

Outcome measures in cleft care: Increasing the sensitivity of the 5 Year Olds' Index.

Principle Applicant: Tarun Mittal

Scientific outline

Summary: As outcomes in Cleft care within the UK have improved there is a need to discriminate more sensitively between good outcomes. Using the 5 Year Olds' Index as a basis, this investigation will refine and modify the existing index to allow increased sensitivity and identification of good outcomes. The 5 Year Olds' Index will be modified in several stages by expert examiners. The CCUK and CSAG records of UCLP will be used. Once the Modified Index has been developed, the reliability will be tested on CCUK models. Should the Modified Index show suitable reliability then it will be proposed as the index of choice for assessing study models taken as part of audit protocol at 5 years of age. The Modified Index will be applied to the CCUK cohort of records by experienced and calibrated examiners. This will provide a baseline dataset for future comparison using the Modified 5 Year Olds' Index.

Keywords: 5 Year Olds' Index, Dentoalveolar outcomes, Reliability, Audit

Plan of investigation: As part of the CSAG and CCUK studies, models of non-syndromic 5-year-old children born with UCLP were obtained. There are 223 models available from the CSAG study and 198 in CCUK study, total of 421 5 year olds' models. All models have been previously scored using the 5 Year Olds' Index. These models will be obtained from the archive at the University of Bristol and provide a pool of models from which a sample of 100 study models (20 from each category) can be selected for development and testing of a Modified 5 Year Olds' Index. Random selection will be performed with computer generated random number tables but stratified on 5 Year Olds' Index categories. All models will be assigned unique identifiers and anonymised from the examiners.

Stage 1 – Subcategory development

The aim of Stage 1 is to develop a Modified 5 Year Olds' Index. The 100 selected models will be independently assessed by an expert panel of three orthodontists experienced in cleft care and calibrated in the use of the current 5 Year Olds' Index.

The panel will independently rank models in order of severity. Ranking orders will be compared and disagreement will be resolved by discussion to give a consensus ranking. Following this examiners will score each model with a 10 cm Visual Analogue Scale (VAS), blinded for original category, for comparison to the consensus ranking. Models will be grouped according to common characteristics into subcategories with verbal descriptors. The number of subcategories created will depend on the outcome of this first stage of development and the consensus opinion of the panel.

Stage 2 – Reliability assessment

Examiners will score 52 models with the Modified 5 Year Olds' Index on two separate occasions. A time period of three weeks between scoring sessions will minimise the effect of memory bias. Intra- and inter-examiner reliability testing will be conducted with weighted kappa analysis (Bland and Altman (1986), Sim and Wright 2005)). .

In the event of the refined index not demonstrating a high level of reliability, the data will be analysed to identify any specific bias or areas of disagreement. The expert panel will be reassembled and shall develop further refinement as in Stage 1. Should the index show acceptable reliability ($\kappa > 0.6$) then the next stage of the study shall be performed.

Stage 3 – Application of refined index to CCUK Cohort

The CCUK cohort of study models (198) will be scored using the Modified Index. This will provide a baseline score for future comparison using the refined index. More discernable trends in the outcome of primary surgery following reorganisation of cleft care in the UK can be identified.

Statistical Analysis: When comparing agreement within and between examiners using an ordinal scale, a weighted kappa analysis is the most appropriate test to use. The type of weights will be chosen depending on the differences between subcategories; for example if the differences between adjacent subcategories are equally important linear weights will be used. Power calculations are not well-established for analysing data with weighted kappa, primarily because the null hypothesis (no agreement between examiners) is not meaningful. Hence the sample size that will be used for the index refinement stage and reliability stages will be 100 and 50 respectively. This number is reasonably achievable and the refinement stage will require a large sample size for fine discrimination between cases.