# STATISTICAL ANALYSIS PLAN

**Version: 1.0**

**Date: November 2024**

**A cluster randomised control trial comparing the effectiveness and cost effectiveness of school vs school and community-based settings in echo screening for undetected Rheumatic Heart Disease (RHD) for year 7 and 8 students**

**Sponsor: Dr Dale Bramley, Director of Planning, Funding and Outcomes, Health New Zealand | Te Whatu Ora**

**Site Counties Manukau and Bay of Plenty/Lakes Districts**

**Principal investigator: Dr Karen Bartholomew**

**Principal Co-investigator: Professor Nigel Wilson**

**Co-ordinating Investigator: Tracey Hale**

**TABLE 1. STUDY STRUCTURE**

| | |
|---|---|
| Sponsor | Dr Dale Bramley<br>Director of Planning, Funding and Outcomes, Health New Zealand \| Te Whatu Ora |
| Contract Research Organisation | N/A |
| Lead Site (New Zealand) | Waitematā district, on behalf of Health New Zealand \| Te Whatu Ora |
| Principal Investigator | Dr Karen Bartholomew<br>Karen.bartholomew@tewhatuora.govt.nz |
| Principal Co-investigator | Professor Nigel Wilson<br>NigelW@adhb.govt.nz |
| Lead analyst investigator | Wendy Bennett<br>Wendy.Bennett@tewhatuora.govt.nz |
| Imaging Storage | In house |
| Image capture | Echocardiogram imaging machines will comprise one or some from the following options:<br>*Portable:*<br>• Lumify (Phillips) handheld machines<br>• Sonos PX<br><br>*Confirmatory machines:*<br>• Vivid IQ<br>• Phillips CX5 or CX50 |

**INTRODUCTION**

The data collected during this pilot will be used to operationalise the study, to manage the screening of participants and their clinical care. It will also be used to meet the research objectives/answer the study questions (see Protocol), largely retrospectively. Interim data analysis may be undertaken during the study to monitor progress and identify any issues with the study protocol. Ethnicity will be an important variable within this analysis.

Analysis components:

1. Capturing all necessary data: identifying all data variables to be captured during the study and linking to any additional data required (eg. test results) will create a comprehensive dataset that includes all the necessary information and that uses a common identifier ie. the participant study number.

2. Data cleaning and pre-processing: Once all data is collated, data will be cleaned and pre-processed to ensure that it is ready for analysis. This involves removing duplicates, missing values, and outliers, and transforming the data as necessary. Data quality (including overall quality of the datasets and the quality of variables of interest) will be taken into account when the analysis is undertaken and when the results are interpreted. Data cleaning may result in identification of quality issues that may require contacting the study participant/their whānau post screening to rectify.

3. Statistical analysis: To assess the differences or associations between the two intervention models, several statistical approaches will be considered. The choice of method will depend on the nature of the data and the specific outcomes being evaluated, ensuring the most appropriate analytical techniques are applied.

   Descriptive Statistics

   - Continuous Variables: Calculate means, medians, standard deviations, and ranges.
   - Categorical Variables: Calculate frequencies and percentages.

   Univariate Analysis:

   - Chi-square Test: To test for significant differences in screening uptake between different hospitals, regions, and ethnic groups.
   - T-test or ANOVA: To compare means of continuous variables between groups.

   Multivariate Analysis

   Logistic Regression:

   - Outcome Variable: Screening status (binary: screened vs. not screened).
   - Outcome Variable: Screening status (binary: screened school based vs. Screened School + Community)
   - Predictor Variables: school, ethnic group, age, gender, etc.
   - Adjust for potential confounders
     - Association with the Outcome: Use univariate analysis to identify variables significantly associated with the outcome.

- o Association with the Predictor: Check if the potential confounder is associated with the predictor variables.
- o Expert Knowledge: Use expert knowledge including literature reviews to identify variables that could influence both the predictors and the outcome.

Linear Regression:

- Outcome Variable: Continuous outcomes related to screening.
- Predictor Variables: Same as above.

Cost-effectiveness Analysis

- Outcome Variable: Net Benefits
- Outcome Variable: Net Costs
- Outcome Variable: Incremental Cost-effectiveness Ratios

4. Robust standard errors will be used to account for potential intra-cluster correlation within schools. By using robust variance estimation (VCE), the model will adjust for the clustering effect, ensuring valid statistical inferences and mitigating the risk of biased standard errors that may arise from the lack of independence between students within the same school. Reporting: to report the findings and conclusions based on the analysis, we may use visualisations such as graphs and charts to help illustrate our findings and make them more accessible to stakeholders. It is also important to provide clear and concise explanations of the statistical tests used and the results obtained. Focus will be on both the statistical as well as the clinical and practical significance of the findings. Some analysis may not be undertaken if the total number available for analysis is small.

5. Software use:
- Excel: Microsoft Corporate. (2016)
- Python: python Software Foundation (2022). Python (version 3.9)
- Stata: StataCorp LLC. (2021)
- **R**: R Core Team. (2023). R (version 4.3.0)
- SAS: SAS Institute Inc. (2022). SAS (version 9.4)
- TreeAge Pro Healthcare: TreeAge Software, LLC . (2024). TreeAge Pro Healthcare 2024 R2.

**Quantitative Analysis**

The following measures will be calculated from the collected data:

***Consent/Assent***

- Number and percentage of consents (parents/guardians) by ethnicity, school (including type of school eg. intermediate) and mode (initial approach, via school's digital platform, via SMS link or other)
- Number and percentage of declines (parents/guardians) by ethnicity and school
- Number and percentage of declines (students - assent) by ethnicity and school
- Number and percentage of e-consents vs paper based consents by ethnicity and school
- Timeliness of consents – comparison of contact dates vs consent dates

### *Screening*

- Number and percentage of positive cases by ethnicity and age via SPLASH + screen
- Number and percentage of positive cases by ethnicity, age and stage via cardiac sonographer screen
- Number and percentage of those with normal screen by ethnicity and age
- Number and percentage of those with incidental findings (including by individual finding) by ethnicity and age
- Analysis of 'other observations' will be undertaken to determine any patterns

*Note: a detailed analysis of the clinical findings from the screens will also be undertaken based on the data collected (see appendix 3)*

### *Coverage by model*

- Uptake (actual screens undertaken) of school based only model by age, ethnicity and school
- Uptake (actual screens undertaken) of school plus community option model by age, ethnicity – including specific uptake analysis by school and community option
- Comparison of above.

### *Quality Assurance – see Appendix 5*

### *Cost Effectiveness*

Analysis of the costs and benefits, compared by setting and perspective

- Settings
    - Baseline: Current Practice
    - Intervention 1: Screening School-based Only
    - Intervention 2: Screening School + Community
- Perspectives
    - Societal (Reference Case 1)
    - Healthcare Sector (Reference Case 2)
    - Patient/Household

### Decision Tree Analysis including Markov Model of Disease Progression

- Outcomes
    - Expected patient counts and costs of disease stages
    - Expected net benefits of interventions
    - Expected net costs of interventions
    - Expected incremental cost-effectiveness ratios for interventions
- Sensitivity Analyses
    - Variable selection depends on results and uncertainty surrounding parameters
    - Single- and multi-variate
- Subgroup analyses will be conducted as needed and when feasible
    - Ethnic profile
    - Age structure
    - Deprivation profile
    - Household characteristics (eg., area of domicile)

***Descriptive Analysis of student baseline characteristics***

Ethnic profile comparing total response ethnicity with prioritised – particularly to ascertain impact of prioritising Māori where more than one ethnicity is recorded.

Description of the cohort of students – compared by prioritised ethnicity:
- Ethnic profile
- Age structure
- Area of domicile
- Deprivation profile

***Other potential analysis***

- Pathway variation for positive cases
- Timeliness of follow up pathways

**Qualitative Analysis**

***Survey Analysis of students, parents/whānau***

- Acceptability of model option, including cultural acceptability
- Experience of approach, consent, screen, information given etc
- Costs incurred eg. transport, out-of-pocket treatment
- Quality of life measures
- Whether they would find AI reading of scans acceptable – in the future (not for this screen)
- Other impacts

***Survey Analysis of Workforce***

There are a range of different roles within the workforce who will be surveyed, including;

- Clinicians – sonographers, doctors and nurses
- Other non-expert screeners
- Kaiāwhina
- Other research staff

Analysis will cover:
- Acceptability of screening by venue (school vs community setting)
- Rating of process
- Elements of best set-up eg. bus for screening vs school facility
- System impacts (gathered from a range of workforce stakeholders beyond the actual screen)
- Specific issues

***Survey Analysis of School***

- Acceptability of pilot at your school
- Time commitment
- Staff commitment
- Specific issues

***Survey Analysis of Community setting/venue***

- Acceptability of pilot in your setting/venue
- Time commitment

- Staff commitment
- Specific issues