when the number of events is very low or trials are small, and that between outcome correlations are rarely reported. In the case of zero events, it is not possible to calculate on odds ratio, so in these cases the correction by Sweeting, Sutton, and Lambert (2004) was used which has been shown to be robust against biases that typical zero-cell corrections suffer from. Following Achana et al. (2014) between outcome correlations were calculated using individual patient data from the IWK Perinatal Follow-up Database. Correlations and uncertainty were calculated using Pearson correlations as described by Achana et al. A strength of this approach is that the IWK Perinatal Follow-up Database is entirely external to the data included the NMA, but a limitation is that it only includes very preterm and extremely preterm infants, which likely does reflect the patient mix of entire units. These potential limitations were deemed acceptable given the relevance of capturing between outcome correlation.

The multivariate model described by Achana et al. (2014) has three important benefits for the decision problem. First, by including a model for between outcome correlation the analysis can "borrow strength" across outcomes when there are substantial missing data. This can both correct potentially biased analyses when missing outcome data are conditional on the performance on other outcomes (i.e. selective outcome reporting) while potentially improving precision. These benefits can become influential for decision making when the amount of missingness is high, which it is for this analysis. Second, the model described by Achana et al. (2014) also allows for borrowing of strength across interventions by assuming "constant potency." In simple terms this implies that active comparators have similar comparative effectiveness across outcomes while also allowing for the difference versus the reference treatment to vary. Of relevance for the problem at hand, this also allows for the imputation of

treatment effects on outcomes that a treatment was never evaluated for. For example, half-wall units are only evaluated in a single study that does not report all outcomes (Altimier et al., 2005). The multivariate model allows for imputation of effectiveness of half-walls into other outcomes, based on both between outcome correlation and the relative effect of half-walls compared to other treatments in outcomes for which it was evaluated. An additional benefit of this assumption is that it can help protect against over-interpretation of large differences in effectiveness across correlated outcomes that may be driven by chance. By assuming constant potency, estimates for any given intervention are "shrunk" together, towards an over-all average effect. This is analogous to approaches recently recommended for the analysis of single outcomes, where treatments within a class have exchangeable treatment effects (Efthimiou & White, 2019).

The impact of the assumption of the multivariate model was explored using sensitivity analysis where no shared potency or correlation between outcomes was assumed and therefore results are equivalent to separate univariate models. However, since the economic model requires inputs for all parameters, assumptions are still required considering how room designs would perform on outcomes that they had not been evaluated in. A simple assumption was made that new open bays would have equivalent efficacy to old open bays in these cases, and that all degrees of private room share the same efficacy. Due to the large number of outcomes and missingness of data, and differences in effect measures (odds ratio vs ratio of means vs mean difference) it was not possible to fit a single multivariate model for every outcome of interest simultaneously. All morbidities were evaluated in one model, and Bayley scores were analyzed in a separate multivariate model using mean difference as the effect measure.

Length of stay was analyzed in a separate univariate model using the ratio of means (Friedrich et al., 2008). This is a departure from other meta-analyses of single rooms (van

Veenendaal et al., 2019) which use mean difference. The decision to use ratio of means was based on a desire to respect that mean length of stay cannot be less than zero and thus it would be more realistic to assume that treatment effects apply on a ratio scale. Importantly, a ratio measure is required in order to get valid results for the economic model, since otherwise probabilistic analyses might select combinations of baseline rates and treatment effects that would imply negative lengths of stay. Since adjusting negative numbers to be zero would change the mean difference and standard error that is estimating, a ratio of means approach offers a practical solution that also more closely reflects the underlying data generating mechanism.

For binary variables (i.e. morbidity and mortality) random effects models were used for all outcomes using informative priors as described by Turner et al (2015). This was even the case when individual evidence networks did not have sufficient evidence to provide data on between trial heterogeneity. This decision reflects the belief that it would be unreasonable to expect all trials to be estimating the same common effect (i.e. fixed effect) and helps to avoid making overly precise estimates. Only fixed effect models were possible to fit for Bayley scores since the network consisted solely of single study connections and it is not possible to derive a set of informative priors that could apply across all outcomes (e.g. the range of possible mean differences are scale dependent). While informative priors have been derived for continuous outcomes assessed on the standardized mean difference (SMD), this also implies that effectiveness acts in terms of standard deviations, requires the use of an estimated parameter (SD) to standardize the outcome, and can distort relative effectiveness in non-homogenous populations (Busse et al., 2015).

A modification was required to the Achana et al. (2014) model in order to allow for convergence and reasonable estimates. In their original model, Achana et al. place a model on

both the within trial correlation between outcomes and the between trial correlation of random effects across outcomes. This allows the model to capture variation between treatment effects on outcomes that is greater than the variation implied by the within study model. In their applied example, Achana et al. found that there was very little data to inform this aspect of the model for a simpler three outcome analysis with less missingness and more studies. Only two studies reported all outcomes of interest suggesting even less data available to estimate an even more sparsely populated network.

Incorporating observational trials. Because of the inherent additional risk of bias in observational trials, naïve or unadjusted syntheses are inappropriate (Schmitz et al., 2013). Four general approaches to combined synthesis have been proposed in the literature: individual adjustment by external experts, incorporation of observational trials as prior information in a Bayesian synthesis, down-weighting of trial contribution to analysis via variance inflation, and Bayesian hierarchical models (Efthimiou et al., 2016). Individual study adjustment was deemed to not be feasible within project timelines and incorporation of observational trials as prior information was not possible since only single-family room had any randomized controlled trials. Bayesian hierarchical models were considered as potentially feasible, but the large amount of missing data across outcomes and the need to account for this as well as correlation of outcomes lead to the choice of variance inflation.

While variance inflation addresses the issue of larger observational trials overwhelming smaller RCTs in meta-analyses, it does not directly address the issue of potential bias. This was particularly problematic for the data at hand as most designs are informed entirely by non-randomized data. Variance inflation alone in these cases would increase the uncertainty (i.e. widen credible intervals) around estimates but would leave their (potentially biased) point

estimates in place. Manual bias adjustment of trials is an option in these cases however it requires both extensive consultation with clinical experts and is limited by the difficulty associated with not only estimating the magnitude of bias but also the direction (Efthimiou et al., 2016). An advantage of conducting analyses within a Bayesian framework is that inferences are based both on data and prior belief. While priors for analyses are often selected to be vague or uninformative, these default priors typically entertain extremely unrealistic treatment effects. For example, typical vague priors would imply that odds ratios greater than 300 are more likely than odds ratios between 1-5 which would be inconsistent with prior meta-analyses of the effects of prophylactic interventions for preterm infants (Askie et al., 2018; Roberts et al., 2017; Subramaniam et al., 2016).

Following these considerations, priors were set that would constrain the model space to consider only those estimates which could be realistically feasible given common treatment effects observed in randomized trials as well as the proposed mechanism of action of room design. Since the influence of priors depend on the precision of the estimate from data (i.e. all priors are overwhelmed by sufficient data), these priors combined with variance inflation results in trials that have greater variance inflation will also be more strongly informed by the prior. This approach avoids issues related to deciding the magnitude or direction of bias.

For dichotomous outcomes, a normal prior centred on zero with a standard deviation of 1 was used. This would imply that a 95% interval on the odds ratio before looking at the data would be 1.00 (0.14 to 7.1) suggesting that room design could feasibly be expected to reduce a common event with probability of 20% to as low as 3% or increase it to 64%. For rarer events (e.g. mortality) a risk of 4% could be reduced to 0.5% or increased to 23%. At the same time, these priors also concentrate 60% of the expected probability between odds ratios of 0.59 and

1.28 which, based on results from reviews of neonatal interventions including environmental changes (Askie et al., 2018; Roberts et al., 2017; Subramaniam et al., 2016; Symington & Pinelli, 2006), was considered to offer sufficient skepticism of large effects while being vague enough to entertain estimates equal to or more beneficial as those observed for antenatal steroids (Roberts et al., 2017). Sensitivity analyses using more vague priors (zero centered with standard deviation of 10) were used to assess sensitivity of conclusions to prior choice.

Priors for treatment effects for length of admission were set on the ratio of means of scale and using a normal distribution centered at 0 with a standard deviation of 0.6. This implies a 95% credible interval of 1 (0.31, 3.24) with 60% of probability concentrated between 0.73 and 1.16. If the control arm mean from Ortenstrand (2010) (mean length of stay = 32.8 days) is used to convert these values to length of stay, then new room designs could be expected to reduce the mean length of stay to as short as 10 days or increase to as much as 106 days, but more credibility is placed on estimates of 23.9 and 38 days. Priors for treatment effects on continuous scales (i.e. mean difference) are more difficult to set as aggressively since they are sensitive to the underlying scales. Since the Bayley scales are large enough (mean 100 with standard deviation of 15) that mean difference is a reasonable scale for the linear predictor. Priors for these outcomes were therefore set to be normal distributions centered on zero with standard deviation of 15. This captures the belief that room could feasibly move a population average Bayley III component score from 70 to 100 (two standard deviations). While this is still quite uninformative, it provides more regularization than the typical default priors and appropriately reflects the decreased certainty in the reasonable bounds of treatment effect relative to morbidities and length of stay.

Missing data. For multivariate models, missing outcome data is entered as "NA" and imputed automatically based on draws from the posterior predictive distribution (Achana et al., 2014). Missing standard errors are modeled as coming from a zero-truncated normal with a standard deviation estimated by the data. For the length of stay model, several trials were missing either means, standard deviations, or both. When medians were available and either quantiles or range were provided, the quantile estimation method developed by McGrath, Sohn, Steele and Benedetti et al (2020) was used to estimate means. When confidence intervals or standard errors were available, the standard deviation was calculated using exact methods (Higgins & Green, 2011). If these were not available then standard errors were imputed using the same approach as used by Achana et al. (2014) for missing standard errors in multivariate network meta-analysis. This is equivalent to a multiple imputation approach.

Outcome measures. Pairwise comparisons of interventions were presented as odds ratios (ORs) for dichotomous outcomes, mean difference (MD) for Bayley III scores, and ratio of means (RoM) for length of stay. Summaries of effects are displayed in forest plots and include 95% credible intervals, which describe the uncertainty in effect estimates conditional on priors and the data and are intended to provide an estimate of credible magnitudes of effect. Each estimate is also accompanied by the probability that a treatment is better than the original open bay, which is calculated by estimating the proportion of posterior samples where the treatment effect estimated is more favourable than the reference treatment. Both estimates are intended to

provide a summary of results by treatment and outcome, however all conclusions are based on the results of the SMAA and health economic model.

Iterations, & convergence. Models were estimated by Markov Chain Monte Carlo samples of the posterior using OpenBUGS (multivariate models) or JAGS (length of stay). Convergence and iterations were determined as outlined in Convergence of Bayesian Models.

Assessment of model fit. Standard recommendations for NMA suggest that fixed and random effect models are fit and compared in terms of absolute fit (e.g. total residual deviance) and relative fit (e.g. deviance information criteria). Absolute fit is used as a measure of how well the model fits the data at hand while deviance information criteria provides an estimate of future out of sample predictive error and can be used to decide between competing models based on principles of parsimony (Dias et al., 2011). These measures were not used in these analyses both because the methods to assess them have not been adapted to these settings (Achana et al., 2014) and because the data underlying the model were assumed to be biased and thus the priors and variance inflation were used in order to shrink estimates closer to zero. This would result in poorer model fit compared to one that did not take these measures and thus would evaluate models against the wrong target (assumed biased trials). Model fit was thus evaluated based primarily on biological plausibility and (in the case of single-family room) concordance with estimates from randomized trials. Information criteria (e.g. deviance information criteria) were not used to select between fixed and random effect models because the fixed effect assumption was assumed to be implausible a priori and because the random effect model collapses to the fixed effect model as a special case (Whiting et al., 2016). This approach thus appropriately considers uncertainty in model selection, which is important since use of deviance information

criteria creates a data-based analysis decision that is otherwise not accounted for (Piironen & Vehtari, 2017a).

Assessment of consistency. Consistency is a statistical phenomenon that arises from the disagreement between direct and indirect sources of evidence. It can be thought of as the quantitative evidence of violation of the transitivity assumption of NMA, which is closely related to the requirement that trials included in a meta-analysis are relatively homogeneous (Dias et al., 2018). The transitivity assumption extends homogeneity within a comparison to between comparisons and thus intransitivity, inconsistency, and heterogeneity are all related. Because there were no closed loops in any networks, inconsistency could not be assessed however the feasibility of the transitivity assumption was evaluated qualitatively, and within comparison heterogeneity was interpreted as a risk for between comparison heterogeneity (i.e. inconsistency).

Incorporation of treatment effects in the economic model. Treatment effects were incorporated into the economic model by applying their effects on the appropriate scale (e.g. log odds for odds ratios) to the estimates from the baseline model. Since only private room and single family room had estimates for the Bayley scales, the following assumptions were made: (1) Since all publications of single family room reviewed in the literature review showed substantial increases in parent presence, the estimates from Lester et al. were used unchanged; (2) Combined pod and single room designs were provided the same treatment effect as single family room; (3) Half-wall and all open bay designs received an effect of zero; (4) Private room infants received the treatment effect from Pineda et al. (2014), if mothers were predicted to have less than 50% presence and zero otherwise. The cut-point of 50% was based on the assumption that the 25 hours of average weekly presence observed in Pineda et al. was split over seven-hour

days. This parameter was specified as a beta distribution with shape and scale of 103 which provides a range of samples from 40% to 60%.

Infant and parent demographics. Analyses above were conducted using a number of different data sources and unbiased estimates required conditioning on variables that were not present in all datasets. Many variables could be estimated through freely available summary documents, but this would ignore correlations between variables. A dataset was therefore created by first combining all individual patient data datasets creating a master dataset with all variables used in any model. Then, missing data were imputed using Bayesian Gaussian copulas (Hoff, 2007) to create a final dataset for sampling. To ensure that patients were representative in terms of gestational age and birthweight the imputed data was then trimmed to only include patients from the IWK costing data. Patients were then sampled from this dataset for evaluation in the health economic model.

Subgroup Analysis

There is limited evidence in the literature to suggest that variability in parent presence may be associated with different effects of room design. Similarly, the ability of the intervention to improve major neonatal morbidity likely depends on the baseline risk for infants of different gestational age. For example, Örtenstrand (2014) found a decrease in severe bronchopulmonary dysplasia in their single rooms (Örtenstrand, 2014). The baseline risk of BPD varies from 3% in infants born at 31-32 weeks gestational age to 61% in infants born less than 25 weeks gestational age (Canadian Neonatal Follow-Up Network, 2019). In addition to the ability to properly represent potential correlated effects between parents and children, the proposed microsimulation model has the additional advantage of simplifying consideration of patient heterogeneity.

Sensitivity Analysis

Deterministic sensitivity analysis. As the base case required the assumption that relative effects estimated in preterm populations are the same in term admissions, a sensitivity analysis was conducted limiting results to preterm infants.

Probabilistic sensitivity analysis. An advantage of the Bayesian methods used throughout this dissertation are that they allow for probabilistic sensitivity analysis to be conducted in a way that maintains correlations between estimated parameters and makes minimal distributional assumptions. Samples from the posterior distributions of all estimated models (e.g. costs, length of stay, disability, morbidities, relative risks) were used for probabilistic sensitivity analysis. For variables that were derived from reports or through assumptions, probabilistic sensitivity analysis were conducted using logical distributions for each parameter: beta distributions for probabilities, log-normal distributions for relative risks, dirichlet distributions for disability states, and gamma distributions for costs and utilities (Appendix Table A 7). The PSA was run as a second order simulation for 10,000 simulations.

Stochastic Multi-criteria Acceptability Analysis. Recognizing that data limitations require a number of assumptions for costs and long-term outcomes, a sensitivity analysis was conducted based on neonatal outcomes only. This took the form of a stochastic multi-criteria acceptability analysis using the estimated probability of survival without disability, and an ordinal indicator for the degree of family centredness of a given design with old open bay as the low anchor and single family room as the upper bound (Tervonen et al., 2011). All outcomes are rescaled to be in the interval [0,1] where higher values are preferred. For survival without disability, the 95% interval hull method was used (Tervonen et al., 2011) which converts estimates to the interval [0,1] using a lower and upper bound defined by the lowest and highest

values of all 95% intervals. This avoids the problem of interpretation of importance weights being a swing across the entire interval of no survival to 100% survival which may imply a highly non-linear partial value function. Weights for survival and family centeredness were sampled with uncertainty with the constraint that any survival swing would be more important than any family centeredness swing. The primary output of the analysis is a vector of preference weights that maximizes utility of each intervention. These weights can therefore be thought of as indicating the implicit value of each outcome based on a choice of one intervention over another.

Validity and Rigour

Best practice guidelines were followed by systematic review and meta-analysis, model development, and model validation (Briggs et al., 2012; CADTH, 2015; CADTH, 2016).

Model development and structure. In any economic assessment, there are two sources of uncertainty that need to be considered before making conclusions regarding the utility and generalization of the model. Structural uncertainty refers to how modifications in time horizons and perspectives can influence results and was assessed using three decision maker perspectives (CADTH, 2016). Parameter uncertainty refers to uncertainty arising from the estimation of input parameters and impact of assumptions on results. Parameter uncertainty was investigated through probabilistic sensitivity analysis. The impacts of key parameters were explored through deterministic sensitivity analysis. Full external validation and calibration of the model against

real-world data is beyond the scope of this proposed project, but may be pursued in the future (CADTH, 2016).

Ethical Considerations

As a non-interventional study, this project was considered low risk. The use of individual participant data from local trials of single room care and the use of individual costing data from the IWK represent the most urgent ethical risks.

Individual participant data. Use of individual participant data from the Chez-NICU Home trial carried the risk of unintentional disclosure of health data through mishandling of electronic files. To mitigate this risk, all data transfer, storage, and analysis was conducted in accordance with IWK information technology guidelines. Files containing personal health information were shared through secure means and were only accessible by the primary investigator and the dissertation supervisor. An additional risk is that it may be possible to identify individuals in the case of rare events. As the study was concerned primarily with major morbidity and mortality for infants, it was not anticipated that this was a concern for infant outcomes since these events are relatively common in the population included.

Administrative data. For administrative data used for costing (IWK) or baseline neonatal complication rates (IWK Perinatal Follow-up), the study followed appropriate guidelines for storage and destruction of data. Eligible records were identified by providing criteria to the appropriate data steward, who conducted all required data linkages to protect privacy. Identifying information was protected through access being limited to data stewards. All investigators have completed training and orientation regarding the privacy of individuals and confidentiality of data.

Chapter 4 – Results

Results of analyses to inform the Baseline History Model

Hospital module.

Major morbidity and mortality base rates. A total of 736 live-born very preterm infants were included in the data-set from 2004-2018 (Figure 9). A total of 73 patients were excluded for not being offered intensive care and 72 patients were excluded for having a major congenital anomaly. The distribution of relevant baseline characteristics is summarized in Table 4. Missing data for individual variables and outcomes were less than 10% for all analyses

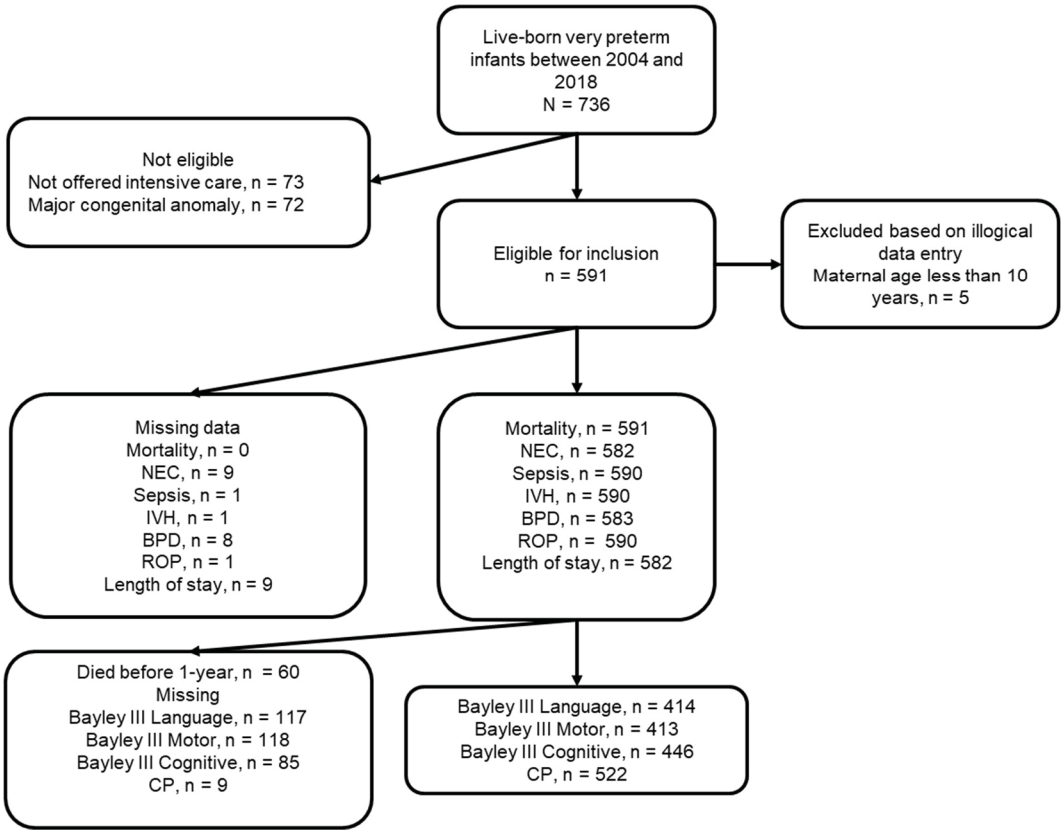


Figure 9. Flow diagram for IWK PNFU patients from 2004-2018. NEC = Necrotizing enterocolitis; IVH = intraventricular hemorrhage; BPD = bronchopulmonary dysplasia; ROP = retinopathy of prematurity; CP = cerebral palsy

Table 4. Characteristics of patients included in the baseline morbidity and mortality model

Variable	Value	Missing (n, %)
N	591	

Variable	Value	Missing (n, %)
Maternal/Family Characteristics		
Maternal age (mean, SD)	29.15 (5.61)	5 (0.8%)
Maternal number of cigarettes smoked per day (Median, IQR)	0 (0, 3)	33 (5.6%)
Maternal use of illicit substance	32 (5.4)	0
Infant characteristics		
Gestational age (mean (SD)	27.53 (2.09)	0
Birthweight (mean (SD))	1101.66 (333.49)	0
Twin	176 (29.8)	0
Male	311 (52.6)	0
Small for gestational age	12 (2.0)	0
Perinatal Characteristics		
Outborn	52 (8.8)	0
Cesarean section	336 (56.9)	0
Any antenatal steroids	515 (87.4)	2 (0.3%)
Ever received surfactant	434 (73.6)	1 (0.2%)
Preterm premature rupture of membranes	145 (24.5)	0
Mechanical ventilation (ever)	449 (76.0)	0
Morbidities/Mortality and Length of Stay		
BPD (Requiring any supplemental oxygen at 36 weeks PMA)	129 (22.1)	8 (1.4%)
NEC	26 (4.5)	9 (1.5%)
Sepsis	130 (22.0)	1 (0.2%)
PDA	168 (28.4)	0
IVH Grade III-IV	76 (12.9)	1 (0.2%)
ROP Stage ≥ 3	58 (9.8)	1 (0.2%)
Mortality	58 (9.8)	0
Length of stay	75.9 (60.2)	9 (1.52%)

Note. SD = standard deviation; IQR = inter-quartile range; SD = standard deviation; BPD = bronchopulmonary dysplasia; NEC = necrotizing enterocolitis; PDA = patent ductus arteriosus; IVH= intraventricular hemorrhage; ROP = retinopathy of prematurity

Bronchopulmonary Dysplasia. Models for bronchopulmonary dysplasia showed good calibration and discrimination. The estimated number of effective parameters that could be reliably estimated was 12. The model has a modest estimated R^2 of 0.15 (95% credible interval = 0.09 to 0.20; adjusted $R^2 = 0.09$) Estimates of calibration are both relatively precise within the range predicted by the data, as evidenced by orange lines representing posterior draws following a similar path, and accurate (Figure 10). The model begins to over-predict at the higher ends of probability, however the accompanying spread in uncertainty suggests that this is an area that is

not supported as strongly by data and would be predicted more rarely. The model is also quite discriminating as indicated with an area under the curve (AUC) of 0.75 (0.73 – 0.76), though this may suggest over-fitting which may not generalize. Estimates from partial dependence plots were generally consistent with expectations (Figure 11). The strongest predictor is gestational age.

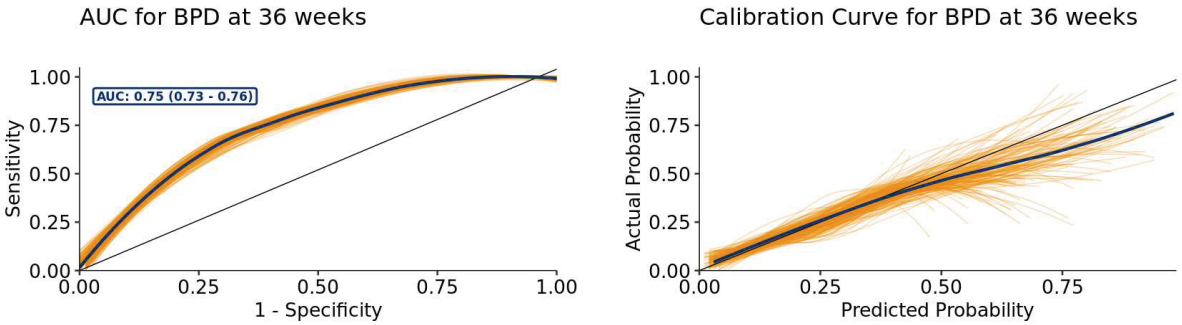


Figure 10. Calibration and discrimination of Bronchopulmonary dysplasia. Orange lines are individual draws from the model posterior and the main blue line shows the posterior mean. For calibration plots, lines below the diagonal suggest the model is over-estimating risk and lines above the diagonal suggest under-prediction. AUC = area under the curve; BPD = bronchopulmonary dysplasia.

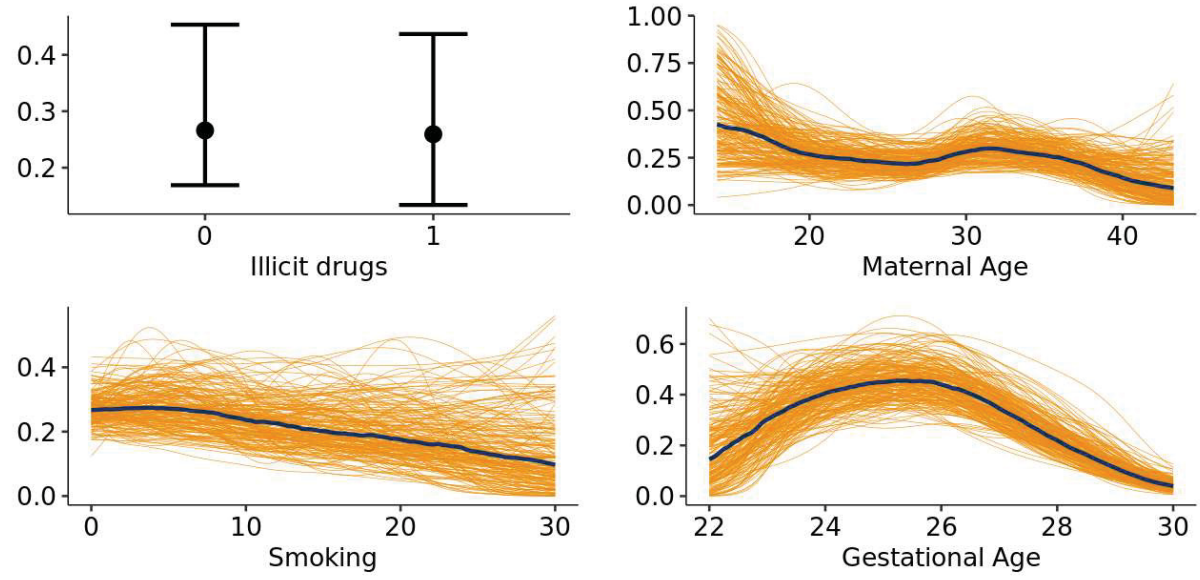


Figure 11. Partial dependence plots for all variables included to predict BPD. Predicted values are on the probability scale. Binary outcomes are coded as zero for no, and one for yes.

IVH III-IV. Predictions for serious IVH are well calibrated for lower values, but rapidly degrade as predicted probability raises above 50% (Figure 12). The estimated number of effective parameters that could be reliably estimated was 12. The model accounted for a small to modest amount of variability in the data with an R^2 of 0.12 (95% credible interval = 0.07 to 0.19; adjusted R2 = 0.07). Similar to findings from bronchopulmonary dysplasia, there is suggestion that this poor calibration is the result of lack of data for patients with the highest predicted risk. This is evidenced by the large increase in uncertainty in the calibration curve, which varies from extreme under prediction to extreme over prediction. As for BPD, the model discriminates well with an area under the curve of 0.78 (95% Credible interval = 0.76 to 0.78) (Figure 12). Partial dependence plots show a similar pattern to those in bronchopulmonary dysplasia, with several apparently weak predictors (Figure 13).

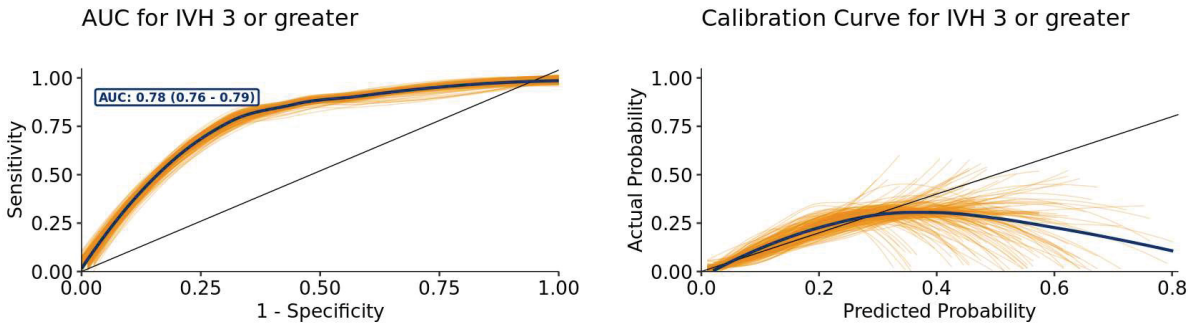


Figure 12. Calibration and discrimination of IVH III-IV. Orange lines are individual draws from the model posterior and the main blue line shows the posterior mean. For calibration plots, lines below the diagonal suggest the model is over-estimating risk and lines above the diagonal suggest under-prediction. AUC = area under the curve; IVH = intraventricular hemorrhage.

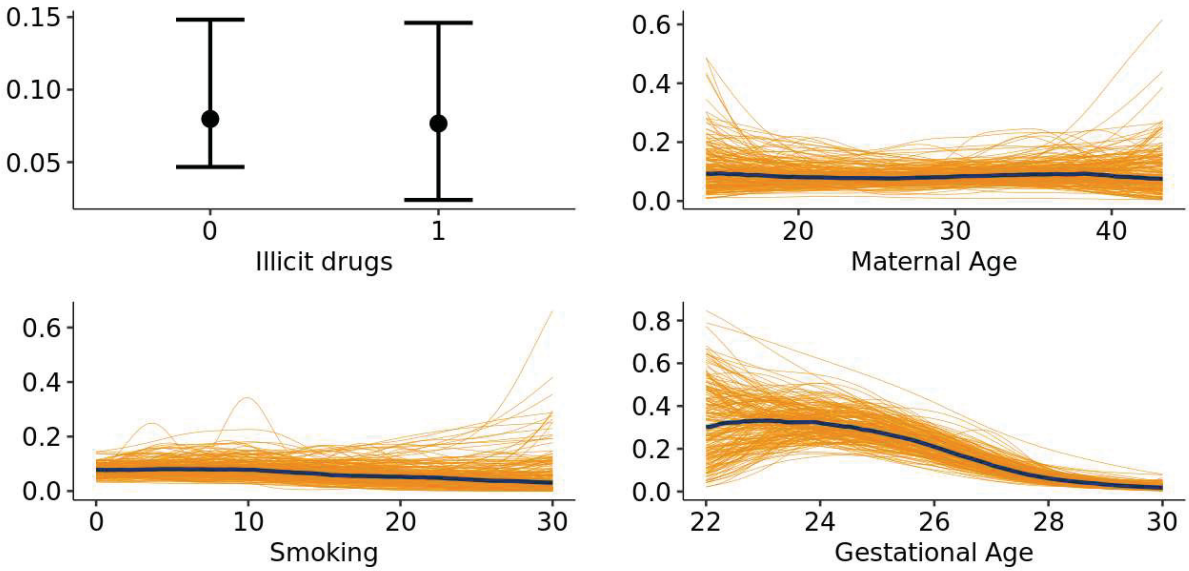


Figure 13. Partial dependence plots for all variables included to predict IVH. Predicted values are on the probability scale. Binary outcomes are coded as zero for no, and one for yes.

ROP Stage 3 or greater. Predictions for serious retinopathy of prematurity are well-calibrated for predictions less than 50% but show increased uncertainty across the entire band compared to other outcomes and over-estimate predictions above 60% (Figure 14). The number of parameters that could be reliably estimated was 12. Interestingly, this uncertainty does not appear to translate into difficulties with discrimination suggesting that low probabilities are predicted more often. Partial dependence plots showed similar trends to other outcomes

(

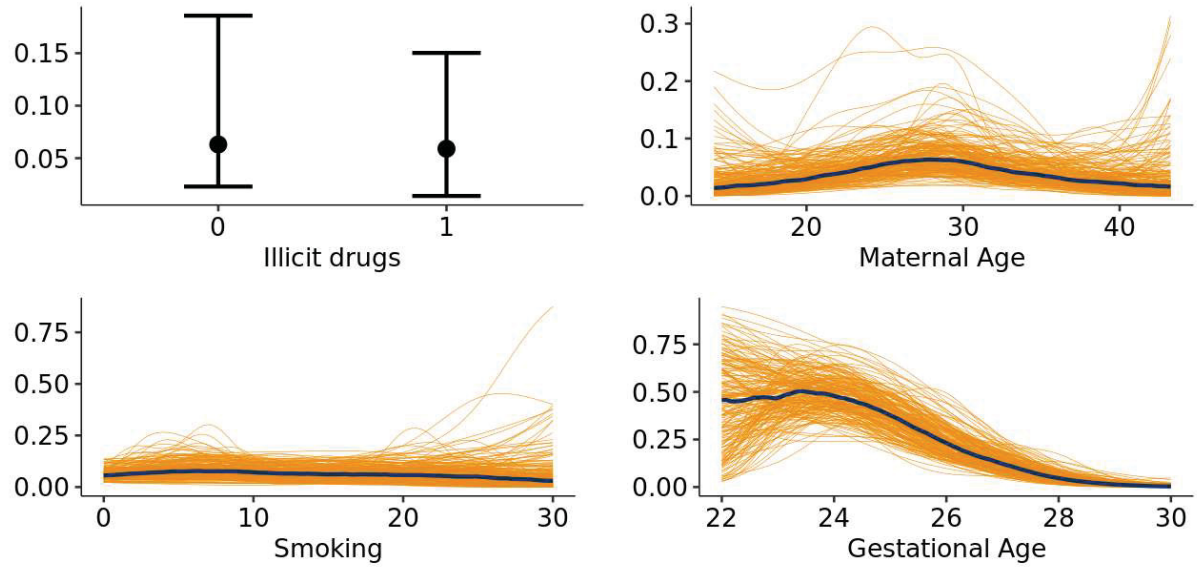


Figure 15). The model explains a modest amount of the variability in the outcome with an R^2 of 0.22 (95% credible interval = 0.13 to 0.30; adjusted R^2 = 0.14).

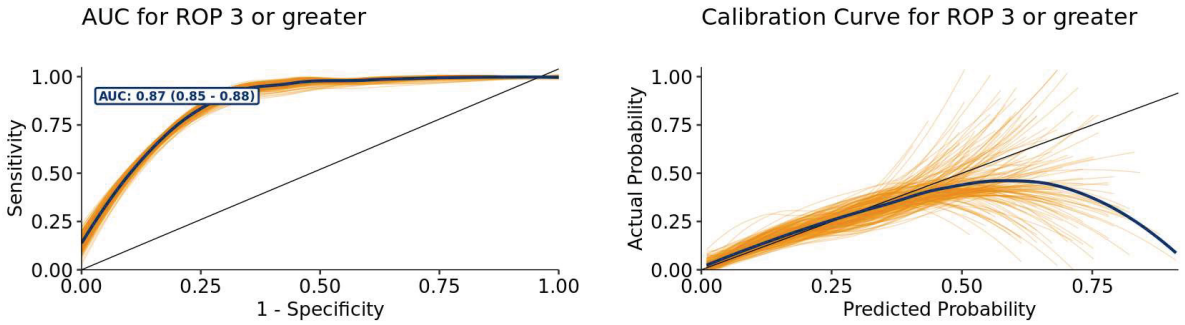


Figure 14. Calibration and discrimination of ROP stage 3 or greater. Orange lines are individual draws from the model posterior and the main blue line shows the posterior mean. For calibration plots, lines below the diagonal suggest the model is over-estimating risk and lines above the diagonal suggest under-prediction. AUC = area under the curve; ROP = retinopathy of prematurity.

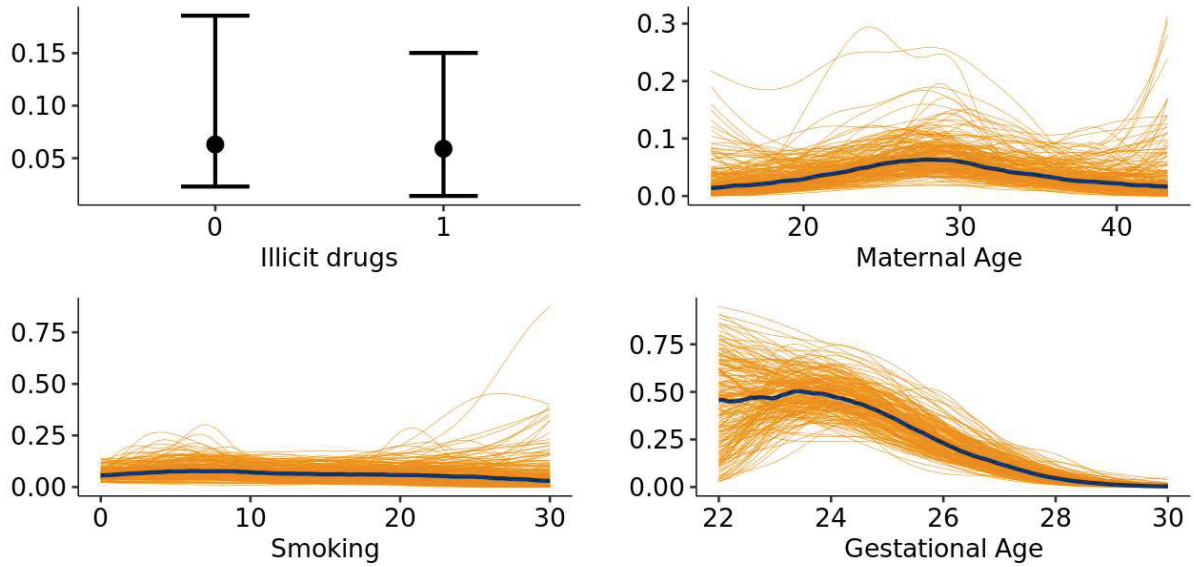


Figure 15. Partial dependence plots for all variables included to predict ROP. Predicted values are on the probit scale. More negative numbers indicate lower probability of ROP.

Sepsis. Sepsis shows evidence of good calibration with high precision in estimates of low predicted probability that grows increasingly imprecise and biased as predictions get more extreme (Figure 16). The number of effective parameters that could be reliably estimated was 12 and the model was poor at explaining variability in the outcome with an R2 of 0.08 (95% credible interval = 0.04 to 0.13; adjusted R2 = 0.04). Unlike for ROP where issues with calibration did not appear in discrimination plots, here the model also performs more poorly than

others in terms of AUC (0.67, 95% Credible interval = 0.65-0.69) (Figure 16). Partial dependence plots showed similar trends to other outcomes (Figure 17).

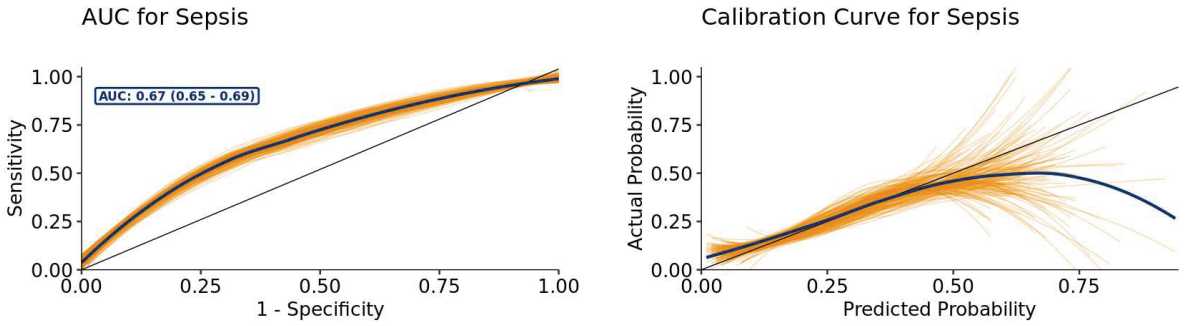


Figure 16. Calibration and discrimination of sepsis. Orange lines are individual draws from the model posterior and the main blue line shows the posterior mean. For calibration plots, lines below the diagonal suggest the model is over-estimating risk and lines above the diagonal suggest under-prediction. AUC = area under the curve.

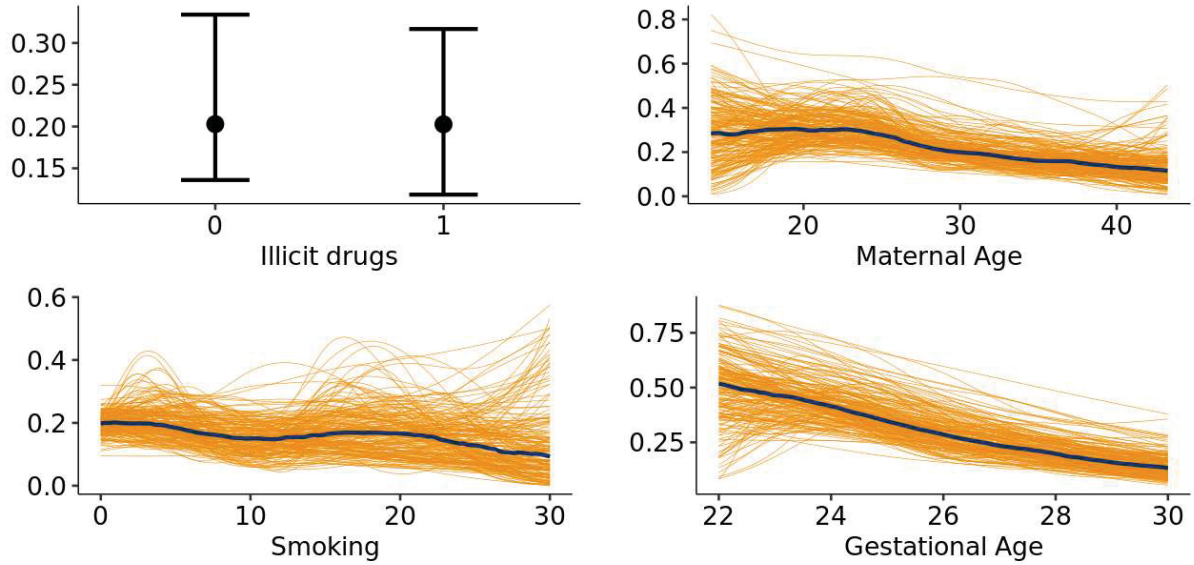


Figure 17. Partial dependence plots for all variables included to predict sepsis. Predicted values are on the probability scale. Binary outcomes are coded as zero for no, and one for yes.

NEC. Necrotizing enterocolitis was poorly calibrated for predictions approaching and beyond 50% with high uncertainty for all but the lowest predicted probabilities (Figure 18). The number of effective parameters that could be reliably be estimated was seven and the R^2 was low (0.08, 95% credible interval = 0.02 to 0.16; adjusted R^2 = 0). This can partly be explained by the

low prevalence in this cohort (4.5%) with the uncertainty likely being driven at least partially by the large number of proposed covariates (27 including splines) used to predict a small total number of events (26). This suggests that the high discrimination (AUC = 0.75, 95% credible interval = 0.68 – 0.78) is the result of predicting very low probabilities for all but a few patients (Figure 18). Partial dependence plots showed similar trends to other outcomes (Figure 19).

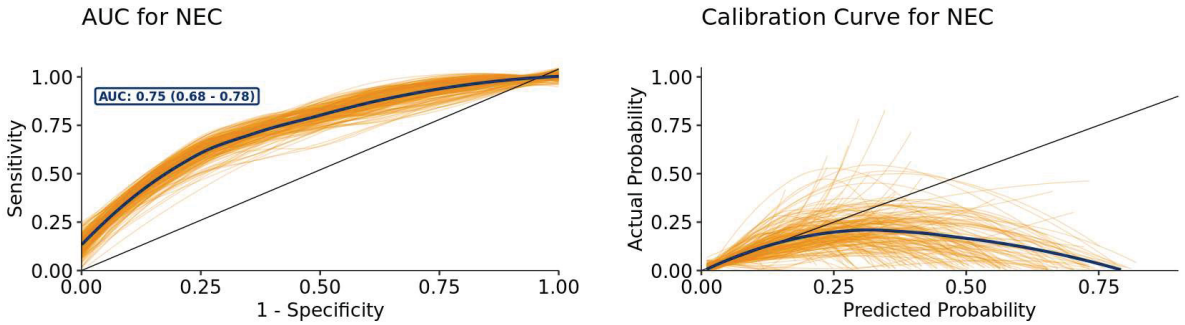


Figure 18. Calibration and discrimination of NEC. Orange lines are individual draws from the model posterior and the main blue line shows the posterior mean. For calibration plots, lines below the diagonal suggest the model is over-estimating risk and lines above the diagonal suggest under-prediction. AUC = area under the curve; NEC = necrotizing enterocolitis.

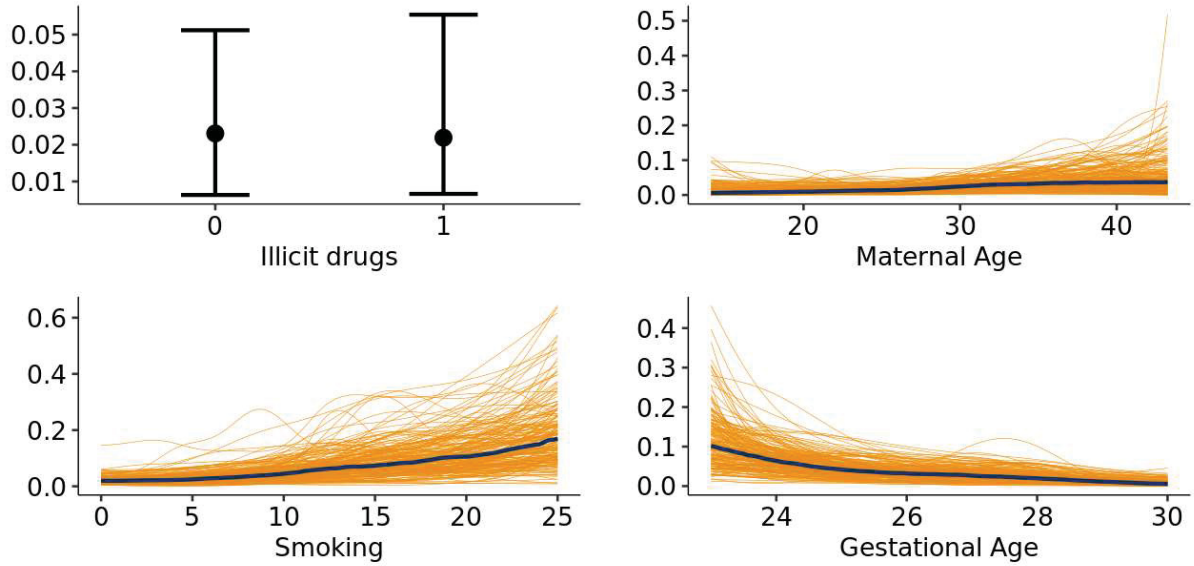


Figure 19. Partial dependence plots for all variables included to predict NEC. Predicted values are on the probability scale. Binary outcomes are coded as zero for no, and one for yes.

Mortality. Mortality showed the strongest calibration on average across the entire range of predictions in the model, although it begins to underestimate mortality beyond predicted probabilities of 65-70% (Figure 20). The number of effective parameters used in models was 12 and the R^2 of 0.26 (95% credible interval = 0.16 to 0.35; adjusted R^2 = 0.19) the highest of all outcomes. Despite this good prediction on average, there is substantial uncertainty in the calibration curve suggesting that lack of data in comparison to the number of model parameters can be causing meaningful over and under predictions beyond even 10%. Unsurprisingly, the most precise predictions are for very low probabilities, suggesting that the majority of the population is at low risk of mortality. This strength of this model is further reflected in the high AUC (0.86, 95% Credible interval = 0.82-0.87), although this raises concerns of the model fitting the data too well (Figure 20). Partial dependence plots showed similar trends to other outcomes (Figure 21). Of note, any receipt of antenatal corticosteroids shows strong effects on this outcome, consistent with prior meta-analysis of randomized trials (Roberts et al., 2017).

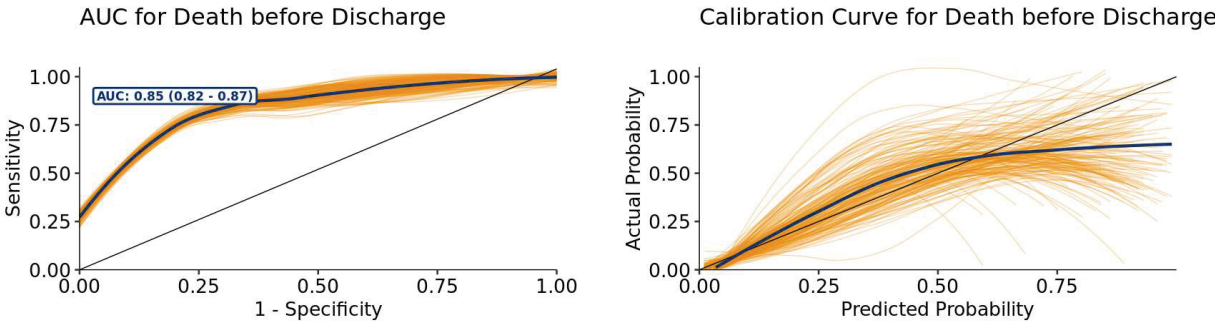


Figure 20. Calibration and discrimination of mortality. Orange lines are individual draws from the model posterior and the main blue line shows the posterior mean. For calibration plots, lines below the diagonal suggest the model is over-estimating risk and lines above the diagonal suggest under-prediction. AUC = area under the curve.

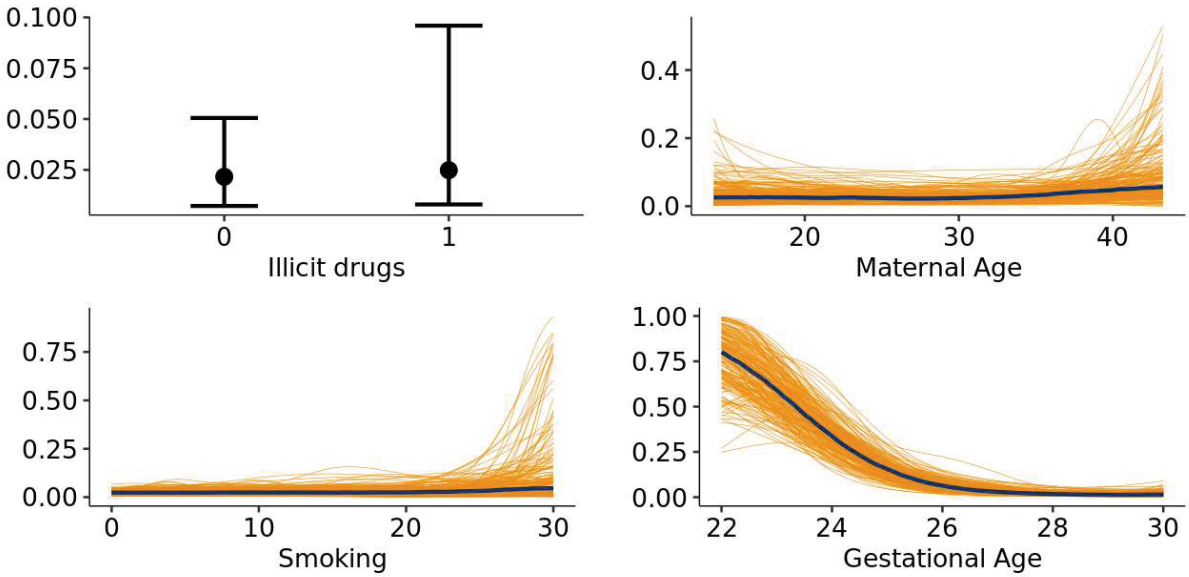


Figure 21. Partial dependence plots for all variables included to predict mortality. Predicted values are on the probability scale. Binary outcomes are coded as zero for no, and one for yes.

Length of stay base rate. Length of stay was analyzed using the same data as described for morbidity and mortality (Table 4). Estimates were consistent with expected direction, with gestational age showing a strong non-linear effect across the range of values. The effective number of parameters that could be estimated was 35. The model only has modest ability to account for variation in length of stay (Bayes $R^{2 =}$ 0.16, 95% Credible interval: 0.10 – 0.25, adjusted R^2 = 0.10). Extrapolation accuracy was determined by using the fitted model to predict length of stay in all infants greater than 30 weeks gestational age in the IWK performance analytics dataset used for costing. The test data included 1461 infants with an average gestational age of 37 completed weeks. The mean predicted length of stay was 8.2 days and the actual average length of stay was 8.67 days. Plotting the difference between actual and predicted length

of stay across gestational age shows a trend from a small overestimation in lower gestational ages to underestimation in term and older infants (Figure 23).

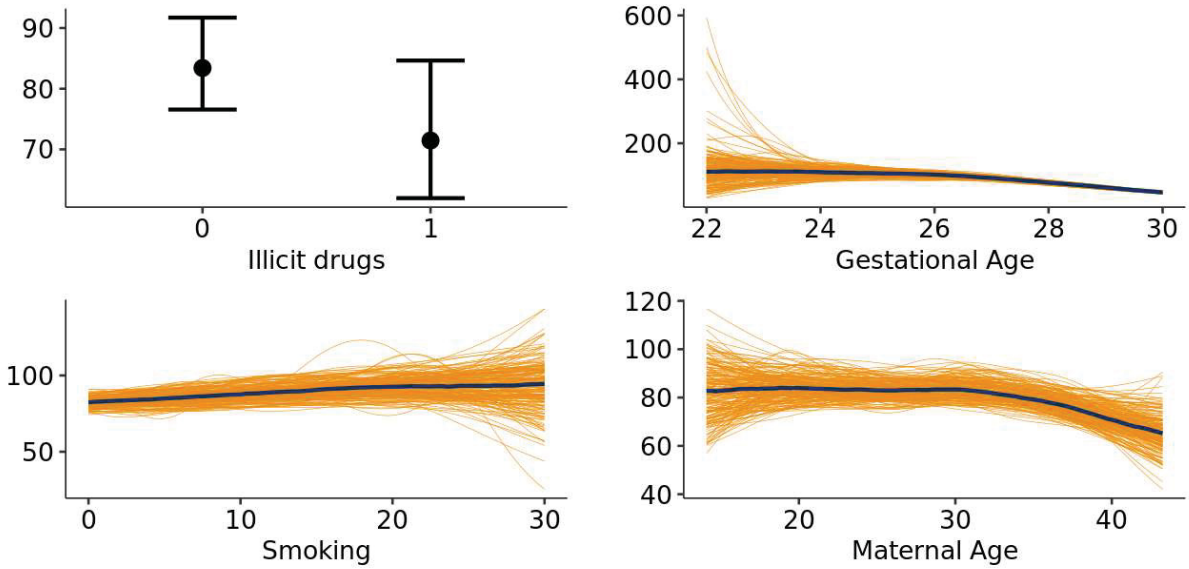


Figure 22. Partial dependence plots for all variables included to predict length of stay. Predicted values are on the length of stay scale. Binary outcomes are coded as zero for no, and one for yes.

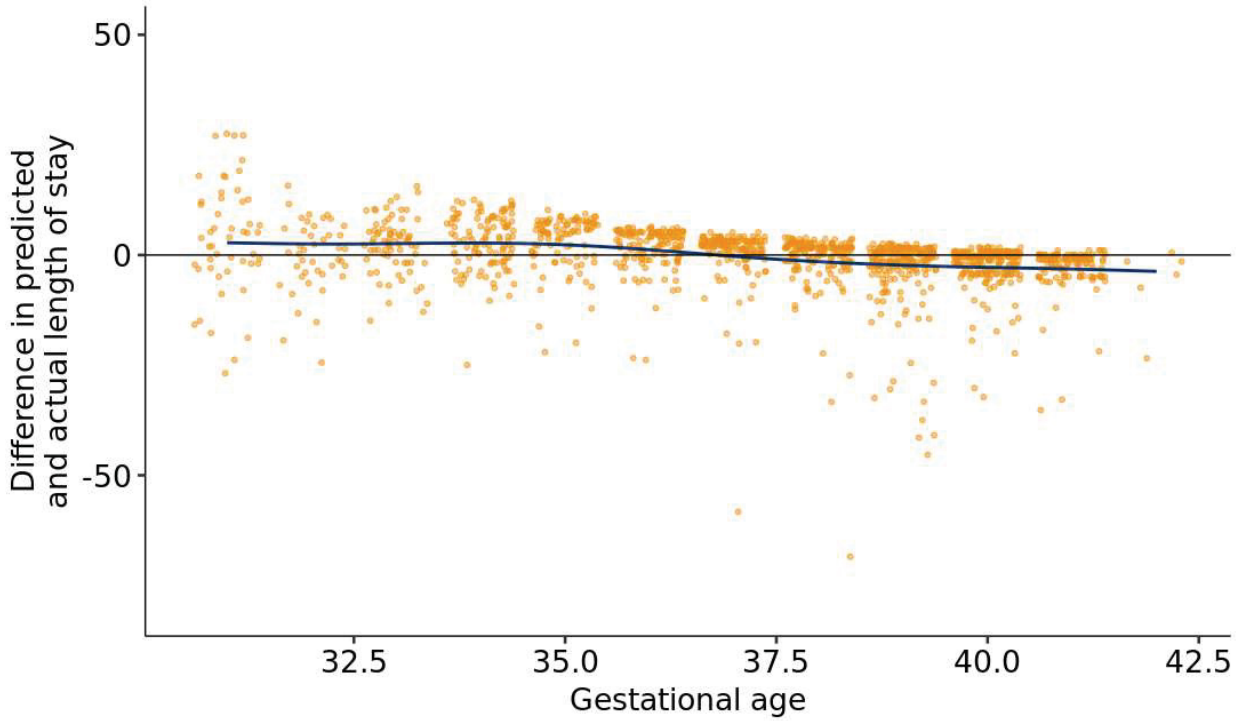


Figure 23. Difference in predicted and actual length of stay by gestational age in the IWK performance analytics database, using the model derived from perinatal follow-up. Axes were limited for visibility and exclude one point at -300.

Infant hospital cost. Costing data were available for 1751 infants admitted to the NICU between September 2013 and March 2016. Of these, twenty-one records were excluded for being the result of a subsequent admission, and 242 records were excluded based on the infant having

record of a major congenital anomaly. This left 1488 patients eligible for inclusion of which 1475 had complete data for all observations of all variables (Figure 24).

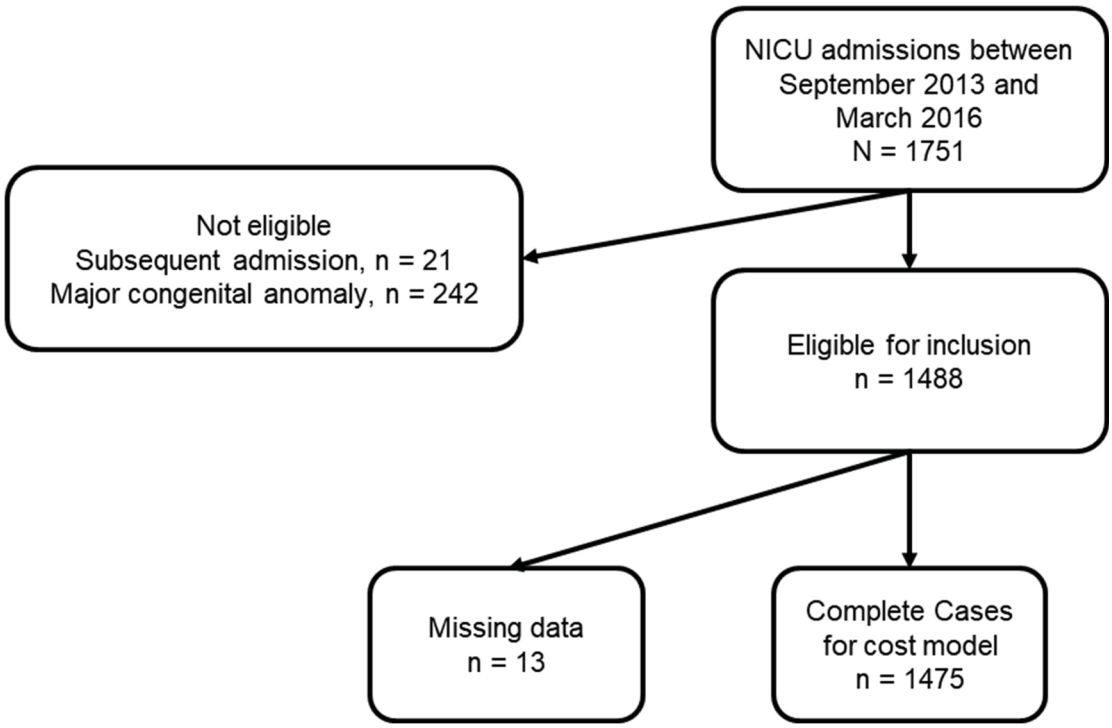


Figure 24. Flow diagram for patients in the cost model

Included infants were close to term on average, with nearly 60% of records being males. Given the high average gestational age, absolute rates of severe morbidities were low (all < 10%). When restricting to very preterm infants, however, rates are similar to those observed in the perinatal follow-up data, suggesting that the data are consistent. The mean length of stay was 13.38 days (range 1 to 345).

Variable	Value	Missing (n, %)
N	1488	
Infant characteristics		
Gestational age, mean (SD)	36.33 (3.81)	13 (0.87%)
Birthweight, mean (SD)	2795.25 (910.40)	15 (1.01%)
Male	888 (59.7%)	0
Morbidities/Mortality		
BPD (diagnosed in chart)	70 (4.7%)	0
NEC	13 (0.9%)	0
Sepsis	108 (7.3%)	0

Variable	Value	Missing (n, %)
IVH Grade III-IV	7 (0.5%)	0
ROP Stage ≥ 3	53 (3.6%)	0
Mortality	14 (0.9%)	0
Length of stay and costs		
Length of stay	13.38 (22.47)	0
Total cost	$25,900.31 (55188.19)	0

The model for cost was able to explain a relatively large portion of the total variation, likely at least partially owing to the underlying simple per-diem model that is used to assign labour costs to infants (Figure 25). The total effective parameters that could be reliable fit was 30 and the Bayesian R^2 for the model was 0.50 (95% Credible Interval = 0.46 to 0.54; adjusted R^2= 0.34). Several of the relationships are unexpected (i.e. having IVH or NEC results in lower per-diem cost) but can be likely explained by a combination of their effect increasing length of stay more than their associated increase in non-personnel costs in addition to the model fit when using a spline on gestational age instead of a linear term. Since the gamma model is non linear, mis-specification of a linear effect of gestational age when the true shape is smooth can result in biased confounder control (Lee, 2017). This hypothesis was tested by fitting a second model predicting total costs by morbidities alone excluding the spline on gestational age term, where the expected relationship is now seen (Figure 26). While this is problematic for interpretation of the analysis on its own, the impact on the economic model is not likely important since the correlation between morbidities and length of stay is captured.

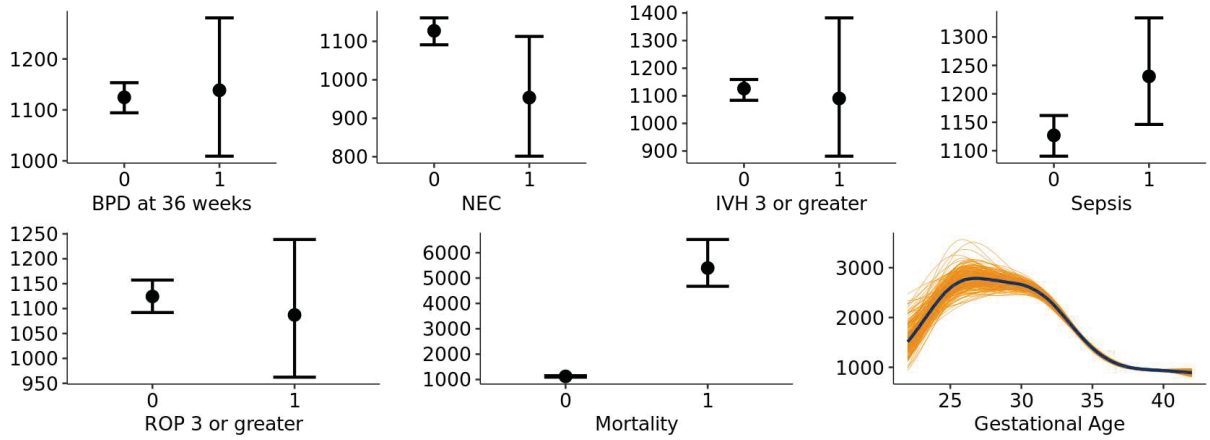


Figure 25. Partial dependence plot for per-diem costs. Binary outcomes are coded as zero for no, and one for yes.

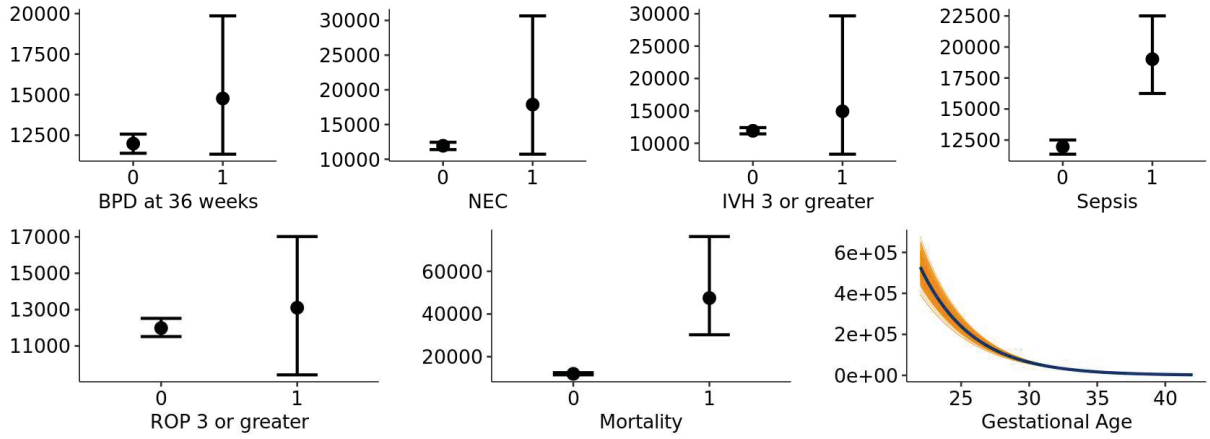


Figure 26. Sensitivity analysis for cost model using total costs as the outcome. Binary outcomes are coded as zero for no, and one for yes.

Predicting medical complexity. Data for the analysis of the prediction of medical complexity data came from recovered pseudo individual patient data from Nassel et al's (Nassel et al., 2019) Table 3 which provides raw counts for patients with an NDI or sNDI with or without medical complexity. A Bayesian logistic regression model was used to predict medical

complexity from NDI status which was entered as a monotonic variable. The effect of increasing disability on probability of medical complexity in Figure 27.

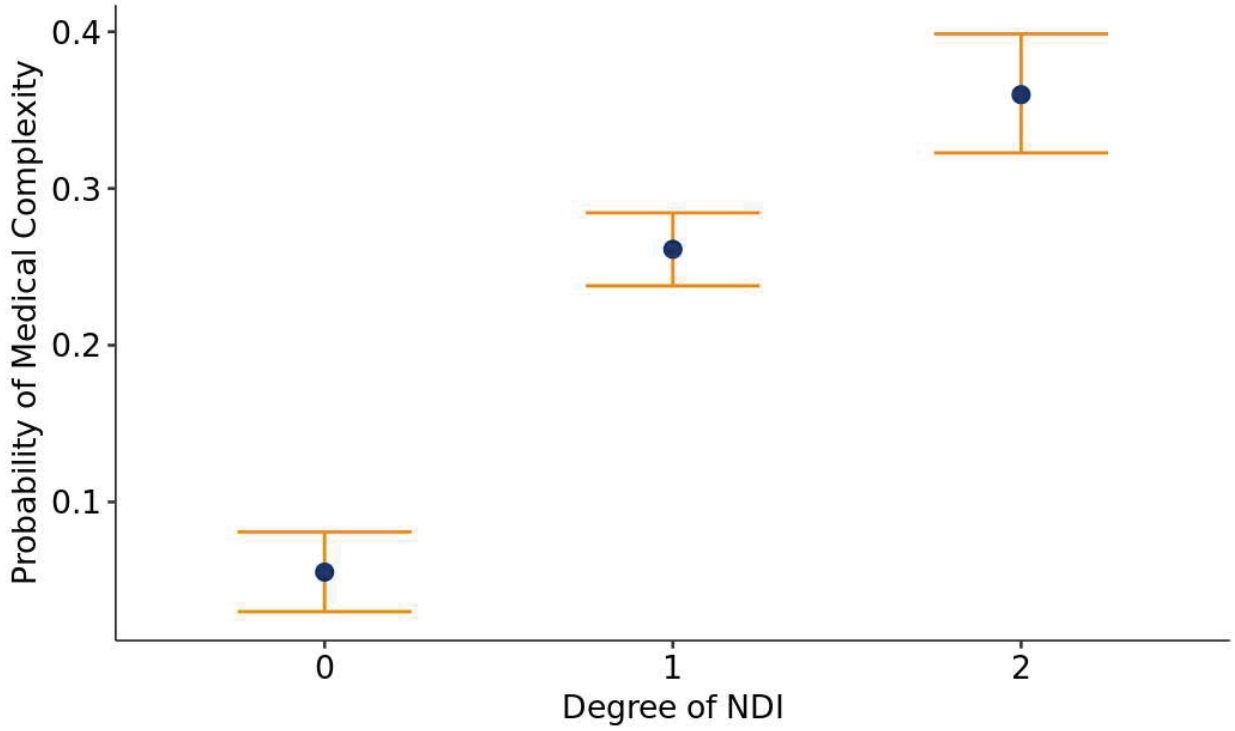


Figure 27. Probability of medical complexity given degree of neurodevelopmental impairment (NDI).

Parent presence, out of pocket costs, and quality of life.

Quality of life. The flow of parents through the quality of life survey data is outlined in Figure 28. Of an original 832 families assessed for eligibility 185 were found to meet all eligibility criteria of which 105 were recruited and 82 completed EQ-5D-5L questionnaires. Distribution of education and income of parents suggested recruited parents were potentially approximately representative of the median household income in Canada of $83,660 2017 dollars for couple families aged 25-34 years (Statistics Canada, 2017) and closely matched 2016 Statistics Canada education breakdowns for Nova Scotia which is also approximately consistent with Canadian education distribution (Statistics Canada, 2018). Occurrence of neonatal

morbidities was rare, as expected given the high average gestational age (Table 5). The number of parameters that could be reliably estimated was eight.

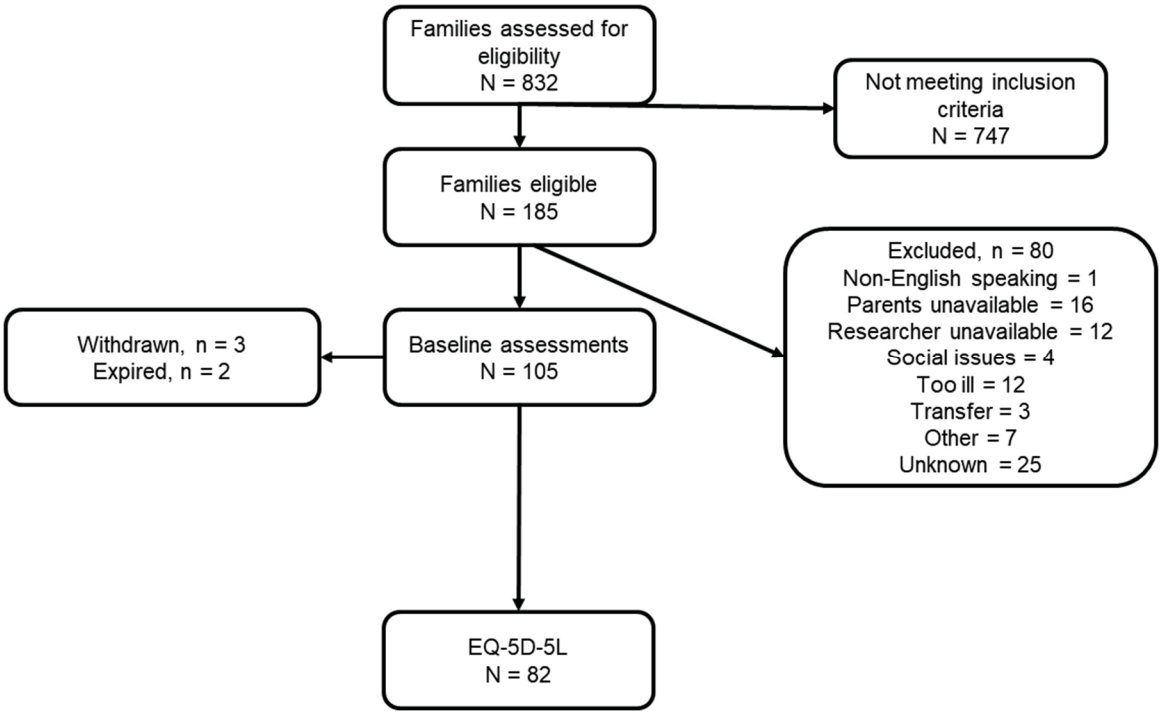


Figure 28. Flow chart for maternal utility

Table 5. Characteristics of families recruited for CHEZ-NICU

Variable	Value	Missing (n, %)
N	100	
Maternal Characteristics		
Maternal age (Mean, SD)		0
Under 20	6 (6%)	
20-29	41 (41%)	
30-39	50 (50%)	
40-49	3 (3%)	
History of depression (n, %)	27 (27%)	0
History of anxiety (n, %)	34 (34%)	0
Highest level of education (n, %)		2 (2%)
High school or less	27 (27.6%)	
College diploma	27 (27.6%)	
Some university	5 (5.1%)	
University graduate	25 (25.5%)	
Postgraduate	14 (14.3)	
Household income (n, %)		7 (7%)
Under $29,000	28 (30.1%)	

Variable	Value	Missing (n, %)
$30,000 - $49,999	10 (10.8%)	
$50,000 - $74,999	18 (19.4%)	
$75,000 - $119,999	18 (19.4%)	
$120,000 - $149,999	13 (14%)	
$150,000 or greater	6 (6.5%)	
Marital status (n, %)		2 (2%)
Single	4 (4%)	
In a relationship	17 (17%)	
Married	49 (49%)	
Living with a partner	28 (28%)	
Illicit substance use	0	
Multiple birth	15 (15%)	
Distance home	1271.1 (142.5)	0
Number of other children at home		0
0	46 (46%)	
1	33 (33%)	
2	13 (13%)	
3	4 (4%)	
4	4 (4%)	
Daily out of pocket costs (Mean, SD)	$25.2 (21.5)	0
Mothers	$25.9 (20.1)	0
Partners	$24.4 (23)	0
Infant characteristics		
Cesarean (n, %)	45 (45%)	0
Gestational age, mean (SD)	32.1 (3.3)	
Birthweight, mean (SD)	159.3 (604.4)	
Male		
Twin (n, %)	15 (15%)	0
Morbidities		
BPD (diagnosed in chart)	9 (9%)	0
NEC	2 (2%)	0
Sepsis	8 (8%)	0
IVH Grade III-IV	5 (5%)	0
ROP Stage ≥ 3	3 (3%)	0

Note. SD = standard deviation; BPD = bronchopulmonary dysplasia; NEC = necrotizing enterocolitis; IVH = intraventricular hemorrhage; ROP = retinopathy of prematurity

The results generally only supported a history of anxiety having potential predictive power to estimate disutility, and R^2 suggested generally poor predictive ability (0.08, 95% Credible interval = 0 to 0.25, adjusted $R^2 = 0$). This is likely the result of high average utility observed in

the sample combined with small standard deviation and the rare frequency of neonatal morbidities. Further, as the average utility increased over time and morbidities are generally diagnosed after recruitment it would be unlikely the data would support an effect of morbidities on utility (Table 6).

Table 6. Average maternal utility over time

Variable	Value	Missing (n, %)
N	100	
Baseline utility	0.80 (0.1)	2 (2%)
Day 14 utility	0.86 (0.1)	18 (18%)
Discharge utility	0.91 (0.1)	52 (52%)
Latest utility	0.9 (0.1)	18 (18%)

Table 7. Results of regression model for disutility

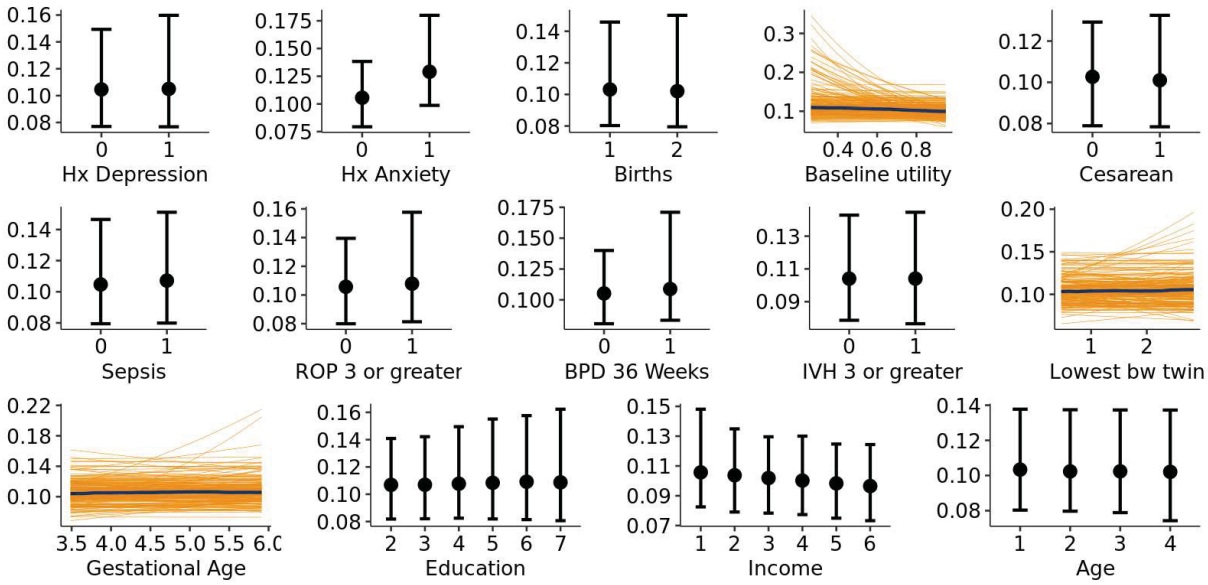


Figure 29. Partial dependence plots for disutility. Binary variables are coded as zero for no and one for yes. Hx = history; ROP = retinopathy of prematurity; BPD = bronchopulmonary dysplasia; IVH = intraventricular hemorrhage; bw = birthweight

Presence and out of pocket costs. Data used for the analysis were the same as those used for utility analysis (Table 5) except for one additional family who were excluded for inconsistent

data entry. This resulted a total of 99 families with data entry from 99 mothers and 96 partners (Figure 30).

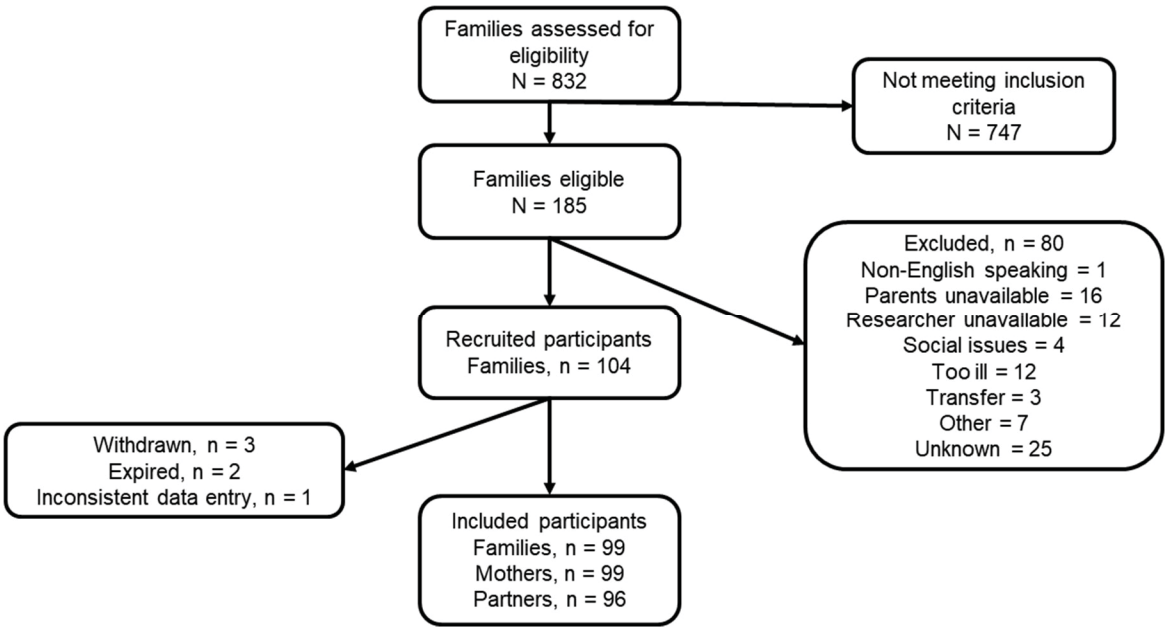


Figure 30. Flow diagram for parent presence and out of pocket costs

Results from the parent out of pocket cost model suggest that the included covariates offer strong predictive power of average daily costs, with an R^2 of 0.42 (95% Credible Interval 0.39 - 0.46, adjusted R^2 = 0.44). The number of effective parameters that could be reliably estimated was 30. Coefficients from the model are consistent with expectations, with the most important predictors being distance from hospital (resulting in more gas, parking, and less access to home kitchen) and the infants gestational age at birth. As splines are difficult to interpret as the effect of coefficients on the average predicted cost are also visualized using conditional effect plots (Figure 31).

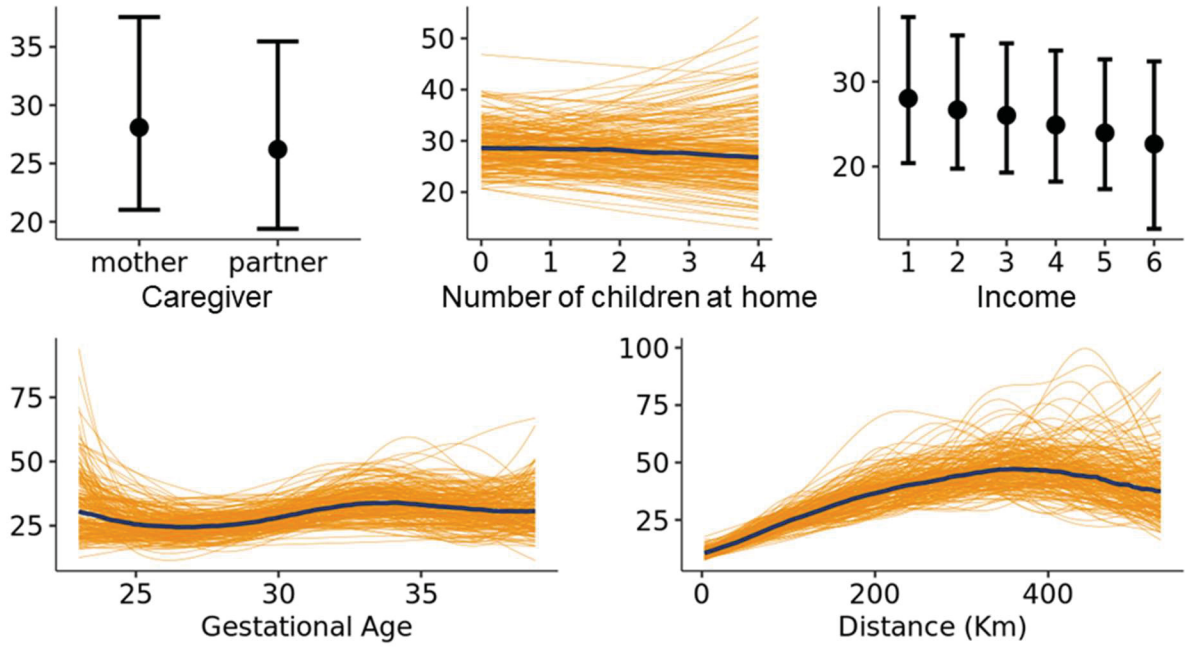


Figure 31. Conditional effect plots for gestational age and distance to home. Km = kilometer

Estimates suggested that 30 parameters could be reliably estimated for presence models, and model estimates showed signs of high overfitting despite use of regularizing priors as suggested by R2 of 0.99 (95% credible interval = 0.99 to 1, adjusted R2 = 0.99). As expected, increased income is associated with increased proportion of days present, and parents are present more often when infants are born at lower gestational ages. The trend observed with distance home likely suggests that those that live within the same neighborhood of the hospital visit more often than those that live within distances up to 400km at which point a large increase is seen, likely because these parents find accommodations nearby and stay in the city. The effect of the number of children at home, while uncertain, is also expected as increased carer responsibilities could be expected to reduce the proportion of days present.

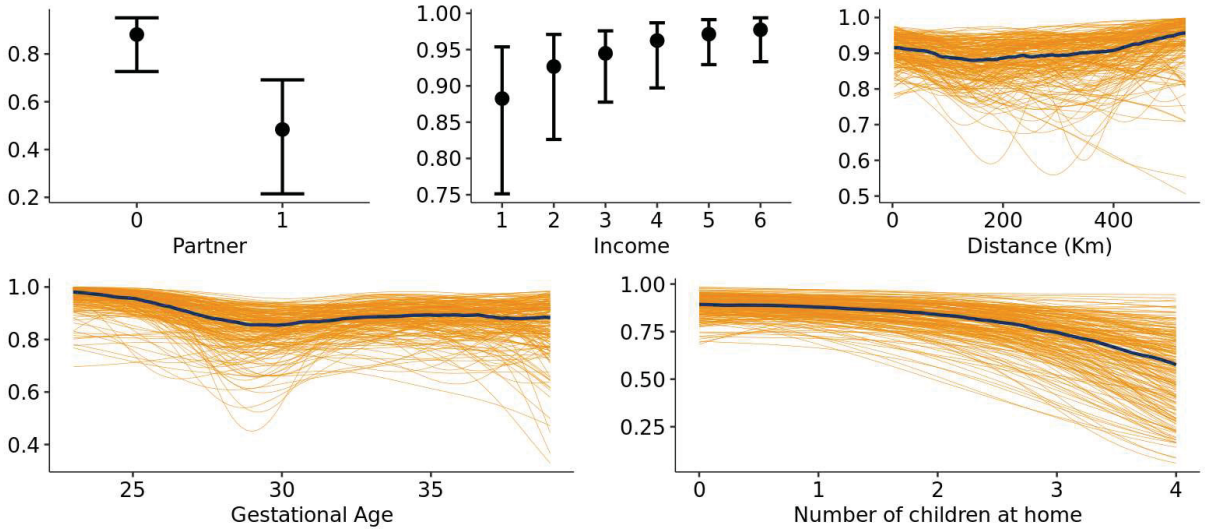


Figure 32. Conditional effects for monotonic effects and smooth variables included in the presence analysis. Dichotomous variables are coded as zero for no and one for yes. Higher income scores indicate higher patient reported income. Km = kilometer

Discharge to end of life.

Probability of disability given hospital course. The flow diagram for probability of disability is summarized in Figure 9. In total, 60 infants died before any assessments were conducted. Patient characteristics are summarized in Table 8.

Table 8. Summary of patients included in disability analyses

Variable	Value	Missing (n, %)
N	591	
Maternal/Family Characteristics		
Maternal age (Mean, SD)	29.2 (5.6)	3 (0.56%)
Maternal number of cigarettes smoked per day (Median, IQR)	0 (0, 3)	26 (4.9%)
Maternal use of illicit substance (n, %)	25 (4.7%)	0
Socio-economic Status		91 (17.14%)
Class I	53 (12%)	
Class II	142 (32.3)	
Class III	134 (30.4%)	
Class IV	76 (17.3%)	
Class V	35 (8.0%)	
Infant characteristics		
Gestational age (Mean , SD)	27.8 (1.9)	0
Birthweight (mean (SD))	1135.9 (319.4)	0
Twin	164 (30.9%)	0

Variable	Value	Missing (n, %)
Male	278 (52.3%)	0
Small for gestational age	11 (2.1%)	0
Perinatal Characteristics		
Outborn	44 (8.3%)	0
Caesarean section	307 (57.8%)	0
Any antenatal steroids	480 (90.6%)	1 (0.2%)
Ever received surfactant	386 (72.7%)	0
Preterm premature rupture of membranes	131 (24.7%)	0
Mechanical ventilation (ever)	397 (74.8%)	0
Morbidities		
BPD (Requiring any supplemental oxygen at 36 weeks PMA)	125 (23.8%)	7 (1.3%)
NEC	20 (3.8%)	0
Sepsis	115 (21.7%)	0
PDA	151 (28.4%)	0
IVH Grade III-IV	53 (10%)	0
ROP Stage ≥ 3	56 (10.5%)	0

Note. SD = standard deviation; BPD = bronchopulmonary dysplasia; NEC = necrotizing enterocolitis; PDA = patent ductus arteriosus; IVH = intraventricular hemorrhage; ROP = retinopathy of prematurity

Bayley III Language Score. Partial dependence plots are consistent with expected relationships (Figure 33), although the model showed a modest Bayesian R^2 of 0.22 (95% Credible interval = 0.16, 0.29).

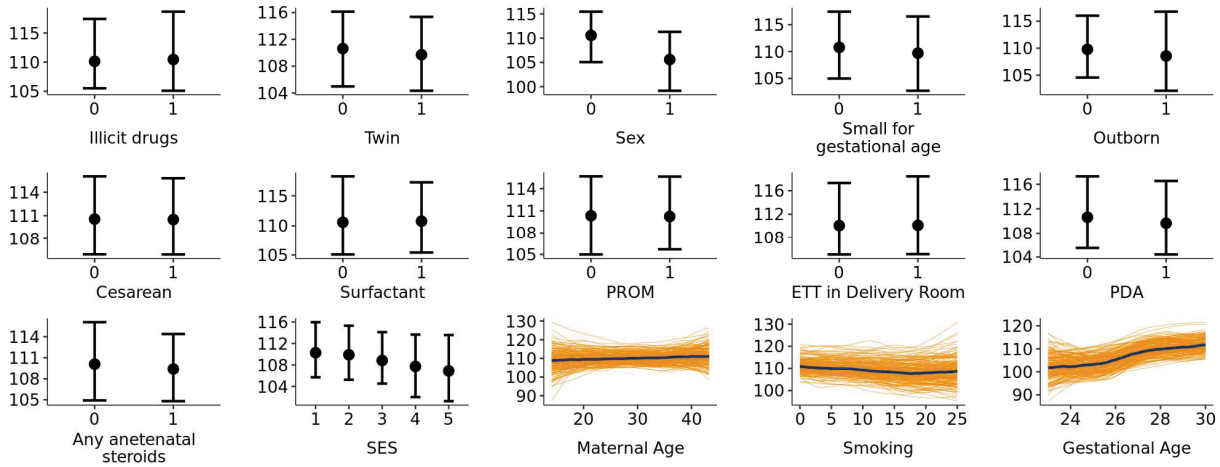


Figure 33. Partial dependence plot for Bayley III language subscale. Lower SES categories indicate higher SES. Binary variables are coded as zero for no and one for yes. PROM = premature rupture of membranes; ETT = endotracheal tube; PDA = patent ductus arteriosus; SES = socio-economic status.

Bayley III Motor Score. Partial dependence plots for the motor score showed generally similar relationship to the language scale but with effects of larger magnitude (Figure 34). Bayesian R2 was also similar to that observed for the language scale (0.2, 95% Credible interval = 0.18-0.30).

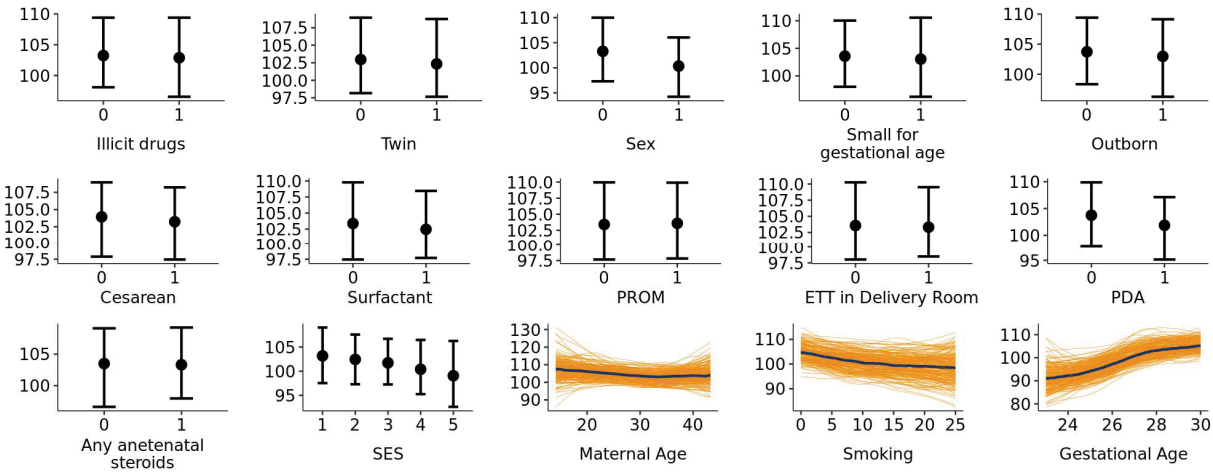


Figure 34. Partial dependence plot for Bayley III motor subscale. Lower SES categories indicate higher SES. Binary variables are coded as zero for no and one for yes. PROM = premature rupture of membranes; ETT = endotracheal tube; PDA = patent ductus arteriosus; SES = socio-economic status.

Bayley III Cognitive Score. Partial dependence plots for cognitive component scores were consistent with expected directions although the effect of receiving surfactant was reversed

compared to the other subscales. The Bayesian R2 was again modest (0.24, 95% Credible interval = 0.18-0.30).

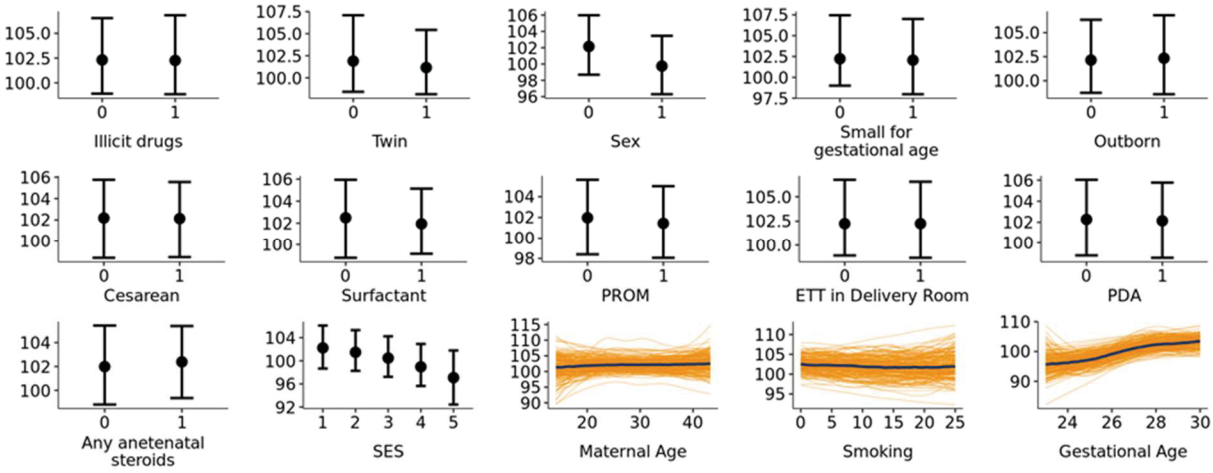


Figure 35. Partial dependence plot for Bayley III cognitive subscale. Predicted scores on y axis Lower SES categories indicate higher SES. Binary variables are coded as zero for no and one for yes. PROM = premature rupture of membranes; ETT = endotracheal tube; PDA = patent ductus arteriosus; SES = socio-economic status

Cerebral Palsy. Partial dependence plots from the model for cerebral palsy were consistent with expected direction of effects. Interestingly, whereas gestational age had a

precisely estimated and strong effect in other analyses, there was substantial uncertainty associated with its direct effect in these analyses (Figure 36).

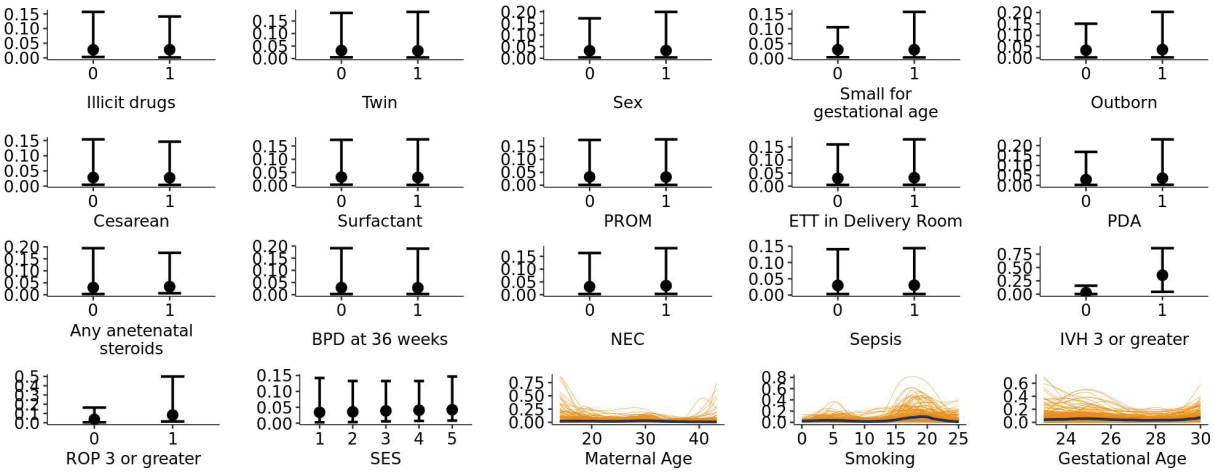


Figure 36. Partial dependence plot of variables for prediction of CP. Outcomes are on the probability scale. Higher values indicate lower SES, binary variables coded as zero for no and one for yes. PROM = premature rupture of membranes; ETT = endotracheal tube; PDA = patent ductus arteriosus; BPD = bronchopulmonary dysplasia; NEC = necrotizing enterocolitis; IVH = intraventricular hemorrhage; ROP = retinopathy of prematurity; SES = socio-economic status.

When evaluated in terms of calibration and discrimination, the model for cerebral palsy performs comparably to those used to predict morbidities (Figure 37). The effective number of parameters that could be reliably fit was estimated as being nine, and the R^2 of 0.29 (95% credible interval = 0.17 to 0.40, adjusted R2 = 0.20) was modest. As expected, the model has the most precision and accuracy in low probability predictions which was unsurprising given the high probability of no cerebral palsy observed in the data. Calibration is well maintained across the curve although there is evidence of increasing uncertainty in predictions beyond approximately 25%.

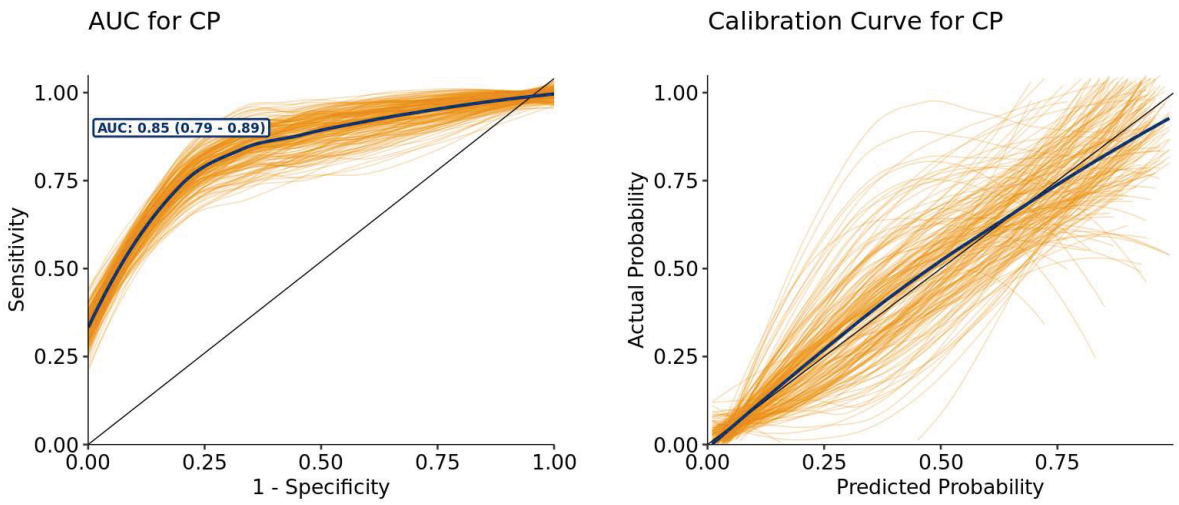


Figure 37. Calibration and discrimination of Cerebral Palsy. Orange lines are individual draws from the model posterior and the main blue line shows the posterior mean. For calibration plots, lines below the diagonal suggest the model is over-estimating risk and lines above the diagonal suggest the model is under-estimating risk.

Treatment Effect Parameters

IWK Single Family Room RCT. In total, 675 infants were admitted to the IWK NICU during the randomization period. Three of these infants had no information regarding their randomization probability and were excluded from the analysis, resulting in 672 patients in the final sample. Detailed information on the sample characteristics were not available, but the average gestational age was consistent with that seen in other datasets that recruited patients from the entire NICU. Unexpectedly, the data were not compatible with meaningful reductions in length of stay for patients in single family room and confidence intervals were consistent with meaningful increased length of stay. This could be partially driven by the smaller number of deaths in the single family room group, although there is too little data to exclude either meaningful benefit or meaningful harm (Table 9).

The single-family room design had a single patient with a length of stay greater than 200 days while the longest length of stay in open bay was 150 days. While it is possible that this

patient had a congenital anomaly resulting in an unavoidably protected length of stay, they were also one of the smallest babies in the dataset in terms of gestational age. A sensitivity analysis excluding this infant had no meaningful effect on the estimated treatment effect or confidence intervals. Based on the theoretical belief that infants of different gestational age may have larger or smaller benefits of treatment a post-hoc sensitivity analyses of this interaction were tested and provided no strong evidence of the presence of an interaction effect (0.00, 95% confidence interval = -0.03 to 0.02).

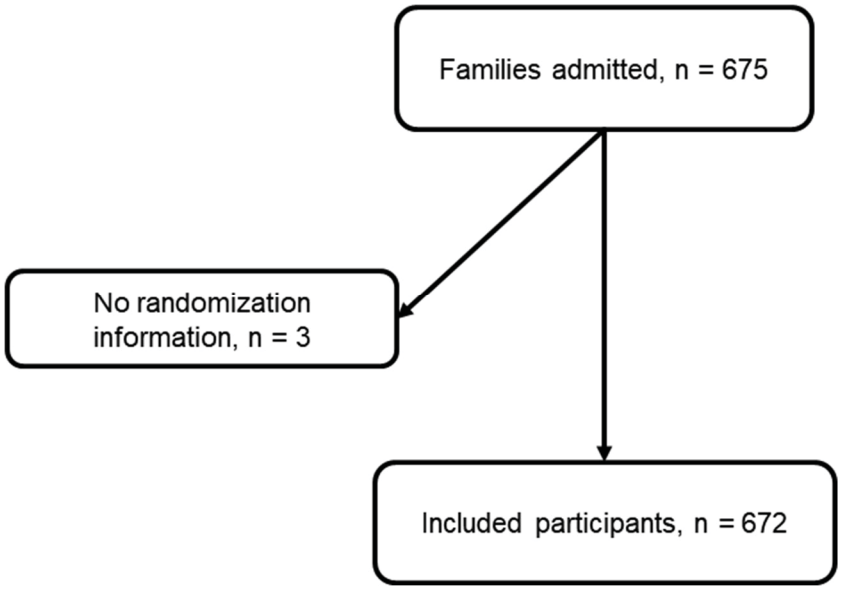


Figure 38. Flow diagram for patients in the IWK SFR RCT

Table 9. Baseline characteristics and outcomes from the IWK SFR RCT

Variable	Open Bay	Single Family Room	Adjusted Rate/Odds Ratio
N	358	317	-
Length of stay (Mean, SD)	13.77 (21.04)	16.22 (25.03)	1.21 (0.97 to 1.51)
Gestational age (Mean, SD)	36.01 (3.91)	35.84 (3.89)	
Death (n, %)	9 (2.5%)	4 (1.3%)	0.47 (0.13 to 1.68)

Note. SD = standard deviation

Chez-NICU single family room phase Of an original 77 families and 84 infants, five withdrawals left 72 families and 79 infants for analysis. For presence and costs, two additional records of expired patients were removed, and two patients did not have any record of a data entered.

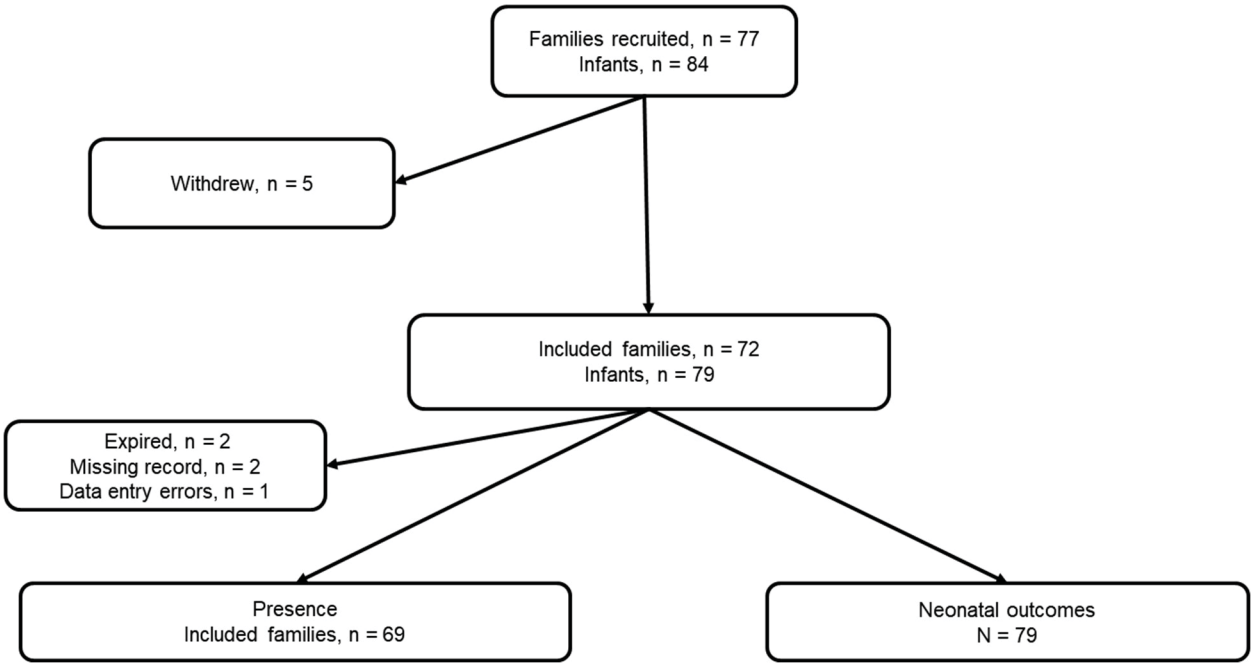


Figure 39. Flow of patients through CHEZ-NICU Phase 2B

Patient characteristics in the second phase of CHEZ-NICU were generally similar to those used in development of the baseline model. There were some numerical imbalances on important prognostic variables included in the baseline model: distance from hospital was skewed by an extreme value, and parents in the single family room units had numerically fewer children at home (Table 10). The analyses initially intended to include these variables as well as family income, however given the smaller than anticipated effect size and substantial missingness in the family income variable the final adjusted analyses were limited to distance to home, gestational age, and the number of other children at home.

Table 10. Patient characteristics from CHEZ-NICU phase 2

	Open Bay	Single Family Room
N	36	36
Maternal age (Mean, SD)	29.9 (5.7)	29.4 (6.1)
History of depression (n, %) (n = 24/33)	8 (33.3%)	6 (18.2%)
History of anxiety (n, %) (n = 24/33)	10 (40%)	7 (21.2%)
Highest level of education (n, %) (n = 35/36)		
High school diploma or less	7 (19.4%)	11 (30.6%)
College diploma	6 (17.1%)	9 (25.0%)
Any University	16 (45.7%)	13 (36%)
Postgraduate	6 (17.1%)	3 (8.3%)
Household income (n, %), (n = 30/32)		
Under $29,000	11 (36.7%)	10 (31.2%)
$30,000 - $49,999	4 (13.3%)	7 (21.9%)
$50,000 - $74,999	5 (16.7%)	9 (28.1%)
$75,000 - $119,999	6 (20%)	5 (15.6%)
$120,000 - $149,999	2 (6.7%)	0 (0.0%)
$150,000 or greater	2 (6.7%)	1 (3.1%)
Marital status (n, %) (n = 35/35)		
Single	2 (5.7%)	4 (11.4%)
In a relationship	5 (14.3%)	2 (5.7%)
Married	20 (57.1%)	17 (48.6%)
Living with a partner	8 (22.9%)	12 (34.3%)
Multiple birth	3 (8.3%)	4 (11.1%)
Distance home (Mean, SD)	95.7 (124.5)	242.8 (790.3)
Number of other children at home (n = 33/35)		
0	10 (30.3%)	19 (54.3%)
1	13 (39.4%)	11 (31.4%)
2	3 (9.1%)	3 (8.6%)
3	6 (18.8%)	2 (5.7%)
Infant characteristics		
Gestational age, mean (SD)	33.4 (3.3)	33.9 (3.2)
Birthweight, mean (SD)	2308.2 (991.3)	2346.5 (942.8)
Male	17 (43.6%)	15 (37.5%)

Note. SD = standard deviation

Crude measures of daily out of pocket costs and proportion of days in which parents were present were consistent with those estimated in the earlier study. Adjusted analyses did not suggest a strong effect room design on any outcome of interest, with 95% credible intervals including meaningful benefit and harm for both costs and IVH. Estimates from the parent

presence analysis are consistent with a small decrease and a small to large increase in in the days present (Table 11).

Table 11. Results from CHEZ-NICU phase 2 analysis

Outcome	Open Bay	Single Family Room	Unadjusted estimate	Adjusted estimate
Parents				
Daily out of pocket costs (Cost ratio)[a]	21.8 (16.9)	23.9 (23.4)	1.05 (0.79 to 1.39)	0.94 (0.68 to 1.28)
Proportion of days present	0.76	0.84	1.11 (0.97 to 1.24)	1.05 (0.95 to 1.19)
Infant morbidities (Odds ratio, unadjusted estimates only)				
BPD (diagnosed in chart)	0 (0%)	1 (2.5%)	NA	
NEC	0 (0%)	1 (2.7%)	NA	
Sepsis	0 (0%)	0 (0.0%)	NA	
IVH Grade III-IV	3 (7.7%)	2 (5.6%)	0.65 (0.10 to 4.12)	
ROP Stage ≥ 3	0 (0.0%)	0 (0.0%)	NA	

Note. Adjusted difference places a smooth spline gestational age, distance home, and number of other children at home. Odds ratios cannot be estimated when zero events

[a] Unweighted average, model estimates differ as a result of partial pooling of families

BPD = bronchopulmonary dysplasia; NEC = necrotizing enterocolitis; IVH = intraventricular hemorrhage; ROP = retinopathy of prematurity

Network meta-analysis.

Evidence identified. The literature search identified 1724 unique citations with an additional four references identified by supplemental searches of reference lists of included studies (Figure 40). Of these, 138 were assessed for eligibility at the full-text stage. In total 25 studies (32 citations) were included in the review. A detailed list of excluded studies with reasons is attached in Appendix Table A 2. Included studies all compared new designs against an old open bay, with comparators including new open bay (Goldmann et al., 1981; Larson et al., 1985; Von Dolinger de Brito et al., 2007), half walls (Altimier et al., 2005), private rooms (Monson et al., 2018; Pineda et al., 2012, 2014; Pineda, Durant, et al., 2017), single family rooms (Chabaud et al., 2012; Chen et al., 2017; Domanico et al., 2011; Erdeve et al., 2008; Erickson et al., 2011; Flacking et al., 2013; Hourigan et al., 2018; Lester et al., 2014, 2016; Ortenstrand et al., 2010; Puumala et al., 2020; Smithgall, 2011; Stevens et al., 2012, 2014; Stevens, Thompson, et al.,

2011; Tandberg, Flacking, et al., 2019; Tandberg, Frøslie, et al., 2019; Vohr et al., 2017; Wataker et al., 2012), and two studies assessing a combined open bay (organized in pods) and single family room (Feeley et al., 2020; Milford et al., 2008). Study sample sizes ranged from very small (N = 32) to very large (N = 9,995) with most studies being of moderate total size. There was substantial variability in the amount and detail of reported patient characteristics, but estimates from the NMA would require the assumption that relative risks are stable over a wide range of average gestational ages (26.3-38.7) and birthweights (937-3155g) in addition to any differences in prognostic variables that were established in analyses of the baseline natural history model. Several studies also exhibited potentially important differences in measured prognostic variables (Chabaud et al., 2012; Wataker et al., 2012). Further, since most studies investigated the effect of room design on outcomes of interest either not at all (e.g. solely

reported in summary characteristics) or as secondary outcomes, those analyses were rarely controlled for using any statistical methods at all.

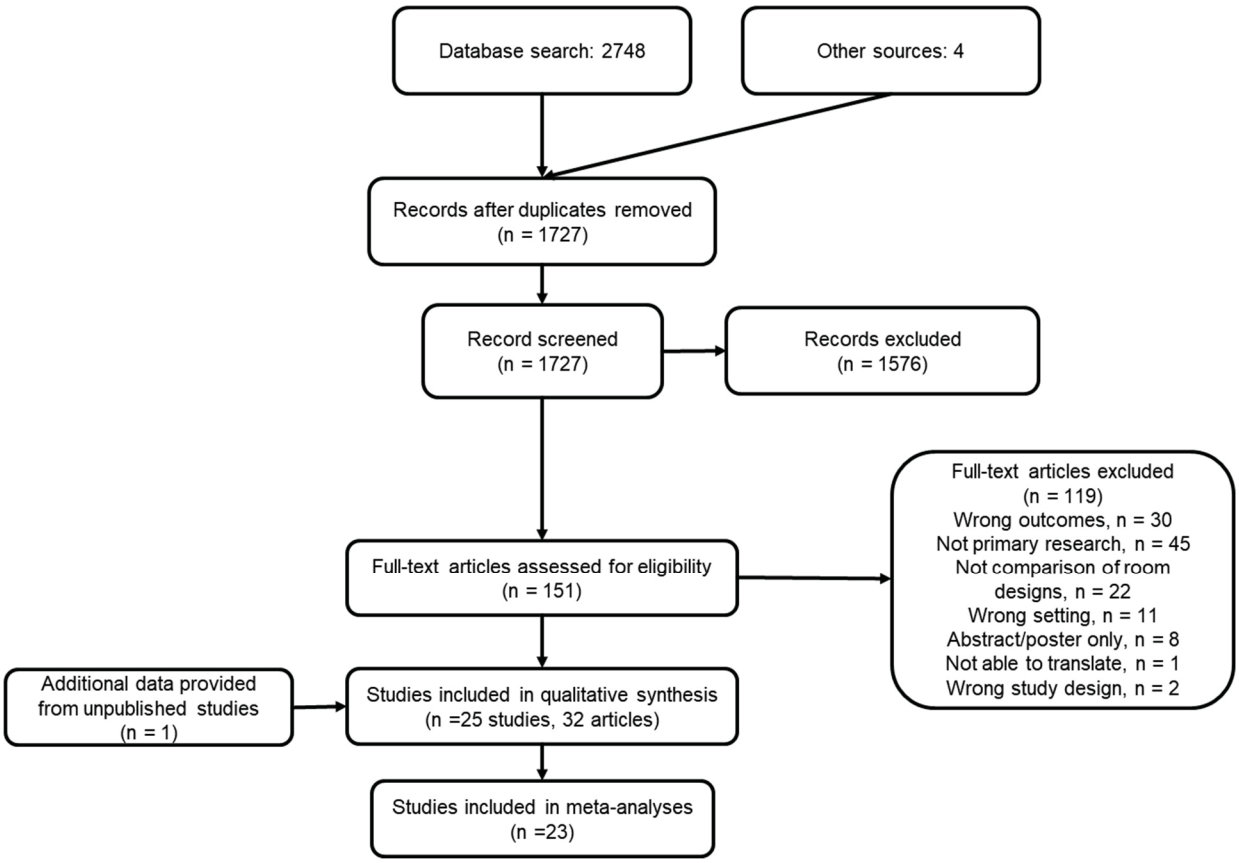


Figure 40. PRISMA flow-diagram of studies in systematic review

A simple approach for quickly assessing how homogenous a group of studies are in terms of prognostic variables (and effect modifiers that are also prognostic) is to look at how placebo response varies across the studies included in an analysis. Results of this exercise are described in Table 13 and show that response rates in open bay vary considerably over most outcomes. For examples, the studies by Pineda et al (2014) show higher than average rates of most outcomes, and the private room studies in general have higher baseline rates of sepsis. This is problematic for three reasons: It suggests that trials may differ on effect modifiers, which can cause both heterogeneity and intransitivity even in networks of randomized controlled trials; It is direct evidence that trials differ in terms of prognostic variables, which is additional risk of bias

considering the evidence base is primarily uncontrolled trials; and the networks are sufficiently sparse in most cases that adjustment is not possible.

Table 12. Characteristics of included studies

Study ID	Comparator	Population	Sample size		Gestational age		Birthweight	
			OPB	Comp	OPB	Comp	OPB	Comp
Larson 1985	New OPB	1	1443	502	-	-	-	-
Goldmann 1981	New OPB	2	642	542	-	-	2333	2314
Jones 2012	New OPB	3	149	152	-	-	-	-
Von Dolinger de Brito 2007	New OPB	4	795	666	-	-	-	-
Chen 2017	SFR	5	242	270	-	-	1827	1798
Altimier 2005	Half Walls	6	419	433	-	-	-	-
Pineda 2014	PR	7	65	71	26.4	26.8	937	954
Pineda 2012	PR	7	39	42	26.31	26.81	-	-
Monson 2018	PR	7	22	24	26.3	27	883	983
Pineda 2017	PR	8	-	-	-	-	-	-
Julian 2015	PR	9	884	912	-	-	-	-
Hourigan 2018	SFR	10	14	18	31.4	33.2	-	-
Erdeve 2008	SFR	11	29	31	30.4	30.8	1413	1452
Ortenstrad- 2010	SFR	12	183	183	-	-	2021	2097
Domanico 2011	SFR	13	133	107	34	34	-	-
Alessio 2011	SFR	14	28	27	-	-	-	-
Stevens 2012	SFR	15	-	-	-	-	-	-
Stevens 2014	SFR	15	269	306	36	35.3	2900	2700
Stevens 2011	SFR	15	1496	1672	35.57	35.25	2710	2660
Erickson 2011	SFR	15	31	42	28.1	26.7	1047	952
Smithgall 2011	SFR	16	52	52	33.4	32.86	2001	2013
Chabaud 2012	SFR	17	31	68	34.1	34.3	1892	2182
Wataker 2012	SFR	18	30	36	34.7	32.7	2600	1900
Lester 2014	SFR	19	151	252	28.2	28.3	1033	1050
Lester 2016	SFR	19	110	183	27.1	26.9	938	914
Vohr 2017	SFR	20	394	257	27.3	27.3	905	934
Flacking 2013	SFR	21	114	186	-	-	-	-
Tanberg 2019a	SFR	22	44	35	30.5	30.1	-	-
Tanberg 2019b	SFR	22	44	35	30.5	30.1	-	-
IWK 2020	SFR	23	358	317	36.01	35.84	-	-
Chez NICU Phase 2	SFR	23	36	236	33.4	33.9	23082	2346.5
Puumala 2020	SFR	24	964	4031	38.71	37.57	3155.02	2829.85
Milford 2008	Combined SFR/OPB	25	-	-	-	-	-	-
Feeley 2020	Combined SFR/OPB	26	70	80	30.27	30.04	1401.36	1422.08

Table 13. Outcomes in control arm of each.

Study	Comparator	IVH	ROP	NEC	Sepsis	Mortality	BPD	Bayley III			Length of Stay
								Cognitive	Language	Motor	
Larson 1985	New OPB	-	-	-	0.1	-	-	-	-	-	21
Goldmann 1981	New OPB	-	-	-	-	-	-	-	-	-	13.8
Jones 2012	New OPB	-	-	0.09	0.3	-	-	-	-	-	-
Von Dolinger de Brito 2007	New OPB	-	-	0	0.13	-	-	-	-	-	-
Chen 2017	SFR	-	-	-	0.08	-	-	-	-	-	27.53
Altimier 2005	Half Walls	0.11	0.14	-	-	-	-	-	-	-	-
Pineda 2014,	PR	-	0.14	0.08	0.35	0.16	0.53	86.8	91.9	86.2	92.69
Julian 2015	PR	-	-	-	0.04	-	-	-	-	-	-
Hourigan 2018	PR	-	-	-	0.86	-	-	-	-	-	33.5
Erdeve 2008	SFR	-	-	-	-	0.14	-	-	-	-	23
Ortenstrand- 2010	SFR	0.04	-	0.03	0.1	0.008	0.06	-	-	-	32.8
Domanico 2011	SFR	-	-	-	0.11	0.08	-	-	-	-	21.16
Alessio 2011	SFR	-	-	-	-	0.14	-	-	-	-	-
Stevens 2012, Stevens 2014, Stevens 2011, Erickson 2011	SFR	0.01	0.01	0.01	0.01	0.02	0.3	-	-	-	11.78
Smithgall 2011	SFR	0.03	-	-	-	-	-	-	-	-	18.23
Wataker 2012	SFR	-	-	-	0.23	-	-	-	-	-	11.3
Lester 2014,	SFR	0.03	0.06	0.03	0.2	0.19	0.21	87.1	84	88.3	79.8
Flacking 2013	SFR	-	-	-	-	-	-	-	-	-	28
Puumala 2020	SFR	-	-	-	0.04	0.01	-	-	-	-	7.79
Ta-berg 2019a, Ta-berg 2019b	SFR	-	-	-	0.04	-	0.06	-	-	-	45
IWK 2020	SFR	0.08	-	0.01	-	0.03	0.01	-	-	-	14.38
Millford 2008	SFR with Pods	-	-	-	-	-	-	-	-	-	14.3
Feeley 2020	SFR with Pods	0.06	0.03	-	-	-	-	-	-	-	54.2

Note. Dichotomous variables entered as proportions. Green shading indicators more favourable outcomes. IVH = intraventricular hemorrhage; ROP = retinopathy of prematurity; NEC = necrotizing enterocolitis; BPD = bronchopulmonary dysplasia; OPB = open bay; SFR = single-family room

Risk of Bias Assessment of Included Studies. Design types included one group pre-test post-test designs, post-test only non-equivalent designs in a single unit or multiple units, and randomized controlled trials. One group pre-test post-test design consist of the evaluation of a unit prior to a move to single family room and again after the switch. The advantage of this design is that the underlying patient population is more likely to be similar and there is a higher likelihood that the same staff and practices are used, although there was evidence that a move to a new design was also taken as an opportunity to introduce new practices such as developmental care or evidence bundles (Table 14). In post-test only non-equivalent designs units that have two design types can compare the new design against an old design. This has the advantage of controlling for policies and practices. A variant of this design that was used for some studies uses another unit as the control, but this is problematic since there is no guarantee that staff, practices, or patient populations are the same.

Table 14. Designs, control, and co-interventions of included studies

New Unit	Population	Included studies	Design	Controlled analysis	Relevant co-interventions
OPB	1	Larson 1985	One group pretest-post-test	X	Increased standardization of nursing tasks through policies/guidelines. Increased use of nursing staff for procedures re: blood draws.
	2	Goldmann 1981	One group pretest-post-test	X	No specific change in care but there was a gradual increase in the intensity and sophistication of infant monitoring, particularly in the increasing use of transducers
	3	Jones 2012	One group pretest-post-test	X	Practice described as unchanged, but hand hygiene compliance jumped from 23/29% to 77-80%

New Unit	Population	Included studies	Design	Controlled analysis	Relevant co-interventions
	4	Von Dolinger de Brito 2007	One group pretest-posttest	X	Percutaneously inserted central catheters introduced in the new unit
	5	Chen 2017	One group pretest-posttest		Three catheter-based bundles were implemented in the new NICU due to the policy for all intensive care units in our hospital; the central line bundle, the ventilator bundle, and the urinary tract bundle
Half-walls	6	Altimier 2005	One group pretest-posttest	X	Moving into the new space also marked the beginning of a large developmental care program. Also involved the purchase of the Phillips Wee Care system
PR	7	Pineda 2014, Pineda 2012, Monson 2018	Post-test only non-equivalent design	X	None described, but units were contemporaneous
	8	Pineda 2017	Post-test only non-equivalent design	X	
	9	Julian 2015	Post-test only non-equivalent	X	Not described
	10	Hourigan 2018	One group pretest-posttest	X	Not described
SFR	11	Erdeve 2008	Post-test only non-equivalent	X	Not described
	12	Ortenstrand 2010	Randomized controlled trial	N/A	Practices described as identical across wards
	13	Domanico 2011	Matched prospective one group pretest-posttest	X	None described
	14	Alessio 2011	One group pretest-posttest	X	Parents in the single family room were encouraged to use Kangaroo care and breastfeeding, this was deemed “nearly impossible” in the open bay
	15	Stevens 2012, Stevens 2014, Stevens 2011, Erickson 2011	One group pretest-posttest	Some analyses appropriate control, some none,	None described, move to more non-neos admitting patients developmental specialist worked with all infants in both designs

New Unit	Population	Included studies	Design	Controlled analysis	Relevant co-interventions
				some over-control	
	16	Smithgall 2011	Matched one group pretest-posttest	Matched on small group of potential prognostic factors	None described
	17	Chabaud 2012	One group pretest-posttest	X	Mother-baby unit is staffed by midwife and lay assistants, NN is staffed by Neos and nurses. Restricts only to population that is premature without any complications
	18	Wataker 2012	Post-test only non-equivalent	X	None described
	19	Lester 2014, Lester 2016, Vohr 2017	One group pretest-posttest	X	Lactation services provided in SFR (16 hours) but not OPB (0 hours), increased use of OT in SFR compared to OPB, 6% less chorio, 7% more antenatal steroids
	20	Flacking 2013	Prospective two group comparison	X	Part of a KC implementation project
	21	Puumala 2020	One group pretest-posttest	Stratified samples were used for NMA	Received baby friendly designation after transitioning but had already put measures in place. Otherwise provider and interprofessional support in the NICU did not significantly change
	22	Tandberg 2019a, Tandberg 2019b	Post-test only non-equivalent	X	Hospital is in Norway where parents have extensive publicly financed social security benefits, 90% of children in kindergarten. Units located in different hospitals.
	23	IWK 2020 Chez-NICU 2	Randomized Controlled Trial	X	Part of ongoing national QI projects but administered in both rooms
Combined SFR and Pod	24	Millford 2008	One group pretest-posttest	X	Part of an initiative to increase developmentally supportive, family centred care

New Unit	Population	Included studies	Design	Controlled analysis	Relevant co-interventions
	25	Feeley 2020	One group pretest-posttest	X	None described

Note. OPB = open bay; PR = private room, SFR = single-family room

Both included RCTs suffered from similar limitations related to the lack of the ability to blind staff, participants, or outcome assessors to intervention (Table 15). This could be expected to have the largest influence on outcomes that can be affected by investigators such as length of stay or suspected sepsis but would have less of an effect on more objective outcomes like mortality and confirmed sepsis. An additional potential limitation of the study by Ortenstrand et al. (2010) was that the units recruited were restricted to level 2 care, which would require that illness severity is not an effect modifying variable if results are to be applied to infants in level 3 units. The study conducted at the IWK used a minimisation-based randomization to maintain approximate balance in nursing requirements across units. This requires that those criteria are accounted for in the analysis in order to ensure proper estimation, which was via propensity score reweighting.

Table 15. Critical appraisal of RCTs

Domain	Ortenstrand 2010	IWK 2020/Chez-NICU 2
Randomization	✓	✓
Allocation concealment	✓	✓
Baseline similarity	✓	✓
Blinding of participants	X	X
Blinding of staff	X	X
Blinding out outcome assessors	X	X
Identical treatment other than intervention	✓	✓
Attrition bias	✓	✓
Intention to treat analysis	✓	✓
Outcomes measurement the same across groups	✓	✓
Reliability of outcome measurement	✓	✓
Appropriate statistical analysis	✓	✓
Was the trial design appropriate and any deviation from standard RCT accounted for	✓	✓

The majority of the included studies used a quasi-experimental design, primarily without a control group. For most studies, morbidities or length of stay were only intended as supportive endpoints and thus typically only crude unadjusted values were available (Table 16). This lack of adjustment is of particular concern in those studies that also showed evidence of imbalance between treatment groups at baseline, since results are potentially biased of these characteristics are either prognostic of effect modifying. Similarly, a number of studies also showed evidence of practice changes which accompanied the move to a new unit. This further threatens the validity of the evidence synthesis since it confounds the effect of room design with the effect of co-interventions. Further, while the JBI critical appraisal tools do not include a domain for selective outcome reporting there is some concern resulting from the relative sparsity with which a complete accounting of all outcomes is seen across studies. In the absence of published protocols or registries for any included studies, this raises potential concern for outcomes to be reported selectively.

Table 16. Critical appraisal of quasi-experimental studies

Publications included in study	Larson 1985	Goldmann 1981	Jones 2012	Von Dolinger de Brito 2007	Chen 2017	Altimier2005	Pineda 2012/2014 Monson 2018	Pineda 2017	Julian 2015	Hourigan 2018	Erdeve 2008	Domanico 2011	Alessio 2011	Stevens 2011/2012/2014	Erickson 2011	Smithgall 2011	Chabaud 2012	Wataker 2012	Lester 2014/2016 Vohr 2017	Flacking 2013	Tandberg 2019a/b	Puumala 2020	Milford 2008	Feeley 2020
Clear temporal cause and effect	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Participant similarity	✓	?	?	?	✓	?	X	?	✓	?	✓	✓	?	X	X	✓	X	X	✓	X	X	X	?	✓
Similar treatment of patients	X	X	?	X	X	✓	✓	✓	✓	?	?	?	X	X	✓	?	X	?	X	?	?	X	X	✓
Existence of control group	X	X	X	X	X	X	✓	✓	✓	X	✓[b]	X	X	X	X	X	✓	✓[b]	X	✓[b]	✓[b]	X	X	X
Multiple measurements pre and post-intervention	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	✓	X	X
Complete follow-up	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓[d]	✓	X	✓	?	✓
Similarity in measurement of outcomes	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Reliability of outcome measurement	✓	✓	✓	✓	✓	?	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Appropriate statistical analysis	X	X	X	X	X	X	X	X	X[a]	X	X	X	X	✓[c]	✓	✓	✓	X	✓[c]	X	X	✓	X	X

Studies excluded from analyses. As the study by Chabaud (2012) only recruited infants born 34 weeks and above and those cared for in a NICU without any pathology other than prematurity in addition to having care managed by two different professions, this study was excluded from any analyses. Pineda et al. (Pineda, Durant, et al., 2017) did not report data stratified by room and those data could not be provided within the timelines required for this study.

Handling multiple reports. Multiple studies were reported across a number of publications, with either complete or partial overlap of study populations. For the purposes of this review any overlap in populations was treated as if the data came from the same patients since it was not possible to determine the true amount of overlap and because there was already concern of large potentially biased trials leading to over-precision in pooled estimates. For studies arising from the Sanford Children's Hospital study (Erickson et al., 2011; Stevens et al., 2012; Stevens, Thompson, et al., 2011), data were primarily used from the Stevens et al. 2014 publication which included a propensity score matched cohort and a larger sample than earlier publications. The exception was NEC, which was extracted from the 2011 study since it was not published elsewhere. For the Lester and Vohr cohort (Lester et al., 2014, 2016; Vohr et al., 2017), morbidities were primarily extracted from Vohr, with Lester 2014 used for ROP, Bayley scores were extracted from Vohr. The St Louis cohort (Pineda et al., 2012, 2014) used data exclusively

from the 2014 report, and only the first of the Tandberg 2019 (Tandberg, Frøslie, et al., 2019) publications were used.

Down-weighting. Based on the results of the critical appraisals, a set of weights were derived that reflect a qualitative judgement of the contribution that a study should be given relative to a well-conducted RCT (Table 17).

Table 17. Weights used for down-weighting

Study	IVH	ROP	NEC	Sepsis	Mortality	BPD	Bayley Components			Length of Stay
							Cognitive	Language	Motor	
Larson 1985	-	-	-	0.4	-	-	-	-	-	0.4
Goldmann 1981	-	-	-	-	-	-	-	-	-	0.3
Jones 2012	-	-	0.5	0.5	-	-	-	-	-	-
Von Dolinger de Brito 2007[a]	-	-	0.5	0.4	-	-	-	-	-	-
Chen 2017	-	-	-	0.4	-	-	-	-	-	0.4
Altimier 2005	0.5	0.5	-	-	-	-	-	-	-	-
Pineda 2014[b]	-	0.7	0.7	0.7	0.7	0.7	0.8	0.8	0.8	0.7
Julian 2015	-	-	-	0.6	-	-	-	-	-	-
Hourigan 2018	-	-	-	0.4	-	-	-	-	-	0.4
Erdeve 2009	-	-	-	-	0.4	-	-	-	-	0.4
Ortenstrand 2010	1	-	1	1	1	1	-	-	-	-
Domanico 2011	-	-	-	0.6	0.6	-	-	-	-	0.6
Alessio 2011	-	-	-	-	0.6	-	-	-	-	-
Stevens 2011/14	0.7	0.7	0.7	0.9	0.7	0.9	-	-	-	0.9
Smithgall 2011	0.6	-	-	-	-	-	-	-	-	0.6
Wataker 2012	-	-	-	0.4	-	-	-	-	-	0.4
Lester/Vohr	0.5	0.5	0.5	0.6	0.5	0.5	0.6	0.6	0.6	0.5
Flacking 2013	-	-	-	-	-	-	-	-	-	0.4
Puumala 2020[d]		-	-	0.5	0.5	-	-	-	-	0.5
Tandberg 2019		-	-	0.4	-	0.4	-	-	-	0.4
IWK	0.8	0.8	0.8	0.8	1	0.8	-	-	-	1
Milford 2008	-	-	-	-	-	-	-	-	-	0.4
Feeley 2020	0.6	0.6	-	-	-	-	-	-	-	-

IVH = intraventricular hemorrhage; ROP = Retinopathy of prematurity; NEC = Necrotizing enterocolitis; BPD = Bronchopulmonary Dysplasia

[a] Vondolinger co-intervention more directly targeted at sepsis reduction

[b] Pineda weights differ because morbidities only known for cohort that was discharged

[c] Weights based on Vohr are lower given lack of equivalence between groups at baseline

[d] Puumala down-weighted because of results from time series as well as large sample

Morbidities and Mortality A multivariate analysis was conducted for morbidities and mortality.

Evidence networks. The evidence networks for these outcomes ranged from very sparse (e.g. ROP) to more robust (e.g. sepsis). No network of evidence included studies of every design type. Only single-family room design had at least one study for every outcome, with private rooms contributing evidence for all outcomes except IVH.

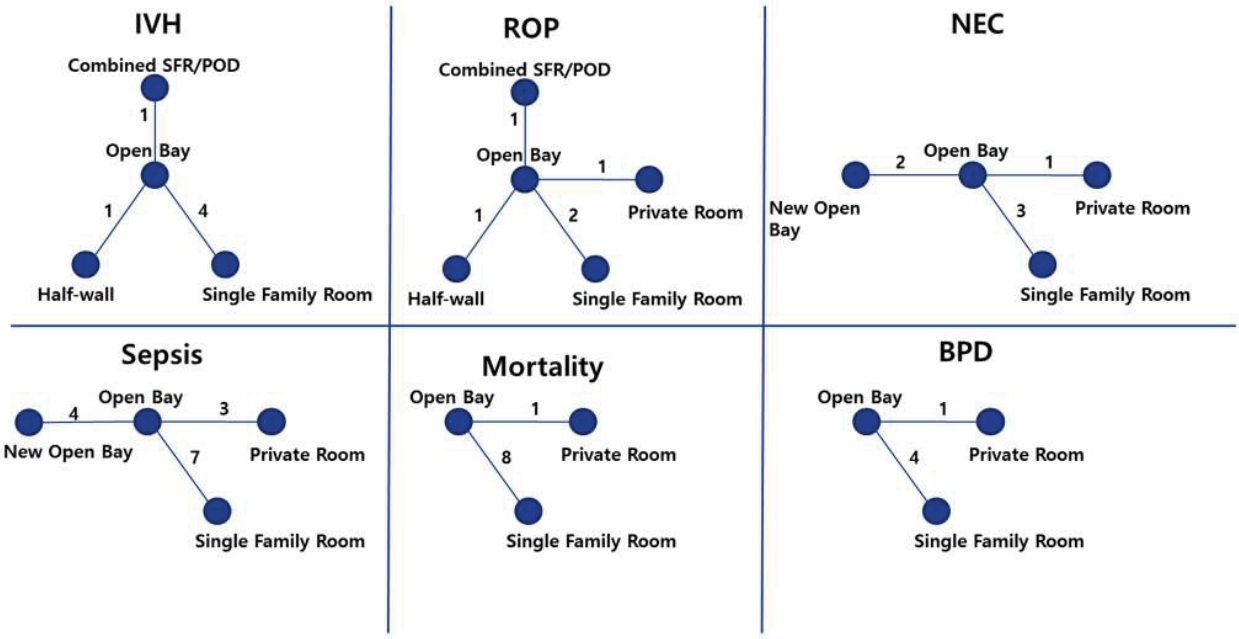


Figure 41. Networks of evidence for mortality and morbidities. IVH = intraventricular hemorrhage; ROP = retinopathy of prematurity; NEC = necrotizing enterocolitis; BPD = bronchopulmonary dysplasia.

Assessment of heterogeneity in pairwise comparisons and the NMA. Results of pairwise meta-analyses are shown in Appendix Figure A. 1 to Appendix Figure A. 8. For IVH, confidence intervals are wide suggesting that the underlying data are consistent with both meaningful benefit and harm for most comparisons. Point estimates between randomized and non-randomized trials do not show a consistent pattern of disagreement for single family room designs but results from half-wall designs are implausibly large based on the underlying theoretical framework of the

analysis. Results for ROP are similar in terms of most design types showing estimates that include both meaningful benefit and harm, and the same study of half-wall designs showing an implausibly large point estimate (OR = 0.52, 95% CI = 0.25 to 1.17) compared to other designs, although there is no strong evidence that any design is superior to another. Estimates from trials including NEC as an outcome also suggest modest effects of room design and there is agreement between randomized and non-randomized evidence. The exception in terms of point estimates comes from the small Chez-NICU trial conducted at the IWK, although this was the result of a single event in the treatment arm, which is reflected by a confidence interval that ranges from implausibly protective to implausibly harmful. The estimate for sepsis is based on the largest number of studies and shows a similar pattern to other outcomes. Prior to down-weighting evidence there was small to modest levels of heterogeneity as evidenced by I^2 ranging from 0% to 58%. After down-weighting trials these values both reduce to 0%. There is no strong evidence of disagreement between randomized and non-randomized trials and effect sizes are generally modest except for those estimated with considerable imprecision. Mortality and BPD show the most variability in estimates between trials. In the case of mortality, this is primarily explained by the imprecision with which odds ratios are estimated and so despite point estimates ranging from 0.47 to 1.20 values of I^2 are 0% in both the raw data and down-weighted analyses. Heterogeneity in pairwise analyses is most pronounced for BPD, where individual studies found effect estimates ranging from meaningful benefit (0.20, 0.26) to meaningful harm (1.58, 3.08). The I^2 value for single family room improves from 59% to 53% after down-weighting.

Crude results from Puumala et al (2020) are consistent when subgroup effects by gestational categories. Over-all results are primarily dominated by results in moderately preterm and term infants where the effects of design are of the opposite sign to those in the extremely

preterm and very preterm groups. Analyses from the authors of this study suggests that these subgroup effects are explained by an over-all trend over time and not by room design. A subgroup effect by gestational age is further contradicted by available subgroup analyses in the Ortenstrand et al (2010) and IWK RCTs, where the ratio of means for length of stay was stable across gestational age groups. The lack of sub-group effects is further corroborated by the general consistency of trial results across outcomes despite wide ranges of average gestational ages. Given these findings and size of this study, it was excluded from the base case and a sensitivity analyses was conducted including it. This led to large increases in the observed heterogeneity in sepsis, mortality and length of stay outcomes.

In summary, down-weighting generally improved the observed between study heterogeneity. There was no clear pattern of disagreement between randomized and non-randomized studies for outcomes in studies assessing single family room, although any conclusions are limited by the existence of only two relatively small randomized trials.

Findings across different methods. Conclusions from multivariate NMAs in terms of robustness of estimates was generally consistent across a range of scenarios including the base case (Figure 42), sensitivities with less informative priors (Appendix Figure A. 9), and less informative priors combined with the assumption of no correlation between outcomes or in terms of the assumption of constant potency Appendix Figure A. 10). In terms of the effect on point estimates, the most important impacts appear to be in adding constant potency effect and then in adding constant potency in combination with informative priors. As can be expected from the pairwise meta-analyses, point estimates vary the most in the univariate case where for example half wall designs show an OR ranging from 0.25 for IVH to 1.03 for mortality. This seems unreliable given the strong correlation observed between IVH and mortality in the PNFU data,

and thus the constant potency assumption appears to result in more consistent estimates across outcomes. When evidence is informed more sparsely, however (for example, combined SFR and POD designs) the resulting point estimates are not consistent with theoretical expectations of increasing benefit as rooms become more parent centred (Appendix Figure A. 9).

Uncertainty intervals in the vague prior scenario are substantial and unrealistic, suggesting only the combined SFR/POD studies are consistent with near complete cure rates of all outcomes and improved point estimates compared to both the new open bay and single-family components of their designs. In the base case, estimates are generally more pessimistic although all designs are associated with improved probability of being better than traditional open bay across most outcomes. Lower bounds of credible intervals still suggest very surprising benefit but are generally more reasonable and comparable across designs. Half-wall designs continue to provide estimates that are inconsistent with theoretical expectations, and the remaining designs are all approximately similar. Including Puumala (2020) in the review results in estimates of decreased benefit for single family room designs that translate across all outcomes as a result of the constant potency assumption (Figure 43). The finding that designs are generally comparable in terms of morbidities is inconsistent with the expectation that greater family-centeredness improves outcomes. One possibility is that this may be the result of studies in other designs being confounded by other additions (e.g. new open bay in sepsis being confounded by sepsis reduction bundles or practice change), but the lack of evidence of disagreement between randomized and non-randomized trials suggests that there could be a common benefit shared across all new designs.

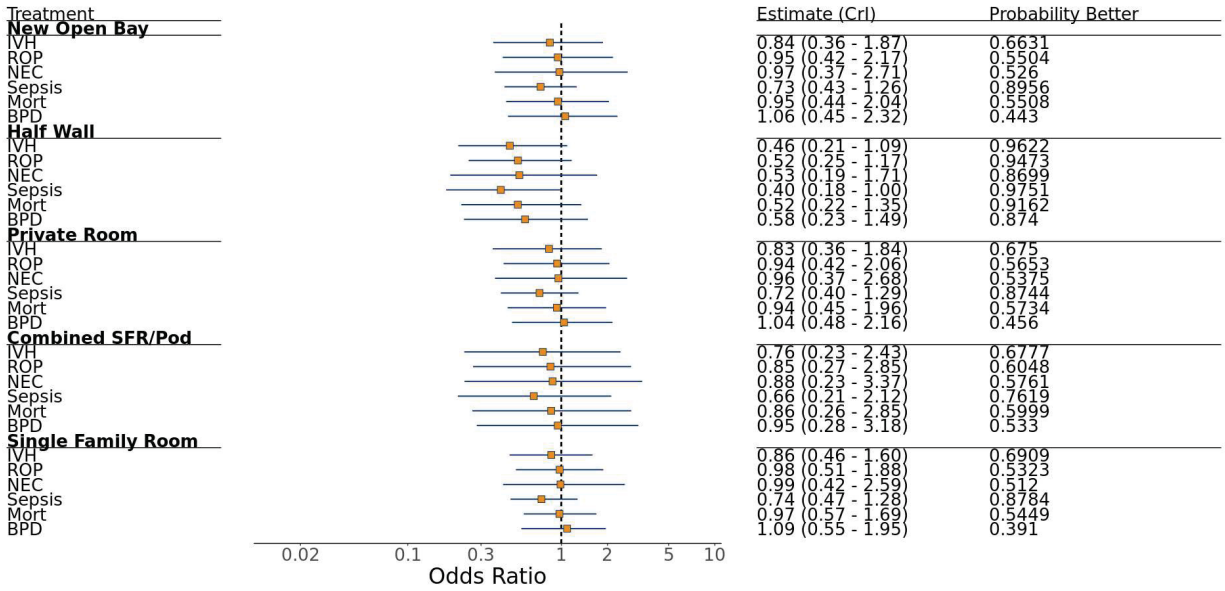


Figure 42. Forest plot from multivariate NMA with correlation structure imputed from PNFU data, the constant potency assumption, and base case priors. Probability better is the probability that a design type is better than open bay. IVH = intraventricular hemorrhage; ROP = retinopathy of prematurity; NEC = necrotizing enterocolitis; BPD = bronchopulmonary dysplasia; SFR = single-family room.

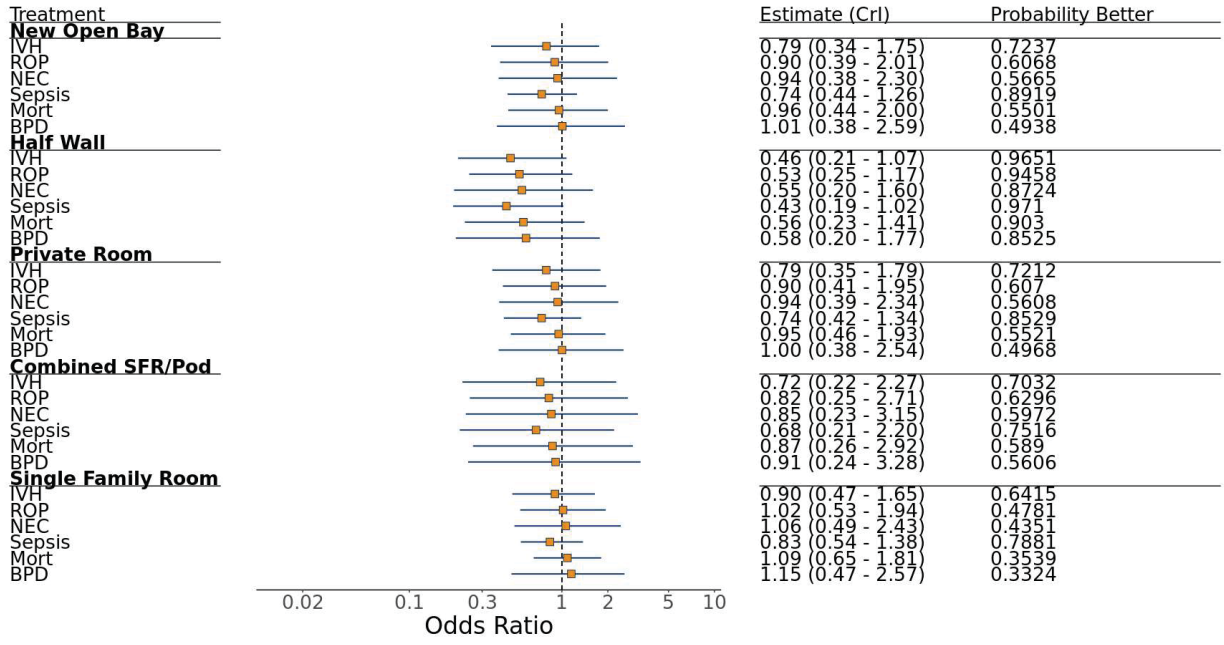


Figure 43. Sensitivity analysis including Puumala. Forest plot from multivariate NMA with correlation structure imputed from PNFU data, the constant potency assumption, and base case priors. Probability better is the probability that a design type is better than open bay. IVH =

intraventricular hemorrhage; ROP = retinopathy of prematurity; NEC = necrotizing enterocolitis; BPD = bronchopulmonary dysplasia; SFR = single-family room.

Length of stay.

Evidence networks. As seen in networks for morbidities and mortality, the largest number of comparisons come from single family room designs. Combined and private rooms both showed two studies each and three studies provided information regarding effectiveness of new open bay designs.

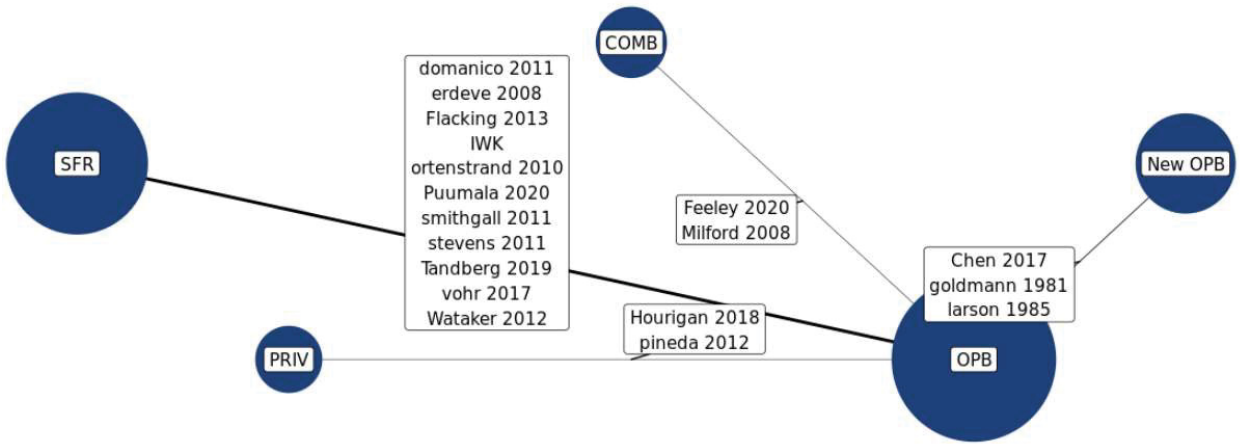


Figure 44. Network plot for length of stay data. Larger nodes indicate a relative increase in the number of patients compared to other treatments. COMB = combined open bay and single-family room; OPB = open bay; PRIV = private room; SFR = single-family room.

Assessment of heterogeneity in pairwise comparisons and the NMA. Results of pairwise meta-analyses are shown in Appendix Figure A. 11 and Appendix Figure A. 12. Only single-family room designs allowed for consideration of heterogeneity owing to missing standard errors for trials comparing other designs. The value of the ratio of means varied from length of stay decreasing (0.64) to length of stay increasing (1.23) with no clear patterns regarding the types of studies that estimated decreased in increased lengths of stay (e.g., randomized vs non-randomized, single vs multi-center). For example, the two included randomized controlled trials offered opposite estimates of the effect on length of stay. This may be partially explained by the

small difference in mortality seen in some studies, which may have resulted in longer average lengths of stay since the majority of deaths occur early. Alternatively, the increased length of stay may be driven partially by increased rates of BPD seen in the morbidities NMA or could also simply be the result of statistical noise since all confidence intervals are consistent with both meaningful benefit and harm. Down-weighting of evidence does help to resolve a portion of the statistical heterogeneity, reducing I^2 from 44% to 14%. One explanation for the high I^2 relative to morbidities may also be that continuous outcomes are estimated much more precisely than dichotomous outcomes in general, leading to less within study variability relative to between study variability. In a sensitivity analysis including Puumala (2020), the estimate of I^2 is increased to 88% (Appendix Figure A. 13) and 79% (Appendix Figure A. 14) and point estimates for single family room flip to being nearly identical.

Findings across different methods. Network meta-analyses were conducted using the base case priors (standard deviation of 0.6 on treatment effects) and vague priors (standard deviation of 100) and shown in Figure 45. Owing to the relatively high precision with which trial results were estimated, the use of the stronger base case priors have no appreciable effect compared to standard vague priors. Results are consistent with those expected from the pairwise meta-analysis, suggesting that differences between room designs are small, with the worst performing design being single family room designs. In a sensitivity analysis excluding Puumala (2020),

estimates for single family room become slightly more precise and point estimates are equal with open bay (Figure 46).

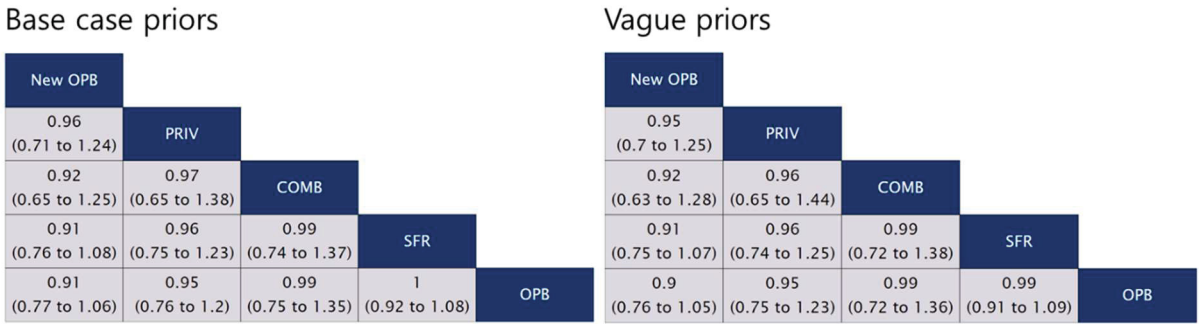


Figure 45. League table for length of stay. Comparisons are column over row, with ratio of means less than 1 indicating a reduction in length of stay compared to the design in the row. OPB = open bay; PRIV = private room; COMB = combined open bay and single family room; SFR = single-family room.

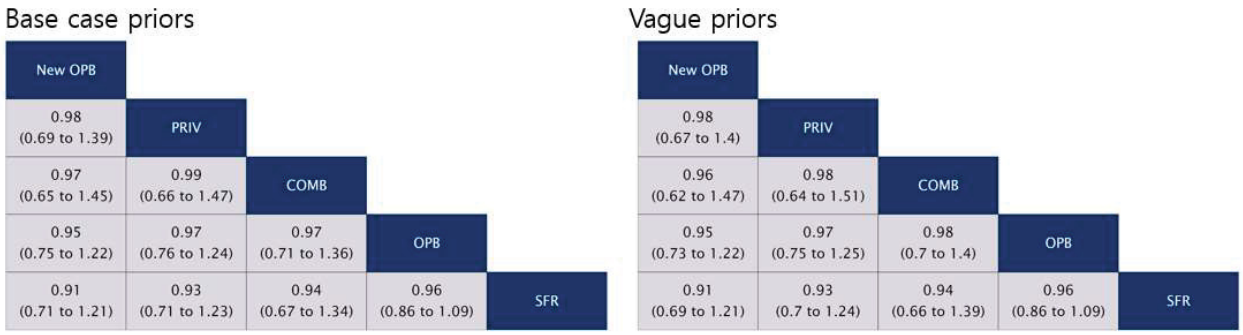


Figure 46. League table for length of stay including Puumala. Comparisons are column over row, with ratio of means less than 1 indicating a reduction in length of stay compared to the design in the row. OPB = open bay; PRIV = private room; COMB = combined open bay and single family room; SFR = single-family room.

Bayley III Scores.

Evidence networks. The evidence network for Bayley III scores was the sparsest of all outcomes assessed, with single studies contributing data for both single family and private room. No other studies were available that assessed Bayley III scores in other room designs.

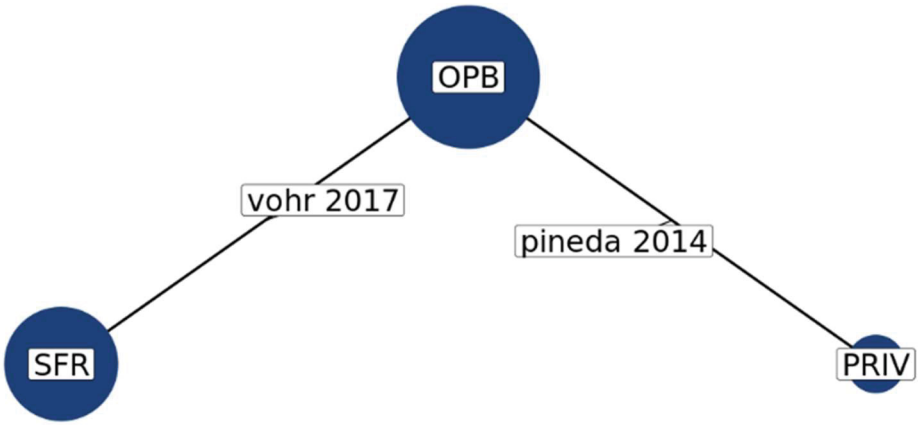


Figure 47. Network diagram for Bayley III scores. OPB = open bay, PRIV = private room; SFR = single-family room.

Assessment of heterogeneity in pairwise comparisons and the NMA. As only one study each for private room and single-family room contributed data to these analyses, no assessments of statistical heterogeneity in pairwise comparisons was possible.

Findings across different methods. Base case priors and vague priors gave similar estimates, with base case priors providing only a small amount of regularization (Figure 48). This is unsurprising given that these priors were still very vague in comparison to the scale of the outcome. While these results would suggest that single family rooms are potentially beneficial and private rooms are potentially harmful, there were important differences between units in

terms of the amount of parent visitation. Further sensitivities of this outcome were explored in the health economic model.

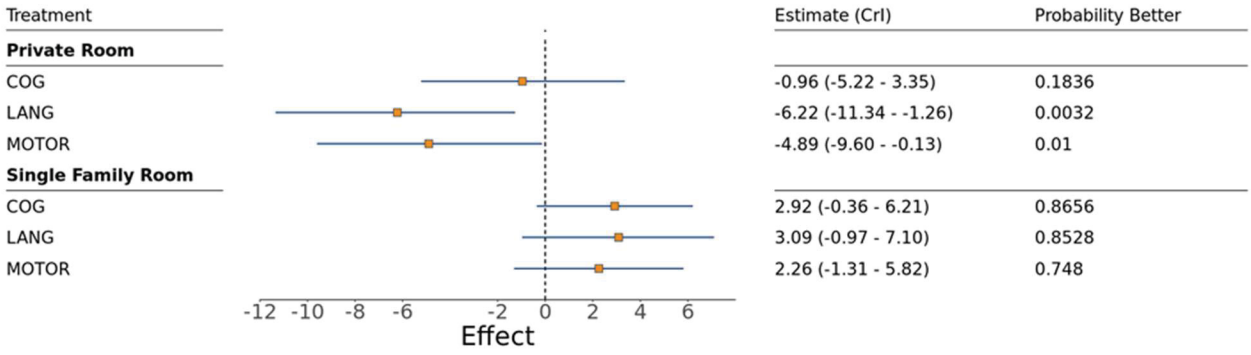


Figure 48. Mean difference compared to open bay for different room designs. COG = Cognitive subscale, LANG = language subscale; MOTOR = motor subscale.

Cost-effectiveness Model

Based on the results from the analyses of the Chez-NICU EQ-5D data, no quality of life decrement was used for the in-hospital portion of the model. This was based on the finding that mothers' QALYs were similar to age and sex adjusted Canadian population averages, in addition to the small portion of time represented by the hospitalization period. Results from the cost-effectiveness model suggest small to modest differences in terms of discounted QALYs, and costs in the over-all population (Figure 49.) Rankings in the overall population are consistent whether a healthcare or societal perspective is used, and results are generally consistent with ranking based off the results of the network meta-analysis. When looking at all treatments in the over all sample, half-wall designs have the highest expected net monetary benefit once the willingness to pay for one QALY crosses $20,000 in both the healthcare (Figure 50) and societal perspective (Figure 51). No treatment has a greater than 50% chance of being cost-effective, and the high probability of being cost-effective observed in open bay designs is primarily the result of high correlation in the effectiveness parameters of competing designs. Results in the over-all

population are unchanged in a sensitivity analysis including the results of the Puumala et al (2020) study (Appendix Figure A. 16).

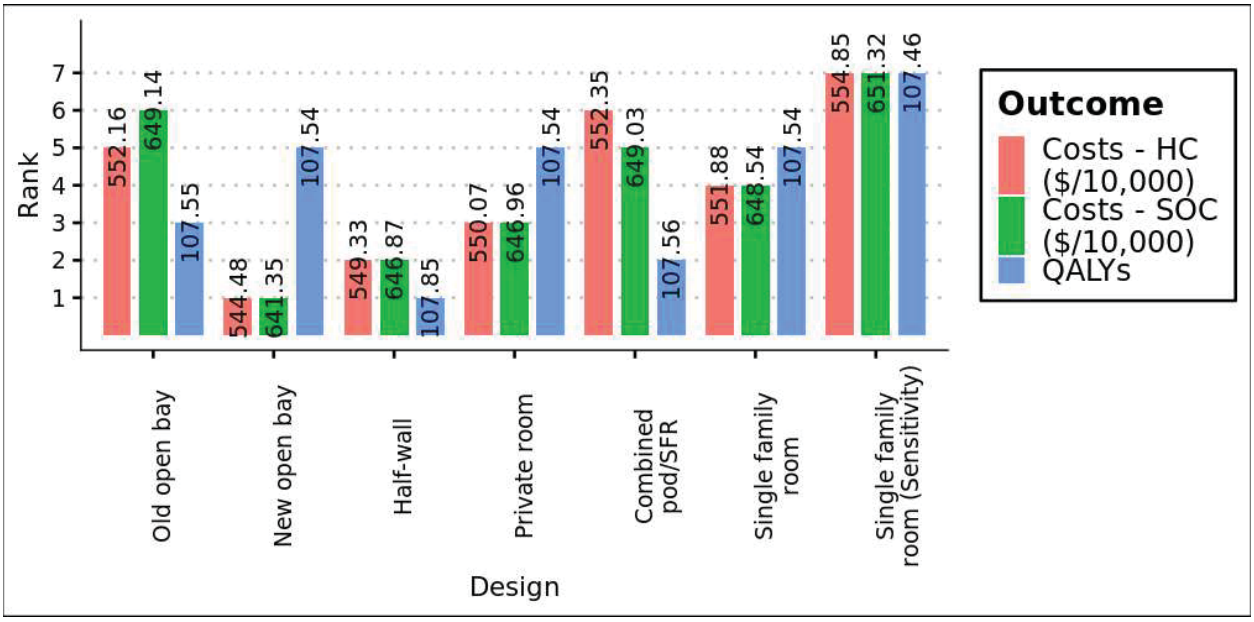


Figure 49. Costs and QALY in the over-all population. SFR = single-family room.

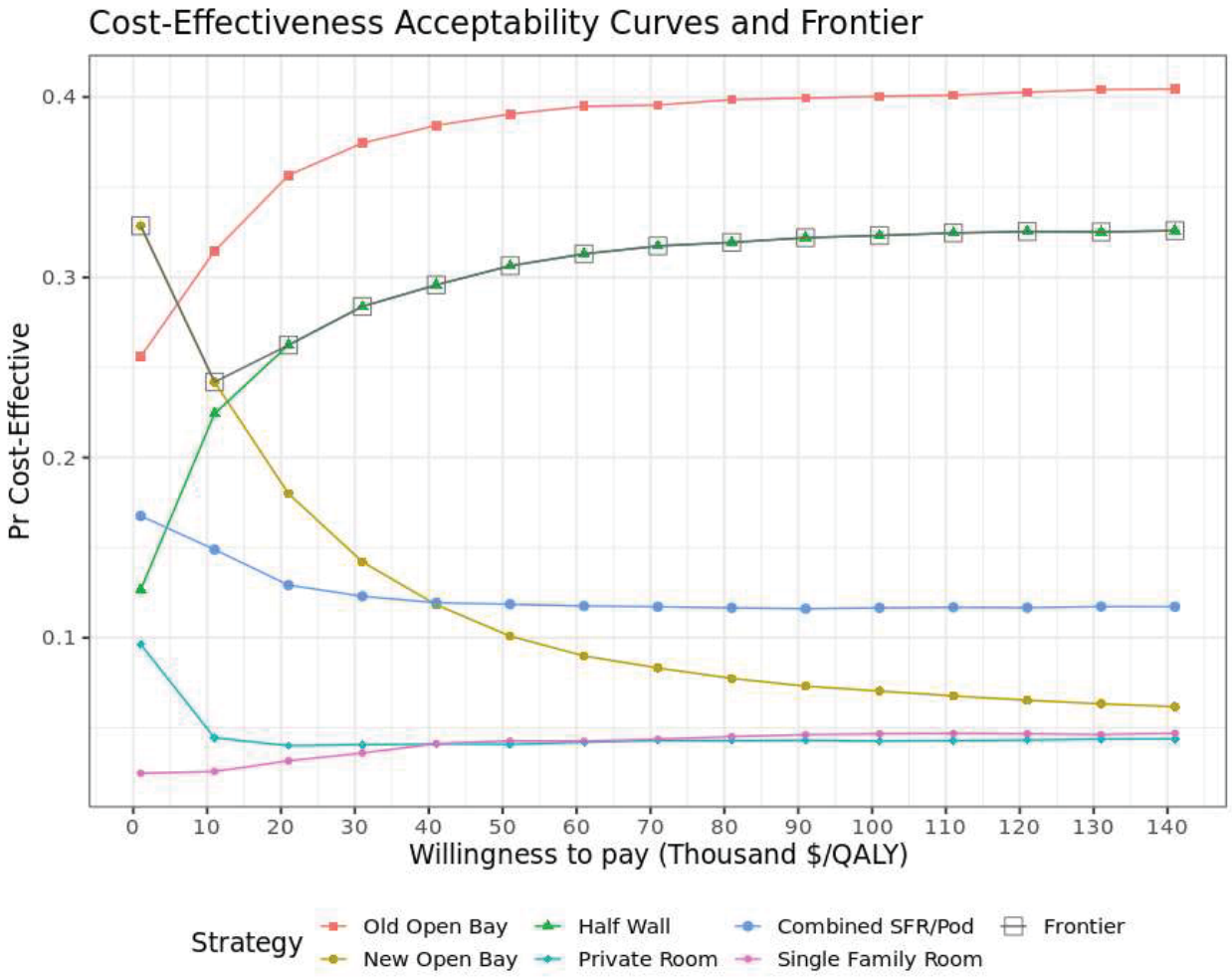


Figure 50. Cost-effectiveness acceptability curves and frontier for the over-all group using the healthcare payer's perspective. Treatments higher on the y axis have a higher probability of being the cost-effective treatment at a given willingness to pay threshold. Gray boxes indicate the treatment with the highest expected value of the net monetary benefit.

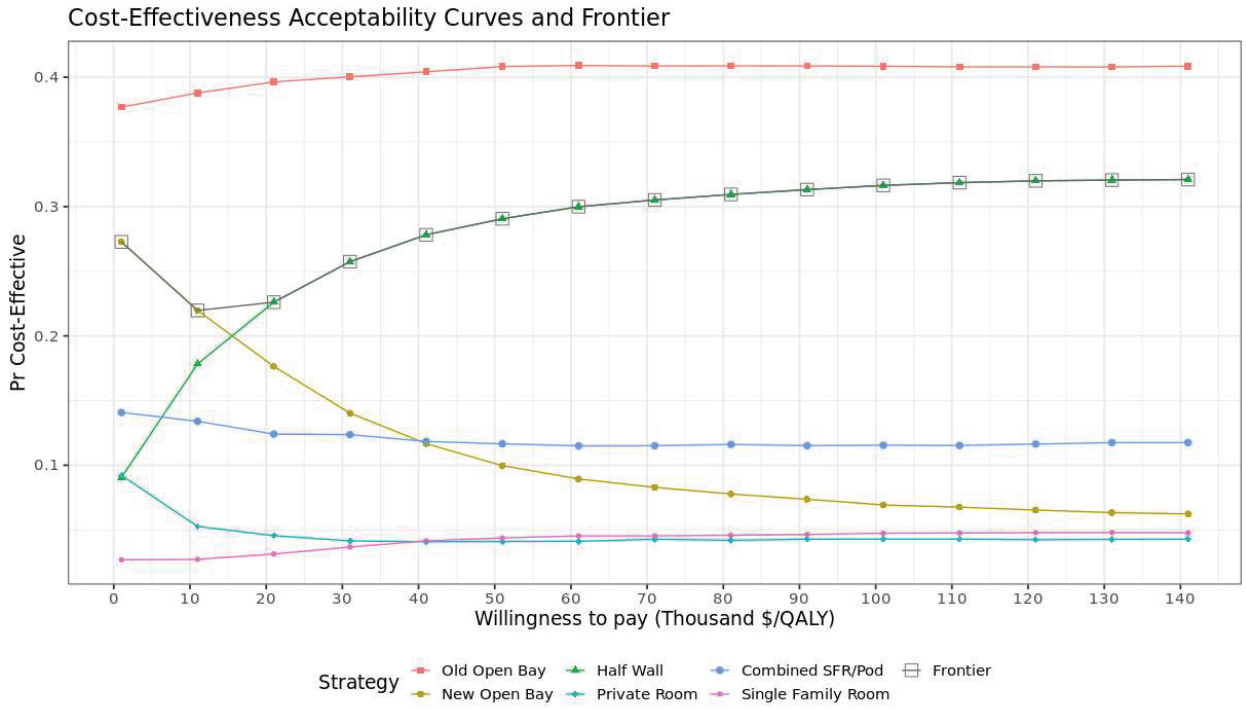


Figure 51. Societal perspective: Cost-effectiveness acceptability curves and frontier for the overall group. Treatments higher on the y axis have a higher probability of being the cost-effective treatment at a given willingness to pay threshold. Gray boxes indicate the treatment with the highest expected value of the net monetary benefit.

When designs are compared across gestational age sub-groups both the optimal treatment and uncertainty in treatment rankings vary (Figure 52). As gestational age increases, a general trend of decreasing importance of clinical performance is observed as treatment effects are applied to smaller and smaller event probabilities and lengths of stay (Figure 53). In extremely preterm infants, cost savings related to treatments with the lowest estimated lengths of stay (e.g. new open bay) are quickly dominated by those with larger benefits on morbidity and mortality. The opposite is observed in term infants, where lengths of stay are so short and events so improbable that old open bay can appear cost-effective. The unexpected trend of open bay becoming the optimal treatment in term infants as the willingness to pay for a QALY increases is likely the result of simulation noise since all designs have 100% intact survival in this population

(Figure 53). Results are consistent in sensitivity analyses including all studies in the efficacy data (Appendix Figure A. 17), and when using the societal perspective (Appendix Figure A. 18).

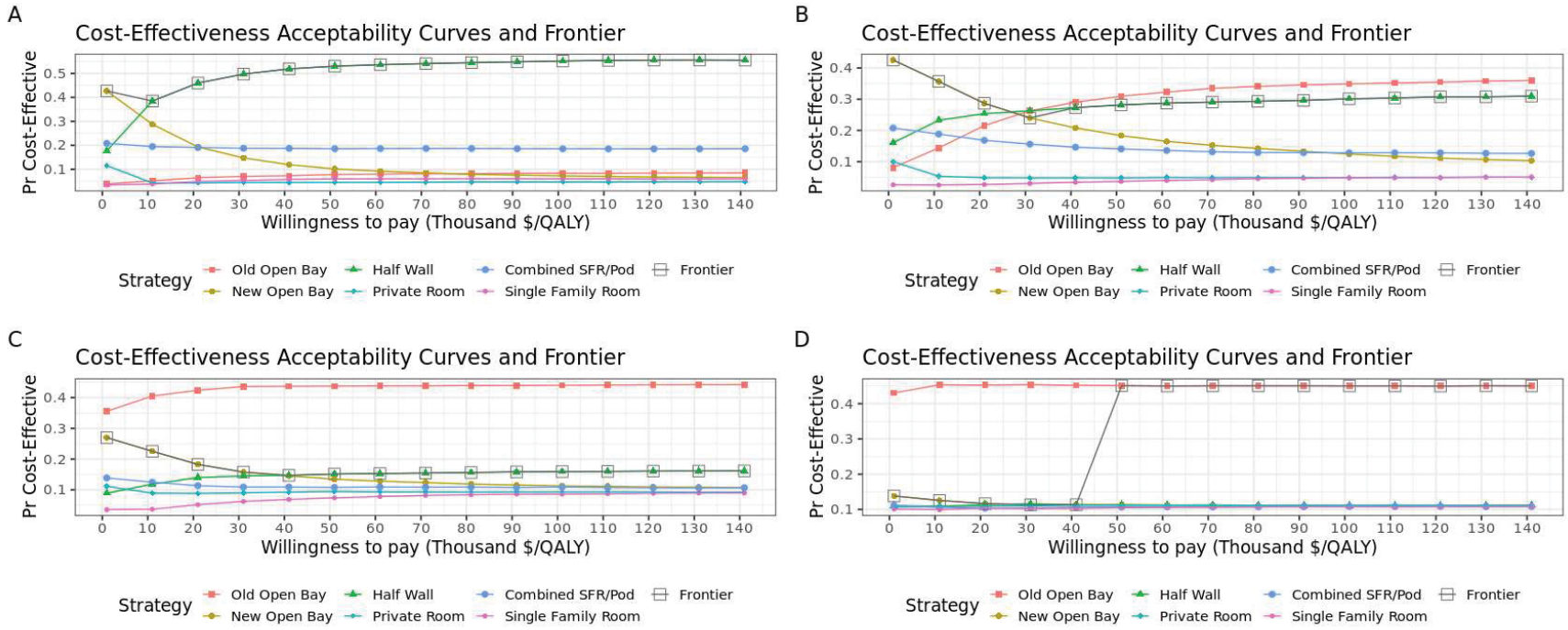


Figure 52. Cost-effectiveness acceptability curves and frontier for extremely preterm (A), very preterm (B), moderately preterm (C), and term (D) gestational age sub-groups. Treatments higher on the y axis have a higher probability of being the cost-effective treatment at a given willingness to pay threshold. Gray boxes indicate the treatment with the highest expected value of the net monetary benefit.

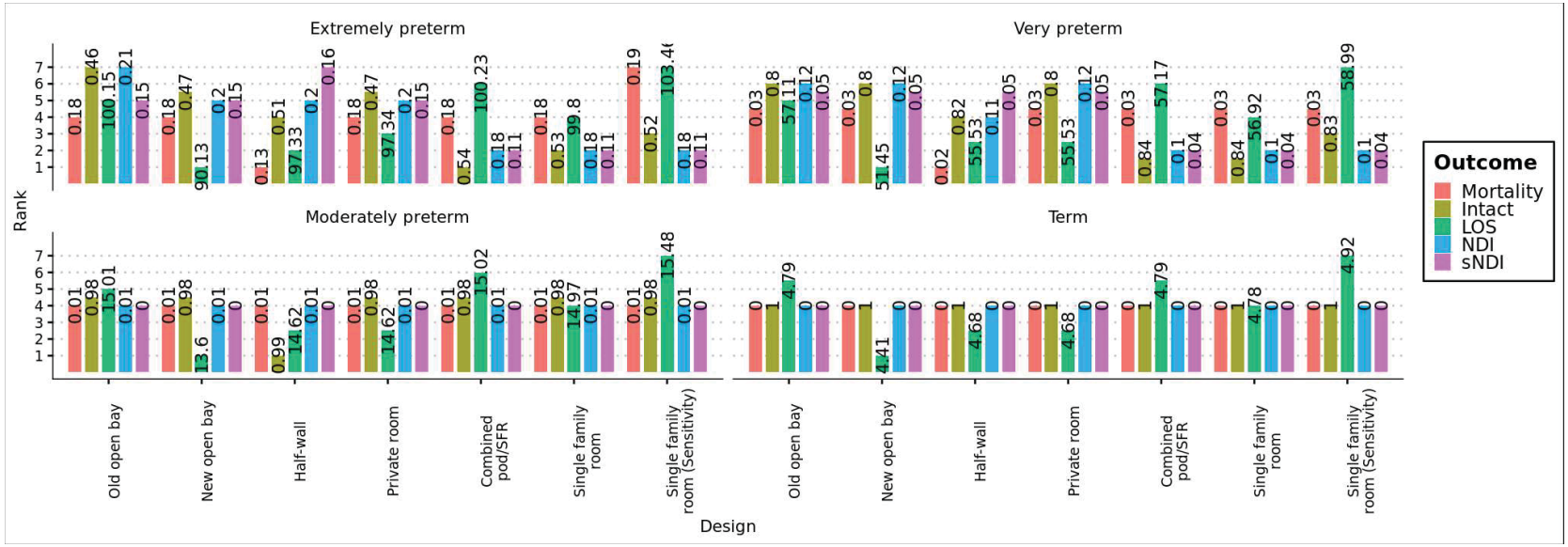


Figure 53. Rates or mortality, intact survival, length of stay, and disability by gestational age sub-groups.

When results are assessed by subgroups of lower income and higher income parents, conclusions are generally unchanged (Figure 54) despite higher rates of mortality and disability and longer lengths of stay in for lower income families (Figure 55). The exception is the small decrease in uncertainty in the point at which half-wall designs are preferred over a new open (Figure 54). This finding is likely the result of the decreased rates of NDI or mortality predicted in higher income families, which decreases the differences between designs.

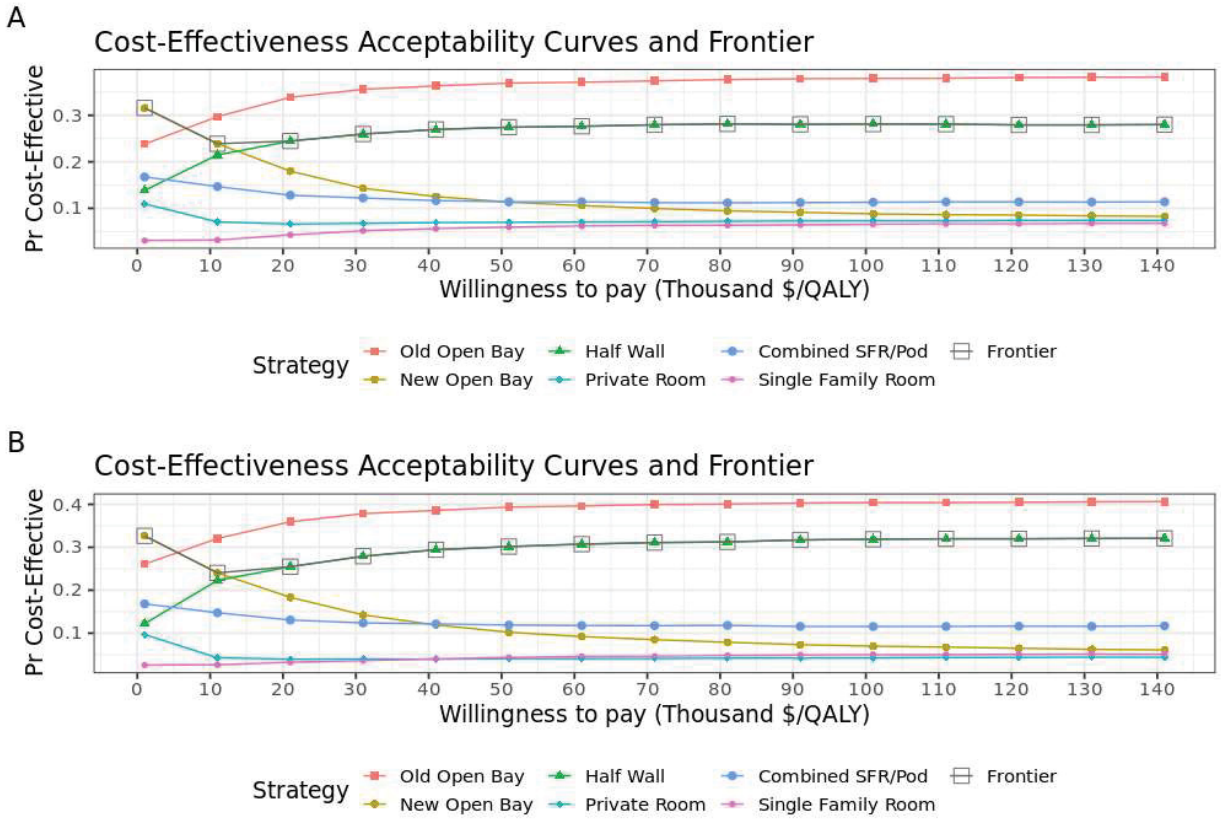


Figure 54. Cost-effectiveness acceptability curves and frontier for lower income (A), and higher income (B) sub-groups. Treatments higher on the y axis have a higher probability of being the cost-effective treatment at a given willingness to pay threshold. Gray boxes indicate the treatment with the highest expected value of the net monetary benefit.

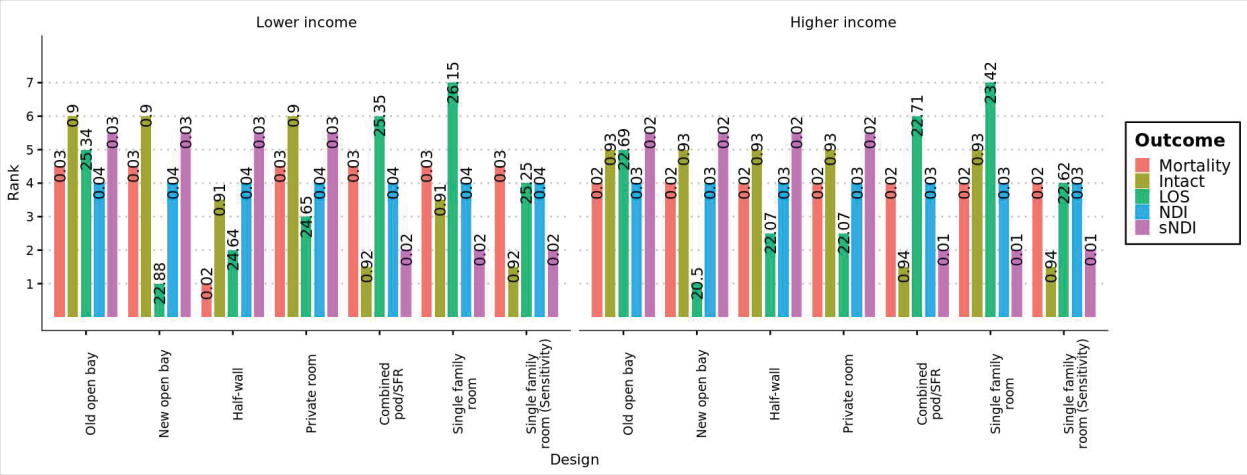


Figure 55. Rates of mortality, intact survival, length of stay, and disability.

Based on the structure of the microsimulation, the most important comparative effectiveness outcomes were determined to be mortality (via influence on QALYs), and the ratio of means for length of stay (via influence on costs). A sensitivity analysis was conducted to assess the changes in net monetary benefit in the over-all sample based on variation in these parameters (Figure 56). Both parameters suggest sufficient sensitivity for single family room (the worst treatment), to jump the being amongst the best treatments when varied over the uncertainty in inputs.

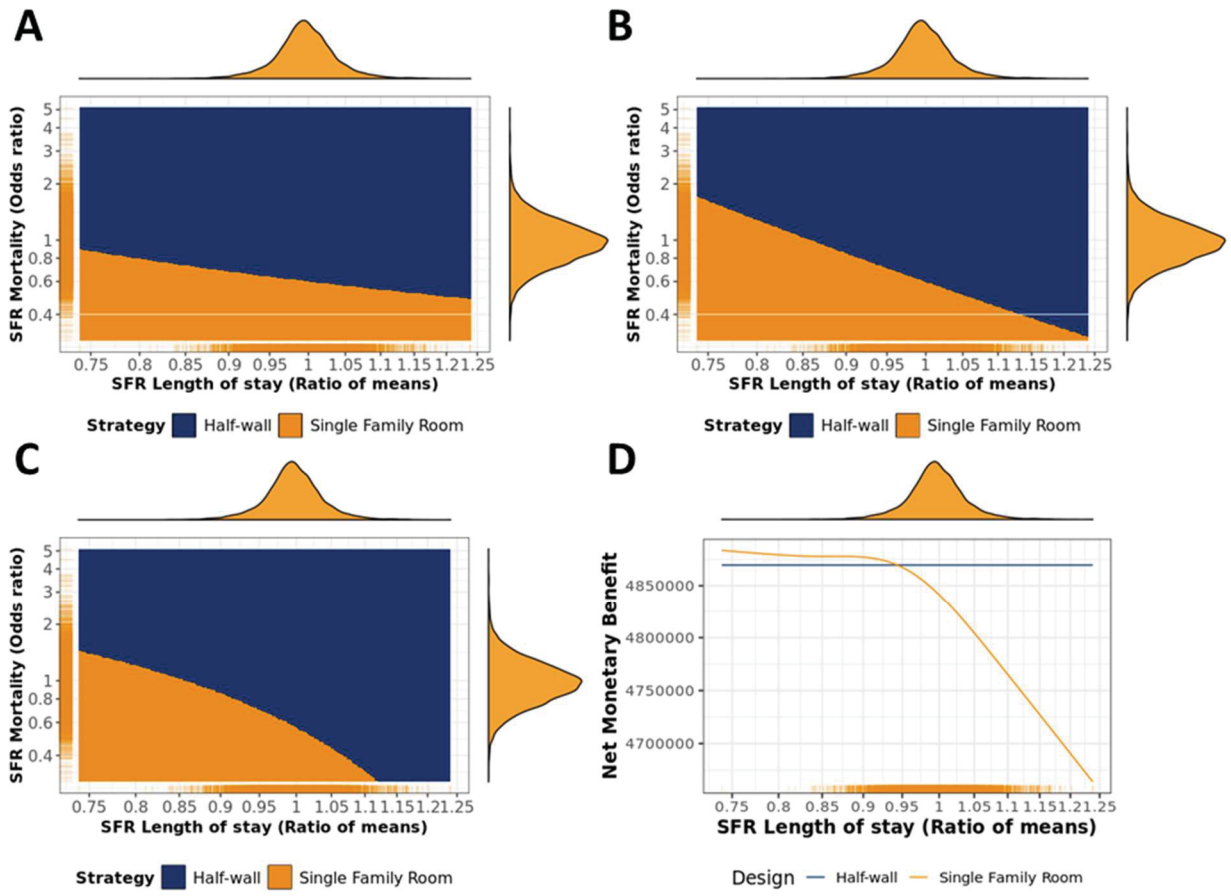


Figure 56. Sensitivity of single-family room rank to mortality, and length of stay efficacy parameters. Horizontal lines indicate the point at which single family room would become best, second best, etc. Comparisons are organized by gestational age subgroups including extremely preterm (A), very preterm (B), moderately preterm (C) and term infants (D). Panels A to C show sensitivity across mortality and length of stay and panel D across length of stay only. Rug plots (orange ticks) and marginal densities show the distribution of uncertainty in inputs for single family room as used in PSAs. SFR = single-family room.

Stochastic Multi-Criteria Acceptability Analysis

The lower and upper bounds of the 95% interval hull for intact survival were 0.85 and 0.90 for the over all sample and 0.1 to 0.46 for the extremely preterm group. This highlights how trade-offs of clinical outcomes are potentially more dramatic in more preterm populations. Findings from the SMAA suggest that most differences in intact survival are small in the overall population, and so single family rooms can be chosen as the optimal design with high confidence

if the intangible benefits of family centeredness and patient preferences are given a modest weight of approximately 0.3 (Figure 57). The only remaining alternatives with a non-negligible probability of being optimal at their central weights are half-wall and combined SFR designs. When considering the extremely preterm population the trend is generally similar, although there is less confidence in the decision and the weights for intact survival now reference the importance of a swing in over 30% of the event rate (Figure 58 and Figure 59).

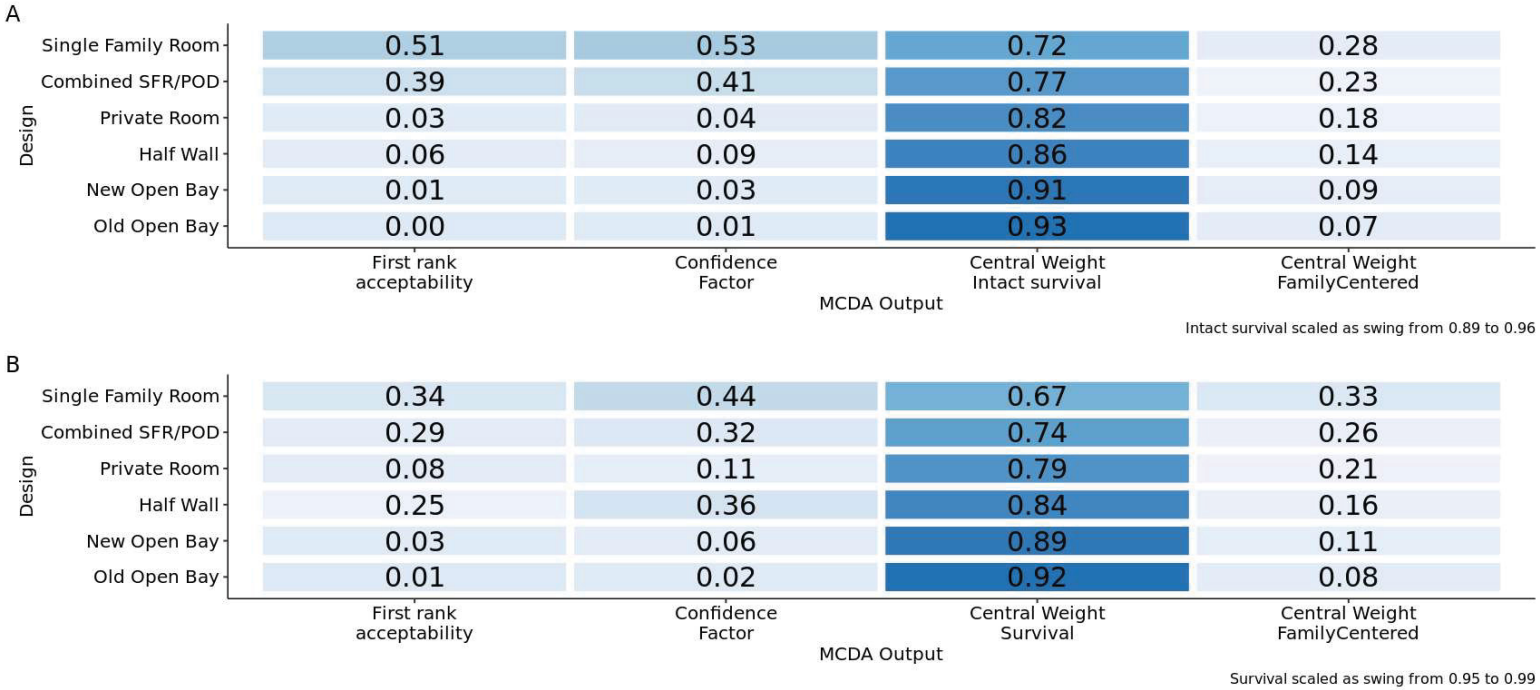


Figure 57. MCDA results for the overall population using intact survival (A) and survival (B) as the outcome of interest. First rank acceptability is the proportion of simulations in which the intervention ranks first. Central weights are the vector of weights at which a given intervention maximizes its probability of being best, and the confidence factor is the proportion of times that a treatment is ranked first at the value of its central weights. Intact survival and survival are scaled to be between 0 and 1 using a 95% confidence interval hull. SFR = single family room.

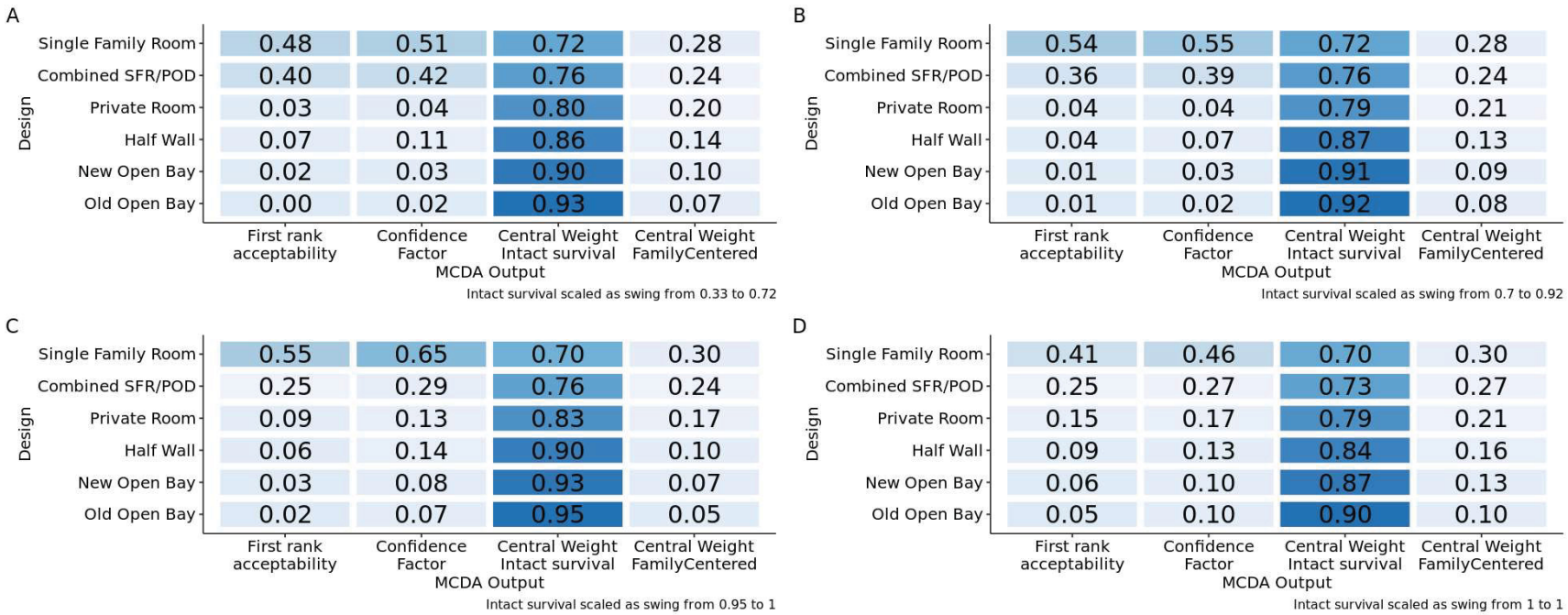


Figure 58. MCDA results for intact survival in the extremely preterm (A) very preterm (B) moderately preterm (C) and term (D) infant populations. First rank acceptability is the proportion of simulations in which the intervention ranks first. Central weights are the vector of weights at which a given intervention maximizes its probability of being best, and the confidence factor is the proportion of times that a treatment is ranked first at the value of its central weights. Intact survival and survival are scaled to be between 0 and 1 using a 95% confidence interval hull. SFR = single family room.

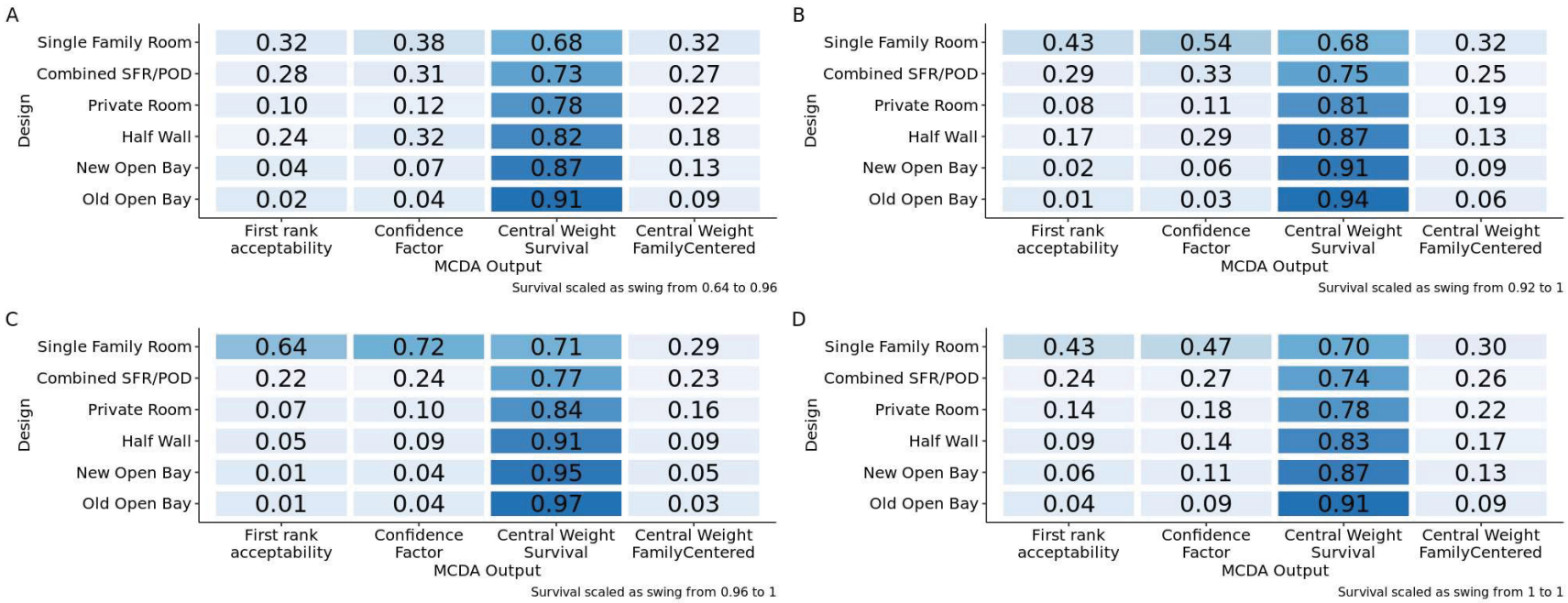


Figure 59. MCDA results for survival in the extremely preterm (A) very preterm (B) moderately preterm (C) and term (D) infant populations. First rank acceptability is the proportion of simulations in which the intervention ranks first. Central weights are the vector of weights at which a given intervention maximizes its probability of being best, and the confidence factor is the proportion of times that a treatment is ranked first at the value of its central weights. Intact survival and survival are scaled to be between 0 and 1 using a 95% confidence interval hull. SFR = single family room.

Chapter 5: Discussion

This chapter includes a discussion of the findings from the study, discussed as outlined in the primary and secondary research questions. This is followed by a discussion of the over all strengths and limitations of the analysis as well as the implications for future research and clinical practice.

When using a life-time horizon and a family-based model for assessing cost-effectiveness of alternative room designs, most designs consistently ranked above an old open bay, but differences in terms of net monetary benefit were small and there is uncertainty in ranking of designs. Performance on clinical outcome measures are relatively more important in more preterm babies and short stay term infants are unlikely to receive sufficient clinical benefit to be a major influence on the decision. While there were not any meaningful differences in conclusions when evaluated based on lower and higher income families, there were differences in the absolute probabilities of negative outcomes suggesting these groups should be considered carefully as evidence of the effect of design on developmental outcomes accrues. Results of the cost-effectiveness analyses and SMAA are consistent with expectations based on the results of the NMA, where half-wall units showed the greatest effectiveness and single-family rooms were generally similar to open bays. This is inconsistent with theoretical expectations that increasingly family-centred or integrated care would result in better outcomes via reduced stress, increased neonatal stability, and increased interaction between family and infant prior to discharge (Als, 1982; Lester et al., 2011; O'Brien et al., 2015). There was limited ability to compare results of non-randomized and randomized trials, and all were restricted to single-family room designs. Despite findings from randomized trials being generally consistent within single family room comparisons, differences across designs are most likely explained by a combination of sparse

data resulting in point estimates that are more favourable by chance alone (as evidenced by wide credible intervals) and bias from the underlying non-randomized evidence base (Verde & Ohmann, 2015). Of further note is that in both the NMA and economic model there was substantial uncertainty in comparisons between design types, which suggests that room design may have a relatively small effect with differences between studies potentially better explained by other aspects of care.

Identifying the Cost-Effective Design Option

The primary research question of the study was to identify the cost-effective NICU design, taking the perspective of the Canadian payer over a life-time horizon. When incorporating uncertainty in the both the baseline natural history model and estimates of treatment effectiveness, the design with the highest probability of being cost-effective is the use of half-walls. Sensitivity analyses suggest that this finding is primarily driven by the large relative decrease in morbidities and mortality leading to increased intact survival alongside comparable performance in terms of length of stay compared to other designs. The treatments with the highest probability of being cost-effective after the half-wall design were the construction of a new open bay, followed by a combined SFR and pod design, private room, the original open bay, and SFR. This trend is not consistent with expectations based on the theoretical justification of private and single family rooms as described by the model of NICU design and infant development and supported by the theories of developmental and family centered or integrated care (Als, 1982; Lester et al., 2011; O'Brien et al., 2015). Single family room NICUs are expected to facilitate family centered, developmental care by providing a space that can be customized to each family's needs. By providing a dedicated family space, single family rooms are the only design that provide parents with the opportunity to truly room-in with

their infant. Studies evaluating the practices of parents in single family rooms compared to traditional open bay units have consistently shown an increase in total parent presence and involvement in care (Aija et al., 2019; Feeley et al., 2020; Pineda et al., 2012; Tandberg, Flacking, et al., 2019).

In addition to the changes in parent presence and involvement expected with single family room care, short-range models also suggest that single family room care is associated with a change in practice philosophy that changes the way that nurses, physicians, and allied health professionals work together to support families (Lester et al., 2011). This change in focus is typically described as a movement towards developmental care, family-centered care, or family-integrated care. Of these, developmental care is the approach most specialized within neonatology with its focus on how care practices and environmental manipulations can be matched with the needs of a given infant (Als, 1982). Single-family room removes a number of barriers to this individualized care, for example by providing increased control over lighting and sound, while providing an environment that encourages parents to be integrated in the care team. Principles of family centered and, more recently, family-integrated care are shared across a wider array of specialties and underlying concepts are central to the practice of family and pediatric nursing (Smith, 2018; Wright & Leahey, 1984). These approaches to care are the subject of a rich literature which is beyond the scope of this project, but their centrality in nursing practice challenges the face validity of the results of the economic model.

One explanation for the findings is that each room design was generally accompanied by both changes in practice and philosophy as well as increased space and in at least some cases, improved access to sinks. For example, all studies evaluating a new open bay design described some change in terms of access to technology, or standardization and improvement in care

practices. The one study evaluating half-wall designs also implemented a large developmental care program which was further supported by the Phillips Wee Care system (Altimier et al., 2005). Many studies therefore set out to evaluate the effect of room design on outcomes but evaluated the effect of both room design and practice change. This explanation is supported by the interrupted time-series analysis conducted by Puumala et al (2020) which showed that apparently large reductions in length of stay were explained by a time trend as opposed to the introduction of single family rooms. When taken as a whole, these trends provide justification that many aspects of the apparent benefits of a change in room design are shared by multiple design types. For example, while construction of a modern open bay does not provide the same ability for fine control of infant environment it does not prevent a family-integrated approach to care (Meredith, 2017). Providing increased access to lactation consultants or occupational therapy is likely more of an issue of budget and commitment to a philosophy of care than the structure of the unit. Similarly, the consistent benefits for reduced infection across room designs could be driven both by quality improvement initiatives that are design agnostic in addition to increased space for infants leading to decreased transmission across sites. Thus, the move from a crowded old unit to a new more spacious one may provide the largest part of the benefit of new design by providing the opportunity to re-think the approach to care and then the incremental benefit of increasingly private rooms is comparatively small. This would result in those incremental benefits being difficult to distinguish amongst the noise of indirect comparisons of varying designs, as reflected by the large uncertainty in comparisons across room types.

Systematic reviews of the benefits associated with developmental care interventions do not identify private rooms as a necessary requirement of interventions, and generally provide the strongest evidence for those aspects of developmental care that are family focused. Notably,

Burke (Burke, 2018) concludes that there is no strong evidence of a dose effect of developmental care. These conclusions are consistent across systematic reviews (Ohlsson & Jacobs, 2013; Symington & Pinelli, 2006), suggesting that the potential incremental benefits of single family room compared to private room or a new open bay may rest largely on the differences in the influence on parent engagement. This has been an under-developed area of research as there is no comparative evidence of the differences in parent engagement associated with changes across designs.

The question of the magnitude of incremental benefit associated with increasingly family-focused room designs is also relevant when discussing the potential impact on the ability to practice family-integrated care. The highest quality evidence to date on the effects of the implementation of a family integrated approach come from a multi-center cluster randomized trial that included 14 centers randomized to family-integrated care and 12 centers randomized to standard-care (O'Brien et al., 2018). Findings from the study included improvement in weight gain (change in Z scores of 1.58 vs 1.45), high frequency exclusive breastmilk feeding (70% vs 63%) and lower mean anxiety scores (70.8 vs 74.2). No meaningful differences were observed in mortality and morbidities or the mean duration of hospital stay or oxygen support. These findings raise questions as to the degree of incremental improvement in family-integrated care that would need to be associated with different designs in order to have a meaningful effect on clinical outcomes.

Alignment of the results of the cost-effectiveness model with theoretical expectations also requires assessment of whether the effect of designs on parents is appropriately captured. The model structure was driven primarily by the effects on infants, with decrements to parent quality of life being tied to the disability and mortality of their children. This decision was justified

based on the observation that parents in the Chez-NICU open bay study were discharged with EQ-5D utility estimates and PDSS scores consistent with the age and sex-adjusted population norms. The decision was further bolstered by the effect of the NICU stay being generally small in reference to the rest of the parent's life, with meaningful differences in quality of life being driven more by infant disability, consistent with the theory of spillover quality of life decrements (Wittenberg et al., 2019). There are currently no systematic reviews of the impact of design on parent outcomes, but studies have generally focused on private and single-family rooms and display mixed results depending on outcomes assessed. Parents in single family room NICUs have generally reported greater support and feelings of closeness to their infants (Shahheidari & Homer, 2012; Tandberg et al., 2018), consistent with expectations of a change in room designs being associated with improved patient centred care. A recent publication from two Norwegian units further suggests a nearly 80% reduction in the number of parents in a single-family room unit scoring as high-risk for post-partum depression based on the Edinburgh Postnatal Depression Scale administered at 14 days when compared to parents cared for in a crowded open bay. When assessed longitudinally across the 14-day timepoint, discharge, term equivalence, and four months after term-equivalence the average difference in EPDS scores was two points lower in the group cared for in single family rooms, suggesting that absolute differences between parents in the two designs decreased over time. Comparisons within a Canadian combined SFR and pod unit also suggested an improvement with a move to a new unit, albeit a smaller 0.35-point difference when measured at time of enrollment. Parents in the two units did not differ on readiness to discharge or breastfeeding self-efficacy, but parents in the combined unit did spend more hours per week present. Scores for parent stress in the open ward and combined pod were

similar in means and standard deviation to those published in by Lester et al in their comparison of open bay to single family rooms (Lester et al., 2014).

The model of NICU design and infant outcome (Figure 1) hypothesizes that the effect of room design is a combination of direct and indirect effects. The mediators of indirect effects include family centered and developmental care, parent and family factors, staff behaviour and practices, and medical practices. Despite variability in relative rankings of treatments in this cost-effectiveness analysis not making intuitive sense, the general trend that new units are better than crowded open bays, that differentiation between new units is difficult, and that results are sensitive to small adjustments in key parameters is consistent with the proposed model. A change to any new unit creates an opportunity to re-evaluate practices and care philosophy to be more supportive and inclusive of families. As a result, any new design can potentially access many of the benefits of mediators of effect, leaving remaining differences to be explained only by those changes that are not possible in one design over another. In terms of increasing parent presence and involvement, single family room units have consistently shown a favourable effect but similar benefits may be possible with a bed on the unit that is not necessarily in the same shared space (Raiskila et al., 2017). Further, there is at least some evidence that a bed alone is not sufficient to increase either presence or involvement (Raiskila et al., 2017), which is also consistent with model of NICU design and infant outcome.

Point estimates from the NMA also showed meaningful sensitivity to the inclusion of Puumala et al (2020). This study was notably different than others as it found strong differences between gestational age subgroups on both length of stay and the rate of sepsis. In both cases, crude and adjusted data suggested a strong benefit of single-family room design in the most preterm infants, with a reversal of effect in moderately preterm and term infants. Since the

primary analysis of NMA used overall data and these were weighted primarily by the inputs from term infants, it is possible that conclusions could differ if the observed subgroup effects were used instead. Findings from this study are further complicated by the results of the interrupted time series analysis, which authors used to conclude that changes in length of stay were explained by trends across time instead of being associated with room design. No time series analysis was conducted for sepsis or mortality although a recent report of the Canadian Neonatal Network (Shah et al., 2019) suggests a numerical trend towards decreasing rates of nosocomial infection over time in extremely preterm infants. The presence of a sub-group effect of design is further refuted by available data from the Ortenstrand et al (2010) and IWK RCTs. Despite observed changes in the magnitude of mean differences across gestational age sub-groups, findings from Ortenstrand et al. are consistent with a single treatment effect on the ratio of means scale. Similarly, when stratified by gestational age subgroups the results of the IWK single family room RCT do not provide evidence of a subgroup effect on the ratio of means scale. These findings are further corroborated by the lack of a clear trend in treatment effects across any outcome by the average gestational age in the included studies. Excluding Puumala et al (2020) from the NMA leads to substantial changes in the estimates for single family room for sepsis, mortality, and length of stay and leads to more favourable performance versus comparators in the cost-effectiveness model.

The results of the cost-effectiveness model can be most directly compared to three economic evaluations of single-family room care, all of which compared against old open bay designs. Stevens et al (Stevens et al., 2014) conducted a strict cost-only comparison based on their unit's transition from open bay to single family room and predicted that direct costs of care were lower in the new unit. Importantly, this result relied on adjustment for post-treatment

variables and length of stay which assumes that design has no effect on these variables. As noted in the network meta-analysis, pooled results across studies are consistent with both clinical benefit and harm of transition to single family room, with no strong patterns consistently favouring one over the other, which suggests that the assumption of similar outcomes may not be warranted.

Shepley, Smith, Sadler and White (Shepley et al., 2014) published a business case for a mixed single family room and semi-private unit and concluded that based on their assumptions a SFR design would be associated with a savings of $1.1 million 2014 USD per year. These figures are markedly different from estimates from the current model, where SFR was the most expensive design. Results in the analysis by Shepley at al are driven by two key assumptions: All infants <30 weeks would have both a 10 day reduction in length of stay, and single family room as well as a 15.5% reduction in overall direct costs, based on the Stevens et al direct cost comparison. Based on the findings of this review, both of those figures are likely overly optimistic for any room design. The final economic analysis was conducted by an author group that overlaps with those in the SFR business case paper (Sadatsafavi et al., 2017), and extends this concept to include a systematic review and probabilistic component. The analysis included three studies of nosocomial infections, one length of stay, and one assessing direct costs of core. As with other economic evaluations, conclusions are primarily driven by use of Ortenstrand et al as the sole source of information on length of stay reduction and the Shepley et al direct cost of care study.

Results from both probabilistic and deterministic sensitivity analyses suggest that the current evidence base results in substantial uncertainty in the optimal decision. In probabilistic analyses only half-wall designs had a probability of being cost effective above 50%, a finding

that was driven by evidence from a single study whose results lack face validity and relied heavily on extrapolation of effects into important unmeasured outcomes including mortality. No remaining designs had a probability of being cost-effective above 20% for typical values of willingness to pay in the overall or any sub analyses. This level of uncertainty is unsurprising given most treatments did not show strong evidence of being better than old open bay across any outcomes in the NMA, and generally had effects and credible intervals in the same range of effects. Findings are further contextualized by the results of deterministic sensitivity analyses which indicate treatment rankings are sensitive to the effect estimates for mortality and length of stay.

The interpretation of the SMAA results is generally consistent with the CEA and provides the benefit of further contextualizing the implied value placed on intangible benefits of room design that could not otherwise be incorporated in the model. A major difference between the results of the SMAA and the CEA is the finding that SFR and combined pod designs would be considered the optimal design, although with substantial uncertainty depending on the type of outcome measured. This is not surprising given that the SMAA ignores the effect of length of stay on costs, and that both of these designs had the best performance on intact survival. Single family room and combined pods are also both given the two highest rankings in terms of family centeredness, providing additional opportunity for them to be ranked highly in the SMAA. As with the CEA, the choice of optimal design is expected to differ across gestational age groups since the additional weight needed to be placed on intangible benefits may be more reasonable when trading against smaller ranges of change in intact survival than larger ones. Uncertainty in the NMA and CEA is also reflected in the SMAA, where confidence factors for most designs indicated that there was uncertainty in optimal treatments even when weights are set at those that

are intended to maximize a given design's probability of being the best treatment. For example, if in the case of extreme preterm infants a decision maker knows that they place more weight on the swing of intact survival from 0.33 to 0.72 over the swing from being least patient centred to most patient centred then the most relevant comparators would be new open bay, half-wall, and private room since those are the designs whose central weights put substantially more weight on intact survival. Only the confidence factor for half-walls is above 10%, suggesting potentially significant decision risk.

In the context of this broader empirical and theoretical knowledge base, it would likely be an overinterpretation of the results to suggest that findings of the economic model should be used as the sole guide for decision making, and should be instead be interpreted with a consideration of the uncertainty captured within the probabilistic and deterministic sensitivity analysis and the potential for biases in the relative effectiveness as well as those considerations that lie outside of those aspects of family outcomes that can be easily measured. When assessed through this lens, the results are broadly consistent theoretical expectations that differences in costs and infant outcomes between alternative designs are small and variable.

Impact of a Societal Perspective

The second research question of interest for this study was whether the optimal design would change as a result of taking a broader societal perspective. The purpose of this question was to capture the potential added benefit of reduced disability on parent out of pocket costs and costs that fall on areas of society outside of the healthcare sector. Following earlier burden of prematurity studies, this consisted primarily of additional costs associated with education as well as parent out of pocket and wage costs to visit the hospital or attend appointments. It was expected that this might lead to changes in treatment rankings favouring those treatments that

were potentially more costly but also had stronger effects on infant outcomes. Unexpectedly, even though this broader perspective did look at larger over-all costs of all designs, there was no meaningful change in the relative ranking of interventions of their probability of being the best design. This is potentially a consequence of the assumption that parent presence was the main driver of Bayley scores between treatments combined with high baseline levels of parent presence. Presence was chosen as the primary influencer of Bayley scores given observations in the literature that the effect of single family room designs on Bayley scores appears to be moderated by parent presence (Lester et al., 2016; Pineda et al., 2014) as well as the lack of estimated treatment effects for most designs. Changes in the assumption may result in changes to treatment ordering, but the since transition probabilities move infants towards healthier states and additional QALY decrements are only applied in the sNDI category, it is not expected that conclusions would change as a result. The impact of these effects is also limited by the most important variables in the model being mortality and length of hospital stay.

Patterns of Cost-Effectiveness Across Gestational Age Groups

A secondary research question was whether groups of infants could be expected to benefit more from a change in room design than others. This is important since the choice to move to a new unit may be different in a NICU that cares for extremely preterm infants that could potentially see large benefits from improved outcomes or reductions in length of stay than in a unit that primarily cares for moderately preterm or term infants with minor complications. As the rates or mortality and serious disability decrease, results begin to move away from the units with the largest clinical benefits, towards those that offer the greatest benefits in terms of costs via decreased costs of operating and decreased length of stay. This leads to new open bay and old open bay gradually rising in the treatment rankings, and alternative designs moving

closer together. Over all conclusions do not change however, since the high degree of uncertainty suggests that results are sensitive to NMA inputs and considerations that are not captured as part of the cost-effectiveness model.

Patterns of Cost-effectiveness Across Parent Income

The baseline natural history model predicts disability partly via parent income and education and therefore lower income families start with a higher risk of NDI and sNDI. This would suggest that if a design is associated with increased disability that lower income families would see larger increases in rates of disability and associated utility decrements and increased costs. This impact is observed in the probabilistic sensitivity analysis via small increase in the probability that combined units, private rooms, and single family rooms are cost-effective. The lack of a strong preference for one design over another in lower or higher income families is comforting given concerns that lower income parents would have less access to transport and supports required high levels of presence. Given higher income parents were predicted to be highly present this analysis would support the assertion that lower income parents may see larger benefits of private or single room care since their amount of presence has the most room to increase as a result of the change in room design. The finding that higher income parents visit more often is somewhat contradictory to early findings that younger parents were present more often in European NICUs (Raiskila et al., 2016) although that analysis was based on measuring the number of hours present instead of the proportion of days, and uncertainty in the difference between two models does not necessarily suggest they are in conflict (Gelman & Stern, 2006).

The lack of a difference in the cost-effectiveness, and the trend towards improved cost-effectiveness of private, combined, and single family rooms will require units to consider the relative trade-offs in small potential long-term improved outcomes and small increases in out of

pocket expenses associated with increased presence as well as the increased per-day spending observed in lower income families. An assessment of the sensitivity of design choice to partial or full reimbursement of out of pocket costs of lower income families was not pursued given the low probability that it could alter conclusions in the cost-effectiveness analysis. Given the results of the model, modest targeted unrestricted reimbursements may help to offset the additional financial costs of lower income families without meaningfully altering the comparative cost-effectiveness of any two designs. A more difficult issue to navigate may be the concern that private and single-family room designs assume that parents will be present, and there is an untested hypothesis that lack of presence may be more harmful in these designs than traditional open bay designs. Building a unit that consists entirely of single-family or private rooms assumes that this would be the preferred design by all families and prevents them from making an informed decision based on the environment in which they would feel the most comfortable. In a qualitative interview study of parents of extremely preterm infants, two themes that are relevant for decisions regarding room design were the importance of feeling connected to other families in the unit, and the feeling of guilt associated with being expected to spend more time at the hospital than was feasible (Bry & Wigert, 2019). These are not new concerns for the discussion of NICU design, as the potential for some parents to feel isolated or to experience increased stress was identified early on in the development the movement towards private and single family room designs (Pineda et al., 2012). This is not to suggest that the potential for some families to experience additional stress or isolation implies that certain designs are not ethical, but there is a need to avoid a paternalistic sense of knowing what is best for patients. Families included in interviews conducted by Bry and Wigert (2019) spoke to how staff could relieve or

amplify the guilt they felt when balancing priorities which suggests that the design of the NICU may not be as important an ethical concern as the approach to care and support offered.

Strengths

The current study has several strengths. Estimates for the baseline natural history model were estimated from a diverse set of high quality individual participant data that allowed for estimation of models that controlled important confounders using the most recent methods for modeling non-linear effects, with variable selection driven primarily by literature review supplemented with modeling approaches that have been shown to have favourable properties when compared to standard approaches to variable selection (Piironen & Vehtari, 2017a). Models were evaluated in terms of discrimination and calibration with performance that ranged from moderate to good. Analyses were guided by a proposed causal structure which clearly communicates underlying assumptions and results from the models in the baseline history model were generally consistent with previous research. The analysis is also the first to incorporate RCT level evidence for the effect of single room care on mortality, morbidities, length of stay, and parent presence and out of pocket costs conducted in a unit that receives a broad range of inborn and outborn infants of varying medical complexity.

The approach to evidence synthesis for efficacy parameters represents the most complete synthesis of studies assessing neonatal outcomes of room design conducted to date. Models were conducted within a multivariate framework that partially account for selective outcome reporting across trials, and treatment effects were imputed when needed to provide an evidence informed estimate of all efficacy parameters needed for the economic model. These models also account for the correlation in uncertainty of estimates of efficacy, which can have important implications in highly non-linear economic models like this one (Davis et al., 2014). The down-weighting

and skeptical prior approach limited the contribution of studies at the highest risk of bias to the NMA.

The microsimulation approach used allowed for further capture of correlations between non-linear prediction models, and leveraged copulas to ensure between multiple binary and continuous variables were maintained. This approach also allowed this model to be the first to incorporate both the effect of potentially improved Bayley scores as well as the effect of morbidities in order to predict cerebral palsy and create a composite outcome of intact survival. Further, this is the first economic evaluation of room design to be conducted within a cost-effectiveness framework, including impact on the entire family unit, incorporating changes in child and parent utility decrements over time, and taking both a healthcare payer and societal perspective.

Limitations

The complex nature of the required CEA lead to several limitations. In the baseline natural history model extrapolations or morbidities beyond week 31 lacked face validity and so an assumption of very low event rates was required. A number of potentially important antenatal variables were not available to control for confounding including cervical shortening, hydrops fetalis, chorioamnionitis, and intra-uterine growth restriction with abnormal testing which are all strong predictors of morbidity (Hamilton et al., 2018). The data used for costing of the NICU stay does not provide a true estimate of the incremental effects of morbidities and mortality since it is primarily driven by a per-diem determined by birthweight. Any incremental effects estimated in the cost model thus only truly capture additional drugs, materials, and testing and not additional nursing or allied health support. This is partially addressed by the correlation between morbidities and length of stay (i.e. babies with more complications have longer length

of stay), but a true resource use based model would be preferred. Costing data are further limited by the lack of granularity in terms of severity of morbidities other than IVH which will lead to under-costing of ROP in particular. Studies included in the NMA also provided estimates of the effect of design on late onset sepsis only, but a portion of the sepsis cases included in both the morbidity and cost models would be early onset. The impact of this misclassification is expected to be small, however, based on the balance of early vs late onset sepsis reported in CNN (Canadian Neonatal Follow-Up Network, 2019).

The use of copulas provided an elegant solution to the problem of double counting of treatment effects that would have occurred if length of stay was predicted by morbidities, but these also assume that correlations between morbidities and length of stay do not change across interventions which may not feasible. Future work should assess the stability of estimated correlations in the presence of neonatal interventions using multiple data sources. Despite best efforts to address confounding from observational studies in the NMA, estimates are primarily based on uncontrolled analyses that lack face validity in some cases. Comparisons between RCTs and observational studies provide some reassurance that biases are unlikely to affect conclusions but the opportunity to conduct them was limited. An analysis restricted to RCTs would be preferable but would not address the decision problem since they would only provide sparse evidence across outcomes for single family room.

The economic model required simplifying assumptions in order to incorporate information on transition probabilities for disability over time and to separate societal from healthcare sector costs. Since relative effects between treatments did not differ significantly when a societal or healthcare perspective was taken, the impact of these assumptions on interpretation is expected to be small. Confidence in results is further bolstered by alignment

between NMA results, estimated effects of interventions on morbidities from the meta-model, CEA results, and SMAA conclusions.

Implications for Research

Based on the findings across trials and their synthesis in the NMA, it is unlikely that the decision between design types will benefit from additional randomized or non-randomized trials intended to assess impact on clinical outcomes. An exception to this statement is the forthcoming long-term assessment of the IWK RCT cohort, which will provide the only randomized evidence of the effect of SFR on developmental outcomes. Higher value research should instead focus on either improving estimation of the underlying model or conducting research to better understand the types of trade offs clinicians and parents are willing to make. In addition, while this analysis considers the costs of additional staff as part of the cost of the unit, it does not capture how construction of a new unit can impact nursing workforce planning and the potential impact of training of new staff or new graduate nurses. There is limited research exploring these impacts, but feelings of isolation reported by nurses in single-family rooms may be experienced differently by new staff or new graduates (Doede et al., 2018). Unit design may have further implications for the pace at which new staff can be trained to practice safely without formal supervision.

Estimates for relative effectiveness may be able to be improved through collaboration of the major centres that form the underlying evidence base. The field of individual participant data NMA is rapidly evolving, and data-sharing could allow for a harmonized approach to adjusting for important clinical confounders that may meaningfully reduce bias. With the exception of half-wall designs however, it is unclear whether bias adjustment would have a meaningful impact on conclusions given the substantial overlap already observed in effect estimates. Another

potential avenue for future research would be in adapting the underlying model from being very specific to being something that could be used across a variety of neonatal CEA applications. This would ideally include efforts to improve efficiency to reduce the need to rely on meta-modeling for reasonable run-times. Similarly, while the SMAA process was used as a sensitivity analysis in this analysis it may have substantial clinical appeal given that conclusions are similar to that of the CEA, it is more accessible to code and embed in evidence synthesis and can flexibly include indicators for intangible benefits. Further development of a full SMAA application could consider a formal swing weighting exercise, with exploration of more complicated value functions (Tervonen et al., 2011).

Implications for Nursing Practice

Nurses involved in the process of transitioning to a newly constructed NICU can use this study to help understand the strengths and weaknesses of the underlying evidence and to aid decision making if interpreted cautiously. The most robust finding from this process has been the uncertainty in choice of the optimal design, which suggests that it would be inappropriate to rely heavily on arguments based on clinical or economic benefit. Secondly, the best design may differ depending on the mix of gestational ages that are cared for on a unit, so while clinical benefits are uncertain it may be important to incorporate flexibility if there is a significant number of extremely preterm admissions in a given unit. This, in addition to the recognition that some parents may prefer a traditional open bay model may suggest combined units as an appropriate compromise (Shepley et al., 2014). The decision to choose a given design type may also depend on how often parents are currently present, and the types of communities the NICU is intended to serve. Lower income families were predicted to have poorer outcomes, and even though this did not affect the outcomes of the CEA there is an ethical requirement to consider whether

commitment to a single design type will result in some of these families seeing a disproportionate amount of excess poor outcomes and the additional life and economic consequences that come with them. Lastly, this study does not capture the potential impact of new construction on workforce planning which also requires careful consideration of local resources in terms of access to experienced NICU nurses not only to staff the unit but also to accommodate potential changes to new staff and new graduate training. Any decision to undergo new construction is thus likely to require a unique solution, based on thorough engagement of all relevant stakeholders with a commitment to identify how any changes can potentially harm some families. These are not expected to be considerations that can be informed solely by the results of this analysis, and will need to be further informed by consideration of the qualitative literature in addition to a commitment to learning what is important to the families being cared for in the community served by the planned NICU.

Conclusions

The results of the cost-effectiveness analysis and SMAA suggest that there is substantial uncertainty in the decision of an optimal NICU room design. Half-wall designs have the highest probability of being cost-effective based on these data, but those results are informed by effectiveness results that lack face validity. No other design type can be confidently recommended over the others. The economic analysis is robust to the perspective chosen but is sensitive to the estimates from key efficacy parameters including the effect on length of stay and mortality. The relative cost-effectiveness of different designs varies by gestational age group but is stable across lower and higher income families. A sensitivity analysis using SMAA had similar conclusions to the main analysis and provided evidence that the optimal design choice is sensitive to the preference that decision makers place on intangible benefits of designs such as

the desire to create a family centred environment. A formal analysis of the potential impact of offering reimbursements to lower income families was not pursued as there was no evidence that conclusions would change. The quantitative evidence base for alternative NICU designs related to effectiveness and cost-effectiveness is highly uncertain. The decision on new construction design should be multi-factorial, which may include a consideration of the general pattern of all units showing improved outcomes relative to old open bay in addition to other considerations such as the patient population served by the new unit, family preferences, family centredness, and impacts on health resource planning and training.

References

Aas, K., & Berg, D. (2009). Models for construction of multivariate dependence - a comparison study. *European Journal of Finance*, *15*(7–8), 639–659. https://doi.org/10.1080/13518470802588767

Achana, F. A., Cooper, N. J., Bujkiewicz, S., Hubbard, S. J., Kendrick, D., Jones, D. R., & Sutton, A. J. (2014). Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes. *BMC Medical Research Methodology*, *14*(1), 92. https://doi.org/10.1186/1471-2288-14-92

Aija, A., Toome, L., Axelin, A., Raiskila, S., & Lehtonen, L. (2019). Parents' presence and participation in medical rounds in 11 European neonatal units. *Early Human Development*, *130*(January), 10–16. https://doi.org/10.1016/j.earlhumdev.2019.01.003

Alessio, L. M. D. (2011). *The impact of neonatal icu single family rooms on select developmental outcomes*. Rhode Island College.

Almadhoob, A., & Ohlsson, A. (2015). Sound reduction management in the neonatal intensive care unit for preterm or very low birth weight infants. *The Cochrane Database of Systematic Reviews*, *1*(100909747, 28135055 [pubmed], 10.1002/cam4.1016 [doi]), CD010333. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=prem&NEWS=N&AN=25633155

Als, H. (1982). Toward a synactive theory of development: Promise for the assessment of infant individuality. *Infant Mental Health Journal1*, *3*, 229–243.

Altimier, L. (2011). Mother and Child Integrative Developmental Care Model: A Simple Approach to a Complex Population. *Newborn & Infant Nursing Reviews*, *11*(3), 105–108

4p. http://ezproxy.library.dal.ca/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=104675172&site=ehost-live

Altimier, L., Eichel, M., Warner, B., Tedeschi, L., & Brown, B. (2005). Developmental care: changing the NICU physically and behaviorally to promote patient outcomes and contain costs. In *Neonatal Intensive Care* (Vol. 18, Issue 4, pp. 12–16). http://0-search.ebscohost.com.ilsprod.lib.neu.edu/login.aspx?direct=true&db=c8h&AN=2005118182&site=ehost-live&scope=site

Andersen, E. W. (2005). Two-stage estimation in copula models used in family studies. *Lifetime Data Analysis*, *11*(3), 333–350. https://doi.org/10.1007/s10985-005-2966-7

Askie, L. M., Darlow, B. A., Finer, N., Schmidt, B., Stenson, B., Tarnow-Mordi, W., Davis, P. G., Carlo, W. A., Brocklehurst, P., Davies, L. C., Das, A., Rich, W., Gantz, M. G., Roberts, R. S., Whyte, R. K., Costantini, L., Poets, C., Asztalos, E., Battin, M., … Simes, R. J. (2018). Association between oxygen saturation targeting and death or disability in extremely preterm infants in the neonatal oxygenation prospective meta-analysis collaboration. *JAMA - Journal of the American Medical Association*, *319*(21), 2190–2201. https://doi.org/10.1001/jama.2018.5725

Battersby, C., Michaelides, S., Upton, M., Rennie, J. M., Michaelides, S., & Upton, M. (2017). *Term admissions to neonatal units in England : a role for transitional care ? A retrospective cohort study*. 1–7. https://doi.org/10.1136/bmjopen-2017-016050

Bayley, N. (2006). *Bayley scales of infant and toddler development* (2nd ed). Harcourt Assessment, Inc.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-*

Analysis. https://doi.org/10.1002/9780470743386.fmatter

Briggs, A., Claxton, K., & Sculpher, M. (2006). *Decision modelling for health economics*. Oxford University Press.

Briggs, A., Weinstein, M. C., Fenwick, E. a L., Karnon, J., Sculpher, M. J., & Paltiel, D. (2012). Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. *Medical Decision Making : An International Journal of the Society for Medical Decision Making*, *32*(5), 722–732. https://doi.org/10.1177/0272989X12458348

Brown, P., & Taquino, L. T. (2001). Designing and delivering neonatal care in single rooms. *The Journal of Perinatal & Neonatal Nursing*, *15*(1), 68–83. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=12095008

Bry, A., & Wigert, H. (2019). Psychosocial support for parents of extremely preterm infants in neonatal intensive care: A qualitative interview study. *BMC Psychology*, *7*(1), 1–12. https://doi.org/10.1186/s40359-019-0354-4

Bujkiewicz, S., Achana, F., Papanikos, T., Riley, R. D., Evidence, P., Unit, D. S., & Court, R. (2019). *NICE DSU Technical Support Document 20: Multivariate meta-analysis of summary data for combining treatent effects on correlated outcomes and evaluating surrogate endpoints*. http://nicedsu.org.uk/wp-content/uploads/2019/10/TSD-20-mvmeta-final.pdf

Burke, S. (2018). Systematic review of developmental care interventions in the neonatal intensive care unit since 2006. *Journal of Child Health Care*, *2011*, 136749351775308. https://doi.org/10.1177/1367493517753085

Bürkner, P. C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *R Journal*, *10*(1), 395–411. https://doi.org/10.32614/rj-2018-017

Bürkner, P. C., & Charpentier, E. (2020). Modelling monotonic effects of ordinal predictors in Bayesian regression models. *British Journal of Mathematical and Statistical Psychology*. https://doi.org/10.1111/bmsp.12195

Busse, J. W., Bartlett, S. J., Dougados, M., Johnston, B. C., Guyatt, G. H., Kirwan, J. R., Kwoh, K., Maxwell, L. J., Moore, A., Singh, J. A., Stevens, R., Strand, V., Suarez-Almazor, M. E., Tugwell, P., & Wells, G. a. (2015). Optimal Strategies for Reporting Pain in Clinical Trials and Systematic Reviews: Recommendations from an OMERACT 12 Workshop. *The Journal of Rheumatology*, *42*(10), 1962–1970. https://doi.org/10.3899/jrheum.141440

CADTH. (2015). *Guidance Document on Reporting Indirect Comparisons*. *October*. https://www.cadth.ca

Canadian Agency for Drugs and Technology in Health (CADTH). (2017). *Guidelines for the Economic Evaluation of Health Technologies*.

Canadian Institute for Health Information. (2011). *Patient Costing Methodology - MIS Supplement - Standards/Data Submission*.

Canadian Institute for Health Information. (2019). *Canadian Patient Cost Database Technical Document: MIS Patient Costing Methodology, January 2019* (Issue January). Canadian Institute for Health Information. https://www.cihi.ca/sites/default/files/document/mis_patient_cost_meth_en_0.pdf

Canadian Neonatal Follow-Up Network. (2019). *Annual Report 2019*.

Canadian Neonatal Network. (2016). *The Canadian Neonatal Network ™ Annual Report*.

Canadian Neonatal Network. (2019). *2019 CNN report final*.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Chabaud, F., David-Tchouda, S., Belin, V., Fau, S., Equy, V., Carraby, S., & Debillon, T. (2012). Influence of hospital location on short-term fate of premature infants born at 34 weeks of gestation. *Influence Du Lieu d'hospitalisation Sur Le Devenir a Court Terme Des Prematures Nes a 34 Semaines d'amenorrhee.*, *19*(4), 391–395. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med7&NEWS=N&AN=22377246

Chen, Y.-C., Lin, C.-F., Rehn, Y.-J. F., Chen, J.-C., Chen, P.-Y., Chen, C.-H., Wang, T.-M., & Huang, F.-L. (2017). Reduced nosocomial infection rate in a neonatal intensive care unit during a 4-year surveillance period. *Journal of the Chinese Medical Association : JCMA*, 4–8. https://doi.org/10.1016/j.jcma.2017.02.006

Claxton, K. (1999). The irrelevance of inference: A decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics*, *18*(3), 341–364. https://doi.org/10.1016/S0167-6296(98)00039-3

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, *83*(403), 596–610. https://doi.org/10.1080/01621459.1988.10478639

Conde-Aguedelo, A., & Díaz-rossello, J. L. (2016). Kangaroo mother care to reduce morbidity and mortality in low birthweight infants (Review). *Cochrane Database of Systematic Reviews*, *8*, 1–149. https://doi.org/10.1002/14651858.CD002771.pub4.www.cochranelibrary.com

Coughlin, M., Gibbins, S., & Hoath, S. (2009). Core measures for developmentally supportive care in neonatal intensive care units: theory, precedence and practice. *Journal of Advanced Nursing*, *65*(10), 2239–2248. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med6&NEWS=N&AN=19686402

Davis, S., Stevenson, M., Tappenden, P., & Wailoo, A. J. (2014). Cost-effectiveness modelling using patient-level simulation. *NICE DSU Technical Support Document*, *15*(April), 1–62. http://www.nicedsu.org.uk/TSD15_Patient-level_simulation.pdf

Degeling, K., Ijzerman, M. J., Koopman, M., & Koffijberg, H. (2017). Accounting for parameter uncertainty in the definition of parametric distributions used to describe individual patient variation in health economic models. *BMC Medical Research Methodology*, *17*(1), 1–12. https://doi.org/10.1186/s12874-017-0437-y

Dias, S., Ades, A. E., Welton, N. J., Jansen, J. P., & Sutton, A. J. (2018). *Network Meta-Analysis for Decision-Making*. Wiley.

Dias, S., Welton, N. J., Sutton, A. J., & Ades, A. (2011). NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. In *Tropical Medicine* (Issue 0).

Doede, M., Trinkoff, A. M., & Gurses, A. P. (2018). Neonatal Intensive Care Unit Layout and Nurses' Work. *Health Environments Research and Design Journal*, *11*(1), 101–118. https://doi.org/10.1177/1937586717713734

Domanico, R., Davis, D. K., Coleman, F., & Davis, B. O. (2011). Documenting the NICU design dilemma: comparative patient progress in open-ward and single family room units. *Journal of Perinatology : Official Journal of the California Perinatal Association*, *31*(4), 281–288.

https://doi.org/10.1038/jp.2010.120

Drummond, M., Sculpher, M., Claxton, K., Stoddart, G., & Torrance, G. (2015). *Methods for the economic evaluation of health care programms* (Fourth Edi). Oxford University Press.

Efthimiou, O., Mavridis, D., Debray, T., Samara, M., Leucht, S., Belger, M., & Salanti, G. (2016). Combining randomized and non-randomized evidence in network meta-analysis. *Statistics in Medicine*.

Efthimiou, O., & White, I. R. (2019). The dark side of the force: multiplicity issues in network meta-analysis and how to address them. *Research Synthesis Methods*, *October 2018*, 1–18. https://doi.org/10.1002/jrsm.1377

Erdeve, O., Arsan, S., Yigit, S., Armangil, D., Atasay, B., & Korkmaz, A. (2008). The impact of individual room on rehospitalization and health service utilization in preterms after discharge. *Acta Paediatrica, International Journal of Paediatrics*, *97*(10), 1351–1357. https://doi.org/10.1111/j.1651-2227.2008.00889.x

Erickson, C., Kattelmann, K., Remington, J., Ren, C., Helseth, C., & Stevens, D. (2011). Traditional open-bay versus single-family room neonatal intensive care unit: a comparison of selected nutrition outcomes. *Research and Reports in Neonatology*, *1*, 15–20. https://doi.org/10.2147/RRN.S15539

Fawcett, J. (1984). The metaparadigm of nursing: Present status and future refinements. *The Journal of Nursing Scholarship*, *16*(3), 84–87.

Feeley, N., Robins, S., Genest, C., Stremler, R., Zelkowitz, P., & Charbonneau, L. (2020). A comparative study of mothers of infants hospitalized in an open ward neonatal intensive care unit and a combined pod and single-family room design. *BMC Pediatrics*, *20*(1), 1–9. https://doi.org/10.1186/s12887-020-1929-1

Flacking, R., Thomson, G., Ekenberg, L., Lowegren, L., & Wallin, L. (2013). Influence of NICU co-care facilities and skin-to-skin contact on maternal stress in mothers of preterm infants. *Sexual & Reproductive Healthcare : Official Journal of the Swedish Association of Midwives*, *4*(3), 107–112. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=24041731

Friedrich, J. O., Adhikari, N. K. J., & Beyene, J. (2008). The ratio of means method as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis: A simulation study. *BMC Medical Research Methodology*, *8*. https://doi.org/10.1186/1471-2288-8-32

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2017). *Visualization in Bayesian workflow*. *2*. http://arxiv.org/abs/1709.01449

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2015). Bayesian data analysis. In *CEUR Workshop Proceedings* (Third Edit, Vol. 1542, Issue 9). CRC Press. https://doi.org/10.1017/CBO9781107415324.004

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y. S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics*, *2*(4), 1360–1383. https://doi.org/10.1214/08-AOAS191

Gelman, A., & Stern, H. (2006). The Difference Between “Significant” and “Not Significant” is not Itself Statistically Significant. *The American Statistician*, *60*(4), 328–331. https://doi.org/10.1198/000313006X152649

Gibbins, S., Hoath, S. B., Coughlin, M., Gibbins, A., & Franck, L. (2008). Foundations in newborn care. The universe of developmental care: a new conceptual model for application

in the neonatal intensive care unit. *Advances in Neonatal Care (Elsevier Science)*, *8*, 141–147. https://doi.org/10.1097/01.ANC.0000342754.18568.d5

Goldmann, D., Durbin, W., & Freeman, J. (1981). Nosocomial Infections in a Neonatal Intensive Care Unit. *The Journal of Infectious Diseases*, *144*(5), 449–459.

Hamilton, E. F., Dyachenko, A., Ciampi, A., Maurel, K., Warrick, P. A., & Garite, T. J. (2018). Estimating risk of severe neonatal morbidity in preterm births under 32 weeks of gestation. *Journal of Maternal-Fetal and Neonatal Medicine*, *0*(0), 1–8. https://doi.org/10.1080/14767058.2018.1487395

Haslam, M. D., Lisonkova, S., Creighton, D., Church, P., Yang, J., Shah, P. S., Joseph, K. S., Synnes, A., Harrison, A., Ting, J., Cieslak, Z., Sherlock, R., Yee, W., Fajardo, C., Aziz, K., Toye, J., Kalapesi, Z., Sankaran, K., Daspal, S., … Murphy, P. (2018). Severe Neurodevelopmental Impairment in Neonates Born Preterm: Impact of Varying Definitions in a Canadian Cohort. *Journal of Pediatrics*, *197*, 75–81.e4. https://doi.org/10.1016/j.jpeds.2017.12.020

Haumont, D. (2013). The environment of the neonatal intensive care unit: Does it make a difference? *Journal of Perinatal Medicine*, *41*(s1). https://doi.org/10.1515/jpm-2013-2001

Hernán, M. A., & Robins, J. M. (2020). *Causal inference: What if*. Chapman & Hall/CRC. https://doi.org/10.1001/jama.1989.03420150114051

Higgins, J. P. T., & Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions* (Version 5.). The Cochrane Collaboration. www.cochrane-handbook.org

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Hoff, P. D. (2007). Extending the rank likelihood for semiparametric copula estimation. *The*

Annals of Applied Statistics, *1*(1), 265–283. https://arxiv.org/pdf/math/0610413.pdf

Hollingsead, A. (1975). *Four Factor Index of Social Status*. Yale University.

Hourigan, S. K., Subramanian, P., Hasan, N. A., Ta, A., Klein, E., Chettout, N., Huddleston, K., Deopujari, V., Levy, S., Baveja, R., Clemency, N. C., Baker, R. L., Niederhuber, J. E., & Colwell, R. R. (2018). Comparison of infant gut and skin microbiota, resistome and virulome between neonatal intensive care unit (NICU) environments. *Frontiers in Microbiology*, *9*(JUN), 1–13. https://doi.org/10.3389/fmicb.2018.01361

Hu, M., Pavlicoa, M., & Nunes, E. (2011). Zero-inflated and hurdle models of count data with extra zeros: Examples from an HIV-risk reduction intervention trial. *Am J Drug Alcohol Abuse*, *37*(5), 367–375. https://doi.org/10.3109/00952990.2011.597280.Zero-inflated

Jefferies, A., & Canadian Paediatric Society: Fetus and Newborn Committee. (2016). Retinopathy of prematurity: An update on screening and management. *Paediatr Child Health*, *21*(2), 101-04. https://doi.org/10.1016/s0031-3955(02)00111-6

Joanna Briggs Institute. (2017a). Checklist for Cohort Studies. *The Joanna Briggs Institute Critical Appraisal Tools for Use in JBI Systematic Reviews*, 1–7. https://joannabriggs.org/ebp/critical_appraisal_tools

Joanna Briggs Institute. (2017b). Checklist for Randomized Controlled Trials. *Joanna Briggs Institute*, 1–9. https://doi.org/10.1371/journal.pgen.1000960

Johnson, T., Patel, A., Jegier, B., Engstrom, J., & Meier, P. (2012). *The cost of morbidities in very low birth weight infants*. *100*(2), 130–134. https://doi.org/10.1016/j.pestbp.2011.02.012.Investigations

Johnston, K. M., Gooch, K., Korol, E., Vo, P., Eyawo, O., Bradt, P., Levy, A., & Johnston K, Gooch K, Korol E, Vo P, Eyawo O, Bradt P, L. a. (2014). The economic burden of

prematurity in Canada. *BMC Pediatrics*, *14*(1), 93. https://doi.org/10.1186/1471-2431-14-93

Jones, A. R., Kuschel, C., Jacobs, S., & Doyle, L. W. (2012). Reduction in late-onset sepsis on relocating a neonatal intensive care nursery. *Journal of Paediatrics and Child Health*, *48*(10), 891–895. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med7&NEWS=N&AN=22897216

Jorge, E. C., Jorge, E. N., & El Dib, R. P. (2013). Early light reduction for preventing retinopathy of prematurity in very low birth weight infants. *Cochrane Database of Systematic Reviews*, *2013*(8). https://doi.org/10.1002/14651858.CD000122.pub2

Julian, S., Burnham, C.-A. D., Sellenriek, P., Shannon, W. D., Hamvas, A., Tarr, P. I., & Warner, B. B. (2015). Impact of neonatal intensive care bed configuration on rates of late-onset bacterial sepsis and methicillin-resistant staphylococcus aureus colonization. *Infection Control & Hospital Epidemiology*, *36*(10), 1–10. https://doi.org/10.1017/ice.2015.144

Kaltenthaler, E., Tappenden, P., Paisley, S., & Squires, H. (2011). *NICE DSU Technical Support Document 13: Identifying and reviewing evidence to inform the conceptualisation and population of cost-effectiveness models*. *0*, 72. http://www.nicedsu.org.uk

Karimi, H., & Alavi, N. M. (2015). Florence Nightingale : The Mother of Nursing. *Nurs Midwifery Stud*, *4*(2), 1–3.

Korvenranta, E., Linna, M., Rautava, L., Andersson, S., Gissler, M., Hallman, M., Hakkinen, U., Leipala, J., Peltola, M., Tammela, O., & Lehtonen, L. (2010). Hospital costs and quality of life during 4 years after very preterm birth. *Archives of Pediatrics and Adolescent Medicine*, *164*(1538–3628 (Electronic)), 657–663.

Krijkamp, E. M., Alarid-Escudero, F., Enns, E. A., Jalal, H. J., Hunink, M. G. M., &

Pechlivanoglou, P. (2018). Microsimulation modeling for health decision sciences using R: A tutorial. *Medical Decision Making*, *38*(3), 400–422. https://doi.org/10.1177/0272989X18754513

Kuhlthau, K., Kahn, R., Hill, K. S., Gnanasekaran, S., & Ettner, S. L. (2010). The well-being of parental caregivers of children with activity limitations. *Maternal and Child Health Journal*, *14*(2), 155–163. https://doi.org/10.1007/s10995-008-0434-1

Larson, E., Hargiss, C., & Dyk, L. (1985). Effect of an expanded physical facility on nosocomial infections in a neonatal intensive care unit. *American Journal of Infection Control1*, *13*(1), 16–20.

Lee, P. H. (2017). Examining the error of mis-specifying nonlinear confounding effect with application on accelerometer-measured physical activity. *Research Quarterly for Exercise and Sport*, *88*(2), 203–208. https://doi.org/10.1080/02701367.2017.1296101

Lester, B., Hawes, K., Abar, B., Sullivan, M., Miller, R., Bigsby, R., Laptook, A., Salisbury, A., Taub, M., Lagasse, L. L., & Padbury, J. F. (2014). Single-family room care and neurobehavioral and medical outcomes in preterm infants. *Pediatrics*, *134*(4), 754–760. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=25246623

Lester, B., Miller, R. J., Hawes, K., Salisbury, A., Bigsby, R., Sullivan, M. C., Padbury, J. F., Abar, B., Sullivan, M. C., Miller, R. J., Bigsby, R., Laptook, A., Salisbury, A., Taub, M., Lagasse, L. L., & Padbury, J. F. (2011). Infant neurobehavioral development. *Seminars in Perinatology*, *35*(1), 8–19. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med7&NEWS=N&AN=21255702

Lester, B., Salisbury, A. L., Hawes, K., Dansereau, L. M., Bigsby, R., Laptook, A., Taub, M., Lagasse, L. L., Vohr, B. R., & Padbury, J. F. (2016). 18-month follow-up of infants cared for in a single-family room neonatal intensive care unit. *The Journal of Pediatrics*, *177*(jlz, 0375410, 28130942 [pubmed], 10.1111/mec.13956 [doi], 28135055 [pubmed], 10.1002/cam4.1016 [doi]), 84–89. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=prem&NEWS=N&AN=27470693

Macho, P. (2017). Individualized developmental care in the NICU. *Advances in Neonatal Care*, *17*(3), 162–174. https://doi.org/10.1097/ANC.0000000000000374

Mangham, L. J., Petrou, S., Doyle, L. W., Draper, E. S., & Marlow, N. (2009). The cost of preterm birth throughout childhood in England and Wales. *Pediatrics*, *123*(2), e312-27. https://doi.org/10.1542/peds.2008-1827

McGrath, S., Sohn, H., Steele, R., & Benedetti, A. (2020). Meta-analysis of the difference of medians. *Biometrical Journal*, *62*(1), 69–98. https://doi.org/10.1002/bimj.201900036

Meredith, J. L. (2017). *The NICU Environment : Infusing Single-Family Room Benefits into the Open-Bay Setting*. *36*(2), 69–76.

Microsoft Corportation. (2019). *Excel*. Microsoft Corportation.

Milford, C. A., Zapalo, B., & Davis, G. (2008). Transition to an individual-room nicu design: Process and outcome measures. *Neonatal Network*, *27*(5).

Monson, B. B., Eaton-Rosen, Z., Kapur, K., Liebenthal, E., Brownell, A., Smyser, C. D., Rogers, C. E., Inder, T. E., Warfield, S. K., & Neil, J. J. (2018). Differential rates of perinatal maturation of human primary and nonprimary auditory cortex. *ENeuro*, *5*(1), 1–12. https://doi.org/10.1523/ENEURO.0380-17.2017

Morag, I., & Ohlsson, A. (2011). Cycled light in the intensive care unit for preterm and low birth weight infants. *The Cochrane Database of Systematic Reviews*, *1*. https://doi.org/10.1002/14651858.CD006982.pub4.www.cochranelibrary.com

Morag, I., & Ohlsson, A. (2017). Cycled light in the intensive care unit for preterm and low birth weight infants. *The Cochrane Database of Systematic Reviews*, *1*. https://doi.org/10.1002/14651858.CD006982.pub4.www.cochranelibrary.com

Nassel, D., Chartrand, C., Doré-Bergeron, M. J., Lefebvre, F., Ballantyne, M., Van Overmeire, B., & Luu, T. M. (2019). Very preterm infants with technological dependence at home: Impact on resource use and family. *Neonatology*, *115*(4), 363–370. https://doi.org/10.1159/000496494

Nguyen, T.-N.-N., Spencer-Smith, M., Zannino, D., Burnett, A., Scratch, S. E., Pascoe, L., Ellis, R., Cheong, J., Thompson, D., Inder, T., Doyle, L. W., & Anderson, P. J. (2018). *Developmental Trajectory of Language From 2 to 13 Years in Children Born Very Preterm*. *141*(5). https://doi.org/10.1542/peds.2017-2831

Nightingale, F. (1974). *Notes on nursing: What it is and what it is not* (2nd ed). Blackie & Son Ltd. (Original work published 1859).

O'Brien, K., Bracht, M., Robson, K., Ye, X. Y., Mirea, L., Cruz, M., Ng, E., Monterrosa, L., Soraisham, A., Alvaro, R., Narvey, M., Da Silva, O., Lui, K., Tarnow-Mordi, W., & Lee, S. K. (2015). Evaluation of the Family Integrated Care model of neonatal intensive care: a cluster randomized controlled trial in Canada and Australia. *BMC Pediatrics*, *15*(1), 210. https://doi.org/10.1186/s12887-015-0527-0

O'Brien, K., Robson, K., Bracht, M., Cruz, M., Lui, K., Alvaro, R., da Silva, O., Monterrosa, L., Narvey, M., Ng, E., Soraisham, A., Ye, X. Y., Mirea, L., Tarnow-Mordi, W., & Lee, S. K.

(2018). Effectiveness of Family Integrated Care in neonatal intensive care units on infant and parent outcomes: a multicentre, multinational, cluster-randomised controlled trial. *The Lancet Child and Adolescent Health*, *2*(4), 245–254. https://doi.org/10.1016/S2352-4642(18)30039-7

Ohlsson, A., & Jacobs, S. E. (2013). NIDCAP: a systematic review and meta-analyses of randomized controlled trials. *Pediatrics*, *131*(3), e881-93. https://doi.org/10.1542/peds.2012-2121

Örtenstrand, A. (2014). The role of single-patient neonatal intensive care unit rooms for preterm infants. *Acta Paediatrica (Oslo, Norway : 1992)*, *103*(5), 462–463. https://doi.org/10.1111/apa.12566

Ortenstrand, A., Westrup, B., Broström, E. B., Sarman, I., Akerström, S., Brune, T., Lindberg, L., & Waldenström, U. (2010). The Stockholm neonatal family centered care study: effects on length of stay and infant morbidity. *Pediatrics*, *125*(2), e278-85. https://doi.org/10.1542/peds.2009-1511

Palisano, R., Rosenbaum, P., Walter, S., Russell, D., Wood, E., & Galuppi, B. (1997). Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Developmenal Medicine & Child Neurology*, *39*, 214–233. https://doi.org/10.1007/978-1-4471-5451-8_152

Pennington, B., & Wong, R. (2019). Modeling carer health-related quality of life in NICE technology appraisals and highly specialised technologies. In *Decision Support Unit*. https://doi.org/10.1016/j.soncn.2009.11.006

Petrou, S., Krabuanrat, N., & Khan, K. (2019). Preference-based health-related quality of life outcomes associated with preterm birth: A systematic review and meta-analysis.

PharmacoEconomics. https://doi.org/10.1007/s40273-019-00865-7

Piironen, J., & Vehtari, A. (2017a). Comparison of Bayesian predictive methods for model selection. *Statistics and Computing*, *27*(3), 711–735. https://doi.org/10.1007/s11222-016-9649-y

Piironen, J., & Vehtari, A. (2017b). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, *11*(2), 5018–5051. https://doi.org/10.1214/17-EJS1337SI

Pineda, R., Durant, P., Mathur, A., Inder, T., Wallendorf, M., & Schlaggar, B. L. (2017). Auditory exposure in the neonatal intensive care unit: Room type and other predictors. *The Journal of Pediatrics*, *183*, 56–66.e3. https://doi.org/10.1016/j.jpeds.2016.12.072

Pineda, R., Guth, R., Herring, A., Reynolds, L., Oberle, S., & Smith, J. (2017). Enhancing sensory experiences for very preterm infants in the NICU: An integrative review. *Journal of Perinatology*, *37*(4), 323–332. https://doi.org/10.1038/jp.2016.179

Pineda, R., Neil, J., Dierker, D., Smyser, C., Wallendorf, M., Kidokoro, H., Reynolds, L., Walker, S., Rogers, C., Mathur, A., Van Essen, D., & Inder, T. (2014). Alterations in brain structure and neurodevelopmental outcome in preterm infants hospitalized in different neonatal intensive care unit environments. *The Journal of Pediatrics*, *164*(1), 52–60.e2. https://doi.org/10.1016/j.jpeds.2013.08.047

Pineda, R., Stransky, K., Rogers, C., Duncan, M., HensleySmith, G., & Neil, J. (2013). The single patient room in the NICU: Maternal and family effects. *Journal of Perinatology*, *32*(7), 545–551. https://doi.org/10.1038/jp.2011.144.The

Pineda, R., Stransky, K., Rogers, C., Duncan, M., Smith, G., Neil, J., & Inder, T. (2012). The single-patient room in the NICU: maternal and family effects. *Journal of Perinatology :*

Official Journal of the California Perinatal Association, *32*(7), 545–551. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med7&NEWS=N&AN=22031044

Plummer, M. (2003). DSC 2003 Working Papers JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Conference on Distributed Statistical Computing*, 1–10. http://www.ci.tuwien.ac.at/Conferences/DSC-2003/

Poley, M. J., Brouwer, W. B. F., Van Exel, N. J. A., & Tibboel, D. (2012). Assessing health-related quality-of-life changes in informal caregivers: An evaluation in parents of children with major congenital anomalies. *Quality of Life Research*, *21*(5), 849–861. https://doi.org/10.1007/s11136-011-9991-7

Puumala, S. E., Rich, R. K., Roy, L., Reynolds, R., Jimenez, F. E., Opollo, J. G., & Brittin, J. (2020). Single-family room neonatal intensive care unit design : do patient outcomes actually change ? *Journal of Perinatology*. https://doi.org/10.1038/s41372-019-0584-6

R Core Team. (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://cran.r-project.org/

Raiskila, S., Axelin, A., Toome, L., Caballero, S., Tandberg, B. S., Montirosso, R., Normann, E., Hallberg, B., Westrup, B., Ewald, U., & Lehtonen, L. (2017). Parents' presence and parent-infant closeness in 11 neonatal intensive care units in six European countries vary between and within the countries. *Acta Paediatrica*, 1–11. https://doi.org/10.1111/apa.13798

Raiskila, S., Lehtonen, L., Tandberg, B. S., Normann, E., Ewald, U., Caballero, S., Varendi, H., Toome, L., Nordhøv, M., Hallberg, B., Westrup, B., Montirosso, R., & Axelin, A. (2016). Parent and nurse perceptions on the quality of family-centred care in 11 European NICUs.

Australian Critical Care, *29*(4), 201–209. https://doi.org/10.1016/j.aucc.2016.09.003

Riley, R. D., Snell, K. I. E., Ensor, J., Burke, D. L., Harrell, F. E., Moons, K. G. M., & Collins, G. S. (2019). Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Statistics in Medicine*, *38*(7), 1276–1296. https://doi.org/10.1002/sim.7992

Roberts, D., Brown, J., Medley, N., & Dalziel, S. R. (2017). Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. *Cochrane Database of Systematic Reviews*, *2017*(3). https://doi.org/10.1002/14651858.CD004454.pub3

Rogers, C. H., Floyd, F. J., Seltzer, M. M., Greenberg, J., & Hong, J. (2008). Long-term effects of the death of a child on parents' adjustment in midlife. *Journal of Family Psychology*, *22*(2), 203–211. https://doi.org/10.1037/0893-3200.22.2.203

Sadatsafavi, H., Niknejad, B., Shepley, M., & Sadatsafavi, M. (2017). Probabilistic return-on-investment analysis of single-family versus open-bay rooms in neonatal intensive care units-synthesis and evaluation of early evidence on nosocomial infections, length of stay, and direct cost of care. *Journal of Intensive Care Medicine*, *bhs*, *8610344*, 885066616689774. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medp&NEWS=N&AN=28118769

Saigal, S., Ferro, M. A., Van Lieshout, R. J., Schmidt, L. A., Morrison, K. M., & Boyle, M. H. (2016). Health-related quality of life trajectories of extremely low birth weight survivors into adulthood. *Journal of Pediatrics*, *179*, 68–73.e1. https://doi.org/10.1016/j.jpeds.2016.08.018

Schmitz, S., Adams, R., & Walsh, C. (2013). Incorporating data from various trial designs into a

mixed treatment comparison model. *Statistics in Medicine*, *32*(17), 2935–2949. https://doi.org/10.1002/sim.5764

Schonhaut, L., Pérez, M., Armijo, I., & Maturana, A. (2020). Comparison between Ages & Stages Questionnaire and Bayley Scales, to predict cognitive delay in school age. *Early Human Development*, *141*(August 2019), 104933. https://doi.org/10.1016/j.earlhumdev.2019.104933

Shah, P., Dunn, M., Aziz, K., Shah, V., Deshpandey, A., Mukerji, A., Ng, E., Mohammad, K., Ulrich, C., Amaral, N., Lemyre, B., Synnes, A., Piedboeuf, B., Yee, W. H., Ye, X. Y., & Lee, S. K. (2019). Sustained quality improvement in outcomes of preterm neonates with a gestational age less than 29 weeks: Results from the Evidence-based Practice for Improving Quality Phase 3 1. *Canadian Journal of Physiology and Pharmacology*, *97*(3), 213–221. https://doi.org/10.1139/cjpp-2018-0439

Shahheidari, M., & Homer, C. (2012). Impact of the design of neonatal intensive care units on neonates, staff, and families: a systematic literature review. *The Journal of Perinatal & Neonatal Nursing*, *26*(3), 260–268. https://doi.org/10.1097/JPN.0b013e318261ca1d

Shepley, M. M., Smith, J. A., Sadler, B. L., & White, R. D. (2014). The business case for building better neonatal intensive care units. *Journal of Perinatology : Official Journal of the California Perinatal Association*, *34*(11), 811–815. https://doi.org/10.1038/jp.2014.174

Siebert, U. (2012). State-transition modeling: A report of the ISPOR-SMDM modeling good research practices task force-3. *Value in Health*, *15*, 812– 820 Available. https://doi.org/10.1016/j.jval.2012.06.012

Smith, W. (2018). Concept analysis of family-centered care of hospitalized pediatric patients. *Journal of Pediatric Nursing*, *42*, 57–64. https://doi.org/10.1016/j.pedn.2018.06.014

Smithgall, L. (2010). *Perceptions of Maternal Stress and Neonatal Patient Outcomes in a Single Private Room Versus Open Room Neonatal Intensive Care Unit Environment A dissertation presented to the faculty of the College of Nursing East Tennessee State University In partial fu*. *December*.

Smithgall, L. (2011). Assessment of neonatal intensive care unit single private room versus open room environment and the impact on maternal and neonatal outcomes. *JOGNN: Journal of Obstetric, Gynecologic & Neonatal Nursing*, *40*, S74–S74. https://doi.org/10.1111/j.1552-6909.2011.01242_103.x

Song, J., Floyd, F., Seltzer, M., Greenberg, J., & Hong, J. (2011). Long-term effects of child death on parents. *Fam Relat*, *59*(3), 269–282. https://doi.org/10.1111/j.1741-3729.2010.00601.x.Long-term

Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., & Way, R. (2012). *OpenBUGS User Manual Contents*. *July*.

Statistics Canada. (2016). Preterm live births in Canada, 2000 to 2013. *Health Fact Sheets*, *Chart 1*, 5. http://www.statcan.gc.ca/pub/82-625-x/2016001/article/14675-eng.htm

Statistics Canada. (2017). *Distribution of total income by census family type and age of older partner, parent, or individual*. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1110001201

Statistics Canada. (2018). *Educational attainment of the population aged 25 to 64, by age group and sex, Organisation for Economic Co-operation and Development (OECD), Canada, provinces and territories*. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3710013001

Stevens, D., Helseth, C. C., Khan, M. A., Munson, D. P., & Reid, E. J. (2011). A comparison of

parent satisfaction in an open-bay and single-family room neonatal intensive care unit. *HERD*, *4*(3), 110–123. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med7&NEWS=N&AN=21866508

Stevens, D., Helseth, C. C., Thompson, P. A., Pottala, J. V, Khan, M. A., & Munson, D. P. (2012). A comprehensive comparison of open-bay and single-family-room neonatal intensive care units at Sanford Children's Hospital. *HERD*, *5*(4), 23–39. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med7&NEWS=N&AN=23224804

Stevens, D., Thompson, P. A., Helseth, C. C., Hsu, B., Khan, M. A., & Munson, D. P. (2014). A comparison of the direct cost of care in an open-bay and single-family room NICU. *Journal of Perinatology*, *34*(11), 830–835. https://doi.org/10.1038/jp.2014.178

Stevens, D., Thompson, P. A., Helseth, C. C., Pottala, J. V, Khan, M. A., & Munson, D. P. (2011). A comparison of outcomes of care in an open-bay and single-family room neonatal intensive care facility. *Journal of Neonatal-Perinatal Medicine*, *4*(3), 189–200. https://doi.org/10.3233/npm-2011-2762

Stevens, D., Thompson, P., Helseth, C., & Pottala, J. (2015). Mounting evidence favoring single-family room neonatal intensive care. *Journal of Neonatal-Perinatal Medicine*, *8*(3), 177–178. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=26485558

Stoll, B. J., Hansen, N. I., Bell, E. F., Walsh, M. C., Carlo, W. a., Shankaran, S., Laptook, A. R., Sánchez, P. J., Van Meurs, K. P., Wyckoff, M., Das, A., Hale, E. C., Ball, M. B., Newman,

N. S., Schibler, K., Poindexter, B. B., Kennedy, K. a., Cotten, C. M., Watterberg, K. L., … Higgins, R. D. (2015). Trends in care practices, morbidity, and mortality of extremely preterm neonates, 1993-2012. *Jama*, *314*(10), 1039. https://doi.org/10.1001/jama.2015.10244

Strauss, D., Brooks, J., Rosenbloom, L., & Shavelle, R. (2008). Life expectancy in cerebral palsy: An update. *Developmental Medicine and Child Neurology*, *50*(7), 487–493. https://doi.org/10.1111/j.1469-8749.2008.03000.x

Strickler, J. (2017). Florence Nightingale: Lighting the way for the future of nursing. *Nursing2017*, *43*, 43–45.

Subramaniam, P., Ho, J., & Davis, P. (2016). Prophylactic nasal continuous positive airway pressure for preventing morbidity and mortality in very preterm infants. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD001243.pub3

Sweeting, M. J., Sutton, A. J., & Lambert, P. C. (2004). What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine*, *23*(9), 1351–1375. https://doi.org/10.1002/sim.1761

Symington, A., & Pinelli, J. (2006). Developmental care for promoting development and preventing morbidity in preterm infants. *Cochrane Database of Systematic Reviews (Online)*, *2*, CD001814. https://doi.org/10.1002/14651858.CD001814.pub2

Synnes, A., Luu, T. M., Moddemann, D., Church, P., Lee, D., Vincer, M., Ballantyne, M., Majnemer, A., Creighton, D., Yang, J., Saigal, S., Shah, P. S., Lee, S. K., Shah, P. S., Harrison, A., Ting, J., Yee, W., Aziz, K., Toye, J., … Sauve, R. (2017). Determinants of developmental outcomes in a very preterm Canadian cohort. *Archives of Disease in Childhood: Fetal and Neonatal Edition*, *102*(3), F235–F243.

https://doi.org/10.1136/archdischild-2016-311228

Tandberg, B. S., Flacking, R., Markestad, T., Grundt, H., & Moen, A. (2019). Parent psychological wellbeing in a single-family room versus an open bay neonatal intensive care unit. *PLoS ONE*, *14*(11), 1–18. https://doi.org/10.1371/journal.pone.0224488

Tandberg, B. S., Frøslie, K. F., Flacking, R., Grundt, H., Lehtonen, L., & Moen, A. (2018). Parent-Infant Closeness, Parents' Participation, and Nursing Support in Single-Family Room and Open Bay NICUs. *Journal of Perinatal and Neonatal Nursing*, *32*(4), E22–E32. https://doi.org/10.1097/JPN.0000000000000359

Tandberg, B. S., Frøslie, K. F., Markestad, T., Flacking, R., Grundt, H., & Moen, A. (2019). Single-family room design in the neonatal intensive care unit did not improve growth. *Acta Paediatrica, International Journal of Paediatrics*, *108*(6), 1028–1035. https://doi.org/10.1111/apa.14746

Tandberg, B. S., Markestad, T., Flacking, R., & Moen, A. (2014). Impact of family centred care on nurses' and parents' assessment of the quality of care in neonatal intensive care. *Archives of Disease in Childhood*, *99*((Tandberg B.) Pediatric, Vestre Viken HT, Drammen, Norway), A75. https://doi.org/10.1136/archdischild-2014-307384.197

Tervonen, T., Naci, H., van Valkenhoef, G., Ades, A. E., Angelis, A., Hillege, H. L., & Postmus, D. (2015). Applying Multiple Criteria Decision Analysis to Comparative Benefit-Risk Assessment. *Medical Decision Making*, *35*(7), 859–871. https://doi.org/10.1177/0272989X15587005

Tervonen, T., van Valkenhoef, G., Buskens, E., Hillege, H. L., & Postmus, D. (2011). A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis. *Statistics in Medicine*, *30*(12), 1419–1428. https://doi.org/10.1002/sim.4194

The Joanna Briggs Institute. (2017). JBI: Checklist for quasi-experimental studies. *The Joanna Briggs Institute*, 1–6. https://doi.org/10.3389/fneur.2015.00202

Turner, R. M., Jackson, D., Wei, Y., Thompson, S. G., & Higgins, J. P. T. (2015). Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Statistics in Medicine*, *34*(6), 984–998. https://doi.org/10.1002/sim.6381

Umlauf, N., & Kneib, T. (2018). A primer on Bayesian distributional regression. *Statistical Modelling*, *18*(3–4), 219–247. https://doi.org/10.1177/1471082X18759140

van Veenendaal, N. R., Heideman, W. H., Limpens, J., van der Lee, J. H., van Goudoever, J. B., van Kempen, A. A. M. W., & van der Schoor, S. R. D. (2019). Hospitalising preterm infants in single family rooms versus open bay units: a systematic review and meta-analysis. *The Lancet Child and Adolescent Health*, *3*(3), 147–157. https://doi.org/10.1016/S2352-4642(18)30375-4

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2019). *Rank-normalization, folding, and localization: An improved Rhat for assessing convergence of MCMC*. 1–23. http://arxiv.org/abs/1903.08008

Verde, P. E., & Ohmann, C. (2015). Combining randomized and nonrandomized evidence in clinical research: A review of methods and applications. *Research Synthesis Methods*, *6*(1), 45–62. https://doi.org/10.1002/jrsm.1122

Vohr, B., McGowan, E., McKinley, L., Tucker, R., Keszler, L., & Alksninis, B. (2017). Differential effects of the single-family room neonatal intensive care unit on 18- to 24-month Bayley Scores of preterm infants. *Journal of Pediatrics*, *185*, 42–48.e1. https://doi.org/10.1016/j.jpeds.2017.01.056

Von Dolinger de Brito, D., Almeida Silva, H., Jose Oliveira, E., Arantes, A., Abdallah, V. O. S., Tannus Jorge, M., & Gontijo Filho, P. P. (2007). Effect of neonatal intensive care unit environment on the incidence of hospital-acquired infection in neonates. *Journal of Hospital Infection*, *65*(4), 314–318. http://ezproxy.library.dal.ca/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=106125451&site=ehost-live

Walsh, M., & Kliegman, R. (1986). Necrotizing enterocolitis: Treatment based on staging criteria. *Pediatrics Clinics of North America*, *33*(1), 180–201.

Walsh, W., McCullough, K. L., & White, R. D. (2006). Room for improvement: nurses' perceptions of providing care in a single room newborn intensive care setting. *Advances in Neonatal Care : Official Journal of the National Association of Neonatal Nurses*, *6*(5), 261–270. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=17045946

Wataker, H., Meberg, A., & Nestaas, E. (2012). Neonatal family care for 24 hours per day: Effects on maternal confidence and breast-feeding. *Journal of Perinatal and Neonatal Nursing*, *26*(4), 336–342. https://doi.org/10.1097/JPN.0b013e31826d928b

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. Psychological Corporation.

Welch, M., Stark, R., Brunelli, S., Austin, J., Fiedor, E., Polin, R., Lorenz, J., Hofer, M., & Myers, M. (2013). Family nurture intervention (FNI) in the NICU: Can we prevent transgenerational effects of adverse rearing in prematurely born infants? *Comprehensive Psychiatry*, *54*(1), E13. https://doi.org/10.1016/j.comppsych.2012.07.055

White, R. D. (2003). Individual rooms in the NICU - an evolving concept. *Journal of*

Perinatology : Official Journal of the California Perinatal Association, *23 Suppl 1*, S22–S24. https://doi.org/10.1038/sj.jp.7210840

White, R. D. (2010). Single-family room design in the neonatal intensive care unit -- challenges and opportunities. *Newborn & Infant Nursing Reviews*, *10*(2), 83–86. https://doi.org/10.1053/j.nainr.2010.03.011

White, R. D. (2014). NICU design. *Archives of Disease in Childhood*, *99*((White R.) Newborn ICU, Memorial Hospital, South Bend, United States), A10. https://doi.org/10.1136/archdischild-2014-307384.28

White, R. D., & Whitman, T. (1992). Design of ICUs. *Pediatrics*, *89*(6 Pt 2), 1267. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med3&NEWS=N&AN=1594388

Whiting, P., Savović, J., Higgins, J. P. T., Caldwell, D. M., Reeves, B. C., Shea, B., Davies, P., Kleijnen, J., & Churchill, R. (2016). ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology*, *69*, 225–234. https://doi.org/10.1016/j.jclinepi.2015.06.005

Winnipeg Regional Health Authority. (2013). *Winnipeg Assessment of Neonatal Nurse Needs Tool* (p. 2). Winnipeg Regional Health Authority. https://cna-aiic.ca/~/media/cna/files/safe-staffing-toolkit/WANNNT-2013.pdf

Wittenberg, E., James, L. P., & Prosser, L. A. (2019). Spillover effects on caregivers' and family members' utility: A systematic review of the literature. *PharmacoEconomics*, *37*(4), 475–499. https://doi.org/10.1007/s40273-019-00768-7

World Health Organisation. (2015). International statistical classificaton of diseases and related health problems (ICD-10). In *International Classification of Diseases and Related Health*

Problems. (10th revis). https://doi.org/10.1016/j.jclinepi.2009.09.002

World Health Organization. (2016). *Kangaroo mother care to reduce morbidity and mortality in low-birth-weight infants*. http://www.who.int/elena/titles/kangaroo_care_infants/en/

World Health Organization. (2018). *Preterm birth*.

Wright, L. M., & Leahey, M. (1984). *Nurses & families: A guide to family assessment & intervention.* (1st ed.). F.A. Davis.

Xie, F., Pullenayegum, E., Gaebel, K., Bansback, N., Bryan, S., Ohinmaa, A., Poissant, L., & Johnson, J. A. (2016). A time trade-off-derived value set of the EQ-5D-5L for Canada. *Medical Care*, *54*(1), 98–105. https://doi.org/10.1097/MLR.0000000000000447

Yi, N., George, V., & Allison, D. B. (2003). Stochastic search variable selection for identifying multiple quantitative trait loci. *Genetics*, *164*(3), 1129–1138.

Zhang, J., Chu, H., Hong, H., Virnig, B. a., & Carlin, B. P. (2015). Bayesian hierarchical models for network meta-analysis incorporating nonignorable missingness. *Statistical Methods in Medical Research*. https://doi.org/10.1177/0962280215596185

Appendix A

MEDLINE Search Strategy
Ovid MEDLINE(R) All 1. Intensive Care Units, Neonatal/ 2. (intensive care adj2 (neonat* or neo nat* or prematur* or pre matur* or newborn)).tw. 3. (icu adj2 (neonat* or neo nat* or prematur* or pre matur* or newborn)).tw. 4. nicu.tw. 5. special care nursery.tw. 6. 1 or 2 or 3 or 4 or 5 7. exp Health Facility Environment/ 8. "Facility Design and Construction"/ 9. "Hospital Design and Construction"/ 10. Rooming-in Care/ 11. ((room? or unit? or bay? or structur* or design* or pod? or ward? or accommodation?) adj2 (family or private or closed or individual or single or shared or open or public)).tw. 12. (open adj (plan or concept)).tw. 13. rooming in.tw. 14. 7 or 8 or 9 or 10 or 11 or 12 or 13 15. 6 and 14

Appendix Table A 1. Electronic Search Strategy

Appendix Table A 2. Reasons for full-text exclusion

Title	Authors	Published Year	Exclusion Reason
Implementation of a Family-Care Suite Model in Nampa, Idaho	Weber, A.; Erickson, A.	2019	Not primary research
Survey of care environment and mortality in a tertiary neonatal intensive care unit	Lee, Y.; Chou, Y.	2005	Not comparison of room designs
Are single family rooms the future for neonatal units?	Banerjee, J.	2019	Wrong outcomes
Room for improvement: nurses' perceptions of providing care in a single room newborn intensive care setting	Walsh, W. F.; McCullough, K. L.; White, R. D.	2006	Wrong outcomes
Effects of the neonatal intensive care unit environment on preterm infant oral feeding	Pickler, R. H.; McGrath, J. M.; Reyna, B. A.; Tubbs-Cooley, H. L.; Best, A. M.; Lewis, M.; Cone, S.; Wetzel, P. A.	2013	Wrong outcomes

Title	Authors	Published Year	Exclusion Reason
Parents' Experiences of Support in NICU Single-Family Rooms	Liu, L. X.; Mozafarinia, M.; Axelin, A.; Feeley, N.	2019	Not comparison of room designs
The heuristics of nurse responsiveness to critical patient monitor and ventilator alarms in a private room neonatal intensive care unit	Joshi, R.; Mortel, H. v d; Feijs, L.; Andriessen, P.; Pul, C. v	2017	Not comparison of room designs
Increased parental satisfaction by unrestricted visiting hours and developmentally supportive care in NICU: results of a German multicenter study	Jannes, C.; Miedaner, F.; Langhammer, K.; Enke, C.; GÃ¶pel, W.; Kribs, A.; Nitzsche, A.; Riedel, R.; Woopen, C.; Kuntz, L.; Roth, B.	2018	Wrong outcomes
The impact of single family room design on patients and caregivers: Executive summary	Harris, D. D.; Shepley, M. M.; White, R. D.; Kolberg, K. J. S.; Harrell, J. W.	2006	Wrong outcomes
Surveillance of central venous catheter bloodstream infections in critical care units in England: Results from the sentinel study May 2016-April 2017	Gerver, S.; Mihalkova, M.; Bion, J.; Wilson, P.; Hope, R.	2018	Not comparison of room designs
Comparison of the Umbilical Cord Bacterial Colonization in Newborn Infants Rooming in with Mothers and Neonates Admitted to Neonatal Intensive Care Unit	Forozeshfard, M.; Ghorbani, R.; Razavi, M.; Danaei, N.; Nooripour, S.	2017	Not comparison of room designs
Journey to mother baby care: Implementation of a combined care/couplet model in a Level 2 neonatal intensive care unit	de Salaberry, J.; Hait, V.; Thornton, K.; Bolton, M.; Abrams, M.; Shivananda, S.; Kiarash, M.; Osiovich, H.	2019	Not comparison of room designs
Rooming-in reduces salivary cortisol level of newborn	De Bernardo, G.; Riccitelli, M.; Giordano, M.; Proietti, F.; Sordino, D.; Longini, M.; Buonocore, G.; Perrone, S.	2018	Wrong setting
Traditional open bay neonatal intensive care units can be redesigned to better suit family centered care application	Al-Motlaq, Mohammad A.	2018	Not comparison of room designs
Patient safety. Private rooms becoming the standard in NICUs	Feldman, L.	2009	Letter to the editor/editorial/commentary
The enhancement of the nursing role in the Single Family Room in NICU	Martina, F.; Motta, D.; Benedetta, A.; Laura, L.; Galbusera, V.; Carlotta, M.; Matteo, R.; Patrizia, S.	2019	Abstract

Title	Authors	Published Year	Exclusion Reason
Differences in the stool and skin microbiome, virulence factor and antimicrobial resistance genes in a private room verses a shared space neonatal intensive care unit	Ta, A.; Hourigan, S.; Klein, E.; Chettout, N.; Hasan, N.; Niederhuber, J.; Colwell, R.	2018	Abstract
Impact of Family-Centered Care on Pediatric and Neonatal Intensive Care Outcomes	Williams, L.	2016	Not primary research
Isolation or interaction: healthcare provider experience of design change	VanHeuvelen, J. S.	2019	Wrong outcomes
Hospitalising preterm infants in single family rooms versus open bay units: a systematic review and meta-analysis	van Veenendaal, N. R.; Heideman, W. H.; Limpens, J.; van der Lee, J. H.; van Goudoever, J. B.; van Kempen, A. A. M. W.; van der Schoor, S. R. D.	2019	Not primary research
Transition to a New Neonatal Intensive Care Unit	van den Berg, Johannes; BÃ¤22ck, Frida; Hed, Zara; Edvardsson, David	2017	Wrong outcomes
Effects of single-family rooms on nurse-parent and nurse-infant interaction in neonatal intensive care unit	Toivonen, M.; Lehtonen, L.; Loyttyniemi, E.; Axelin, A.	2017	Wrong outcomes
Parent-Infant Closeness, Parents' Participation, and Nursing Support in Single-Family Room and Open Bay NICUs	Tandberg, B. S.; Froslie, K. F.; Flacking, R.; Grundt, H.; Lehtonen, L.; Moen, A.	2018	Wrong outcomes
A new infrastructure for patient empowerment through Family Integrated Obstetric and Neonatal Healthcare in Single Family Rooms. A case study	Stelwagen, Mireille; Westmaas, Alvin; Kempen van, Anne; Blees, Yvonne; Scheele, Fedde	2018	Abstract
Impacts of the design of a neonatal intensive care unit (single-family room care and open-ward care) on clinical and environmental outcomes	Soleimani, F.; Rostami, F. F.; Nouri, J. M.; Hatamizadeh, N.; Sajedi, F.; Norouzi, M.	2020	Not primary research
Determining Appropriate Sensory Exposures in the NICU: Too Much, Too Little, or Just Right?	Smith, J. R.; Pineda, R. G.	2016	Letter to the editor/editorial/commentary
Neonatal outcomes in a modified NICU environment	Renaud, M. T.	2007	Not comparison of room designs
Neonatal Intensive Care Units Design: Implementation Strategies in South Africa	Rakhetla, Mabatho Mapoeng Elsina; Lubbe, Welma	2016	Not comparison of room designs
Parents' presence and parent-infant closeness in 11 neonatal intensive	Raiskila, S.; Axelin, A.; Toome, L.; Caballero, S.;	2017	Wrong outcomes

Title	Authors	Published Year	Exclusion Reason
care units in six European countries vary between and within the countries	Tandberg, B. S.; Montirosso, R.; Normann, E.; Hallberg, B.; Westrup, B.; Ewald, U.; Lehtonen, L.		
Evidence-based design for neonatal units: a systematic review	O'Callaghan, N.; Dee, A.; Philip, R. K.	2019	Not primary research
Recommendations on the environment for hospitalised newborn infants from the French neonatal society: rationale, methods and first recommendation on neonatal intensive care unit design	Kuhn, P.; Sizun, J.; Casper, C.; Society, G. s g f t F. N.	2018	Not primary research
Hospitals should try to keep mothers and babies together	Kendall-Raynor, Petra	2017	Letter to the editor/editorial/commentary
Determinants of Preterm Infant's Language Environment in the Neonatal Intensive Care Unit	Zauche, Lauren Head	2018	Not comparison of room designs
Staff Nurse Perceptions of Open-Pod and Single Family Room NICU Designs on Work Environment and Patient Care	Winner-Stoltz, R.; Lengerich, A.; Hench, A. J.; O'Malley, J.; Kjelland, K.; Teal, M.	2018	Wrong outcomes
The Relationship Between Hospital Construction and High-Risk Infant Auditory Function at NICU Discharge: A Retrospective Descriptive Cohort Study	Willis, V.	2018	Not comparison of room designs
Does the architectural layout of a NICU affect alarm pressure? A comparative clinical audit of a single-family room and an open bay area NICU using a retrospective study design	Joshi, R.; Straaten, H. v; Mortel, H. v d; Long, X.; Andriessen, P.; Pul, C. v	2018	Wrong outcomes
Sensory deprivation in private rooms in the NICU	Jobe, A. H.	2014	Letter to the editor/editorial/commentary
The Single-Family Room Neonatal Intensive Care Unit-Critical for Improving Outcomes?	Jobe, A. H.	2017	Letter to the editor/editorial/commentary
The impact of special care nursery design on neonatal nurses	Hogan, Christy; Jones, Liz; Saul, Julie	2016	Wrong outcomes
Surface Finish Materials: Considerations for the Neonatal Intensive Care Unit (NICU)	Harris, Debra	2016	Not primary research
Care practices and neonatal survival in 52 neonatal intensive care units in	Hanson, Claudia; Singh, Samiksha; Zamboni, Karen;	2019	Not comparison of room designs

Title	Authors	Published Year	Exclusion Reason
Telangana and Andhra Pradesh, India: A cross-sectional study	Tyagi, Mukta; Chamarty, Swecha; Shukla, Rajan; Schellenberg, Joanna		
Retrospective surveillance of antibiotic use in maternity wards and neonatal intensive care units in Saint Petersburg, Russia	Galankin, T. L.; Kolbin, A. S.; Sidorenko, S. V.; Kurylev, A. A.; Malikova, E. A.; Lobzin, Y. V.; Ivanov, D. O.; Shabalov, N. P.; Mikhailov, A. V.; Klimko, N. N.; Dolgov, G. V.	2018	Not comparison of room designs
NICU Nurses' Stress and Work Environment in an Open Ward Compared to a Combined Pod and Single-Family Room Design	Feeley, N.; Robins, S.; Charbonneau, L.; Genest, C.; Lavigne, G.; Lavoie-Tremblay, M.	2019	Wrong outcomes
Open plan and two cot nicu design: Comparing neonatal neurodevelopmental outcomes	Evans, M.; Broom, M.; Abdel-Latif, M. E.	2019	Abstract
Neonatal Intensive Care Unit Layout and Nurses' Work	Doede, M.; Trinkoff, A. M.; Gurses, A. P.	2018	Not primary research
Through the Eyes of the User: Evaluating Neonatal Intensive Care Unit Design	Denham, M. E.; Bushehri, Y.; Lim, L.	2018	Wrong outcomes
The kangaroo-mother method: mothers' living experience in the rooming-in	Davim, R. M. B.; Enders, B. C.; Dantas, J. C.; da Silva, R. A. R.; da NÃ³brega, E. J. P.	2009	Not English
Early environment and long-term outcomes of preterm infants	Cheong, J. L. Y.; Burnett, A. C.; Treyvaud, K.; Spittle, A. J.	2020	Not primary research
Exploring the Impact of a Dual Occupancy Neonatal Intensive Care Unit on Staff Workflow, Activity, and Their Perceptions	Broom, M.; Kecskes, Z.; Kildea, S.; Gardner, A.	2019	Wrong outcomes
Transition from an open-plan to a two-cot neonatal intensive care unit: a participatory action research approach	Broom, M.; Gardner, A.; Kecskes, Z.; Kildea, S.	2017	Wrong outcomes
Impact of NICU design on environmental noise	Szymczak, S. E.; Shellhaas, R. A.	2014	Wrong outcomes
The effect of open plan versus single family room nursery design on length of stay, weight at discharge and breast feeding in preterm infants	Feary, A; Wiliams, A; Jones, L.; Hont, T	2015	Abstract
Neonatal outcomes in a private room versus open room neonatal intensive care unit (NICU) design	Rosenblum, D. A.	2004	Abstract
Two contrasting NICU environments	Zahr, L. K.	1998	Not comparison of room designs

Title	Authors	Published Year	Exclusion Reason
Addressing noise in the NICU. Forward	Witt, C. L.	2008	Letter to the editor/editorial/commentary
Adolescents and rooming-in	Winkelstein, M. L.; Carson, V. J.	1987	Wrong setting
The potential for harm from alarm fatigue in single-room NICUs	Walsh, M. C.; Powers, E.; Fanaroff, J.	2015	Letter to the editor/editorial/commentary
Building blocks: how one hospital designed the core components of a new NICU	Vestal, R.	1999	Wrong outcomes
Safe patient monitoring is challenging but still feasible in a neonatal intensive care unit with single family rooms	Van Pul, C.; Mortel, H. P. M. E. V. D.; Bogaart, J. J. L. V. D.; Mohns, T.; Andriessen, P.	2015	Not comparison of room designs
Single-family room NICU influences infant outcomes	Stokowski, L. A.	2013	Letter to the editor/editorial/commentary
The impact of architectural design upon the environmental sound and light exposure of neonates who require intensive care: an evaluation of the Boekelheide Neonatal Intensive Care Nursery	Stevens, D. C.; Akram Khan, M.; Munson, D. P.; Reid, E. J.; Helseth, C. C.; Buggy, J.	2007	Wrong outcomes
Making the most of the single-family-room NICU	Steinhorn, R. H.	2016	Letter to the editor/editorial/commentary
Diabetic mothers and their newborn infants - rooming-in and neonatal morbidity	Stage, E.; Mathiesen, E. R.; Emmersen, P. B.; Greisen, G.; Damm, P.	2010	Wrong setting
Private NICU rooms a real pleaser	Spader, C.	2009	Not primary research
All sick newborns should receive single rooms	Solevag, A. L.; Borge, A. K.; Olsen, M.; Lie, H.; Nakstad, B.	2014	Letter to the editor/editorial/commentary
A comparative study of occupancy and patient care quality in four different types of intensive care units in a children's hospital1	Soares, Marcelo M.; Jacobs, Karen; Smith, Thomas J.	2012	Duplicate
Assessment of Neonatal Intensive Care Unit Single Private Room Versus Open Room Environment and the Impact on Maternal and Neonatal Outcomes	Smithgall, Lisa	2011	poster of published include trial
A comparative study of occupancy and patient care quality in four	Smith, T. J.	2012	Wrong outcomes

Title	Authors	Published Year	Exclusion Reason
different types of intensive care units in a children's hospital			
Clinically speaking: issues in designing the NICU	Smith, J.	1994	Letter to the editor/editorial/commentary
Transition to the private room NICU	Schoenbeck, K.	2006	Wrong outcomes
Structure of neonatal intensive care units	Ramirez, R.	1992	Not primary research
NICU environment--a need for change	Raman, T. S.	1997	Not primary research
Effects of Neonatal Intensive Care Unit Environmental Characteristics on Preterm Infant Oral Feeding	Pickler, Rita H.; Tubbs-Cooley, Heather; Cone, Sharon; McGrath, Jacqueline; Wetzel, Paul; Best, Al; Lewis, Marty; Reyna, Barbara	2012	poster of published include trial
Long live rooming-in	Perreault, E.; Lavandier, K. A.; Venne, M.	2008	Not primary research
The role of single-patient neonatal intensive care unit rooms for preterm infants	Ortenstrand, A.	2014	Letter to the editor/editorial/commentary
Beyond technology: meeting developmental needs of infants in NICUs	Oehler, J. M.; Strickland, M.; Nordlund, C.	1991	Not comparison of room designs
Rooming-in care for infants of opioid-dependent mothers: Implementation and evaluation at a tertiary care hospital	Newman, A.; Davies, G. A.; Dow, K.; Holmes, B.; Macdonald, J.; McKnight, S.; Newton, L.	2015	Wrong setting
Medical alarm management in a single-room neonatal intensive care unit (NICU): Evaluation after one year	Mortel, H. V. D.; Van Pul, C.; Ploem, E.; Bogaart, J. V. D.; Mohns, T.; Andriessen, P.	2014	Wrong outcomes
Preventing methicillin-resistant staphylococcus aureus (MRSA) transmission - a private room helps but is not the solution	Merheb, O. A.	2010	Not comparison of room designs
Rooming-in for Infants at Risk of Neonatal Abstinence Syndrome	McKnight, S.; Coo, H.; Davies, G.; Holmes, B.; Newman, A.; Newton, L.; Dow, K.	2016	Wrong setting
Single-room design in the NICU: making it work for you	McGrath, J. M.	2005	Letter to the editor/editorial/commentary
Design, Implementation, and Early Outcome Indicators of a New Family-Integrated Neonatal Unit	Mann, Donna	2016	Not comparison of room designs

Title	Authors	Published Year	Exclusion Reason
Rooming-in care of newborn infants	Mangili, G.; Formica, I. C.	2013	Letter to the editor/editorial/commentary
NICU environment -- What should it be like?	Lubbe, Welma; van der Walt, Christa; Klopper, Hester	2012	Not primary research
Individualised care rooms: The future of neonatal care	Liew, E.; Cane, C.	2015	Not primary research
Comparison of neonatal nighttime sleep-wake patterns in nursery versus rooming-in environments	Keefe, M. R.	1987	Wrong setting
NICU Open Ward vs Single Room: Do Hospital Room Layouts Affect Infant Brain Development?	Jackson, Yamile; LaCoursiere, Jasmine	2016	Not primary research
A Rooming-in Program to Mitigate the Need to Treat for Opiate Withdrawal in the Newborn	Hodgson, Z. G.; Abrahams, R. R.	2012	Wrong setting
Rooming-in: a preventative health care measure in the neonatal intensive care unit	Hayward, E. A.; Janes-Kelley, S.; Sikora, M.	1988	Not primary research
Designs for the delicate: a look at evolving NICU design standard	Harrell, J. W.; Moon, R. G.	2008	Not primary research
Challenges and Successes: The Baby-Friendly Initiative in Norway	Hansen, Mette Ness; BÃ¦rug, Anne; Nylander, Gro; HÃ¤ggkvist, Anna-Pia; Tufte, Elisabeth; Alquist, Ragnhild; StÃ¸re, Elisabeth Gahr	2012	Not comparison of room designs
Related settings: a family-centered approach to neonatal & pediatric intensive care unit design	Hall, J. H.	1994	Not primary research
Care of preterm infants in the neonatal intensive care unit	Hack, M.	2009	Letter to the editor/editorial/commentary
Role of the "rooming-in" on efficacy of universal neonatal hearing screening programmes	Grasso, D. L.; Hatzopulos, S.; Cossu, P.; Ciarafoni, F.; Rossi, M.; Martini, A.; Zocconi, E.	2008	Wrong setting
Challenging designs of neonatal intensive care units	Floyd, A. M.	2005	Not primary research
'Being in a womb' or 'playing musical chairs': the impact of place and space on infant feeding in NICUs	Flacking, R.; Dykes, F.	2013	Wrong outcomes
Private rooms becoming the standard in NICUs: staffing, other resource allocation need to be taken into account	Feldman, L.	2009	Letter to the editor/editorial/commentary

Title	Authors	Published Year	Exclusion Reason
Research abstracts	Eschiti, V. S.	2007	Wrong outcomes
Single Family Rooms for the NICU: Pros, Cons and the Way Forward	Dunn, Michael S.; MacMillan-York, Elizabeth; Robson, Kate	2016	Not primary research
Risk factors and prevalence of newborn hearing loss in a private health care system of Porto Velho, Northern Brazil	de Oliveira, J. S.; Rodrigues, L. B.; AurÃ©lio, F. S.; da Silva, V. B.	2013	Not comparison of room designs
Rooming-in to reduce neonatal abstinence syndrome	Davies, G.; Newman, A.; Newton, L.; Holmes, B.; Macdonald, J.; Connelly, R.; McKnight, S.; Dow, K.	2015	Wrong setting
Changing the NICU environment: the Boston City Hospital model	Cole, J. G.; Begish-Duddy, A.; Judas, M. L.; Jorgensen, K. M.	1990	Not primary research
Challenges in design and transition to a private room model in the neonatal intensive care unit	Carlson, B.; Walsh, S.; Wergin, T.; Schwarzkopf, K.; Ecklund, S.	2006	Wrong outcomes
Neuroprotective Core Measures 1-7: A Developmental Care Journey: Transformations in NICU Design and Caregiving Attitudes	Cardin, A. D.; Rens, L.; Stewart, S.; Danner-Bowman, K.; McCarley, R.; Kopsas, R.	2015	Not primary research
Single-Family Room NICUs and Neurodevelopmental Outcomes	Caldwell, Curtis D.	2015	Letter to the editor/editorial/commentary
Focus on a unit. A new way of rooming in	Bruhn, G. P.; Hansen, C. C.	2005	Not primary research
Designing and delivering neonatal care in single rooms	Brown, P.; Taquino, L. T.	2001	Wrong outcomes
Transition from an open plan to a two cot neonatal intensive care unit: A participatory action research approach	Broom, M.; Gardner, A.; Kecskes, Z.; Kildea, S.	2016	Wrong outcomes
Single-room infant care: future trends in special care nursery planning and design	Bowie, B. H.; Hall, R. B.; Faulkner, J.; Anderson, B.	2003	Wrong outcomes
Single Patient Room Design in the Neonatal Intensive Care Unit - Parent Perceptions of Open Ward vs. Single Patient Room Units	Bodack, E.; Schenk, O.; Karutz, H.	2016	patients all exposed to both design types
Environmental impact of the NICU on developmental outcomes	Blackburn, S.	1998	Not primary research
Changing units for changing times: the evolution of a NICU	Altimier, L.; Lutes, L.	2000	Wrong outcomes
High-tech, high-touch care	Altimier, L.	2001	Not comparison of room designs

Title	Authors	Published Year	Exclusion Reason
An evaluation of rooming-in among substance-exposed newborns in British Columbia	Abrahams, R. R.; MacKay-Dunn, M. H.; Nevmerjitskaia, V.; MacRae, G. S.; Payne, S. P.; Hodgson, Z. G.	2010	Wrong setting
Rooming-in for preterm infants: How far should we go? Five-year experience at a tertiary hospital	Abecasis, F. D. G.; Gomes, A.	2006	Wrong setting
Private NICU Better		2015	Wrong study design
Single-Family Room NICUs Better for Preterm Infants, Parents		2015	Letter to the editor/editorial/commentary
Rethinking private neonatal intensive care unit rooms		2014	Letter to the editor/editorial/commentary
Breast milk expression in the NICU: location matters		2014	Letter to the editor/editorial/commentary
The newborn intensive care unit		2003	Not comparison of room designs

Appendix Table A 3. Data used for the analysis of morbidities and mortality

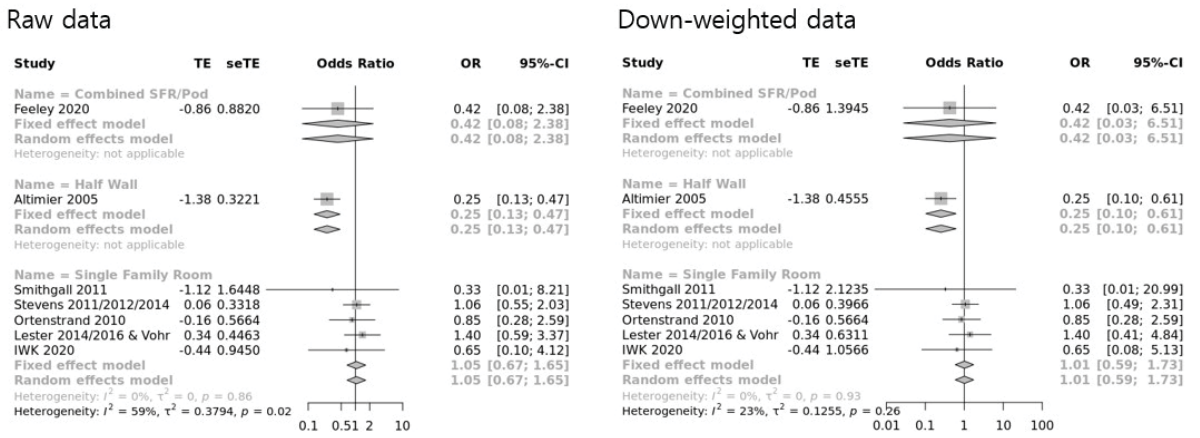
Name	Design	IVH		ROP		NEC		Sepsis		Mortality		BPD	
		Log OR	se	Log OR	se	Log OR	se	Log OR	se	Log OR	se	Log OR	se
larson 1985	Open Bay	NA	NA	NA	NA	NA	NA	-2.22	0.09	NA	NA	NA	NA
larson 1985	New Open Bay	NA	NA	NA	NA	NA	NA	-2.34	0.16	NA	NA	NA	NA
jones 2012	Open Bay	NA	NA	NA	NA	-2.27	0.28	-0.87	0.18	NA	NA	NA	NA
jones 2012	New Open Bay	NA	NA	NA	NA	-2.29	0.28	NA	NA	NA	NA	NA	NA
vondolingerdebrito 2007	Open Bay	NA	NA	NA	NA	-5.98	0.71	-1.92	0.11	NA	NA	NA	NA
vondolingerdebrito 2007	New Open Bay	NA	NA	NA	NA	-5.81	0.71	-2.24	0.13	NA	NA	NA	NA
Chen 2017	Open Bay	NA	NA	NA	NA	NA	NA	-2.41	0.23	NA	NA	NA	NA
Chen 2017	New Open Bay	NA	NA	NA	NA	NA	NA	-3.26	0.32	NA	NA	NA	NA
altimier 2005	Open Bay	-2.09	0.16	-1.81	0.14	NA	NA	NA	NA	NA	NA	NA	NA

Name	Design	IVH		ROP		NEC		Sepsis		Mortality		BPD	
		Log OR	se	Log OR	se	Log OR	se	Log OR	se	Log OR	se	Log OR	se
altimier 2005	Half Wall	-3.48	0.28	-2.43	0.18	NA	NA	NA	NA	NA	NA	NA	NA
pineda 2014	Open Bay	NA	NA	-1.79	0.41	-2.42	0.52	-0.63	0.30	-1.69	0.36	0.12	0.29
pineda 2014	Private Room	NA	NA	-2.16	0.43	-2.60	0.52	-1.05	0.30	-1.66	0.33	-0.14	0.26
julian 2015	Open Bay	NA	NA	NA	NA	NA	NA	-3.16	0.17	NA	NA	NA	NA
julian 2015	Private Room	NA	NA	NA	NA	NA	NA	-3.19	0.17	NA	NA	NA	NA
Hourigan 2018	Open Bay	NA	NA	NA	NA	NA	NA	1.79	0.76	NA	NA	NA	NA
Hourigan 2018	Private Room	NA	NA	NA	NA	NA	NA	-0.22	0.47	NA	NA	NA	NA
ortenstrand 2010	Open Bay	-3.22	0.39	NA	NA	-3.38	0.42	-2.22	0.25	-4.9	0.83	-2.75	0.31
ortenstrand 2010	Single Family Room	-3.38	0.42	NA	NA	-3.57	0.45	-2.57	0.29	-5.9	1.43	-4.09	0.58
domanico 2011	Open Bay	NA	NA	NA	NA	NA	NA	-2.14	0.28	-2.41	0.31	NA	NA

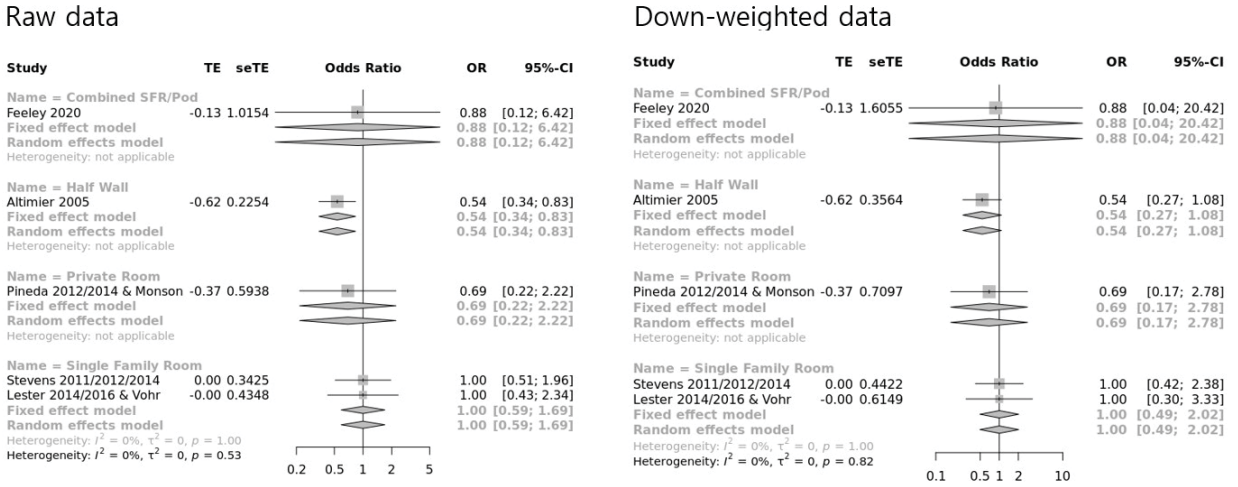
Name	Design	IVH		ROP		NEC		Sepsis		Mortality		BPD	
		Log OR	se	Log OR	se	Log OR	se	Log OR	se	Log OR	se	Log OR	se
domanico 2011	Single Family Room	NA	NA	NA	NA	NA	NA	-2.82	0.42	-2.39	0.35	NA	NA
alessio 2011	Open Bay	NA	NA	NA	NA	NA	NA	NA	NA	-1.79	0.54	NA	NA
alessio 2011	Single Family Room	NA	NA	NA	NA	NA	NA	NA	NA	-1.25	0.46	NA	NA
stevens 2011_2014	Open Bay	-4.46	0.24	-4.50	0.25	-4.19	0.50	-4.33	0.24	-4.18	0.21	-0.85	0.06
stevens 2011_2014	Single Family Room	-4.40	0.22	-4.50	0.24	-4.32	0.50	-4.82	0.31	-4.00	0.18	-0.39	0.06
smithgall 2011	Open Bay	-3.53	0.83	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
smithgall 2011	Single Family Room	-4.64	1.42	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Wataker 2012	Open Bay	NA	NA	NA	NA	NA	NA	-1.19	0.43	NA	NA	NA	NA
Wataker 2012	Single Family Room	NA	NA	NA	NA	NA	NA	-0.96	0.37	NA	NA	NA	NA
lester_vohr	Open Bay	-3.34	0.31	-2.76	0.34	-3.54	0.34	-1.39	0.20	-1.46	0.13	-1.35	0.14

Name	Design	IVH		ROP		NEC		Sepsis		Mortality		BPD	
		Log OR	se	Log OR	se	Log OR	se	Log OR	se	Log OR	se	Log OR	se
lester_vohr	Single Family Room	-3.00	0.32	-2.76	0.27	-3.71	0.45	-1.93	0.19	-1.50	0.16	-1.11	0.16
Puumala 2020	Open Bay	NA	NA	NA	NA	NA	NA	-3.24	0.14	-4.42	0.12	NA	NA
Puumala 2020	Single Family Room	NA	NA	NA	NA	NA	NA	-3.08	0.14	-4.07	0.12	NA	NA
Tandberg 2019	Open Bay	NA	NA	NA	NA	NA	NA	-3.29	0.82	NA	NA	-2.77	0.65
Tandberg 2019	Single Family Room	NA	NA	NA	NA	NA	NA	-4.36	1.49	NA	NA	-4.36	1.49
IWK	Open Bay	-2.40	0.60	NA	NA	-4.29	1.42	NA	NA	-3.53	0.30	-4.29	1.42
IWK	Single Family Room	-2.80	0.73	NA	NA	-3.16	0.83	NA	NA	-4.29	0.13	-3.16	0.83
Erdeve 2008	Open Bay	NA	NA	NA	NA	NA	NA	NA	NA	-1.79176	0.44	NA	NA
Erdeve 2008	Single Family Room	NA	NA	NA	NA	NA	NA	NA	NA	-2.48491	0.6	NA	NA
Feeley 2020	Open Bay	-2.80	0.52	-3.53	0.72	NA	NA	NA	NA	NA	NA	NA	NA

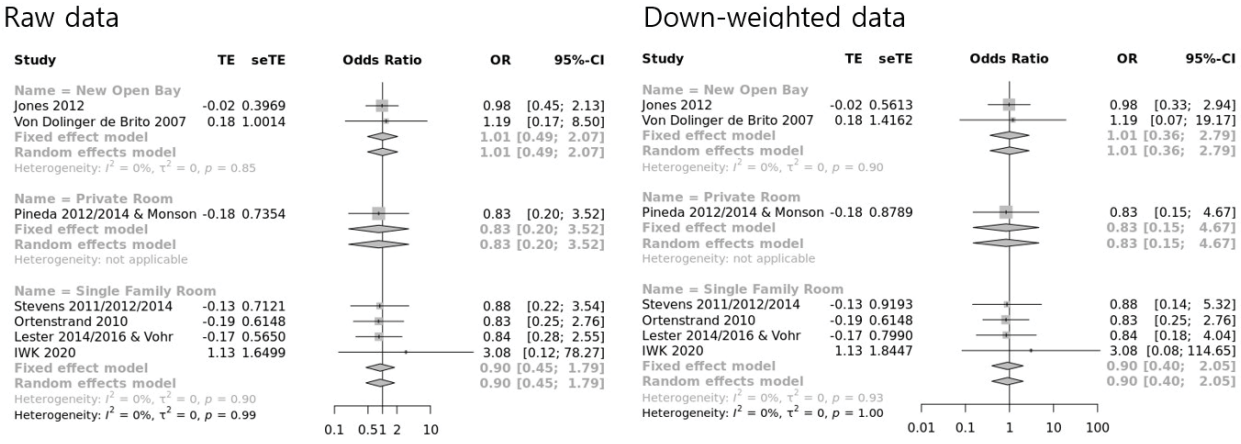
Name	Design	IVH		ROP		NEC		Sepsis		Mortality		BPD	
		Log OR	se	Log OR	se	Log OR	se	Log OR	se	Log OR	se	Log OR	se
Feeley 2020	Combined SFR/Pod	-3.66	0.72	-3.66	0.72	NA	NA	NA	NA	NA	NA	NA	NA



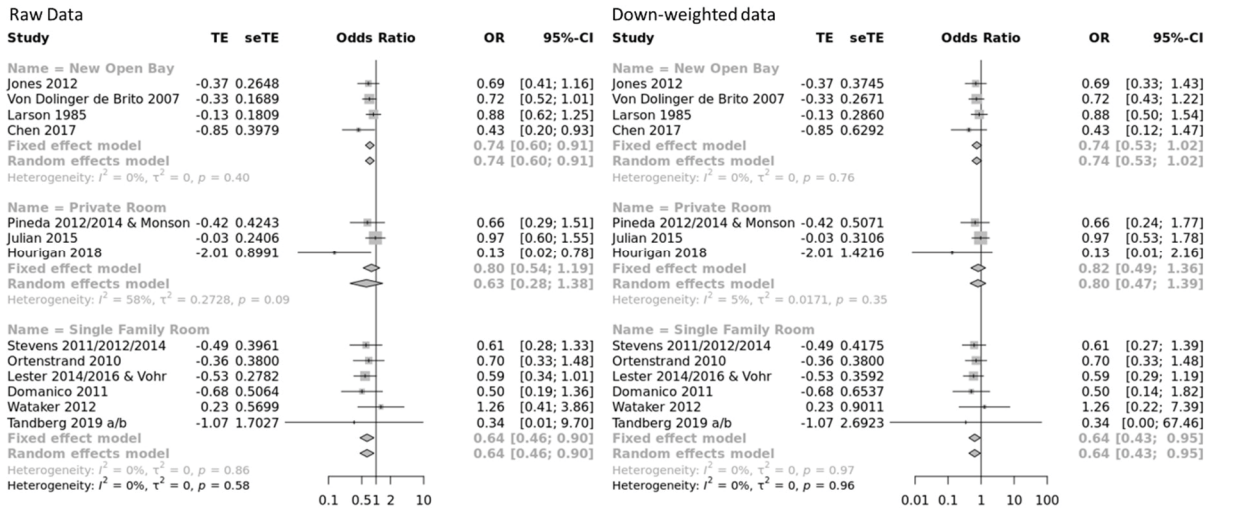
Appendix Figure A. 1. Forest plot of pairwise meta-analysis for IVH



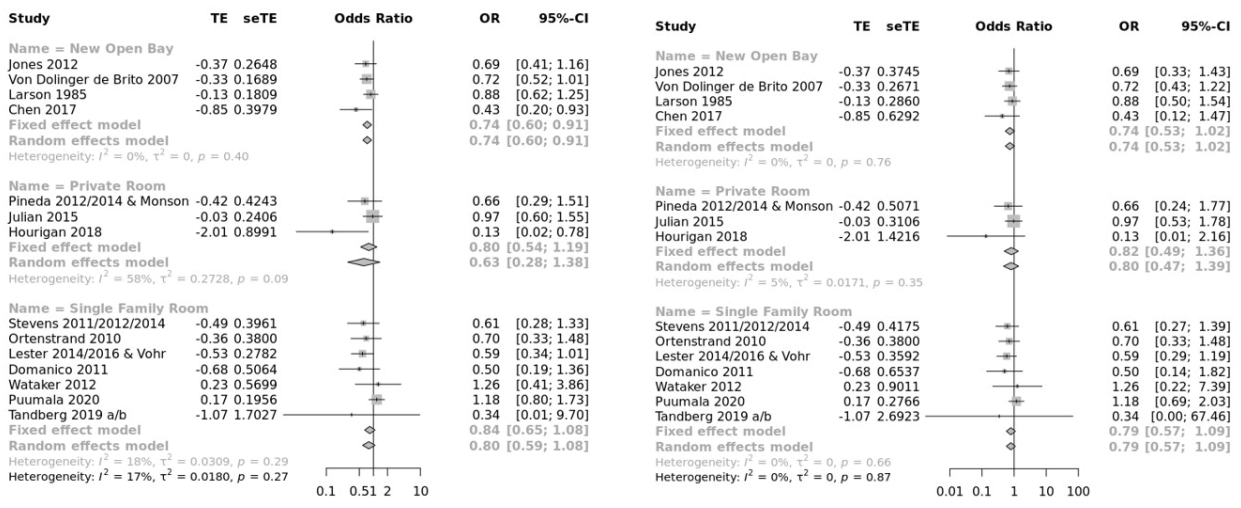
Appendix Figure A. 2 Forest plot of pairwise meta-analysis for ROP



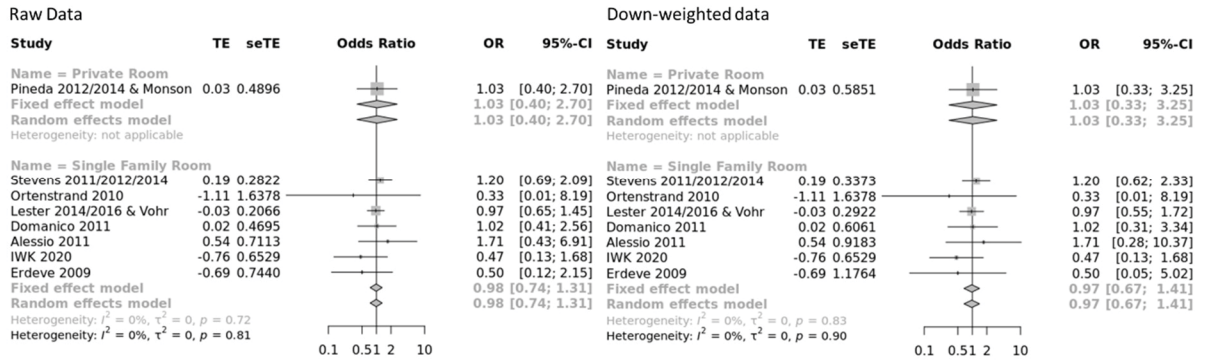
Appendix Figure A. 3. Forest plot for pairwise meta-analysis for NEC



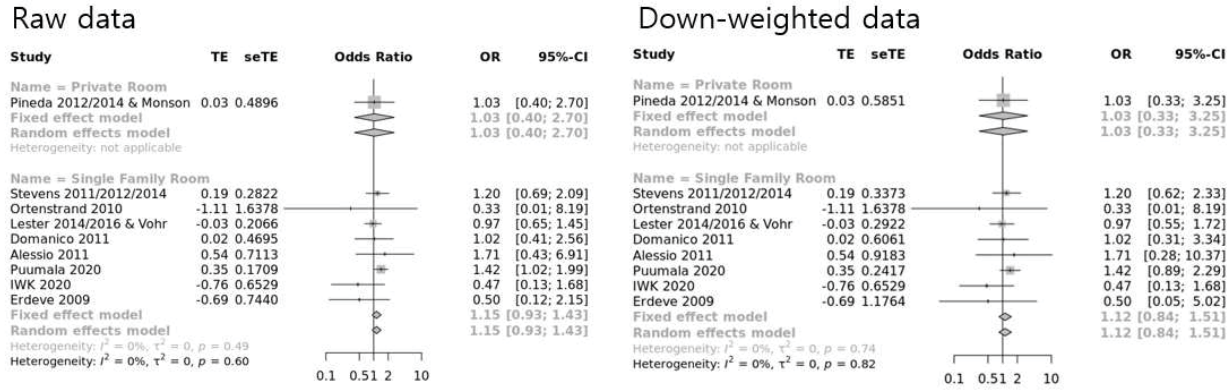
Appendix Figure A. 4. Forest plot of pairwise meta-analyses for sepsis excluding Puumala. Data prior to down-weighting is on the left, and after down-weighting on the right.



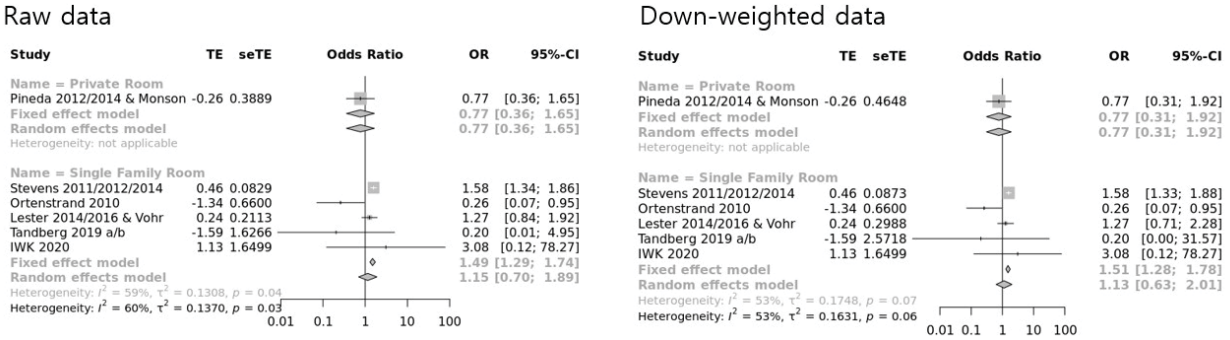
Appendix Figure A. 5. Forest plot of pairwise meta-analyses for sepsis including Puumala et al. Data prior to down-weighting is on the left, and after down-weighting on the right.



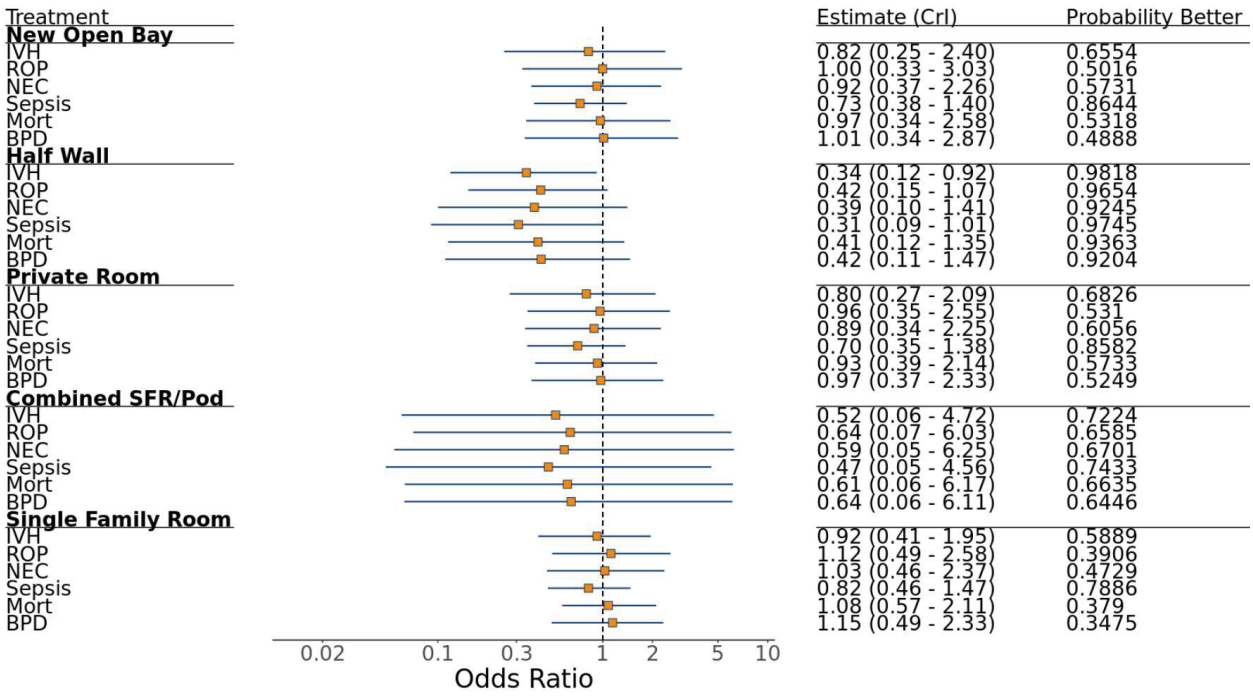
Appendix Figure A. 6. Forest plot of pairwise meta-analysis for mortality. Data prior to down-weighting is on the left, and after down-weighting on the right.



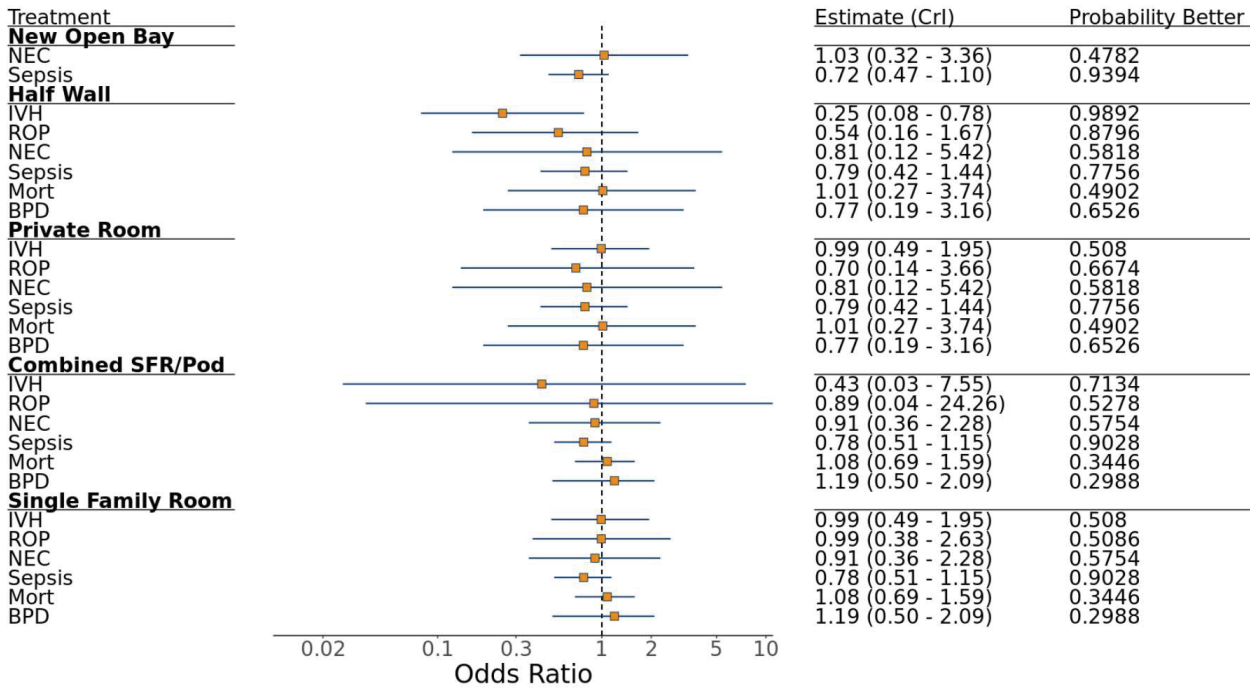
Appendix Figure A. 7. Forest plot of pairwise meta-analysis for mortality. Data prior to down-weighting is on the left, and after down-weighting on the right. Sensitivity analysis including Puumala et al.



Appendix Figure A. 8. Forest plot of pairwise meta-analysis for BPD



Appendix Figure A. 9. Forest plot for multivariate model estimating the effect of room designs on morbidity and mortality. Correlation estimated from individual patient data, default priors with constant potency assumption. Probability better refers to the probability that a design is better than an old open bay. Includes Puumala et al.



Appendix Figure A. 10. Forest plot for multivariate model estimating the effect of room designs on morbidity and mortality. Univariate sensitivity (no correlation), default priors without constant potency assumption. Probability better refers to the probability that a design is better than an old open bay. Includes Puumala et al.

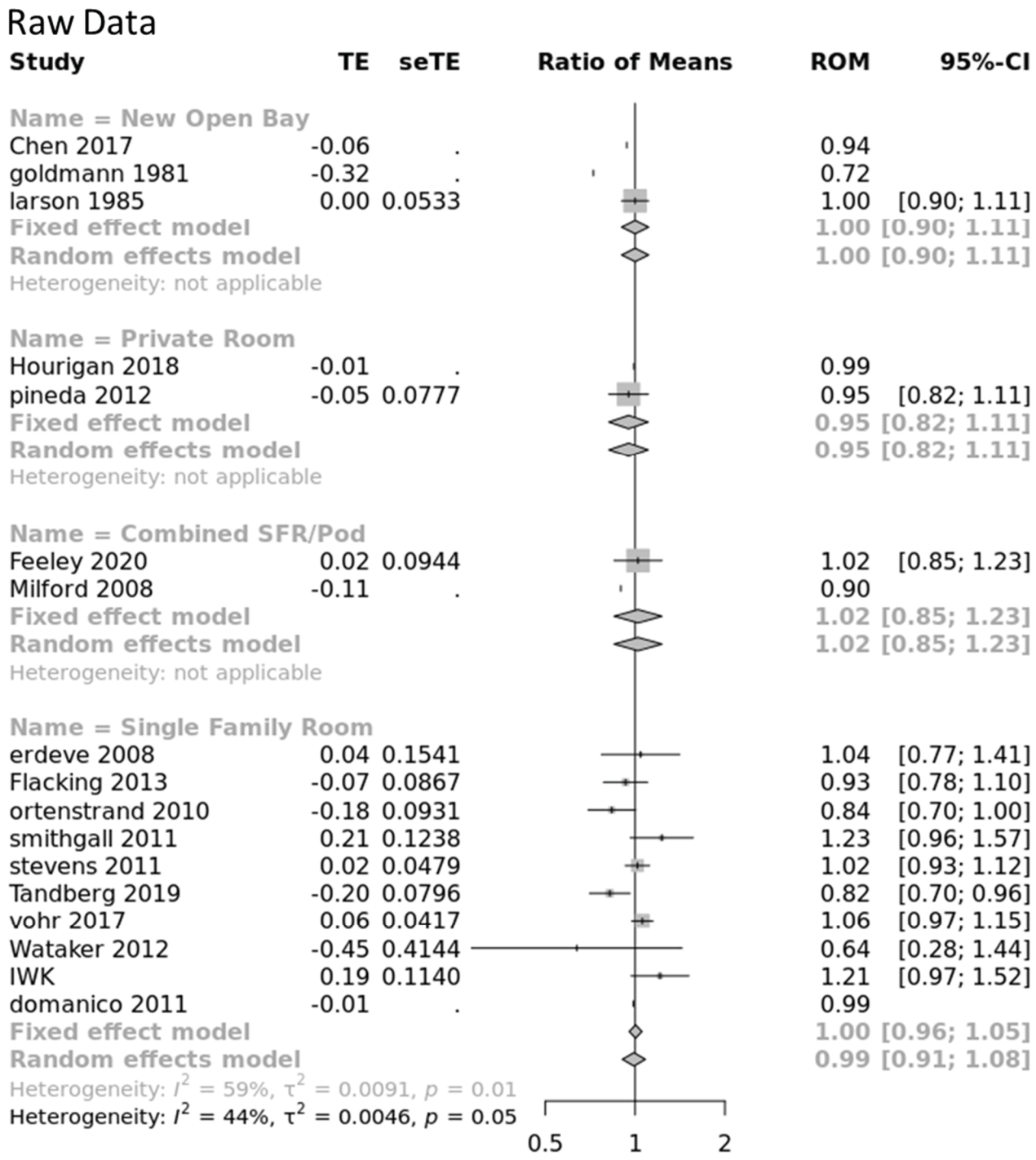
Appendix Table A 4. Data for arm based data entry portion of length of stay model

Study	Design	N	Mean	SD
goldmann 1981	OPB	642	13.8	14
goldmann 1981	New OPB	542	10	15
Larson 1985	OPB	1443	21	21.6
Larson 1985	New OPB	502	21	21.6
Chen 2017	Open Bay	242	27.53	21.6
Chen 2017	New OPB	270	25.82	21.6
Pineda 2014, Pineda 2012, Monson 2018	OPB	39	92.69	29.67
Pineda 2014, Pineda 2012, Monson 2018	PR	42	88.12	33.375
Hourigan 2018	Open Bay	14	33.5	21.6
Hourigan 2018	PR	18	33.2	21.6
Erdeve 2008	OPB	29	23	14
Erdeve 2008	SFR	31	24	14
ortenstrand 2010	OPB	182	32.8	21.68154
ortenstrand 2010	SFR	183	27.4	29.33313

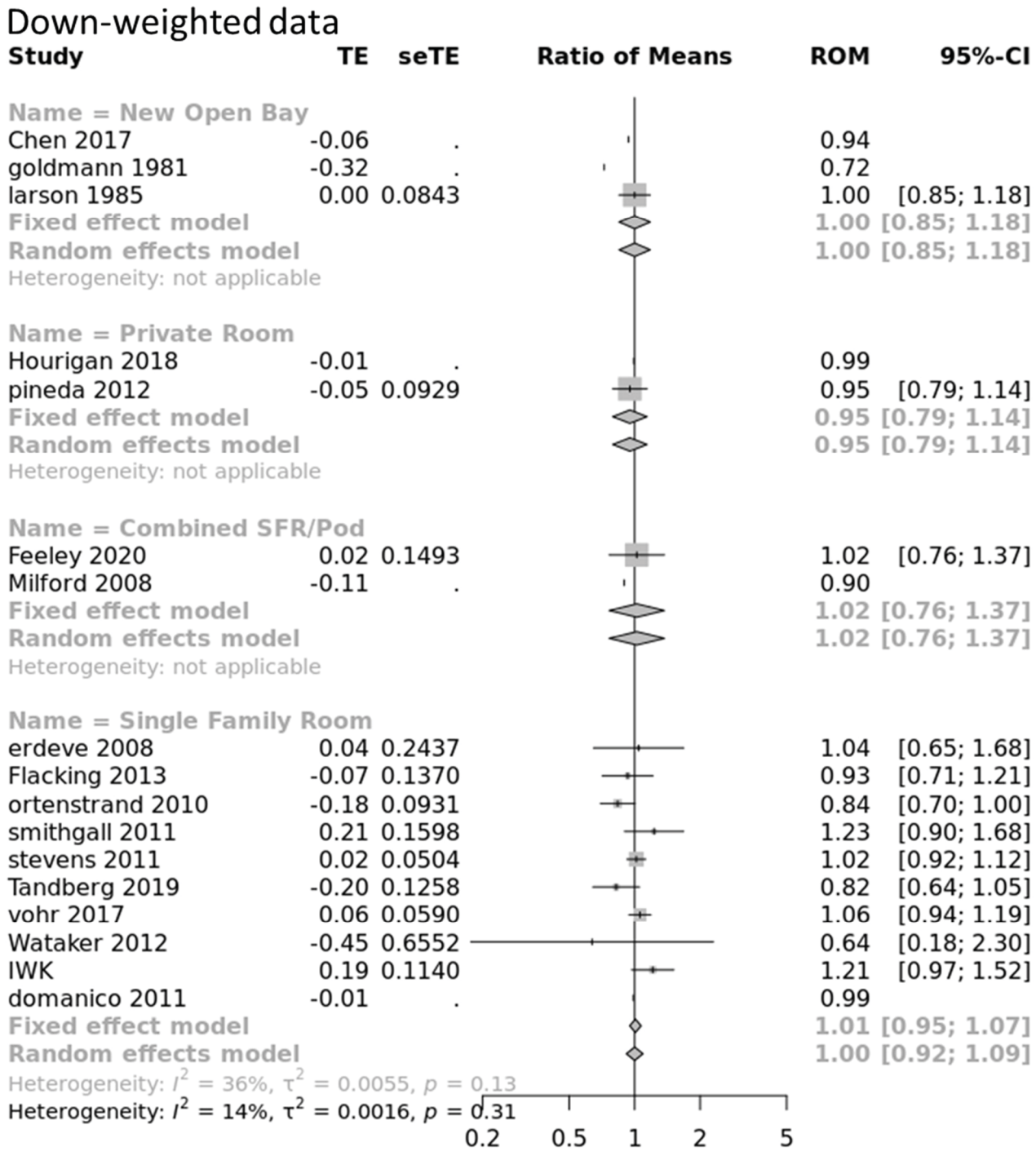
Stevens 2012, Stevens 2014, Stevens 2011	OPB	1300	11.78	14
Stevens 2012, Stevens 2014, Stevens 2011	SFR	1300	11.99	15
Smithgall 2011	OPB	52	18.23	10.82
Smithgall 2011	SFR	52	22.4	14.93
Wataker 2012	OPB	30	11.3	14
Wataker 2012	SFR	36	7.2	15
Lester 2014, Lester 2016, Vohr 2017	OPB	320	79.8	41
Lester 2014, Lester 2016, Vohr 2017	SFR	210	84.4	37
Flacking 2013	Open Bay	114	28	21
Flacking 2013	SFR	186	26	18
Tandberg 2019	Open Bay	42	45	18
Tandberg 2019	Single Family Room	35	37	11
Feeley 2020	Open Bay	70	54.27	30.48
Feeley 2020	Combined SFR	80	55.5	32.97

Appendix Table A 5. Data for contrast based data entry for length of stay

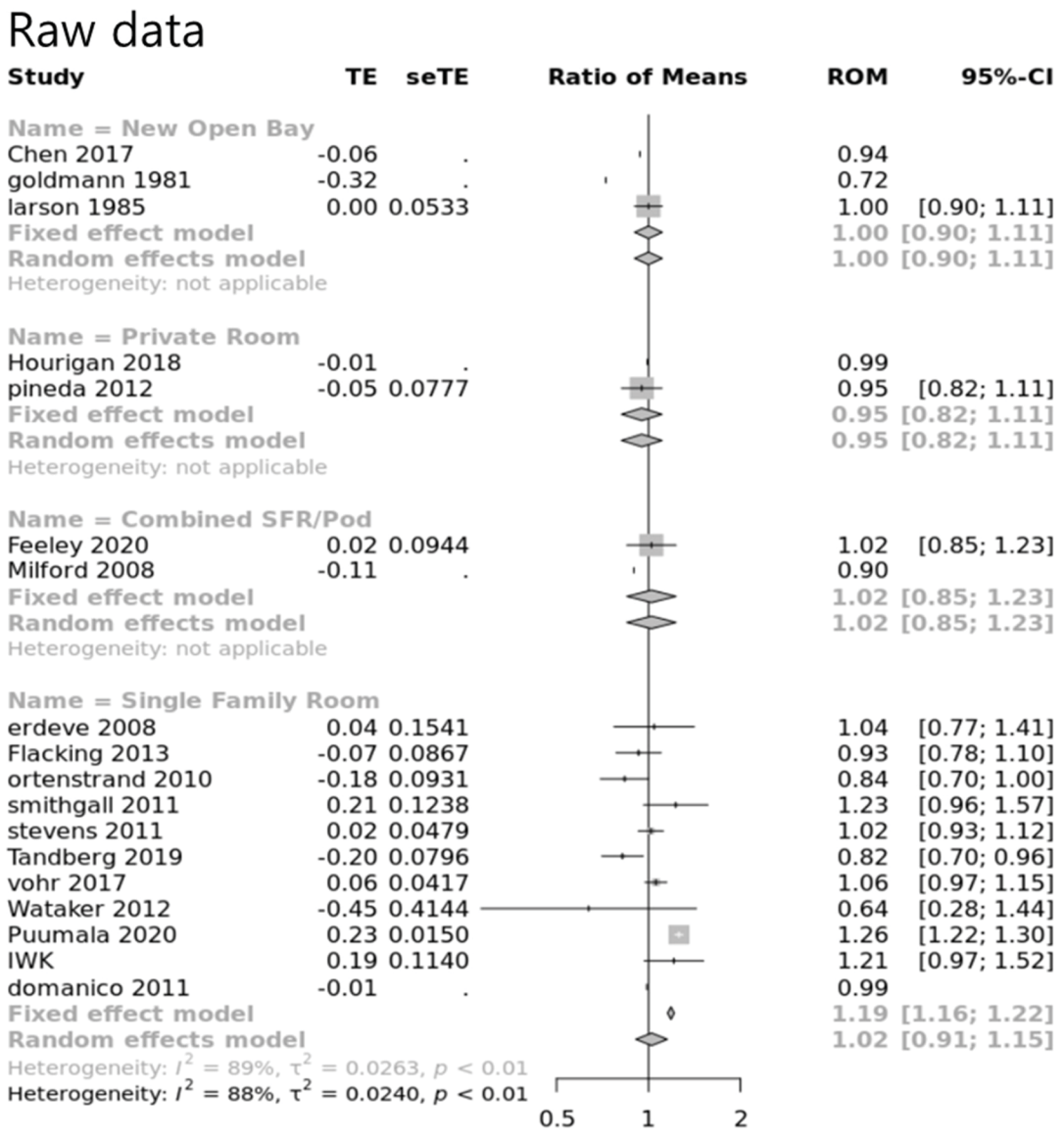
Study	Treatment 1	Treament 2	Mean 1	Diff 2	SE 1	SE 2	# Arms	N 1	N 2
Puumala 2020	OPB	SFR	7.79	0.23	0.12	0.015	2	5964	4031
IWK	OPB	SFR	14.38	0.192	0.81	0.114	2	358	317
domanico 2011	OPB	SFR	21.16	-0.01	0.75	0.013	2	81	81



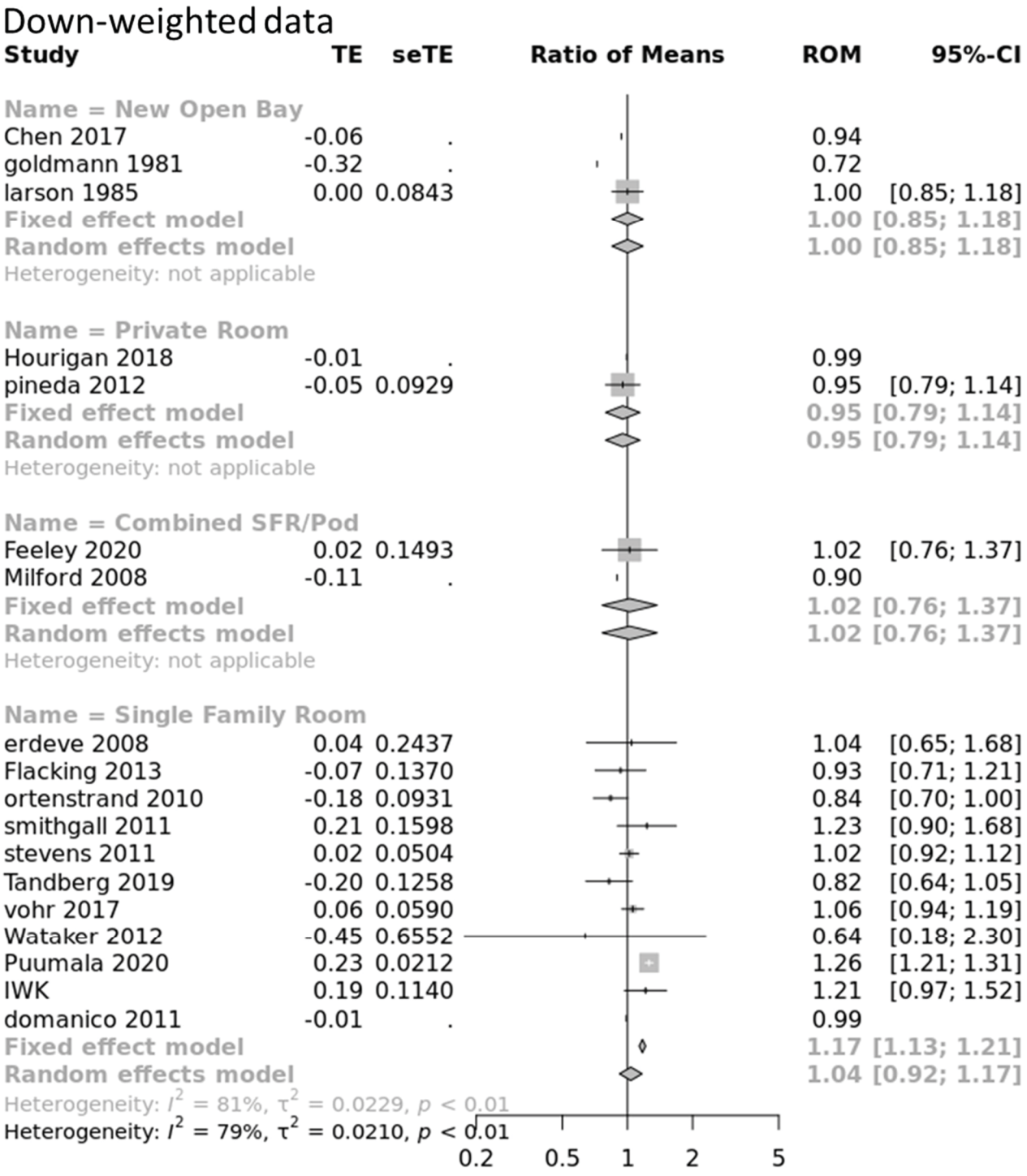
Appendix Figure A. 11. Forest plot for length of stay data. Missing standard errors are imputed in final Bayesian model but entered as NA. No down-weighting applied.



Appendix Figure A. 12. Forest plot for length of stay data. Missing standard errors are imputed in final Bayesian model but entered as NA. Down-weighting applied.



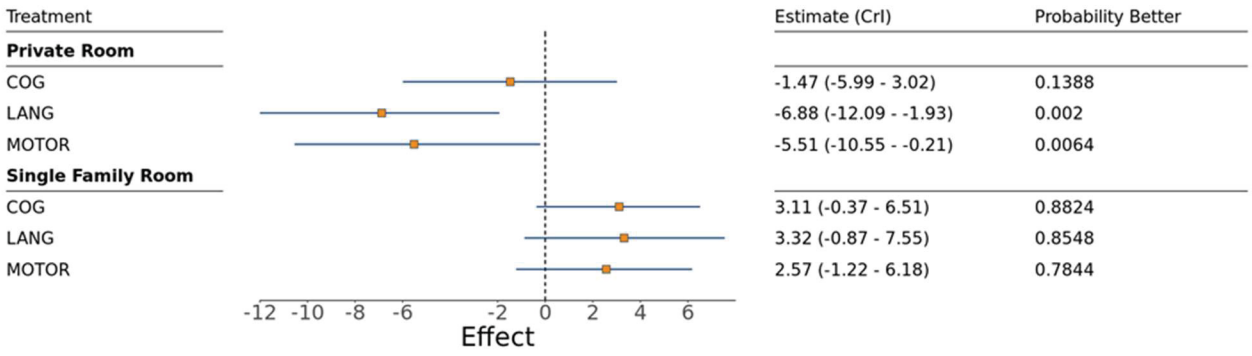
Appendix Figure A. 13. Forest plot for length of stay data. Sensitivity analysis including Puumala. Missing standard errors are imputed in final Bayesian model but entered as NA. No down-weighting applied.



Appendix Figure A. 14. Forest plot for length of stay data. Sensitivity analysis including Puumala. Missing standard errors are imputed in final Bayesian model but entered as NA. Down-weighting applied.

Appendix Table A 6. Data used for analysis of Bayley III scores

Study	Design	Bayley III Cognitive		Bayley III Language		Bayley III Motor	
		Mean	SE	Mean	SE	Mean	SE
Pineda 2014, Monson 2018	OPB	86.8	1.58	91.9	1.80	86.2	1.66
Pineda 2014, Monson 2018	PR	85.3	1.30	84.9	1.57	80.7	1.61
Lester 2016, Vohr 2017	OPB	87.1	0.79	84	0.98	88.3	0.88
Lester 2016, Vohr 2017	SFR	90.2	1.10	87.3	1.34	90.8	1.18



Appendix Figure A. 15. Univariate model with vague priors

Appendix Table A 7. Costs and their sources in the post-discharge model. Alpha and lambda parameters used to provide estimates of uncertainty for probabilistic sensitivity analysis.

Description	**Expected Value and standard error**	**Distribution for PSA**	**Source**
Cost of admission	$53	NA	Nova Scotia Medical Services Insurance Physician's manual
Daily rounds	$7	NA	Nova Scotia Medical Services Insurance Physician's manual
Palliative care consult	$62	NA	Nova Scotia Medical Services Insurance Physician's manual
One hour prolonged palliative care consult	$50	NA	Nova Scotia Medical Services Insurance Physician's manual
Discharge cost	$10	NA	Nova Scotia Medical Services Insurance Physician's manual
Cost of interpretation of an X-Ray	$6.25	NA	Nova Scotia Medical Services Insurance Physician's manual
Cost of ultrasound interpretation	$33.49	NA	Nova Scotia Medical Services Insurance Physician's manual
Initial surgical consult	$24	NA	Nova Scotia Medical Services Insurance Physician's manual
Repeat surgical consult	$15	NA	Nova Scotia Medical Services Insurance Physician's manual

Description	Expected Value and standard error	Distribution for PSA	Source
Laparoscopic surgery	$180	NA	Nova Scotia Medical Services Insurance Physician's manual
Initial ROP screen	$25	NA	Nova Scotia Medical Services Insurance Physician's manual
Subsequent ROP screen	$15	NA	Nova Scotia Medical Services Insurance Physician's manual
Cost of laser treatment	$171	NA	Nova Scotia Medical Services Insurance Physician's manual
Cost of avastin injection	$25	NA	Nova Scotia Medical Services Insurance Physician's manual
Cost per gross square foot (uninflated 2018 USD)	$550	NA	Hessam Sadatsafavi, Niknejad, Shepley, & Sadatsafavi, 2019
Number of beds	40	NA	assumption
Square footage new open bay	120	NA	assumption
Square footage half wall	130	NA	assumption
Square footage private room	150	NA	assumption

Description	Expected Value and standard error	Distribution for PSA	Source
Square footage single family room	165	NA	assumption
Square footage combined SFR and pod	Blend of open bay (30%) and sfr (70%)	NA	assumption
Yearly cost from discharge to 2 years for less than 28 weeks (Uninflated 2012 CDN)	$9280	NA	Johnston et al., 2014
Yearly cost from discharge to 2 years for less than 28-32 weeks	$6573	NA	Johnston et al., 2014
Yearly cost from discharge to 2 years for less than 33-36 weeks	$2228	NA	Johnston et al., 2014
Yearly cost no disability years 2-4 (uninflated 2006 GBP)	315 (59)	Gamma	Mangham et al., 2009
Yearly cost no disability years 5-10 (uninflated 2006 GBP)	3467 (59)	Gamma	Mangham et al., 2009
Yearly cost no disability years 11-18 (uninflated 2006 GBP)	4388 (59)	Gamma	Mangham et al., 2009

Description	Expected Value and standard error	Distribution for PSA	Source
Yearly cost mild disability years 2-4 (uninflated 2006 GBP)	611 (95)	Gamma	Mangham et al., 2009
Yearly cost mild disability years 5-10 (uninflated 2006 GBP)	3763 (311)	Gamma	Mangham et al., 2009
Yearly cost mild disability years 11-18 (uninflated 2006 GBP)	4684 (95)	Gamma	Mangham et al., 2009
Yearly cost sev disability years 2-4 (uninflated GBP)	933 (188)	Gamma	Mangham et al., 2009
Yearly cost sev disability years 5-10 (uninflated GBP)	8601 (847)	Gamma	Mangham et al., 2009
Yearly cost sev disability years 11-18 (uninflated GBP)	9201 (794)	Gamma	Mangham et al., 2009

Description	Expected Value and standard error	Distribution for PSA	Source
Canadian Consumer Price index percent change 2006-2020	1.25	NA	http://www.bankofcanada.ca/rates/related/inflation-calculator/
Canadian Consumer Price index 2012-2020	1.12	NA	http://www.bankofcanada.ca/rates/related/inflation-calculator/
Purchasing power 1 USD in GBP 2006	0.697	NA	https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm
Purchasing power parity CDN/USD 2006	1.206	NA	https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm
Male hourly wage (2019 CDN)	29.61	NA	2019 estimate of hourly wages all industries statistics Canada. Table 14-10-0064-01 Employee wages by industry, annual

Description	Expected Value and standard error	Distribution for PSA	Source
Female hourly wage (2019 CDN)	26.02	NA	2019 estimate of hourly wages all industries statistics Canada. Table 14-10-0064-01 Employee wages by industry, annual
Probability of movement from no disability to healthy, NDI, and sNDI for ages 2-5	List(88, 27, 2)	Dirichlet	Mangham et al., 2009
Probability of movement from NDI to healthy, NDI, and sNDI for ages 2-5	List(22,18,11)	Dirichlet	Mangham et al., 2009
Probability of movement from sNDI to healthy, NDI, and sNDI for ages 2-5	List(2, 10, 30)	Dirichlet	Mangham et al., 2009
Probability of movement from no disability to healthy, NDI, and sNDI for ages 6-18	List(100, 12, 0)	Dirichlet	Mangham et al., 2009

Description	Expected Value and standard error	Distribution for PSA	Source
Probability of movement from NDI to healthy, NDI, and sNDI for ages 6-18	List(20, 29, 6)	Dirichlet	Mangham et al., 2009
Probability of movement from sNDI to healthy, NDI, and sNDI for ages 6-18	List(1, 7, 35)	Dirichlet	Mangham et al., 2009

Appendix Table A 8. Utility decrements for disability states across the lifespan. Decrements calculated by subtracting relevant means. Standard errors calculated by converting confidence intervals to variance and taking the square root of the sum.

Description	Expected Value (95% interval)	Decrement and standard error	Source
Normal Birthweight control			(Saigal et al., 2016)
Birth-16 years	0.884 (0.862, 0.909)	-	
17-26 years	0.893 (0.867, 0.920)	-	
27 years to death	0.853 (0.809, 0.896)	-	
No NDI and NDI			
Birth-16 years	0.831 (0.794, 0.868)	**-0.05 (0.02)**	
17-26 years	0.829 (0.784, 0.874)	-0.06 (0.03)	
27 years to death	0.766 (0.704, 0.827)	-0.09 (0.04)	
Severe NDI			
Birth-16 years	0.680 (0.576, 0.783)	-0.20 (0.05)	
17-26 years	0.654 (0.560, 0.747)	-0.24 (0.05)	
27 years to death	0.600	-0.25	

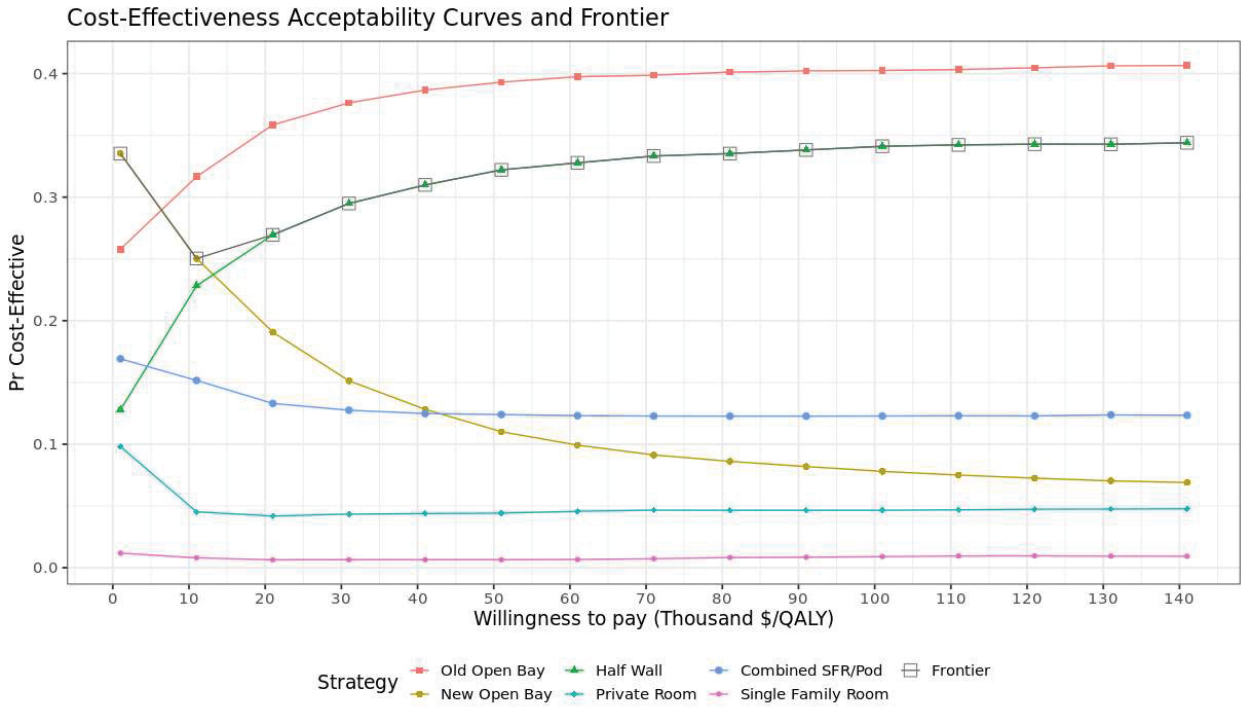
	(0.502, 0.699)	(0.05)	

Appendix Table A 9. Statistics Canada 2016-2018 Life tables for both sexes

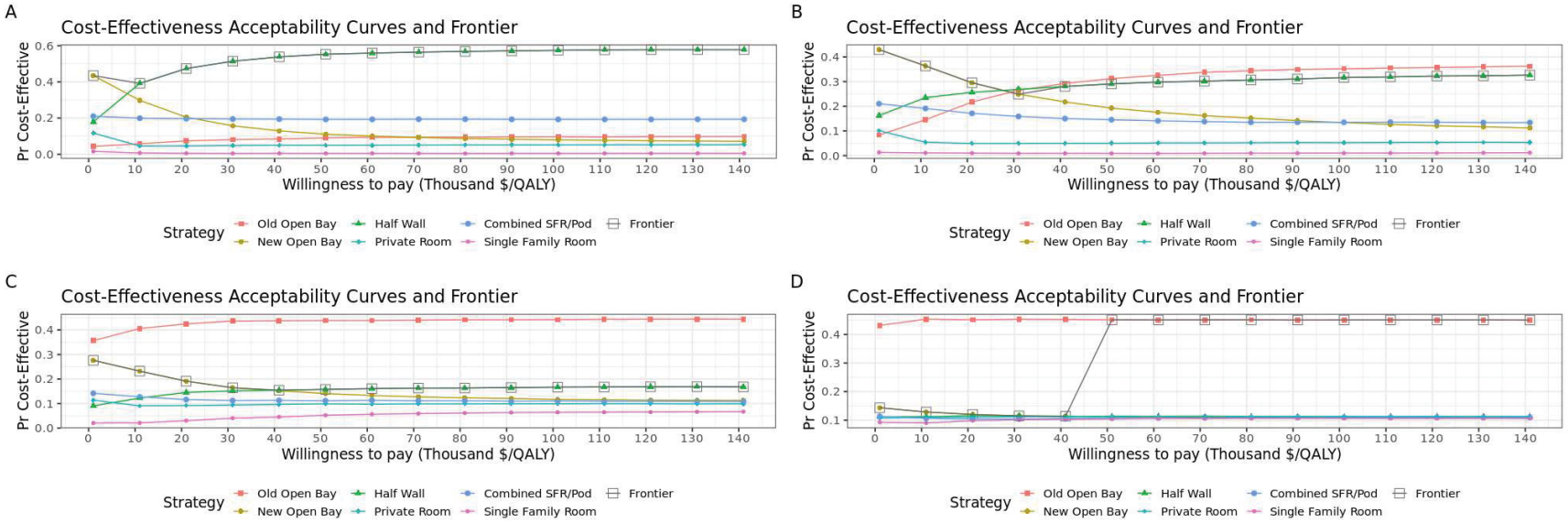
Age	Alive	Died	Probability of Death
0 year	100,000	453	0.45
1 year	99,547	25	0.03
2 years	99,522	18	0.02
3 years	99,504	14	0.01
4 years	99,490	11	0.01
5 years	99,479	9	0.01
6 years	99,470	8	0.01
7 years	99,462	8	0.01
8 years	99,454	7	0.01
9 years	99,447	8	0.01
10 years	99,439	8	0.01
11 years	99,431	9	0.01
12 years	99,422	10	0.01
13 years	99,412	12	0.01
14 years	99,399	16	0.02
15 years	99,384	20	0.02
16 years	99,363	27	0.03
17 years	99,337	33	0.03
18 years	99,303	39	0.04
19 years	99,264	45	0.05
20 years	99,220	50	0.05
21 years	99,169	55	0.06
22 years	99,114	59	0.06
23 years	99,056	62	0.06
24 years	98,994	63	0.06
25 years	98,930	65	0.07
26 years	98,866	66	0.07
27 years	98,800	68	0.07
28 years	98,732	70	0.07
29 years	98,662	72	0.07
30 years	98,590	75	0.08
31 years	98,515	78	0.08
32 years	98,438	80	0.08
33 years	98,358	82	0.08
34 years	98,276	84	0.09
35 years	98,192	85	0.09
36 years	98,107	87	0.09
37 years	98,020	90	0.09
38 years	97,931	94	0.10
39 years	97,837	99	0.10
40 years	97,738	106	0.11

Age	Alive	Died	Probability of Death
41 years	97,632	115	0.12
42 years	97,517	124	0.13
43 years	97,394	134	0.14
44 years	97,260	144	0.15
45 years	97,116	156	0.16
46 years	96,959	169	0.17
47 years	96,790	183	0.19
48 years	96,607	198	0.21
49 years	96,409	215	0.22
50 years	96,194	233	0.24
51 years	95,962	253	0.26
52 years	95,709	274	0.29
53 years	95,435	298	0.31
54 years	95,137	323	0.34
55 years	94,814	352	0.37
56 years	94,462	382	0.41
57 years	94,080	416	0.44
58 years	93,664	452	0.48
59 years	93,212	492	0.53
60 years	92,720	536	0.58
61 years	92,184	584	0.63
62 years	91,600	636	0.69
63 years	90,964	692	0.76
64 years	90,272	754	0.84
65 years	89,517	822	0.92
66 years	88,696	895	1.01
67 years	87,801	975	1.11
68 years	86,826	1,061	1.22
69 years	85,765	1,155	1.35
70 years	84,611	1,256	1.48
71 years	83,355	1,365	1.64
72 years	81,991	1,482	1.81
73 years	80,508	1,608	2.00
74 years	78,900	1,742	2.21
75 years	77,158	1,885	2.44
76 years	75,273	2,036	2.71
77 years	73,236	2,196	3.00
78 years	71,041	2,362	3.32
79 years	68,679	2,533	3.69
80 years	66,146	2,710	4.10
81 years	63,436	2,888	4.55
82 years	60,548	3,065	5.06
83 years	57,483	3,239	5.63
84 years	54,244	3,404	6.28
85 years	50,840	3,556	6.99
86 years	47,285	3,688	7.80

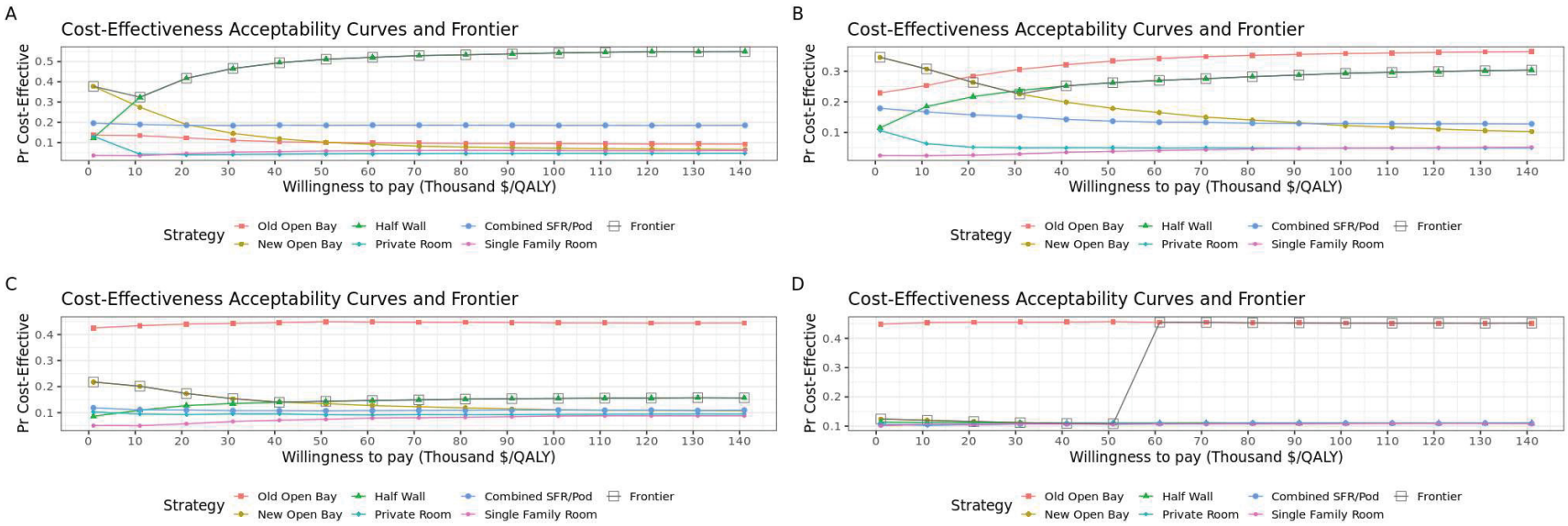
Age	Alive	Died	Probability of Death
87 years	43,597	3,796	8.71
88 years	39,801	3,870	9.72
89 years	35,931	3,905	10.87
90 years	32,025	3,894	12.16
91 years	28,132	3,819	13.58
92 years	24,312	3,671	15.10
93 years	20,642	3,451	16.72
94 years	17,190	3,170	18.44
95 years	14,020	2,849	20.32
96 years	11,171	2,482	22.22
97 years	8,689	2,102	24.19
98 years	6,587	1,728	26.24
99 years	4,859	1,377	28.34
100 years	3,482	1,062	30.49
101 years	2,420	790	32.66
102 years	1,630	568	34.84
103 years	1,062	393	37.00
104 years	669	262	39.14
105 years	407	168	41.22
106 years	239	103	43.25
107 years	136	61	45.19
108 years	74	35	47.05
109 years	39	19	48.81
110 years and over	20	20	100.00



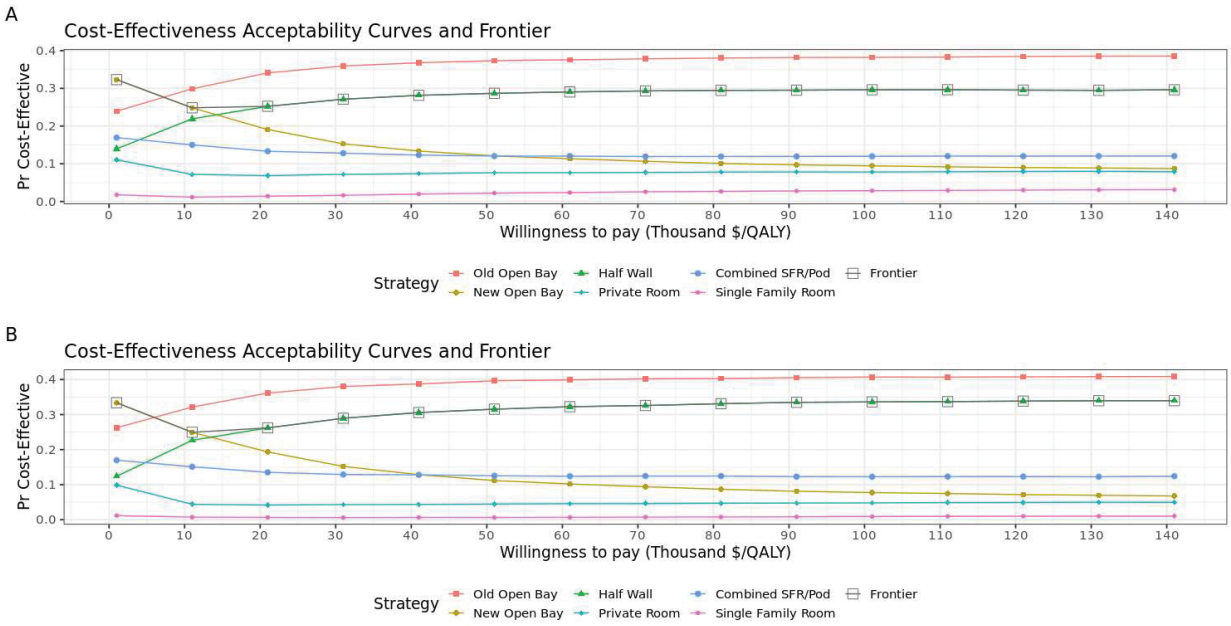
Appendix Figure A. 16. Cost-effectiveness acceptability curves and frontier for the over-all group using the healthcare payer's perspective. Sensitivity analysis including Puumala et al in the efficacy data. Treatments higher on the y axis have a higher probability of being the cost-effective treatment at a given willingness to pay threshold. Gray boxes indicate the treatment with the highest expected value of the net monetary benefit.



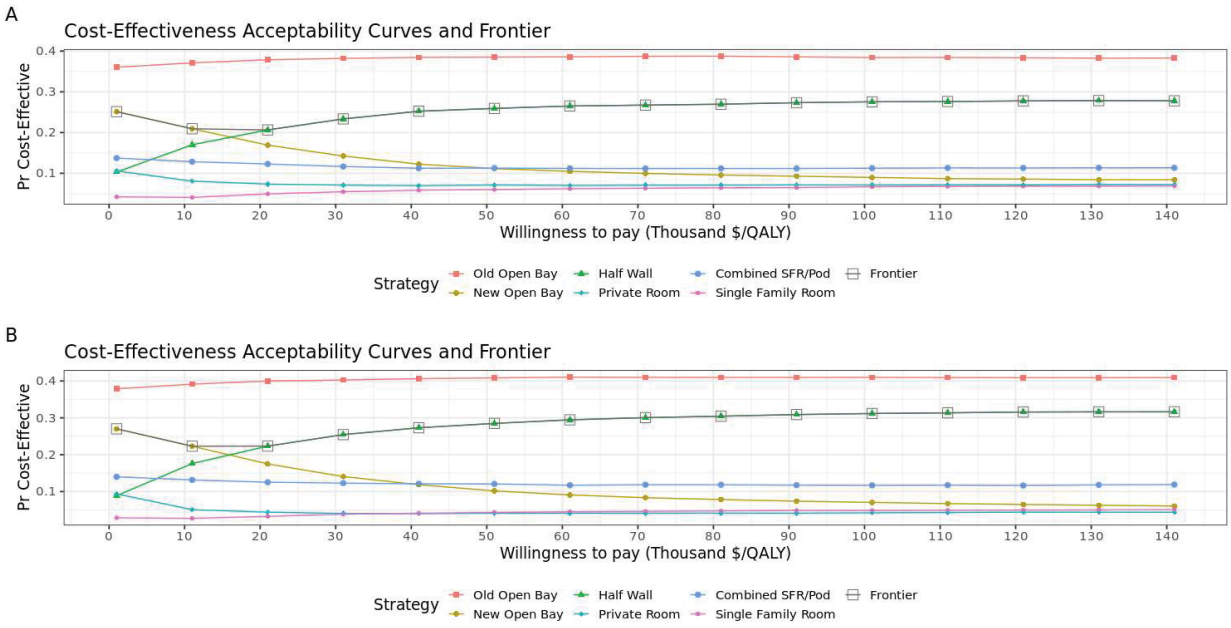
Appendix Figure A. 17. Sensitivity analysis including Puumala et al in efficacy data. Cost-effectiveness acceptability curves and frontier for extremely preterm (A), very preterm (B), moderately preterm (C), and term (D) gestational age sub-groups. Treatments higher on the y axis have a higher probability of being the cost-effective treatment at a given willingness to pay threshold. Gray boxes indicate the treatment with the highest expected value of the net monetary benefit.



Appendix Figure A. 18. Societal perspective: Cost-effectiveness acceptability curves and frontier for extremely preterm (A), very preterm (B), moderately preterm (C), and term (D) gestational age sub-groups. Treatments higher on the y axis have a higher probability of being the cost-effective treatment at a given willingness to pay threshold. Gray boxes indicate the treatment with the highest expected value of the net monetary benefit.



Appendix Figure A. 19. Sensitivity analysis including Puumala et al in efficacy data. Societal perspective: Cost-effectiveness acceptability curves and frontier for lower income (A), and higher income (B) sub-groups. Treatments higher on the y axis have a higher probability of being the cost-effective treatment at a given willingness to pay threshold. Gray boxes indicate the treatment with the highest expected value of the net monetary benefit.



Appendix Figure A. 20. Societal perspective: Cost-effectiveness acceptability curves and frontier for lower income (A), and higher income (B) sub-groups. Treatments higher on the y axis have a higher probability of being the cost-effective treatment at a given willingness to pay threshold. Gray boxes indicate the treatment with the highest expected value of the net monetary benefit.