Discrete choice experiment to develop a summary metric to determine successful treatment of paediatric surgery across multiple conditions: a study protocol

Oliver Rivero-Arias, Benjamin Allin, Marian Knight

Version 1.0 01 Dec 2020

Table of contents

1 Introduction

Around half a million children need surgery in England and Wales every year [1]. In 2011, the National Confidential Enquiry into Patient Outcomes and Death (NCEPOD) conducted a review of children's surgery in the UK and concluded that outcomes were not appropriate and challenges in the surgical decision-making process were noted as one reason for the poor outcomes [2]. These challenges reflect variation in the management of key conditions as reported in different studies leading to within-country variation in outcomes [3-10]. Variation in the management of conditions inevitably exposes children to poorer outcomes, but can also affect the wellbeing and quality of life of carers of children treated by paediatric surgeons [11]. Reducing unwanted variation in the management of paediatric surgical conditions is possible, but three barriers complicate this process for UK policy makers: 1) an inability to detect variation in meaningful outcomes between centres, 2) an inability to develop evidence-based management guidelines and 3) a failure to implement the recommendations of developed guidance. Our NIHR funded study "Improving unwarranted variation in outcomes of children's surgery through a new Children's Surgery Outcome Reporting system using routinely available data (C-SOR)" has been funded to investigate whether one unified system is capable of addressing these three issues and therefore reduce unwanted variation in surgical care in the National Health Service (NHS). In this manuscript, we describe the methodology of one of the C-SOR sub-studies tackling the first barrier.

To understand outcome variation of children's surgeries across centres within a jurisdiction, national data with appropriate information is needed. In England, the Getting It Right First Time (GIRFT) initiative is a national programme designed to improve medical care within the NHS by reducing unwarranted variations (https://www.gettingitrightfirsttime.co.uk). The programme links routine national data from different sources of a number of medical and surgical specialties with the aim of detecting variation in outcomes, and publishing provider- and national-specific reports. One of the surgical specialities within GIRFT is general paediatric and neonatal surgery. However, recent work has identified limitations in the information available in the linked datasets to assess variation in outcomes in that area. In particular, it has been indicated that there is a lack of parent-reported outcomes and long-term outcome data, and that currently routinely captured data in the NHS does not adequately include information to identify management differences across centres [12].

The gold standard approach for comparing outcomes of interest to patients between centres within a jurisdiction is to use a core outcome set (COS) [13]. In the context of paediatric surgery, a COS is a collection of standardised outcomes agreed by relevant stakeholders that should be measured and reported, as a minimum, in all published reports falling within the scope of that COS. For example, where the outcomes of two interventions or centres are compared for a condition in which a COS has been developed, the COS should be used as the starting point to identify the outcomes against which the interventions/centres are compared. Several COS have recently been developed that are relevant to children's surgery, including one for Hirschsprung's disease, one for gastroschisis, one that is relevant to all infants receiving neonatal care, and one for paediatric appendicitis[7, 14-16], and development of these COS has opened a window of opportunity to improve the measurement of outcomes of paediatric surgery. Whilst the nature of paediatric surgical conditions suggests that many outcomes within a COS will be specific to that condition, some universality exists, with several outcomes being repeated across the identified paediatric surgical core outcome sets. These common outcomes could be compared across conditions to understand between-centre variation. To make such comparison meaningful, a summary metric that categorised a child's outcome into, for instance, `successful` or `unsuccessful` from a combination of common core outcomes is needed. The aim of this study is to develop an algorithm to assist in defining this summary metric, and therefore determining successful treatment of children undergoing surgery for a variety of conditions.

Whether a certain combination of common core outcomes across conditions indicates a successful or unsuccessful surgical result depends on the value that relevant stakeholders place on the different core outcomes. Economists employ a wide range of preference elicitation techniques to determine the value of health, treatments or health care services [17]. These are broadly classified into revealed preferences (observed choices by individuals in a real-life scenario) and stated preferences (observed choices by individuals presented with a realistic hypothetical scenario) [18]. It is very difficult (or virtually impossible) to estimate the value parents or health care professionals place on core health outcomes observing only their behaviour. Stated preference techniques such as discrete choice experiments (DCE) are better suited to understand the value of potential combinations of core outcomes of paediatric surgery [19].

A DCE is a technique for eliciting preferences that provides information about how individuals value different attributes in a scenario [19]. The theoretical underpinnings supporting DCEs specifies that the value of a scenario depends on its attributes, which are the characteristics of health, treatments or health care services being evaluated [19]. In paediatric surgery these characteristics may include the survival of the infant or child, adverse events associated with surgery, or the child or carer's quality of life. DCEs are based on the assumption that a potential outcome after surgery can be described by key characteristics, and that an individual's valuation depends on the levels of the attributes. During a DCE, participants are presented with a number of vignettes, are asked to make trade-offs among the attributes described and make a choice regarding their preference. Vignettes in a DCE can take different formats, but pair comparisons, where the participant is presented with 2 or more scenarios and asked to choose one are the most widely used in the literature [20]. Participants choices are analysed using discrete choice analysis and econometric techniques, where patterns of responses are associated with combinations of attributes and levels to determine participant's preferences and their relative importance.

In this manuscript we describe a protocol for a DCE designed to estimate the value parents and health care professionals place on different combinations of health outcomes following surgery in childhood. Attributes of a potential outcome after surgery are identified from core health outcomes common across paediatric surgical conditions. Stakeholders will be presented with a pair comparison non-adaptive task and a novel DCE adaptive task to overcome shortcomings of traditional formats as identified in recent studies. The product of the discrete choice analysis will be an algorithm to determine successful treatment of paediatric surgery categorised as "successful" or "unsuccessful" outcome. The feasibility of implementing such an algorithm in real practice will be evaluated in a subsequent phase of the C-SOR programme. The use of stakeholder preferences to help in health care decision-making by policy makers has increased considerably across most developed jurisdictions [20]. This has been accompanied by best practice guidance that provide detailed information about what methodology to use for each of the different steps involved in a DCE [21-23]. We will employ this guidance to inform the conduct and statistical analysis of this DCE.

1.1 Aims

This study aims to understand how parents and health care professionals value common health outcomes following surgery in childhood across multiple conditions using a stated preference exercise. The specific objectives of the study are:

A. To estimate the relative importance of key health outcomes following surgery in childhood for multiple conditions using a discrete choice experiment.
B. To compare a novel DCE format called adaptive tasks with a standard non-adaptive task in the context of paediatric surgery.
C. To estimate an algorithm using weights derived from the relative importance estimates in A to derive a summary metric that categorises outcomes following surgery in childhood into "successful" or "unsuccessful" outcome.

2 Methods and analysis

2.1 Overview of framework for the discrete choice experiment (DCE)

Figure 1 describes the framework and different phases that will be followed to conduct this DCE. This protocol describes the following sections: 1) identification and description; 2) experimental design, 3) survey instrument and 4) statistical analysis.

2.2 Identification and description

2.2.1 Decision model and descriptive framework

Our research question explores how to best summarise successful outcomes following surgery in childhood from the values that relevant stakeholders place on key core outcomes across paediatric surgical conditions. Our decision model hypothesises that a successful outcome following surgery in childhood can be represented by a combination of characteristics or attributes. In this study, attributes are defined as core health outcomes included in available core outcome sets relevant to paediatric surgery. The attributes and associated levels that describe potential outcomes following surgery in childhood defines our descriptive framework.

Conceptual attributes for use in the DCE have been identified through a review of published COS relevant to paediatric surgery, as well as focussed discussions with a Parent Advisory Group consisting of parents of children who have undergone early surgery for conditions including Hirschsprung's disease, gastroschisis and necrotising enterocolitis. Relevant COS have been developed for children undergoing surgery for Hirschsprung's disease, gastroschisis, and appendicitis, and for children receiving neonatal care in a high-income setting[7, 14-16]. Each of these COS were developed using a combination of literature reviews, an online Delphi process, and consensus meetings, and included in their stakeholder groups, clinicians, allied health professionals, parents, and children or adults previously treated for the target condition. The core outcomes identified in these four conditions are presented in Table 1.

An iterative process to the identification of attributes was followed. In a first step, overlap in outcomes of importance was identified between the four COS. Each COS also identified outcomes that were relevant only to the condition of interest, and not represented in the other COSs. In the case of Hirschsprung's disease for example, problems associated with bowel function are not present in the other three conditions. Similarly, necrotising enterocolitis or brain injury on imaging are issues primarily associated with infants receiving neonatal care. In a second step, we reviewed these condition-specific outcomes with the Parent Advisory Group in order to determine how best to represent them within the DCE. The group concluded that these condition-specific outcomes were highly likely to impact the child's overall quality of life and would therefore be adequately represented in the DCE through the attribute quality of life. Three outcome categories, survival, adverse events, and health-related quality of life were common to all four relevant COS and therefore selected as the initial set of attributes for the DCE. Similar outcomes have also been identified as important in other developed paediatric COS [24, 25]. The specific adverse events identified from each COS were discussed with members of the Parent Advisory Group and paediatric surgeons on the CSOR steering committee. Both groups agreed that the main adverse events could be summarised as readmission, significant infection, and reoperation. A description of each attribute is given next:

1. Survival

Although death is relatively uncommon following most surgery in childhood, it is such a significant outcome that all four core outcome sets relating directly to childhood surgery, and the majority of

paediatric core outcome sets include it. For the purposes of the DCE, this outcome will be presented positively as survival, as opposed to negatively as death.

2. Health-related quality of life

Each of the core outcome sets included the outcome quality of life, whilst some also specifically included outcomes relating to psychological wellbeing. There are multiple definitions of health-related quality of life, and multiple tools, including the EQ5D, PedsQL and SEIQoL-DW that can be used to measure it. These tools generally describe multiple domains, including social functioning, physical functioning and psychological wellbeing, with their output generally reported in a continuous manner. However, for the purposes of the DCE, only overall quality of life will be described, and this will be done in a categorical fashion. The impact of key condition specific outcomes such as faecal incontinence, need for parenteral nutrition, and liver disease will be reflected in the child's overall health related quality of life.

3. Readmission

Readmission following paediatric surgery occurs for a wide variety of reasons. Some readmissions are unplanned, some are planned, some will require medical intervention only, and some will require surgery. Duration of readmission can vary widely dependent on the reason for readmission, and many children will require multiple readmissions. Overall, readmission to hospital as a direct or indirect result of children's surgery reflects the development of many of the condition-specific adverse events identified in the core outcome sets. For the purposes of the DCE, readmission to hospital will be described according to length of stay and frequency of readmission.

4. Significant infection

Each of the core outcome sets identified as relevant to paediatric surgery include condition specific significant infective complications, such as Hirschsprung's associated enterocolitis, necrotising enterocolitis, and intra-abdominal abscess. Some also include a more generalised measure of significant infection, *sepsis.* For the purposes of the DCE, the infective complications included in each COS will be represented by the attribute significant infection, and the levels will more specifically define the broad cause and frequency of infections.

5. Reoperations

Each of the surgical COSs includes a measure of the operative burden for the child, either as a total number of operations performed, or as a measure of whether the child had any unplanned reoperations. Operations are classified by the National Institute for Health and Care Excellence

(NICE) as minor, intermediate or major. For the purposes of the DCE, this will be simplified to describe operations as minor or major and will additionally categorise them as emergency or planned. A measure of the total number of operations performed will also be included in the reoperation's levels. The importance of operations that are not directly linked to the condition for which the child initially underwent surgery, for example a fractured limb, are not intended to be captured by this attribute.

The identification of the attribute levels also employed an iterative process. In a first step, we reviewed the epidemiological data available for each of the attributes, to guide the range that could be presented to participants. Existing large scale cohort studies describing the outcomes for children with any of the six conditions for which CSOR will initially collect data were reviewed [4, 6, 8, 9, 26-35]. One of our researchers (BA) extracted point estimates and associated measures of uncertainty (if reported) for each of the attributes. In discussions with a second researcher (OR-A) initial deterministic ordinal levels for each attribute were developed. This initial list of ordinal levels was presented to the paediatric surgeons collaborating on CSOR to ensure their clinical appropriateness. Surgeons suggested changes to the wording which were implemented and a first draft of the ordinal levels for each attribute was created. The language used to describe these will be refined following review by members of the Parent Advisory Group in a final step. The proposed attributes and preliminary attribute levels are described in the descriptive framework in Table 2.

2.2.2 Elicitation task and format

Two elicitation formats will be used in this study to estimate preferences: a standard non-adaptive task and a novel adaptive task. The standard task will include two scenarios describing outcomes of paediatric surgery without an opt-out option and an example is presented in Figure 2. This type of pair comparison is the most widely used DCE format in health care [20] and it has been used previously to elicit preferences associated to different aspects of health care or outcomes of surgery [36-38].

Recent work has reported that common approaches to elicit preferences from adults, including standard DCE, are problematic in the context of child health. Eliciting values to inform decision making at the start of life or early childhood requires stakeholders to complete tasks from someone

else's point of view. There is some evidence that when the tasks in an elicitation exercise refers to someone else instead of our own preferences, individuals find the exercise strenuous. This is accentuated when the perspective is that of a new-born or a young individual [39]. In the example of Figure 1, it is not difficult to find arguments that the decision to choose between two undesirable outcomes of paediatric surgery is not an easy one. Individuals from the general public completing a task like that one would end up making a choice if the scenario does not involve an opt-out. However, it is a task they do not enjoy as participants find difficult to relate to the decision context [40]. This can also affect the preferences elicited in the DCE. It is not clear whether stakeholders familiar with the decision context find standard DCE easier than members of the general public.

An alternative to a non-adaptive DCE task is to ask participants to complete an adaptive design (Figure 3). In an adaptive design, participants are presented first with a single profile of an outcome following surgery in childhood, and are asked to rate the profile as poor, fair or good. Then, participants are asked to make improvements in the profile based on available attribute changes that are available from a second outcome following surgery in childhood. Attribute changes are added to the profile and at the end of the task, participants are asked to complete a standard DCE with the original profile alongside the improved one, and to rate the new profile. If the task has been understood, then all participants would select the new improved profile over the original one, and provide additional information for analysis about what constitutes a poor, fair or good outcome. The mathematical formulation of adaptive designs are similar to rank-ordered sequential stated preference elicitation methods such as first (best), second (best), etc. or repeated best-worst, widely used in marketing and transport economics [41, 42]. In the context of this study, an adaptive DCE design mirrors the clinical decision-making process by paediatric surgeons where they prioritise outcomes in terms of need.

2.3 Experimental design

Two separate experimental designs will be generated for each of the elicitation tasks (non-adaptive, adaptive). The non-adaptive design will generate a set of choice tasks using a D-efficiency approach to maximise efficiency and minimise standard error of coefficients. Choice tasks will be divided into blocks in which participants will be randomly allocated. Each task will include a choice set with two

outcomes of paediatric surgery and no opt-out. We will exclude implausible and dominated outcomes and include overlaps (i.e., similar attribute levels outcomes of surgery in childhood). The adaptive design will employ a fractional factorial that allows the estimation of main effects and maintain the properties of near level balance and near orthogonality. It is unlikely that complete orthogonality will be possible because we will exclude implausible outcomes of surgery in childhood.

2.4 Survey instrument

The survey will be administered online and will be programmed in Oxford University servers with the open source platform LimeSurvey (https://community.limesurvey.org). The survey will consist of an initial consent form, followed by a general welcome and an introduction to the research question. Next, participants will be asked to provide background information, age, sex, and experience with neonatal/childhood surgical conditions. For health care professionals we will also ask their job title and level of professional experience with neonatal/childhood surgical conditions. This will be followed by the non-adaptive tasks including a warm-up question and then the adaptive tasks that will also include a warm-up question. At the end of the survey participants will be given the opportunity to clarify whether they found difficulties completing the non-adaptive and adaptive tasks and state their preference towards one of the methods. The face validity of our survey instrument will be discussed with our Parent Advisory Group.

2.5 Statistical evaluation

2.5.1 Data collection, recruitment strategy and sampling

The survey instrument will be completed by a sample of parents of children who have undergone early surgery, and health professionals caring for those who undergo surgery in childhood, Mothers/fathers/carers will be invited to complete the survey separately but will be categorised as the same household in our data collection. Recruitment of parents and carers will be conducted through parent support groups including the CHAMPS appeal, Avery's Angels, TOFs, CDH-UK, and NEC UK, and our Parent Advisory Group. We will make use of the registry and mailing lists of the British Association of Paediatric Surgeons to recruit surgeons to the study. Both these recruitment strategies have been successfully used in previous quantitative and qualitative studies[7, 14, 43-45]

Sample size calculations for DCE remains an area of controversy without consensus and several approaches have been proposed. The most recent ISPOR guidance suggests that statistical precision (i.e. smaller standard errors) increases with sample sizes of 150 and after 300 observations the marginal increase in statistical precision is almost negligible [46]. Lancsar et al. suggests a minimum of 20 observations per choice set to estimate a model capable of capturing statistical significance in the coefficients if they provide important information about preferences [47]. Another widely used technique was developed by Orme that suggests a rule of thumb based on the number of levels for any of the attributes, and the number of alternatives in the choice set [48]. In this study we aim to collect 200 responses from parents and 200 responses from clinicians (total of 400) over a three-month period. Participants will all be recruited from high-income Western countries.

2.5.2 Data analysis

DCE responses to the non-adaptive and adaptive tasks will be analysed within the random utility model (RUM). Separate analyses using this model will be conducted for each task. In this protocol, we describe the RUM nomenclature for the non-adaptive tasks and refer the reader to the mathematical formulation of ordinal rank models for the adaptive one [19].

The random utility model assumes that the different paediatric outcomes presented to participants generate a level of preference (also known as utility) that constitutes a successful or unsuccessful outcome and that the individual selects the alternative that represents a better outcome following childhood surgery. The overall utility (U_i) for the *i*th alternative is divided into a component that is observed and can be explained by the analyst (V_i) and contributions that are not observed (ε_i). The relationship between the explainable and unexplainable components is assumed to be independent and additive so the level of utility for the *i*th alternative is described as:

$$U_i = V_i + \varepsilon_i$$

V_i is also known as the “representative component of utility” because it can be explained through the attributes that are observed in the DCE. The relative contribution of each attribute to the overall utility

can be represented by a weight (i.e. a coefficient or parameter) that in its simplest form can take the form of a linear expression:

$$V_i = \beta_{oi} + \beta_{1i}X_{1i} + \beta_{2i}X_{2i} + \cdots + \beta_{ki}X_{ki}$$

where

β_{1i} is the parameter associated with the attribute X_{1i} and alternative i

β_{oi} is a parameter that is not associated to any of the observed attributes and represents an *alternative-specific constant* indicating on average the role of all the unobserved sources of utility

Different assumptions can be made about the error term ε_i but often for simplicity and a good starting point for the selection of choice models, it is assumed to be independent and with the exact same distribution (identically distributed) among alternatives. These sets of assumptions are known as IID (independent and identically distributed). Under the random utility model, the individual evaluates each alternative represented as $U_j; j = 1, \ldots J$ alternative, and compares $U_1, U_2, U_3, \ldots, U_j$ selecting the alternative with the highest utility (better outcome of paediatric surgery), i.e. max (U_j). Therefore, the probability of selecting a specific alternative i compared to an alternative j can be expressed as:

$$Prob_i = Prob\big(U_i \geq U_j\big) \; \forall \; j \; \in j = 1, \ldots, J : i \neq j$$

In words, the probability of an individual choosing alternative i is equal to the probability that the utility of alternative i is greater than (or equal to) the utility associated with alternative j after evaluating each and every alternative in the choice set of $j = 1 \ldots i \ldots J$ alternatives.

This is equivalent to:

$$Prob_i = Prob\big[(V_i + \varepsilon_i) \; \geq \big(V_j + \varepsilon_j\big) \; \forall \; j \; \in j = 1, \ldots, J : i \neq j\big]$$

Also equivalent to:

$$Prob_i = Prob\big[\big(\varepsilon_j + \varepsilon_i\big) \; \leq \big(V_i + V_j\big) \; \forall \; j \; \in j = 1, \ldots, J : i \neq j\big]$$

In words, the probability of an individual choosing alternative i is equal to the probability that the difference in the unobserved sources of utility of alternative i compared to j is less than (or equal to) the difference in the observed sources of utility associated with alternative i compared to alternative j after evaluating each and every alternative in the choice set of $j = 1 \ldots i \ldots J$ alternatives.

This final expression indicates that to estimate the probability of an alternative i being selected compared to an alternative j, we need information on V_i and V_j (we can directly observe information on attributes and levels) and information on ϵ_i and ε_j (that we do not observe and in fact we have no idea what this looks like). Therefore, to estimate the probability of i being selected, we need to impose some structure for ε that helps us in identifying a practical choice model. The structure of the random component takes the form of a statistical distribution and a common distribution used in discrete choice analysis is the extreme value type 1 (EV1). The final selected choice model creates a relationship between the observed attributes, the unobserved attributes and the stated choice outcome. Under EV1 and IID assumptions we can derive the most widely used choice model known as multinomial logit (MNL), which will be the initial discrete model used in this study to estimate preferences. The predicted probabilities of an alternative i being selected from the complete set of alternatives $j = 1, \ldots, J$ in a MNL are given by:

$$Prob_i = \frac{expV_i}{\sum_{j=1}^{j} expV_i}; j = 1, \ldots, i, \ldots, J: i \neq j$$

Unobservable preference heterogeneity in each of the samples will be explored using latent class models. We will compare preferences between the two samples and two format strategies as summarised in Table 3. Preferences eliciting using non-adaptive and adaptive formats will be compared within each sample of parents and clinicians (AvC and BvD) and within each DCE format (AvB and CvD). Preferences will be compared examining the relative attribute importance (RAI) scores by dimension. This approach involves estimating the utility range for each attribute and subsequently applying a normalisation to enable sample comparisons. We will implement an attribute-based normalisation and will follow best practice guidance to report the results of this component of the analysis [49].

A final pooled model combining responses from parents and clinicians will be estimated for each format separately (A+B and C+D). Non-adaptive and adaptive responses will not be combined or estimated together and the final choice model for the algorithm will be based on the feedback obtained from participants and the face validity of the preferences obtained. Using the coefficients from the preferred elicitation strategy, we will predict the combination of attributes and levels with the highest level of utility, in other words, the combination of attributes and levels that will have higher probabilities of being selected as “better outcome” by parents and health care professionals. We will evaluate then the distribution of predicted utilities using quantiles to determine the cut-offs points that will divide such distribution into successful and unsuccessful outcomes. If the adaptive design is selected as the preferred source of preference information, we will also make use of the rating information provided by participants to determine what pool of outcomes are considered successful and unsuccessful combinations.

2.6 Ethics and dissemination

2.6.1 Ethical considerations

Ethics approval to conduct this study will be obtained from the Medical Sciences Inter-Divisional Research Ethics Committee (IDREC) at the University of Oxford. Informed consent will be obtained for all participants at the start of the survey. All participants will be informed that the survey is completely anonymous, and that no identifiable data will be collected at any time.

2.6.2 Dissemination

The outcome of the DCE exercise and the resulting algorithm will be disseminated through peer-review publications and scientific presentations. A lay summary of the findings will be created using our Parent Advisory Group and circulated to parent support networks and the British Association of Paediatric Surgeons.

3 References

1. Tanner, S., *Trends in children's surgery in England.* Arch Dis Child, 2007. **92**(8): p. 664-7.
2. Mason, D.G., et al., *Are we there yet- a review of organisational and clinical aspects of children's surgery*, in *National Confidential Enquiry into Patient Outcomes and Death (NCEPOD)*. 2011: London.
3. Owen, A., et al., *Gastroschisis: a national cohort study to describe contemporary surgical strategies and outcomes.* J Pediatr Surg, 2010. **45**(9): p. 1808-16.
4. Bradnock, T.J., et al., *Gastroschisis: one year outcomes from national cohort study.* BMJ, 2011. **343**: p. d6749.
5. Burge, D.M., et al., *Contemporary management and outcomes for infants born with oesophageal atresia.* Br J Surg, 2013. **100**(4): p. 515-21.
6. Allin, B., et al., *Outcomes at One-Year Post Anastomosis from a National Cohort of Infants with Oesophageal Atresia.* Plos One, 2014. **9**(8).
7. Allin, B.S.R., et al., *NETS(1HD) study: development of a Hirschsprung's disease core outcome set.* Arch Dis Child, 2017. **102**(12): p. 1143-1151.
8. Allin, B.S.R., et al., *One-year outcomes following surgery for necrotising enterocolitis: a UK-wide cohort study.* Arch Dis Child Fetal Neonatal Ed, 2018. **103**(5): p. F461-F466.
9. Long, A.M., et al., *Oesophageal atresia with no distal tracheoesophageal fistula: Management and outcomes from a population-based cohort.* Journal of Pediatric Surgery, 2017. **52**(2): p. 226-230.
10. Wilkinson, D.J., et al., *Hypospadias surgery in England: Higher volume centres have lower complication rates.* Journal of Pediatric Urology, 2017. **13**(5).
11. Hinton, L., et al., *What can make things better for parents when babies need abdominal surgery in their first year of life? A qualitative interview study in the UK.* BMJ Open, 2018. **8**(6): p. e020921.
12. Allin, B.S.R., et al., *Variability of outcome reporting in Hirschsprung's Disease and gastroschisis: a systematic review.* Scientific Reports, 2016. **6**(1): p. 38969.
13. Williamson, P.R., et al., *The COMET Handbook: version 1.0.* Trials, 2017. **18**(Suppl 3): p. 280.
14. Sherratt, F.C., et al., *Core outcome set for uncomplicated acute appendicitis in children and young people.* Br J Surg, 2020. **107**(8): p. 1013-1022.
15. Allin, B.S.R., et al., *Development of a gastroschisis core outcome set.* Arch Dis Child Fetal Neonatal Ed, 2019. **104**(1): p. F76-F82.
16. Webbe, J.W.H., et al., *Core outcomes in neonatology: development of a core outcome set for neonatal research.* Arch Dis Child Fetal Neonatal Ed, 2020. **105**(4): p. 425-431.
17. Weernink, M.G.M., et al., *A Systematic Review to Identify the Use of Preference Elicitation Methods in Healthcare Decision Making.* Pharmaceutical Medicine, 2014. **28**(4): p. 175-185.
18. Harrison, G.W., *Handbook of Choice Modelling*, in *Real Choices and Hypothetical Choices*. 2014.
19. Hensher, D.A., J.M. Rose, and W.H. Greene, *Applied Choice Analysis*. 2nd edition ed. 2015, Cambridge: Cambridge University Press.
20. Soekhai, V., et al., *Discrete Choice Experiments in Health Economics: Past, Present and Future.* Pharmacoeconomics, 2019. **37**(2): p. 201-226.
21. Bridges, J.F., et al., *Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force.* Value Health, 2011. **14**(4): p. 403-13.
22. Hauber, A.B., et al., *Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force.* Value in Health, 2016. **19**(4): p. 300-315.
23. Lancsar, E., D.G. Fiebig, and A.R. Hole, *Discrete Choice Experiments: A Guide to Model Specification, Estimation and Software.* Pharmacoeconomics, 2017. **35**(7): p. 697-716.
24. Sinha, I.P., et al., *Development of a core outcome set for clinical trials in childhood asthma: a survey of clinicians, parents, and young people.* Trials, 2012. **13**: p. 103.
25. Fink, E.L., et al., *Development of a core outcome set for pediatric critical care outcomes research.* Contemp Clin Trials, 2020. **91**: p. 105968.
26. Allin, B., et al., *A UK wide cohort study describing management and outcomes for infants with surgical Necrotising Enterocolitis.* Scientific Reports, 2017. **7**.
27. Allin, B.S.R., et al., *Outcomes at five to eight years of age for children with Hirschsprung's disease.* Archives of Disease in Childhood, 2020: p. archdischild-2020-320310.

28. Allin, B.S.R., et al., *Management of Gastroschisis: Results From the NETS2G Study, a Joint British, Irish, and Canadian Prospective Cohort Study of 1268 Infants.* Ann Surg, 2020.
29. Abellan-Perpiñan, J.M., H. Bleichrodt, and J.L. Pinto-Prades, *The predictive validity of prospect theory versus expected utility in health utility measurement.* Journal of Health Economics, 2009. **28**(6): p. 1039-1047.
30. Long, A.M., et al., *Early population-based outcomes of infants born with congenital diaphragmatic hernia.* Arch Dis Child Fetal Neonatal Ed, 2018. **103**(6): p. F517-F522.
31. Long, A.M., et al., *One-year outcomes of infants born with congenital diaphragmatic hernia: a national population cohort study.* Arch Dis Child Fetal Neonatal Ed, 2019. **104**(6): p. F643-F647.
32. Bradnock, T.J., et al., *Hirschsprung's disease in the UK and Ireland: incidence and anomalies.* Arch Dis Child, 2017. **102**(8): p. 722-727.
33. Brownlee, E., et al., *Current epidemiology and antenatal presentation of posterior urethral valves: Outcome of BAPS CASS National Audit.* J Pediatr Surg, 2019. **54**(2): p. 318-321.
34. Wragg, R., et al., *The postnatal management of boys in a national cohort of bladder outlet obstruction.* J Pediatr Surg, 2019. **54**(2): p. 313-317.
35. Wang, Y., et al., *One-year outcomes for congenital diaphragmatic hernia.* BJS Open, 2019. **3**(3): p. 305-313.
36. Ahmadi, S., et al., *Patient Preferences Around Extent of Surgery in Low-Risk Thyroid Cancer: A Discrete Choice Experiment.* Thyroid, 2020. **30**(7): p. 1044-1052.
37. van Dijk, J.D., et al., *An Empirical Comparison of Discrete Choice Experiment and Best-Worst Scaling to Estimate Stakeholders; Risk Tolerance for Hip Replacement Surgery.* Value in Health, 2016. **19**(4): p. 316-322.
38. Szawlowski, S., et al., *How do surgeons' trade-off between patient outcomes and risk of complications in total knee arthroplasty? a discrete choice experiment in Australia.* BMJ Open, 2019. **9**(7): p. e029406.
39. Rowen, D., et al., *Review of Valuation Methods of Preference-Based Measures of Health for Economic Evaluation in Child and Adolescent Populations: Where are We Now and Where are We Going?* Pharmacoeconomics, 2020. **38**(4): p. 325-340.
40. Craig, B.M., et al., *Valuation of Child Health-Related Quality of Life in the United States.* Health Econ, 2016. **25**(6): p. 768-77.
41. Lancsar, E., et al., *Best worst discrete choice experiments in health: methods and an application.* Soc Sci Med, 2013. **76**(1): p. 74-82.
42. Louviere, J.J., T.N. Flynn, and A.A.J. Marley, *Best-worst scaling: methods and applications*. 2015, Cambridge: Cambridge University Press.
43. Allin, B.S.R., et al., *Development of a gastroschisis core outcome set.* Arch Dis Child Fetal Neonatal Ed, 2019. **104**(1): p. F76-f82.
44. Allin, B.S.R., et al., *Outcomes at five to eight years of age for children with Hirschsprung's disease.* Arch Dis Child, 2020.
45. Long, A.-M., *Short and long-term outcomes of children born with abdominal wall defects* in *National Perinatal Epidemiology Unit*. 2017, University of Oxford.
46. Johnson, F.R., et al., *Constructing Experimental Designs for Discrete-Choice Experiments: Report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force.* Value in Health, 2013. **16**(1): p. 3-13.
47. Lancsar, E. and J. Louviere, *Conducting discrete choice experiments to inform healthcare decision making: a user's guide.* Pharmacoeconomics, 2008. **26**(8): p. 661-77.
48. Orme, B.K., *Getting started with conjoint analysis : strategies for product design and pricing research*. 2006, Madison, WI: Research Publishers, LLC.
49. Gonzalez, J.M., *A Guide to Measuring and Interpreting Attribute Importance.* Patient, 2019. **12**(3): p. 287-295.

Figure 1: A framework for discrete choice experiments

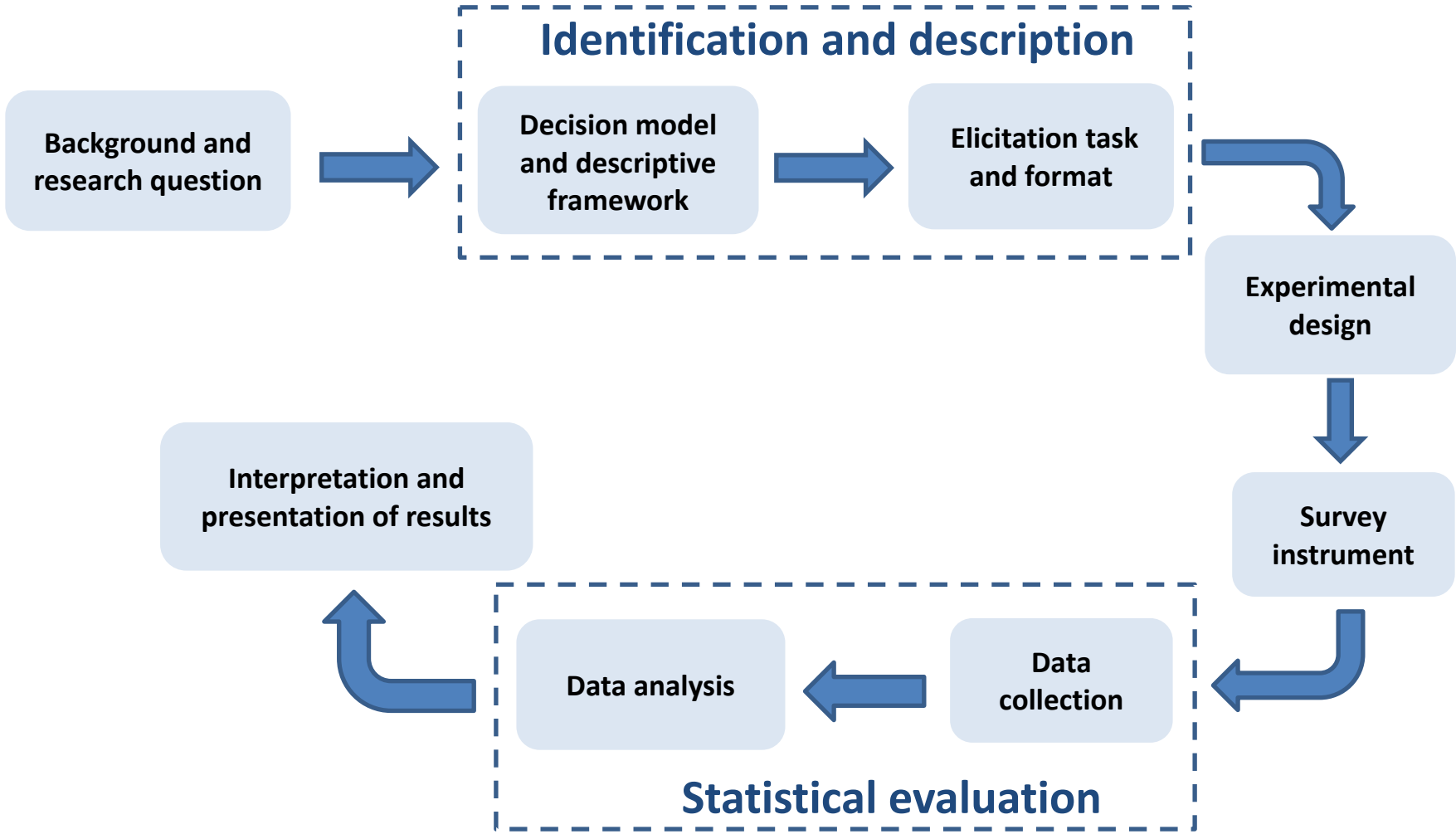

Table 1: Summary of identified core outcomes in neonatal conditions

Category	Core outcome	Included in core outcome set?			
		Hirschsprung's Disease	**Gastroschisis**	**Neonatal care**	**Appendicitis**
Survival	Survival	x	x	x	x
Quality of life	Quality of life	x	x	x	x
	Psychological stress	x			x
	Time away from full activity				x
Adverse events	Unplanned reoperation	x			x
	Number of operations		x		
	Severe gastrointestinal complication		x		
	Retinopathy of prematurity			x	
	Chronic lung disease			x	
	Bowel obstruction				x
	Readmission				x
	Length of hospital stay				x
	Significant infection		x	x	
	Hirschsprung's Associated Enterocolitis	x			
	Necrotising enterocolitis			x	
	Wound infection				x
	Wound complication				x
	Intra-abdominal abscess				x
Condition specific	Faecal incontinence	x			
	Bowel function score	x			
	Voluntary bowel movements	x			
	Urinary incontinence	x			
	Permanent stoma	x			
	Growth		x		
	Time on parenteral nutrition		x		
	Liver disease		x		
	Brain injury on imaging			x	
	Motor/cognitive/visual/hearing ability			x	
	Antibiotic failure				x
	Negative appendicectomy				x
	Recurrent appendicitis				x

Table 2: Preliminary descriptive framework

Attributes	**Attribute Levels**
Survival	- Low - High
Health-related quality of life at year 1	- Very good/Excellent - Good - Fair - Poor - Very poor
Significant infection	- Single significant infection during the child's initial admission, occurring secondary to a complication of the child's operation (e.g. anastomotic leak) - Single significant infection during the child's initial admission, but not as a result of a complication of the operation (e.g. a chest infection or central line infection) - Multiple significant infections during the child's initial admission, occurring secondary to a complication of the child's operation (e.g. anastomotic leak) - Multiple significant infections during the child's initial admission, but not as a result of a complication of the operation (e.g. a chest infection or central line infection) - Single significant infection occurring after the child was initially discharged (e.g. Hirschsprung's associated enterocolitis, lower respiratory tract infection, urinary tract infection) - Multiple episodes of significant infection occurring after the child was initially discharged (e.g. Hirschsprung's associated enterocolitis, lower respiratory tract infection, urinary tract infection)
Readmission	- Single unplanned readmission of <72 hours - Single unplanned readmission of 72 hours - 7 days - Single unplanned readmission of 8 days - 1 month - Single unplanned readmission of > 1 month - Multiple unplanned readmissions of any length - Single planned admission of any length

	- Multiple planned admissions of any length
Reoperation	- Single unplanned (emergency) minor operation - Single unplanned (emergency) major (operation - Multiple unplanned (emergency) minor operations - Multiple unplanned (emergency) operations of a mix of complexities - Single planned minor operation - Single planned major operation - Multiple planned minor operations - Multiple planned operations of a mix of complexities

Figure 1: A non-adaptive standard DCE task example

Question: Which of these two options represent a better outcome of paediatric surgery?

Attributes	**Outcome A**	**Outcome B**
Survival	High	High
HRQoL at year 1	Fair	Poor
Significant infection	Multiple significant infections	None
Readmission	Single unplanned readmission of 8 days - 1 month	Single unplanned readmission of 72 hours - 7 days
Reoperation	None	Single unplanned (emergency) major operation
Please select one	☐	☒

Figure 2: An adaptive DCE task example

Step 1: Presentation of single outcome of paediatric surgery and rating

Attributes	Outcome
Survival	Low
HRQoL at year 1	Fair
Significant infection	Multiple significant infections
Readmission	Single unplanned readmission of > 1 month
Reoperation	Single unplanned (emergency) major operation
Would you consider this profile	Poor? ☒ Fair? ☐ Good? ☐

Step 2: Which potential change do you choose first?

Attributes	Outcome	Potential change	Please select one
Survival	Low	High	☒
HRQoL at year 1	Fair	Very good/Excellent	
Significant infection	Multiple significant infections	None	
Readmission	Single unplanned readmission of > 1 month	Single unplanned readmission of 8 days - 1 month	
Reoperation	Single unplanned (emergency) major operation	Single unplanned (emergency) minor operation	

Step 3: Which potential change do you choose second?

Attributes	Outcome	Potential change	Please select one
Survival	High		
HRQoL at year 1	Fair	Very good/Excellent	☒
Significant infection	Multiple significant infections	None	
Readmission	Single unplanned readmission of > 1 month	Single unplanned readmission of 8 days - 1 month	
Reoperation	Single unplanned (emergency) major operation	Single unplanned (emergency) minor operation	

Step 4: Which potential change do you choose third?

Attributes	Outcome	Potential change	Please select one
Survival	High		
HRQoL at year 1	Very good/Excellent		
Significant infection	Multiple significant infections	None	☒
Readmission	Single unplanned readmission of > 1 month	Single unplanned readmission of 8 days - 1 month	
Reoperation	Single unplanned (emergency) major operation	Single unplanned (emergency) minor operation	

Step 5: Which of these two options represent a better outcome of paediatric surgery?

Attributes	Outcome A	Outcome B
Survival	High	Low
HRQoL at year 1	Very good/Excellent	Fair
Significant infection	None	Multiple significant infections
Readmission	Single unplanned readmission of > 1 month	Single unplanned readmission of > 1 month
Reoperation	Single unplanned (emergency) major operation	Single unplanned (emergency) major operation
Please select one	☒	☐
Would you consider this profile	Poor? ☐ Fair? ☐ Good? ☒	

Table 3: Summary of comparison of preference weights that will be conducted across samples and elicitation formats

	Parents	**Clinicians**	**Pool model**
Non-adaptive	A	B	A+B
Adaptive	C	D	C+D