Cover Letter

Official title: Developing a Down Syndrome Health Instrument

NCT Number: NCT04631237

Document Date: 12/16/2024

Developing a Down Syndrome Health Instrument (DHI)

Original Proposed Statistical Analysis Plan:

Power calculation: Sample sizes are based on several criteria. The basic rule of thumb for psychometric analysis (primarily correlation analyses) is 10 subjects per survey item.[47] The current conceptual model includes 95 items, with refining in SA1, I plan a final item pool of 70 items. The current target sample sizes will support up to 70 variables (items) available for inclusion in psychometric analyses. For instance, if the final instrument contains 70 items, 700 subjects are sufficient for 80% power to detect a difference of ≥ .25 standard deviations using type I error of 0.10.

Statistical Analyses: A critical aspect of instrument validation, evaluations include:

Distribution evaluation

Ceiling and floor effects: Item response will be evaluated for ceiling and floor effects. Items that exhibit greater than 85% response in the tail of the distribution within severe respondents will be removed. Comparison will be both between and within sample frames.

Internal Validity

- Reliability: This will include comparison between and within sample frames; retest 2-4 weeks later.
- Internal Consistency: Items that are expected to measure a domain or construct should be related to each other in a systematic way. Internal consistency of item groupings will be evaluated with Cronbach's alpha[58]. Values >0.9 will be interpreted to mean an item is simply being asked over and over again and values <0.4 will be viewed as an unacceptable cut-point relative to item differentiation.
- Test-retest Reliability. To test differences among sample means, Tukey's HSD test with threshold alpha ≤0.05 will be used to evaluate all items.[59]
- Internal Dimensional validity: The goal is to identify dimensions of DS health, indicated by factors. Factor analytic methods will be used to identify a factor structure of the DHI. While we may hypothesize what the dimensions of DS health are, a priori this cannot be known. Therefore, we will initially run a confirmatory factor analysis (FA) using our hypothesized dimensions. If the initial CFA does not reveal emergent factors corresponding to the hypothesized factor structure, we will employ exploratory factor analysis (EFA) and principal components analysis (PCA) within hypothesized dimensions or emergent factors. Several criteria will be applied in evaluation of dimensions and factor retention: the Kaiser-Guttman rule (Eigenvalues>1.0), factor loading thresholds of 0.60/.40, scree plots, Kaiser-Meyer-Olkin (KMO) residuals off-diagonal partial correlations measure of sampling adequacy >0.70.[52,60,61]

External Criterion Validity: In this proposal, we are seeking to establish validity to draw a valid inference about a distribution and establish some basic inferential validity. External validity will be operationalized using a multitrait-multimethod matrix (MTMM) approach, with the "gold standard" being establishment of forms of both convergent and divergent (discriminatory) validity.[62] For validation of the DHI, each domain is considered a trait. Method refers to a measurement procedure, which includes consideration of both data source as well as data collection method. The referent method is a caregiver-administered survey. The data source is DHI respondent proxy-report for an individual with DS. The data collection methods include survey instruments and medical record review of clinically-significant metrics, and will be

collected as outlined above in “Population:” (Table 1). Generally, correlation coefficients between DHI domains and data external to the DHI will be used as evaluative criteria, with the specific correlation statistic being determined by the test value variable type. Utilization of the MTMM allows evaluation of correlation within and between methods and sources as well as within and between dimensions or traits.

Final Statistical Analysis Plan:

Planned Method of Analysis

The primary aim of this research project will be to develop a survey instrument valid to measure health of individuals with Down syndrome ages 0-21 years. As such, the focus is on the psychometric properties of the survey instrument; the study outcomes are the psychometric assessments of the instrument.

Pre-determined Sample Size Calculation and Recruitment Goal

Sample size was determined *a priori* based on the sample size needed to conduct the psychometric analysis. The basic rule of thumb is that (5-10) ten subjects per item are needed for analyses.[39] For instance, if the final instrument contains 50 items, 500 subjects are sufficient for 80% power to detect a difference of ≥ .25 standard deviations using type I error of 0.10. For our DHI with 113 core items, a sample size of 565-1130 caregivers is needed based on using an alpha of .05, a targeted power of .80, and an assumed change of 20% in the score between baseline and follow-up.

Determining "Complete" Surveys

We will closely follow progression through the components of our survey from screening to consent to survey. We considered a survey "Complete" if a participant had a timestamp for completing the DHI section. This occurred when individuals advanced from the final item in the DHI section. Although 683 individuals started the DHI section of our survey, 542 were completed and included in our study.

Item-level Evaluation

Item Distributions (including Missing data)

Ensure no "out of range values" including confirming branching logic followed flow to subsequent questions or skipped as we anticipated / planned.

Basic descriptive statistics are used to report distributions of continuous variables with mean, standard deviation, IQR and frequency for categorical variables. Item distributions evaluated missingness as well as floor and ceiling effects, defined at greater than 85% of responses located in item response anchor. For those with ≥85% response at the tails, our research team reviewed the corresponding item and discussed if item should be eliminated. For example, we did not eliminate items related to developmental skills / milestones based on floor and ceiling effects as more than half of individuals are expected to attain some of these skills across the age span studied.

Handling of missing data

Item missingness was evaluated.

Survey-level Evaluation

Scale scoring for all scales/measures:
1. Develop the scoring algorithm, including minimum number of items requiring response to score. DHI scores are continuous and based on summed means of scale items with higher values indicating better health.
2. Evaluate scale score distributions; evaluate for floor and ceiling effects ≥10% response at the tails.

Analysis of Key Psychometric Properties:

1. Factor Structure: principle components (PCA) and principle factor analysis (PFA) methods that included orthogonal (varimax) and oblique rotation (promax).[39,40] Eigenvalue statistics are used to evaluate the strength of these relationships; a high eigenvalue indicates that an underlying subscale exists and low values indicate that the underlying subscale might not exist. If a subscale's eigenvalue meets the threshold for retention, the next step in the analysis is to evaluate how items relate to a particular factor using factor loading scores. We used standard criteria (eigenvalues > 1) and items within factors (factor loading scores > 0.40)[39,40] in our iterative scale development process.

2. Internal Validity: Internal dimensional validity of scales was evaluated using factor analysis (FA). Initially, exploratory FA of item grouping within the 3 hypothesized domains (physical, mental, and social health) was conducted to identify factors within each domain, followed by a confirmatory FA of all items within those factors of all domains. The final factor loadings are a result of the confirmatory analysis. Internal consistency was evaluated with Cronbach's alpha. Items with factor loadings < 0.40 were dropped for low convergent validity; items with loadings >0.9 were reviewed for redundancy and dropped as appropriate.

3. Test-retest Reliability: Reliability analyses included individuals who completed the retest survey. Test-retest reliability of each subscale score was assessed through the interclass correlation coefficient (ICC). An item was used to screen out individuals who experienced major changes in status between the test and the retest. Test-retest reliability was evaluated with paired t-tests and chi square tests at the item level and 2-way mixed-effects intraclass correlation coefficient (ICC) models at the scale level, with a 95% confidence interval of the estimate using the following guideline: poor (< 0.5), moderate (0.5-0.75), good (0.75-0.9), and excellent (>0.90).

4. External Criterion / Construct Validity: In the criterion validity evaluation, Pearson correlations served as the primary evaluation technique to evaluate how scores on our new measure correlated with scores on other external measures of components of health status (the PEDSQL, the SDQ, the CSHQ). Published scoring algorithms were used whenever available. General MOS items were scored using a simple mean.

We are seeking to establish validity to draw a valid inference about a distribution and establish some basic inferential validity. External validity will be operationalized using a multitrait-multimethod matrix (MTMM) approach, with the "gold standard" being establishment of forms of both convergent and divergent (discriminatory) validity.[62] For validation of the DHI, each domain is considered a trait. Method refers to a measurement procedure, which includes consideration of both data source as well as data collection method. The referent method is a caregiver-administered survey. The data source is DHI respondent proxy-report for an individual with DS. The data collection methods include survey instruments and medical record review of clinically-significant metrics. Generally, correlation

coefficients between DHI domains and data external to the DHI will be used as evaluative criteria, with the specific correlation statistic being determined by the test value variable type. Utilization of the MTMM allows evaluation of correlation within and between methods and sources as well as within and between dimensions or traits.

Subgroup Analyses

We anticipated analysis by mode (PAPI vs CASI) but found that nearly all individuals elected to complete CASI.

All analyses were conducted using SAS© (v. 9.4).

All analyses were conducted using SAS, version 9.4 (SAS Institute, Cary, NC) and R, version 2.1.9 (R Core Team, Vienna, Austria), with R package psych. The protocol was registered (ClinicalTrials.gov; NCT04631237).

This study was approved by the MGB Institutional Review Board. De-identified supporting data are available upon request from the authors.

Changes to Statistical Analysis Plan

Section	Original	Final
Power Calculation	Planned 70 items in final pool	113 core items in final pool
Determining Complete surveys	-	Added information regarding determination of "Complete" survey
Item-level evaluation	-	Added details regarding missingness and handling missing data
Survey-level evaluation	-	Rotation analysis incorporated
Test-retest reliability	Tukey's HSD test with threshold alpha ≤0.05 will be used to evaluate all items	Paired t-tests and chi square tests at the item level and 2-way mixed-effects intraclass correlation coefficient (ICC) models at the scale level
External validity	Description of MTMM	Details added on the specific external criterion measures used and analysis method (Pearson correlation)