TMR Statistical Analysis Plan (SAP)

Guidance for Study Staff

Ian Hambleton

13 Jan 2026

Table of contents

ToDo: Expand inclusion of controls into secondary analytics

---

1. Study Identification

This Statistical Analysis Plan (SAP) details the planned analyses for the Saint Helena Total Meal Replacement (TMR) diabetes remission study. The intervention is a structured, low-energy Total Meal Replacement diet followed by food reintroduction and weight-maintenance support, delivered within the routine health system of Saint Helena.

The study is designed as an implementation science evaluation rather than a conventional explanatory randomised trial. Adults with type 2 diabetes who are not using insulin at baseline constitute the primary analytic population, reflecting the group in which remission is most plausible based on prior trials. Adults who are using insulin at baseline form a secondary exploratory cohort, recognising their greater clinical complexity and the more limited evidence base for remission in this subgroup.

---

2. Background and Rationale

Total Meal Replacement programmes, typically prescribing around 800 kilocalories per day in the form of nutritionally complete soups and shakes, have been shown to drive large and rapid improvements in metabolic health. They reduce liver and pancreatic fat, improve insulin sensitivity and can restore first-phase insulin secretion in people with relatively recent-onset type 2 diabetes. The DiRECT trial demonstrated that such a programme, delivered in primary care with accompanying medication withdrawal and behavioural support, achieved remission of diabetes in a large proportion of participants.

Saint Helena has a high burden of obesity and diabetes in a small, remote health system. The island context presents both a pressing need and a real-world test of whether this type of intervention can be implemented at scale, sustained, and adapted to local circumstances. This study therefore aims to generate not only efficacy-type information about metabolic outcomes but also implementation-relevant insight into feasibility, safety, acceptability and system fit.

The evidence for using TMR in insulin-treated type 2 diabetes is encouraging but less extensive. Trials have shown that TMR can reduce insulin requirements and improve HbA1c, but long-term remission in this group is less well documented. Insulin-treated participants are therefore included with explicit recognition that they form an innovative, higher-risk group whose data will be interpreted in a more exploratory way.

---

3. Objectives

3.1 Primary Objectives (Non-Insulin Cohort)

- **To estimate the proportion of non-insulin participants who achieve 15 kg weight loss at 12 months.**

- **To estimate the proportion of non-insulin participants who achieve diabetes remission at 12 months**, defined as HbA1c <48 mmol/mol with no glucose-lowering medication for at least three months.

Explanatory text

The principal aim is to determine whether a TMR programme, implemented in the real-world setting of Saint Helena, can reproduce the kind of large weight losses and remission rates observed in prior trials. The first primary outcome, loss of at least 15 kilograms, captures the level of weight reduction associated with substantial metabolic benefit. The second primary outcome, remission, uses a widely accepted definition based on HbA1c and the absence of glucose-lowering medication. Together, these outcomes directly address whether the intervention can meaningfully alter the clinical course of type 2 diabetes for non-insulin-treated participants.

3.2 Secondary Objectives (Non-Insulin Cohort)

The secondary objectives of the study, applicable primarily to the non-insulin cohort, are:

- **To describe changes in key metabolic and clinical measures**, including weight, HbA1c, blood pressure, waist and hip circumference, lipids, renal function and liver function, at both 12 and 24 months.
- **To estimate the proportion of participants achieving more stringent glycaemic targets**, such as HbA1c <42 mmol/mol, acknowledging that improvement short of remission may still be clinically meaningful.
- **To characterise changes in diabetes medication burden**, including the number and classes of glucose-lowering medications and their doses, and to quantify reductions in medication use.
- **To describe relapse from remission** during follow-up, defined as subsequent HbA1c 48 mmol/mol after remission has been achieved.
- **To evaluate changes in quality of life**, using the EQ-5D-3L descriptive system, EQ-VAS, and UK tariff–based utility scores, with careful interpretation in the absence of Saint Helena–specific tariffs.
- **To evaluate changes in diabetes-related emotional distress**, using the PAID scale, including both continuous scores and the proportion exceeding accepted distress thresholds (e.g. PAID 40).
- **To explore basic health-economic consequences of the intervention**, including potential reductions in insulin expenditure, glucose-lowering medication costs, and service utilisation that may be attributable to improved weight, glycaemia and medication de-intensification.

Explanatory text

These secondary objectives collectively reflect the multiple domains through which a TMR programme may affect health and wellbeing. Improvements in metabolic markers, medication burden and psychosocial wellbeing each represent meaningful progress even where full remission is not achieved. Measuring relapse rates also provides important information about durability and the need for maintenance strategies. The inclusion of basic health-economic evaluation acknowledges the unique constraints of a small island health system, where reductions in medication expenditure—particularly expensive insulin formulations—may yield substantial system-level benefits.

3.3 Exploratory Objectives (Insulin Cohort)

- **To estimate the extent to which insulin-treated participants can safely reduce or discontinue insulin** while maintaining acceptable glycaemic control.
- **To evaluate a "stepped improvement" outcome**, combining changes in HbA1c and medication burden into a pragmatic ordinal measure of benefit.
- **To describe changes in weight, HbA1c, quality of life and diabetes distress** among insulin-treated participants.
- **To assess the safety and feasibility of a structured insulin down-titration protocol** during TMR.

Explanatory text

Insulin-treated participants are unlikely to mirror the remission rates seen in non-insulin populations, but may still experience large clinical benefits through improved glycaemic control on fewer medications. The exploratory objectives therefore emphasise de-intensification, stepped improvements and safety rather than remission alone. Understanding how far insulin can be reduced, and under what conditions, is a key question for service planning.

3.4 Process and Behavioural Objectives (EMA Substudy)

Note: EMA substudy is a potential not definitive inclusion at this point. For discussion.

Bulleted formulation

- **To explore day-to-day adherence, mood, cravings and environmental context**, using brief repeated EMA prompts in a purposive subsample.
- **To examine how EMA-derived behavioural patterns relate to clinical outcomes**, such as weight and HbA1c changes.
- **To inform refinement of the programme**, by identifying behavioural and contextual facilitators of success and barriers to adherence.

Explanatory text

EMA adds a dynamic, real-time dimension to the evaluation by capturing experiences that may not be adequately recalled in retrospective interviews. These data will help explain why the intervention works well for some participants and less well for others, and will support future tailoring of the programme.

---

4. Study Design and Populations

The study is a non-randomised implementation evaluation of a TMR programme within the Saint Helena health system. All eligible and willing participants will be offered TMR. For analytic purposes, outcomes in TMR participants will be compared to outcomes in a registry-based cohort of individuals with type 2 diabetes who do not participate in TMR but are otherwise clinically similar.

Participants are stratified at baseline according to insulin use:

- **Stratum A (primary)**: adults with type 2 diabetes not using insulin.
- **Stratum B (exploratory)**: adults with type 2 diabetes using insulin.

Within each stratum, TMR participants will be matched to comparison individuals using propensity scores, and longitudinal data will be analysed using Difference-in-Difference methods. These methods are chosen to approximate some of the benefits of randomisation despite the pragmatic design.

4.1 Analysis Sets

- **Primary analysis set**: all TMR participants in Stratum A (non-insulin) with baseline data, analysed on an intention-to-treat basis.
- **Exploratory insulin set**: all TMR participants in Stratum B (insulin users) with baseline data.
- **Per-protocol set**: TMR participants in either stratum who complete at least 12 weeks of TMR and attend the 12-month assessment.
- **EMA subsample**: TMR participants who agree to EMA and qualitative interviews.

---

5. Endpoints

5.1 Primary Endpoints (Non-Insulin, 12 Months)

- **15 kg weight loss**, defined as a reduction of at least 15.0 kg from baseline body weight at the 12-month assessment.
- **Diabetes remission**, defined as:
  - HbA1c <48 mmol/mol at the 12-month assessment, **and**
  - No glucose-lowering medication for at least three months prior to the 12-month assessment.

Explanatory text

The weight-loss endpoint captures the degree of weight reduction that has been linked with major improvements in metabolic status in earlier trials. The remission endpoint follows international consensus definitions, linking HbA1c and withdrawal of pharmacological treatment. These two endpoints together provide a clear, clinically meaningful summary of the programme's impact in the non-insulin cohort.

The analysis will follow a hierarchical approach: the weight-loss endpoint will be examined first; only if this shows evidence of benefit, as judged by confidence intervals and overall pattern of results, will the remission endpoint be subjected to formal inferential testing at the conventional 5% significance level.

5.2 Secondary Endpoints (Non-Insulin)

Secondary endpoints for the primary (non-insulin) cohort will include:

- **Continuous changes** in weight, HbA1c, blood pressure, waist/hip circumference, lipids and renal/liver function between baseline and 12-month and 24-month assessments.
- **Binary glycaemic thresholds**, including HbA1c <42 mmol/mol and HbA1c <48 mmol/mol irrespective of medication use.
- **Measures of medication de-intensification**, including the reduction in the number of agents, reduction in medication classes, and reductions in cumulative medication dosage. *These information will be extracted from clinic notes rather than double-collected in the TMR database environment.*
- **Relapse from remission**, defined as a subsequent HbA1c 48 mmol/mol after remission.
- **Quality-of-life outcomes**, including EQ-5D-3L index scores (using UK tariffs), the five EQ-5D dimensions individually, and EQ-VAS.
- **Diabetes distress outcomes**, including PAID total score and the proportion of participants exceeding the PAID 40 threshold.
- **Health-economic indicators**, such as:
  - Changes in insulin and oral therapy costs,
  - Changes in total medication burden cost, and
  - Descriptive summaries of shifts in healthcare utilisation (e.g., GP visits for glycaemic instability, emergency hyperglycaemia management).

Explanatory text

These endpoints recognise the broader scope of TMR beyond remission alone. They allow the study to describe clinical, psychological and economic consequences of participation. In particular, in a remote island system with limited fiscal capacity, changes in the cost and utilisation of diabetes medications—including reduction or cessation of insulin—are essential components of evaluating system-wide benefit.

5.3 Exploratory Endpoints (Insulin Cohort)

- **Off-insulin with acceptable control**, defined as:
  - No insulin use for at least three months prior to 12 months, and

  - HbA1c <53 mmol/mol at 12 months.

- **Stepped-improvement outcome (ordinal)**:
  - **Level 1 (Best)**: HbA1c <48 mmol/mol, off all glucose-lowering medications 3 months.
  - **Level 2 (Good de-intensification)**:
    * HbA1c baseline + 3 mmol/mol or HbA1c <53 mmol/mol, **and**
    * Insulin dose reduced by at least 50% or insulin stopped, with no net increase in other glucose-lowering drug classes.
  - **Level 3 (Unfavourable)**:
    * HbA1c worse than baseline by >5 mmol/mol or 64 mmol/mol, **or**
    * An increase in medication burden (e.g. higher insulin dose or additional agents).

- **Continuous changes** in insulin dose, HbA1c, weight, quality of life and distress.

Explanatory text

The insulin cohort is too heterogeneous and too small for remission to serve as the only meaningful endpoint. Instead, we use an ordinal stepped-improvement outcome that allows us to distinguish between substantial de-intensification with good control (Level 1 and Level 2) and poorer outcomes (Level 3). This approach acknowledges that for many insulin-treated participants, stabilising or slightly improving HbA1c while significantly reducing medication exposure may be a major success in its own right.

The off-insulin endpoint provides a straightforward and clinically intuitive indicator of success, while continuous measures of insulin dose change and metabolic parameters provide additional nuance.

---

6. Statistical Analysis

6.1 Overall Analytic Strategy and Implementation-Science Rationale

This is a pragmatic implementation study without randomisation. In such settings, no single analytic approach can fully replicate the strengths of a randomised trial. For this reason, the study employs a dual analytic strategy consisting of:

1. **Matched cohort analysis**, and
2. **Difference-in-Difference (DiD) analysis**.

Both methods are considered part of the primary analytic framework, used in parallel and interpreted jointly.

From an implementation-science perspective:

- The **matched cohort analysis** aims to construct a comparison group that looks as similar as possible to the TMR group at baseline. This helps address confounding by indication and allows intuitive comparisons ("like with like").

- The **DiD analysis** focuses on changes over time and is robust to unmeasured confounding that is stable over time. It asks whether the trajectory of outcomes (such as weight and HbA1c) differs between the TMR group and the comparison group.

The use of both approaches together embodies the principle of triangulation: if matched analyses and DiD analyses point in the same direction with similar magnitudes, confidence in the findings increases, even in the face of small sample sizes and potential residual confounding. If they diverge, this suggests that underlying assumptions may not hold, and the results must be interpreted with more caution.

Although conventional statistical significance at $p<0.05$ will be reported, the emphasis will be on estimation and confidence intervals. For each outcome, we will present point estimates (e.g. mean differences, risk differences, risk ratios) with 95% confidence intervals, and interpret their magnitude and precision in light of prior evidence and clinical relevance. P-values will be considered supportive but not definitive in isolation, particularly given the limited sample size.

6.2 Sensitivity Analyses and Robustness

Given the relatively small number of TMR participants and the real-world nature of the comparison cohort, sensitivity analyses are essential to assess the robustness of findings. Planned sensitivity analyses include:

- Alternative propensity score specifications and matching calipers.

- Different matching ratios (e.g. 1:1 vs 1:2 or 1:3).

- Analyses restricted to more tightly matched subgroups.

- Alternative DiD model specifications, including additional covariates or different functional forms.

- Per-protocol analyses restricted to participants who complete a minimum TMR exposure.

- Alternative assumptions about missing data (e.g. multiple imputation vs complete-case analysis).

The purpose of these sensitivity analyses is not to produce a multitude of competing p-values, but to examine whether the estimated effects are broadly stable across reasonable analytic choices. Consistency of effect direction and order of magnitude across these analyses will strengthen confidence in the conclusions.

6.3 Matching Strategy

Within each stratum (non-insulin and insulin), a **propensity score** will be estimated using logistic regression, where the dependent variable is TMR participation (yes/no). The propensity score covariates will be pragmatically restricted to those available available in the Diabetes Registry, specifically:

- Age, sex, diabetes duration, and BMI will be the key propensity score matching criteria, all sourced from the St Helena Diabetes Registry.

Matched sets will be formed using nearest-neighbour matching within a pre-specified caliper (e.g. 0.2 standard deviations of the logit of the propensity score). Matching will be done without replacement where possible, and with replacement if necessary to achieve acceptable balance. Balance will be examined using standardised mean differences; values below 0.1 will be considered indicative of good balance.

6.4 Difference-in-Difference (DiD) Models

For continuous outcomes such as weight and HbA1c, the primary DiD model will be:

The key parameter is (_3), which represents the Difference-in-Difference effect, that is, the extra change over time associated with being in the TMR group as opposed to the comparison group. For binary outcomes, a similar structure will be used in a generalised linear mixed model with an appropriate link function.

We will explore the parallel trends assumption by examining pre-baseline data where available. If strong deviations are detected, DiD results will be interpreted more cautiously and may be treated as exploratory.

6.5 Primary Analysis (Non-Insulin Cohort)

For the non-insulin cohort, the primary analysis will focus on the two primary endpoints and key secondary outcomes.

- For 15 kg weight loss and remission, risk ratios and risk differences with 95% confidence intervals will be estimated using log-binomial models or Poisson regression with robust standard errors. Logistic regression may be used if convergence issues arise, with odds ratios carefully interpreted.

- For continuous outcomes such as weight and HbA1c change, linear mixed-effects models with random intercepts and fixed effects for group, time and group-by-time interaction will be used, adjusting for baseline values.

Results will be presented with a clear focus on effect sizes and confidence intervals, and interpreted in the context of the broader TMR evidence base. Formal hypothesis tests at the 5% significance level will be reported, but not over-emphasised.

6.6 Exploratory Analysis (Insulin Cohort)

Analyses in the insulin cohort will be primarily descriptive but will use formal models where numbers allow.

The stepped-improvement outcome will be summarised as the proportion of participants in each level (Level 1, Level 2, Level 3) at 12 months. Where cell counts allow, ordinal logistic regression or related models may be used to explore associations with baseline characteristics and to compare outcomes across subgroups. The focus, however, will be on interpretable descriptive summaries: for example, the proportion of insulin users who move from Level 3 at baseline to Level 1 or 2 at follow-up.

The off-insulin endpoint will be examined by estimating the proportion of participants who are off insulin at 12 months with acceptable HbA1c levels, together with confidence intervals. Comparisons with matched registry controls may be made to contextualise these findings.

Continuous changes in insulin dose, HbA1c, weight and quality of life will be analysed with linear mixed-effects models. Again, emphasis will be placed on effect sizes and confidence intervals rather than statistical significance alone, given the smaller sample and exploratory nature.

6.7 EMA Analysis

For the EMA subsample, the first step will be to describe compliance with EMA prompts, typical patterns of mood, cravings and self-reported adherence, and their variability over time. We will then explore relationships between EMA-derived variables and key clinical outcomes.

For example, we may create individual-level summaries such as the average daily adherence score or the proportion of EMA time points with high craving. These summaries can then be related to weight loss or HbA1c change using linear regression or mixed models. Time-varying EMA measures may also be modelled against outcomes using more complex structures if data volume permits.

EMA analyses are explicitly exploratory; their purpose is to generate hypotheses and insight rather than formal confirmatory tests.

6.8 Analysis of EQ-5D and PAID Outcomes

EQ-5D-3L

EQ-5D-3L data will be analysed through several complementary approaches:

1. **Descriptive analysis**
   We will summarise the distribution of responses across the five dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) at baseline and follow-up. Transition matrices may be used to illustrate improvement or deterioration in each dimension.

2. **EQ-VAS analysis**
   The EQ-VAS (0–100 scale) will be analysed as a continuous outcome using linear mixed-effects models, adjusting for baseline values.

3. **Utility score estimation using UK tariffs**
   In the absence of a validated Saint Helena EQ-5D value set, we will derive utility scores using the established UK MVH tariff. This approach is widely used in international settings where local

tariffs are unavailable. Although the UK tariff may not perfectly reflect preferences of a Saint Helena population, it provides:

- A consistent, standardised method for converting EQ-5D states into utility values
- Compatibility with international health-economic evaluation conventions
- Interpretability across studies

The limitations of applying UK tariffs to a distinct cultural and demographic context will be acknowledged. However, given the size of the island population and the resource implications of developing a bespoke tariff, the UK tariff is considered the most feasible and defensible option. Results will be interpreted with caution, emphasising relative within-person change rather than absolute utility levels.

4. **Statistical modelling**
   Utility scores will be analysed using linear mixed-effects models, with focus on estimated mean differences and their confidence intervals rather than p-values. Analyses will explore associations between utility change and weight/HbA1c change as part of secondary modelling.

PAID (Problem Areas in Diabetes)

PAID scores will be treated as both:

- **A continuous scale (0–100)** capturing intensity of diabetes-related emotional distress, and
- **A categorical indicator** of high distress using the conventional PAID 40 threshold.

Analysis will proceed as follows:

1. **Continuous analysis**
   Changes in PAID score from baseline to 12 and 24 months will be modelled using linear mixed-effects models, controlling for baseline scores. Interpretation will centre on the magnitude and direction of change, with 95% confidence intervals.
2. **High-distress prevalence**
   The proportion exceeding PAID 40 at each time point will be compared using risk differences and risk ratios, again with emphasis on estimation.
3. **Associations with clinical outcomes**
   Exploratory analyses will examine whether improvements in PAID are correlated with reductions in weight, HbA1c or medication burden.

Implementation-science justification
EQ-5D and PAID outcomes enrich the study by capturing dimensions of benefit that matter to patients and health systems beyond glucose and weight. These measures help identify whether TMR improves lived experience, emotional burden, and daily functioning—central concerns when assessing the real-world value of a behavioural/metabolic intervention.

6.9 Physical activity outcomes (IPAQ short form)

Physical activity will be assessed using the International Physical Activity Questionnaire (IPAQ) short form, administered at baseline, post–total diet replacement, and during follow-up. The IPAQ short form captures self-reported physical activity over the preceding 7 days across walking, moderate-intensity, and vigorous-intensity domains, as well as sedentary time.

6.9.1 Data processing and scoring

IPAQ responses will be processed in accordance with the IPAQ scoring protocol. For each assessment, total physical activity will be expressed as MET-minutes per week, calculated as:

- Walking MET-minutes = walking minutes × walking days × 3.3
- Moderate MET-minutes = moderate minutes × moderate days × 4.0
- Vigorous MET-minutes = vigorous minutes × vigorous days × 8.0

Total physical activity will be computed as the sum of walking, moderate, and vigorous MET-minutes per week. Sedentary time (minutes per weekday spent sitting) will be summarised descriptively and analysed separately.

In addition, participants will be categorised into low, moderate, or high physical activity categories using standard IPAQ thresholds for descriptive reporting.

6.9.2 Analytical role

Physical activity outcomes are considered secondary and supportive, intended to characterise behavioural change associated with the intervention rather than as primary drivers of remission outcomes.

Analyses will include:

- Descriptive summaries of MET-minutes per week and activity categories at each timepoint
- Within-participant change from baseline using linear mixed-effects models with random intercepts
- Between-group comparisons (intervention vs control) using adjusted regression models where data availability permits

Physical activity measures may also be included as covariates in exploratory models examining weight loss and glycaemic outcomes, recognising the potential for bidirectional relationships between activity and metabolic change.

6.9.3 Missing data and interpretation

IPAQ data are self-reported and may be subject to recall and social desirability bias. Missing IPAQ items will not be imputed; analyses will be conducted on available data only. Results will be interpreted descriptively and cautiously, and triangulated with other behavioural indicators (e.g. EMA adherence measures) where relevant.

Physical activity outcomes will be reported as contextual and supportive findings, not as definitive causal mediators of clinical outcomes.

6.10 Health-Economic Data Collection and Analysis

This evaluation includes **basic health-economic components** intended to inform decision-makers about the financial implications of implementing the TMR programme at scale.

6.10.1 Data Collection

The following health-economic data will be collected:

- **Medication costs**, including insulin (by formulation and dose), oral glucose-lowering therapies, and other diabetes-related medications. Unit prices will be obtained from Saint Helena procurement systems.
- **Changes in medication usage**, as available - including reductions or cessation of insulin and oral agents, *with medication dosage being extracted from the Diabetes Registry or St Helena medical records.*
- **Programme delivery costs**, including meal-replacement products, staff time and materials.

6.10.2 Analytic Approach

The purpose of this health-economic work is not to develop a full cost-effectiveness model but to provide pragmatic economic justification for the intervention in a small-island health context.

Analyses will include:

1. **Medication cost changes**
   For each participant, we will calculate monthly medication costs at baseline and follow-up:
2. **Reduction in insulin expenditure**
   Because insulin is expensive to procure and challenging to store and transport to Saint Helena, reducing insulin reliance may provide substantial financial benefit. Analyses will quantify:
   - The proportion of insulin-treated participants who discontinue insulin
   - Average reductions in total daily insulin dose
   - Associated pharmaceutical cost savings
3. **Simple cost-offset calculations**
   Estimated cost offsets will be compared directly to estimated programme costs (primarily meal-replacement product and staff time), yielding a basic cost-neutrality or cost-saving assessment.

6.10.3 Interpretation and Limitations

Given the small sample size, the economic analyses will rely primarily on descriptive statistics and wide confidence intervals, with transparent assumptions. The intention is not definitive cost-effectiveness modelling but a decision-support narrative for local health planners.

Even small reductions in insulin reliance may represent meaningful savings for Saint Helena due to:

- High procurement costs
- Supply-chain vulnerabilities
- Storage and transport requirements
- Small population scale, where any system-level change produces noticeable financial effects

This aligns strongly with the objectives of implementation science: understanding not only whether an intervention works, but whether it is **practical, affordable and scalable** within a given system.

---

7. Missing Data

For primary binary outcomes, missing outcome data at 12 months will be treated as non-response in the main analysis, which is conservative for weight loss and remission. For continuous outcomes, mixed-effects models inherently accommodate missing follow-up data under the assumption that data are Missing At Random, given observed covariates and prior outcomes.

Sensitivity analyses will explore alternative assumptions, including complete-case analysis and multiple imputation, especially if missingness exceeds modest levels or differs meaningfully between TMR and comparison groups.

---

8. Safety Analyses

Safety analyses will summarise the incidence of adverse events and serious adverse events, including episodes of hypoglycaemia and symptomatic hyperglycaemia, as well as clinically important changes in blood pressure and renal function. Particular attention will be paid to safety in the insulin cohort during the early phase of TMR and insulin down-titration.

Narrative descriptions will be provided for any serious or unexpected events, and counts and proportions will be presented with confidence intervals.

---

9. Sample Size and Power

The planned TMR intervention cohort comprises approximately $\mathbf{n = 50}$ participants, with the expectation that most will fall in the non-insulin stratum. This sample size is driven by service capacity and budget rather than by formal power calculations. Nevertheless, it is important to understand what the study can reasonably detect.

9.1 Power for Binary Outcomes (Remission and 15 kg Loss)

Assume that in the non-insulin stratum we have approximately:

If we assume a control remission rate (p_0 0.05) and an intervention remission rate (p_1 0.30), then (d = 0.25). Under these values, the standard error is small enough that the 95% confidence interval for (d) is likely to exclude zero in most samples, corresponding to power in the range of roughly 80–90%. If (p_1) is closer to 0.25 (difference of 0.20), approximate power remains reasonable but somewhat lower.

For the 15 kg weight-loss endpoint, similar reasoning applies. If we expect a low rate in the comparison group and a moderate to high rate in TMR participants, the study is well positioned to detect a large absolute difference.

These calculations are illustrative rather than exact but show that with moderate sample sizes, the study is powered to detect large, practice-changing differences in remission and substantial weight loss.

9.2 Power for Continuous Outcomes (Weight and HbA1c)

For continuous outcomes compared between two groups of size (n_1) and (n_0), the standardised effect size (Cohen's d) is:

In other words, for continuous outcomes, the study is powered to detect moderate to large differences in mean change between TMR participants and comparisons.

9.3 Benefits of Larger Sample Sizes (n = 60 and n = 70)

Consider increasing the TMR sample size to **n = 60** or **n = 70**. If the number of comparison participants increases proportionally, the standard errors for estimated differences shrink approximately in proportion to the inverse square root of the sample size.

If we denote the standard error with N participants per group as (SE*N), and the standard error with 50 per group as (SE*{50}), then:

Narrower confidence intervals imply greater precision and higher power to detect the same underlying effect size. This means that with **n = 60** or **n = 70**, the study would be better able to distinguish moderate effects from zero, and the estimates would be more convincing for policy-makers, even if point estimates remain similar.

9.4 Implementation-Science Interpretation

Because this is an implementation study with a limited sample, the aim is not to detect very small effects with high certainty but to determine whether the programme produces clear, clinically important differences that justify wider adoption, and to understand the conditions under which those differences arise. The sample size of around 50, and the potential to extend towards 60 or 70, is arguably adequate for this purpose, especially when combined with:

- A dual analytic approach (matching and DiD),
- A focus on effect sizes and confidence intervals,
- Sensitivity analyses to examine robustness.

Taken together, these elements support credible, context-sensitive inference despite the inherent constraints of working in a small island system.

End of Statistical Analysis Plan (Version 3.2)