TMR Data Dictionary

Guidance for Data Handling Staff

Ian Hambleton

13 Jan 2026

Table of contents

Data Collection Schedule Summary

Longitudinal Timepoints

Phases	Event label	Approx weeks
Phase 1: Screening	screening	-1
Phase 1: Pre-baseline	prebaseline	-∞ to 0
Phase 1: Baseline	baseline	0
Phase 2: TMR phase	wk00–wk12	1–12 (visit every 2 weeks)
Phase 3: Food reintroduction	wk13–wk19	13–19 (visit every 2 weeks)
Phase 4: Weight maintenance (Year 1)	wk23–wk52	23–52 (every 8 weeks)
Phase 4: Weight maintenance (Year 2)	wk56–wk104	56–104 (every 8 weeks)

Participant Visits / Longitudinal Events

The differing visit schedules of cases and controls means that we will probably work with two database arms

Phases	Participant Visit / REDCap Event (Cases)	Participant Visit / REDCap Event (Controls)
Phase 1:	pre-baseline, screening, baseline (wk00)	pre-baseline, screening, baseline
Phase 2: (TMR)	wk02, wk04, wk06, wk08, wk10, wk12	–
Phase 3: (Reintroduce Food)	wk14, wk16, wk18, wk20	–
Phase 4: (Maintain Weight, Year 1)	wk24, wk28, wk36, wk44, wk52 (**1yr**)	wk28 **(6m)**, wk52 **(1 yr)**
Phase 4: (Maintain Weight, Year 2)	wk65, wk78, wk91, wk104 (**2 yr**)	wk78 **(18m)**, wk104 **(2 yr)**

Data Collection Instrument Summary

Instrument	Repeating	Assigned events (Cases)
eligibility_screening	No	Screening *(This will be maintained in the Diabetes Registry)*
participant_admin	No	Baseline only
weight	No	Baseline, every follow-up visit
anthropometry	No	baseline, wk12, wk20, wk36, wk52, wk78, wk104
blood_pressure	No	baseline, every follow-up visit
glucose_monitor	No	baseline, every follow-up visit

Instrument	Repeating	Assigned events (Cases)
hba1c	No	baseline, wk12, wk20, wk27, wk36, wk52, wk78, wk104
liver_labs	No	baseline, wk12, wk52, wk104
==fibroscan==	==No==	==baseline, wk12, wk52, wk104==
medication	No	baseline, every follow-up visit
eq5d3l_items	No	baseline, wk12, wk52, wk104
paid_items	No	baseline, wk12, wk52, wk104
ipaq_sf	No	baseline, wk12, wk52, wk104
health_economics	No	baseline, wk52, wk104
prebaseline_registry	**Yes**	prebaseline *(This will be extracted directly from the Diabetes Registry)*

Instrument: Recruitment & Screening

1. For each new diabetes intervention, we create a new registry instrument. The logic is mainly to have a registry record of the number of intervention involvements for each individual - for some protection against over-studying
2. This will not be possible for TMR - as the registry needs further updating before it can be populated with additional data.
3. **This means - for TMR - recruitment & screening will occur in the main REDCap TMR database.** The suggested TMR eligibility instrument for the registry is tabulated below.
4. The TMR eligibility criteria are:
    - Age 65 years or younger
    - Duration of diabetes 6 years or less
    - 1 or more visits in past 12 months
    - IF (and only IF) we don't recruit enough participants (50 cases, 50 controls in both locations combined), we will consider diabetes duration 10-years or less.
    - BMI of 25 or greater
5. **Screening Part 1.** Using data from the diabetes registry, auto populate the TMR database screening form (Ascension = manual entry):
    - Date of Birth
    - Date of Diabetes Diagnosis
    - Date of Last Clinic Visit
6. Auto-Calculate:
    - Age
    - Duration of Diabetes
    - Time since last visit
7. Flag Eligible / Non-Eligible
8. Eligible participants are randomly reordered, then pushed to a REDCap "potentially eligible" report for use by the *community health workers*.

Eligibility Screening Part 1. Draw registry data into new registry instrument. Calculate derived variables.

Variable	Label	Field type	Coding	Required
dob_registry	**(auto-populate)** Date of birth	YYYY-MM-DD	–	Y
dodiag_registry	**(auto-populate)** Date of diabetes diagnosis	YYYY-MM-DD	–	Y
dolv_registry	**(auto-populate)** Date of last clinic visit	YYYY-MM-DD	–	Y
ddur_registry	***(derived)*** Duration of diabetes	text (decimal)	Present as Years and Weeks	Y
diab6_registry	***(derived)*** Has the potential participant been diagnosed for 6 years or less	yesno	0/1	Y
diab10_registry	***(derived)*** Has the potential participant been diagnosed for 10 years or less	yesno	0/1	Y
visit12_registry	***(derived)*** 1 or more visits in the past year	text (decimal)	–	Y
tmr_eligible1	***(auto-calculated)*** Is participant eligible so far for TMR pilot?	yesno	–	Y

Eligibility Process Part 2. If eligible from *Screening Part 1*, create report / list of these eligible participants, along with their eligible values. This list will be held in the TMR database, randomly reordered. It will be used by the *community health workers* when contacting potential participants for interest, and if interested to attend an in-person screening visit.

The Table below documents the additional variables collected at this in-person screening visit.

If the participant is still eligible after Part 2, they should be asked if they are interested in hearing more about the pilot, or interested in taking part.

- If they are interested - schedule a baseline visit
- If they are not interested - ask about using their registry data in the control group.

Variable	Label	Field type	Coding	Required
dob_screen	Confirm date of birth	YYYY-MM-DD	–	Y
dov_screen	Date of screening visit	YYYY-MM-DD	–	Y
ddur_screen	***(derived)*** Calculate duration of diabetes	text (decimal)	Present as Years and Weeks	Y
diab6_screen	***(derived)*** Has the potential participant been diagnosed for 6 years or less	yesno	0/1	Y
diab10_screen	***(derived)*** Has the potential participant been diagnosed for 10 years or less	yesno	0/1	Y

Variable	Label	Field type	Coding	Required
visit12_screen	***(derived)*** 1 or more visits in the past year	text (decimal)	–	Y
height_screen	Height	text (decimal)	cm	Y
weight_screen	Weight	text (decimal)	kg	Y
bmi_screen	***(derived)*** Body Mass Index	text (decimal)	kg/m2	Y
tmr_eligible2	***(auto-calculated)*** Is participant eligible for TMR pilot?	yesno	–	Y
interested_screen	Are you interested in taking part in the TMR pilot?	yesno	–	Y
visit_scheduled	If Yes to taking part, has a baseline visit been scheduled?	yesno	–	Y
control	If No to taking part, might we use their registry data in the control group?	yesno	–	Y
fu6m	If Yes to being a control, are they happy to visit for 1 baseline visit, then every 6 months for diabetes support?	yesno	–	Y

NOTE 1 (13-Jan-2026): As part of this screening process, there will now be some screening questions exploring eating habits. These are to be finalised.

NOTE 2 (13-Jan-2026): From here, instruments are post-screening, meaning that they are collected only if participants have agreed to join the study

Instrument: participant_admin

Collects core administrative identifiers and group allocation information required for participant linkage, stratification, and governance.

Variable	Label	Field type	Coding	Required
pid	Participant study ID	text	–	Y
site	Study site	radio	1=St Helena; 2=Ascension	Y
group	Study group **(defined by study arms)**	radio	1=Intervention (TMR); 2=Control	Y
registry_id2	Diabetes Registry ID	text	–	Y
dob	Date of birth	date	YYYY-MM-DD	Y

Variable	Label	Field type	Coding	Required
sex	Participant sex	radio	1=female2=male	Y
fname	First name	text	–	Y
lname	Last name	text	–	Y
phone	Phone number	text	–	Y
email	Email address	text	–	Y
consent_date	Date informed consent obtained	date	YYYY-MM-DD	Y

Instrument: weight

Records repeated measurements of body weight over time to quantify weight change / trajectories.

Variable	Label	Field type	Units	Required
visit_date	Date of measurement	date	–	Y
weight	Weight	text (decimal)	kg	Y

Instrument: anthropometry

Records repeated measurements of body circumferences over time to quantify body composition trajectories.

Variable	Label	Field type	Units	Required
visit_date	Date of measurement	date	–	Y
waist	Waist circumference	text (decimal)	cm	N
hip	Hip circumference	text (decimal)	cm	N

Instrument: blood_pressure

Captures systolic and diastolic blood pressure measurements over follow-up to assess secondary cardiometabolic outcomes.

Variable	Label	Field type	Units	Required
visit_date	Date of measurement	date	–	Y
sbp	Systolic blood pressure	integer	mmHg	Y
dbp	Diastolic blood pressure	integer	mmHg	Y

Instrument: glucose_monitor

Records glucose monitor readings over time to assess glycaemic control, remission, and relapse. These non-fasting values are primarily for clinical monitoring.

Variable	Label	Field type	Units	Required
visit_date	Date of measurement	date	–	Y
glucose	Blood glucose measurement	text (decimal)	mmol/L	Y

Instrument: hba1c

Records HbA1c measurements over time to assess glycaemic control, remission, and relapse.

Variable	Label	Field type	Units	Required
visit_date	Date of measurement	date	–	Y
hba1c	HbA1c	text (decimal)	mmol/mol	Y

Instrument: liver_labs

Records HbA1c measurements over time to assess glycaemic control, remission, and relapse.

Variable	Label	Field type	Units	Required
visit_date	Date of measurement	date	–	Y
alt	Alanine aminotransferase (ALT)	text (decimal)	U/L	Y
ast	Aspartate aminotransferase (AST)	text (decimal)	U/L	Y
alp	Alkaline phosphatase (ALP)	text (decimal)	U/L	Y
ggt	Gamma-glutamyl transferase (GGT)	text (decimal)	U/L	N
bilirubin_total	Total bilirubin	text (decimal)	µmol/L	Y
albumin	Albumin	text (decimal)	g/L	Y
protein_total	Total protein	text (decimal)	g/L	N

Instrument: fibroscan

Records liver disease variables using the Echosens fibroscanner. Variables to be finalised

Variable	Label	Field type	Coding / Units	Required
fibroscan_date	Date of FibroScan assessment	date	dd/mm/yyyy	y
lsm_kpa	Liver stiffness measurement	numeric	kilopascals (kPa)	y

Variable	Label	Field type	Coding / Units	Required
cap_dbm	Controlled attenuation parameter	numeric	decibels per metre (dB/m)	y
fibrosis_stage	Derived liver fibrosis stage	categorical	F0, F1, F2, F3, F4	n
steatosis_grade	Derived liver steatosis grade	categorical	S0, S1, S2, S3	n
lsm_iqr	Interquartile range of liver stiffness	numeric	kilopascals (kPa)	n
lsm_iqr_median_pct	IQR-to-median ratio for liver stiffness	numeric	percentage (%)	n
probe_type	FibroScan probe type used	categorical	M, XL	y
valid_measurements_n	Number of valid FibroScan measurements	integer	count	n
success_rate_pct	FibroScan measurement success rate	numeric	percentage (%)	n

Instrument: medication

Captures high-level information on insulin use, insulin dose, and overall glucose-lowering medication burden over time.

Variable	Label	Field type	Coding / Units	Required
visit_date	Date of measurement	date	–	Y
diabetes_meds	Are you on diabetes medication	1=monotherapy oral2=monotherapy injectable3=combination	1-3	Y
hypertension_meds	Are you on hypertension medication	yesno	0/1	Y

Instrument: eq5d3l_items

Collects EQ-5D-3L health-related quality of life responses to assess changes in generic health status over time.

Go to EuroQol website. REDCap instrument available - must complete registration form

Variable	Label	Field type	Coding	Required
eq_mob	Mobility	radio	1–3	Y
eq_sc	Self-care	radio	1–3	Y
eq_ua	Usual activities	radio	1–3	Y
eq_pd	Pain / discomfort	radio	1–3	Y
eq_ad	Anxiety / depression	radio	1–3	Y
eq_vas	We would like to know how good or bad your health is TODAY	integer	0–100	Y

Instrument: paid_items

Problem Areas in Diabetes (PAID) scale. Collects PAID questionnaire items to quantify diabetes-related emotional distress longitudinally.

Instructions: Which of the following diabetes issues are currently a problem for you? Tick the box that gives the best answer for you. Please provide an answer for each question

Variable	Label	Field type	Coding	Required
paid01	Not having clear and concrete goals for your diabetes care?	radio	Not a problem (0), Minor problem (1), Moderate problem (2), Somewhat serious problem (3), Serious problem (4)	Y
paid02	Feeling discouraged with your diabetes treatment plan?	radio	0–4	Y
paid03	Feeling scared when you think about living with diabetes?	radio	0–4	Y
paid04	Uncomfortable social situations related to your diabetes care (e.g. people telling you what to eat)?	radio	0–4	Y

Variable	Label	Field type	Coding	Required
paid05	Feelings of deprivation regarding food and meals?	radio	0–4	Y
paid06	Feeling depressed when you think about living with diabetes?	radio	0–4	Y
paid07	Not knowing if your mood or feelings are related to your diabetes?	radio	0–4	Y
paid08	Feeling overwhelmed by your diabetes?	radio	0–4	Y
paid09	Worrying about low blood glucose reactions?	radio	0–4	Y
paid10	Feeling angry when you think about living with diabetes?	radio	0–4	Y
paid11	Feeling constantly concerned about food and eating?	radio	0–4	Y
paid12	Worrying about the future and the possibility of serious complications?	radio	0–4	Y
paid13	Feelings of guilt or anxiety when you get off track with your diabetes management?	radio	0–4	Y
paid14	Not 'accepting' your diabetes?	radio	0–4	Y
paid15	Feeling unsatisfied with your diabetes physician?	radio	0–4	Y

Variable	Label	Field type	Coding	Required
paid16	Feeling that diabetes is taking up too much of your mental and physical energy every day?	radio	0–4	Y
paid17	Feeling alone with your diabetes?	radio	0–4	Y
paid18	Feeling that your friends and family are not supportive of your diabetes management efforts?	radio	0–4	Y
paid19	Coping with complications of diabetes?	radio	0–4	Y
paid20	Feeling 'burned out' by the constant effort needed to manage diabetes?	radio	0–4	Y

Instrument: ipaq_sf

Collects the IPAQ short form to quantify physical activity and sedentary behaviour across study follow-up.

Note. For use with young and middle aged adults.

Variable	Label	Field type	Units	Required
ipaq_walk_days	Walking days (last 7 days)	integer	days/week	Y
ipaq_walk_mins	Walking minutes per day	integer	min/day	Y
ipaq_mod_days	Moderate activity days	integer	days/week	Y
ipaq_mod_mins	Moderate activity minutes	integer	min/day	Y
ipaq_vig_days	Vigorous activity days	integer	days/week	Y
ipaq_vig_mins	Vigorous activity minutes	integer	min/day	Y

Variable	Label	Field type	Units	Required
ipaq_sit_mins	Sitting time per weekday	integer	min/day	Y

Instrument: health_economics

Captures programme product supply and medication cost inputs to support a pragmatic assessment of drug-related cost savings.

Confirm / Improve this variable set with Chief Economist

Variable	Label	Field type	Units	Required
he_currency	Currency used for costing	radio	GBP / USD / Other	Y
he_insulin_supplied	Number of weeks of insulin issued since the previous study visit	integer	weeks	Y
he_metformin_supplied	Number of weeks of metformin issued since the previous study visit	integer	weeks	Y
he_antihyp_supplied	Number of weeks of antihypertensive issued since the previous study visit	integer	weeks	Y
he_notes	Costing notes (e.g. supply interruptions or special circumstances)	notes	–	N

Instrument: prebaseline_registry (REPEATING)

Pre-baseline measurements maintained IN THE REGISTRY DATABASE

I'm suggesting that this information is NOT stored as a separate instrument. We extract directly from the Registry Database using an API script

Variable	Label	Field type	Units	Required
pre_date	Date of registry measurement	date	–	Y
pre_weight	Historical weight	text (decimal)	kg	N
pre_hba1c	Historical HbA1c	text (decimal)	mmol/mol	N
pre_sbp	Historical systolic BP	integer	mmHg	N
pre_dbp	Historical diastolic BP	integer	mmHg	N

Variable	Label	Field type	Units	Required
pre_diabetes_meds	Diabetes medication use at the time	1=monotherapy oral2=monotherapy injectable3=combination	1-3	Y
pre-hypertensives	antihypertensive use at that time	yesno	0/1	N

Appendices. Variables Uses. Derived Variables.

Appendix: Variable Collection Timing and Analytical Use

This appendix provides a **variable-by-variable enactment of the Statistical Analysis Plan (SAP)**.
For each variable, it documents: - when it is collected, - its analytical role, - model families in which it is used, - and its contribution to tables, figures, and narratives.

Unless explicitly stated, all derived variables are calculated **outside REDCap** in analysis code.

Appendix A: participant_admin

Variable	Collection timepoints	Detailed analytical use
pid	Baseline	Unique identifier for data management, dataset linkage, and longitudinal modelling; never analysed
site	Baseline	Fixed-effect covariate and stratification variable; appears in baseline descriptive tables and sensitivity analyses
group	Baseline	Primary exposure indicator (TMR vs control); used in matched cohort analyses, difference-in-differences (DiD) models, and all outcome tables
registry_id2	Baseline	Linkage to diabetes registry and construction of pragmatic control cohort; not analysed
dob	Baseline	Used to derive age; age enters propensity score models, regression covariate sets, and baseline tables
consent_date	Baseline	Governance and audit only

Appendix B: baseline_eligibility

Variable	Collection timepoints	Detailed analytical use
elig_reviewed	Baseline	Governance only
elig_t2d	Baseline	Defines analytic population (type 2 diabetes)

Variable	Collection timepoints	Detailed analytical use
elig_insulin_exclusion	Baseline	Stratifier for insulin vs non-insulin subgroup analyses
elig_notes	Baseline	Audit / quality assurance only

Appendix C: baseline_anthropometry

Variable	Collection timepoints	Detailed analytical use
height	Baseline	Used to derive BMI at all timepoints; baseline covariate in descriptive tables

Appendix D: anthropometry

Variable	Collection timepoints	Detailed analytical use
visit_date	All visits	Defines time scale for longitudinal and DiD models
weight	Baseline; all follow-ups	**Primary continuous outcome**; analysed using linear mixed-effects models and DiD regression; binary transformation used for 15 kg weight-loss endpoint; contributes to all primary outcome tables and trajectory figures
waist	Baseline; selected follow-ups	Secondary adiposity outcome; descriptive statistics and exploratory linear models
hip	Baseline; annual	Used to derive waist-hip ratio; descriptive cardiometabolic profiling

Appendix E: blood_pressure

Variable	Collection timepoints	Detailed analytical use
sbp	Baseline; all follow-ups	Secondary outcome; analysed using linear mixed-effects models; appears in metabolic outcome tables
dbp	Baseline; all follow-ups	Secondary outcome; analysed using linear mixed-effects models

Appendix F: glycaemia_labs

Variable	Collection timepoints	Detailed analytical use
hba1c	Pre-baseline; baseline; follow-ups	**Primary glycaemic outcome**; analysed as a continuous outcome in linear mixed-effects and DiD models; dichotomised for remission (<48 mmol/mol) and tighter control (<42 mmol/mol); used to define relapse; appears in all primary and secondary outcome tables

Appendix G: liver_labs

Variable	Collection timepoints	Detailed analytical use
alt	Baseline; wk12; 12m; 24m	Secondary metabolic outcome; analysed descriptively and using exploratory linear mixed-effects models
ast	Baseline; wk12; 12m; 24m	Interpreted alongside ALT; supports exploratory AST:ALT ratio analyses
alp	Baseline; follow-ups	Descriptive liver function context
ggt	Baseline; follow-ups	Exploratory marker of hepatic and metabolic stress
bilirubin_total	Baseline; follow-ups	Safety and liver function descriptor
albumin	Baseline; follow-ups	Liver synthetic function; safety descriptor
protein_total	Baseline; follow-ups	Descriptive only

Liver panel analytical notes: 1. Analyses are **secondary and exploratory**, intended to describe metabolic improvement rather than confirm liver disease outcomes.
2. ALT is treated as the primary liver enzyme of interest, reflecting hepatic fat reduction and insulin sensitivity.
3. AST:ALT ratio may be derived descriptively to aid interpretation but is not a formal endpoint.
4. No diagnostic claims regarding NAFLD/NASH are made; interpretation is contextual and supportive.

Appendix H: medication_use / medication_detail

Variable	Collection timepoints	Detailed analytical use
insulin_use	Pre-baseline; baseline; follow-ups	Binary outcome; stratifier; component of remission definition; logistic and DiD analyses
insulin_dose	Pre-baseline; baseline; follow-ups	Continuous outcome; analysed descriptively and via linear models; input to economic analyses
metformin_use	Pre-baseline; baseline; follow-ups	Binary covariate; remission definition; drug cost modelling
metformin_dose	Baseline; follow-ups	Dose reduction metric; economic analysis input
other_glu_use	Pre-baseline; baseline; follow-ups	Medication burden indicator; remission definition

Variable	Collection timepoints	Detailed analytical use
antihyp_use	Pre-baseline; baseline; follow-ups	Secondary outcome and covariate in BP models
lipid_use	Pre-baseline; baseline; follow-ups	Contextual cardiovascular prevention covariate

Medication analytical notes: - REDCap captures **summary medication status** sufficient for primary analyses.
- Where greater detail is required (e.g. specific agents, dose changes, dates), these may be extracted from **clinic notes or the EMR** and linked at analysis stage.
- REDCap is not treated as the authoritative medication prescribing system.

Appendix I: eq5d3l_items (EuroQol)

Variable	Collection timepoints	Detailed analytical use
eq_mob–eq_ad	Baseline; wk12; 12m; 24m	Ordinal health-related quality-of-life domains; descriptive distributions and transition matrices; sensitivity ordinal models
eq_vas	Baseline; wk12; 12m; 24m	Continuous self-rated health outcome; analysed using linear mixed-effects models

EuroQol analytical notes: 1. EQ-5D-3L health states are converted to **utility values** using the **UK value set (tariff)**.
2. Utility scores are analysed as continuous outcomes to assess change over time and between groups.
3. Results are interpreted as **health-related quality-of-life change**, not preference-based cost-utility outcomes.
4. Utilities may be reported descriptively and used to contextualise clinical outcomes; formal QALY modelling is **out of scope**.

Appendix J: paid_items

Variable	Collection timepoints	Detailed analytical use
paid01–paid20	Baseline; wk12; 12m; 24m	Item-level checks; scale construction

Derived (analysis only): - PAID total score (0–100): analysed using linear mixed-effects models
- PAID 40: binary high-distress outcome analysed using logistic or Poisson models
Appears in psychosocial outcome tables.

Appendix K: ipaq_sf

Variable	Collection timepoints	Detailed analytical use
ipaq_*	Baseline; wk12; 12m; 24m	Used to derive MET-minutes/week; included as descriptive outcome and covariate in adjusted models

Appendix L: health_economics_type1

Variable	Collection timepoints	Detailed analytical use
he_currency	Baseline	Costing metadata
he_tmr_weeks_supplied	At each supply visit	Interval-based programme exposure; multiplied by unit cost/week to estimate programme cost
he_notes	All	Interpretation of costing anomalies

Health-economic analytical notes: - Drug use and dose variables from REDCap are combined with **unit prices held centrally** to estimate medication costs.
- Where needed, detailed medication histories and quantities may be extracted from **clinic notes or EMR systems**.
- Analyses focus on **budget impact and cost offsets**, not full cost-effectiveness modelling.

Appendix M: prebaseline_registry

Variable	Collection timepoints	Detailed analytical use
pre_date	3 pre-baseline	Time alignment; parallel-trends diagnostics
pre_weight	3 pre-baseline	DiD outcome; pre-intervention slope estimation
pre_hba1c	3 pre-baseline	DiD outcome; pre-intervention slope estimation
pre_sbp	3 pre-baseline	Secondary DiD outcome
pre_dbp	3 pre-baseline	Secondary DiD outcome
pre_metformin_use	3 pre-baseline	Medication stability assessment

Appendix N: Derived Outcomes (Analysis Only)

Outcome	Exact calculation	Model family	SAP role
15 kg weight loss	Baseline weight − 12m weight 15 kg	Log-binomial / Poisson	Primary endpoint
Diabetes remission	HbA1c <48 mmol/mol **and** no glucose-lowering medication for 3 months	Log-binomial / Poisson	Primary endpoint
Tight glycaemic control	HbA1c <42 mmol/mol	Logistic / descriptive	Secondary
HbA1c change	HbA1c(time) − HbA1c(baseline)	Linear mixed-effects	Secondary

Outcome	Exact calculation	Model family	SAP role
Insulin cessation	insulin_use = 0 at 12m	Logistic	Exploratory (insulin cohort)
Stepped improvement	Ordered composite of HbA1c reduction and medication reduction	Ordinal regression	Exploratory
Relapse	HbA1c 48 mmol/mol after remission	Descriptive	Secondary
EQ-5D utility	UK tariff applied to EQ-5D-3L health states	Linear mixed-effects	Secondary
PAID high distress	PAID total 40	Logistic / Poisson	Secondary
MET-minutes/week	IPAQ scoring protocol	Linear / descriptive	Secondary

Appendix O: Derived Variables

Appendix P: Definitions, Derivation Rules, and Analytical Use

This appendix defines all derived variables required for analysis of the Saint Helena Total Meal Replacement (TMR) Diabetes Study. It is intended to be **fully implementation-ready**: an independent analyst should be able to recreate every derived variable exactly as specified and understand how and where each is used within the Statistical Analysis Plan (SAP).

Unless otherwise stated, all derivations are performed in analysis code (e.g. Stata, R, Python) using cleaned, analysis-ready datasets.

General Principles

- All dates are treated as calendar dates (ISO format recommended: YYYY-MM-DD).
- Durations are calculated in **decimal years** unless otherwise stated.
- Baseline is defined as the measurement closest to (and not after) the TMR start date.
- Follow-up windows align with the visit structure defined in the Data Dictionary and SAP.
- Binary variables are coded: 1 = Yes / achieved, 0 = No / not achieved.

P1. Demographic and Eligibility Derivations

P1.1 Age at Baseline (`age_bl`)

Source variables `dob`, `baseline_date`

Derivation

$$\text{age\_bl} = \frac{\text{baseline\_date} - \text{dob}}{365.25}$$

Format Continuous, years (1 decimal place)

Analytical use Baseline description; propensity score models; adjusted analyses.

P1.2 Duration of Diabetes (`ddur_yrs`)

Source variables `dodiag_registry`, `baseline_date`

Derivation

$$\text{ddur\_yrs} = \frac{\text{baseline\_date} - \text{dodiag\_registry}}{365.25}$$

Format Continuous, years (2 decimal places)

Analytical use Eligibility documentation; covariate adjustment; subgroup analyses.

P1.3 Short-Duration Diabetes Indicators

- `diab_le6`: 1 if `ddur_yrs` `6.0`, else 0
- `diab_le10`: 1 if `ddur_yrs` `10.0`, else 0

Analytical use Eligibility confirmation; sensitivity and subgroup analyses.

P2. Anthropometric Derivations

P2.1 Body Mass Index (`bmi`)

Source variables `weight` (kg), `height` (cm)

Derivation

$$\text{bmi} = \frac{\text{weight}}{(\text{height}/100)^2}$$

Format Continuous, kg/m² (1 decimal place)

Analytical use Baseline description; covariate adjustment.

P2.2 Absolute Weight Change (`wt_change_kg_t`)

Derivation

$$\text{wt\_change\_kg}_t = \text{weight}_t - \text{weight}_{bl}$$

Negative values indicate weight loss.

Analytical use Primary continuous outcome.

P2.3 Percentage Weight Change (`wt_change_pct_t`)

Derivation

$$\text{wt\_change\_pct}_t = \frac{\text{weight}_t - \text{weight}_{bl}}{\text{weight}_{bl}} \times 100$$

Analytical use Secondary descriptive outcome.

P2.4 15 kg Weight Loss Indicator (`wt_loss15_t`)

Derivation 1 if (`weight_bl - weight_t`) 15.0 kg; else 0.

Primary timepoint 12 months

Analytical use Primary binary endpoint.

P3. Glycaemic Derivations

P3.1 HbA1c Change (`hba1c_change_t`)

Derivation

$$\text{hba1c\_change}_t = \text{hba1c}_t - \text{hba1c}_{bl}$$

Analytical use Primary continuous glycaemic outcome.

P3.2 HbA1c Threshold Indicators

- `hba1c_lt48_t`: 1 if HbA1c < 48 mmol/mol
- `hba1c_lt42_t`: 1 if HbA1c < 42 mmol/mol

Analytical use Secondary binary outcomes.

P3.3 Diabetes Remission (`remission_t`)

Definition (non-insulin cohort) All criteria must be met:

1. HbA1c < 48 mmol/mol at time t
2. No glucose-lowering medication for 3 months prior

Medication status is determined from clinic notes and/or EMR extraction.

Primary timepoint 12 months

Analytical use Primary binary endpoint.

P3.4 Relapse from Remission (`relapse_t`)

Among participants previously in remission:

1 if HbA1c 48 mmol/mol at a subsequent follow-up; else 0.

Analytical use Durability analyses.

P4. Medication Burden Derivations

P4.1 Insulin Use Indicator (`insulin_use_t`)

1 if any insulin prescribed in the preceding 3 months; else 0.

Analytical use Stratification; remission definition component.

P4.2 Insulin Dose Change (`insulin_dose_change_t`)

Derivation

$$\text{dose}_t - \text{dose}_{bl}$$

Units: international units/day.

Analytical use Exploratory insulin-cohort analyses.

P4.3 50% Insulin Reduction (`insulin_red50_t`)

1 if `dose_t  0.5 × dose_bl`; else 0.

Analytical use Stepped improvement summaries.

P4.4 Medication Class Count (`med_class_n_t`)

Count of glucose-lowering medication classes in use at time t.

Analytical use Medication de-intensification summaries.

P5. Insulin-Cohort Stepped Improvement Outcome

P5.1 Ordinal Stepped Improvement (`step_improve_t`)

Level 1 – Best outcome HbA1c $<$ 48 mmol/mol **and** no glucose-lowering medication 3 months.

Level 2 – De-intensification HbA1c baseline + 3 mmol/mol or <53 mmol/mol *and* insulin dose reduced 50% or stopped *and* no net increase in other agents.

Level 3 – Unfavourable HbA1c worsened by >5 mmol/mol or 64 mmol/mol or increased medication burden.

Analytical use Primary exploratory insulin-cohort outcome.

P6. Liver and FibroScan Derivations (Exploratory)

P6.1 AST:ALT Ratio (`ast_alt_ratio`)

$$\text{AST/ALT}$$

Analytical use Descriptive metabolic context only.

P6.2 Fibrosis and Steatosis Staging

Derived using manufacturer-recommended cut-points applied to:

- Liver stiffness (kPa) → fibrosis stage (F0–F4)
- CAP (dB/m) → steatosis grade (S0–S3)

Analytical use Exploratory descriptive summaries only.

P7. Quality of Life and Distress Scores

P7.1 EQ-5D Utility Index (`eq5d_utility_t`)

Derived using UK MVH tariff applied to EQ-5D-3L health states.

Analytical use Secondary outcome; health-economic interpretation.

P7.2 PAID Total Score (`paid_total_t`)

Sum of 20 items × 1.25 → scale 0–100.

P7.3 High Diabetes Distress (`paid_ge40_t`)

1 if PAID 40; else 0.

Analytical use Secondary psychosocial outcome.

P8. Physical Activity (IPAQ) Derivations

P8.1 MET-Minutes per Week (`ipaq_metwk_t`)

- Walking: minutes × days × 3.3
- Moderate: minutes × days × 4.0
- Vigorous: minutes × days × 8.0

Total = sum of above.

P8.2 Activity Category (`ipaq_cat_t`)

Low / Moderate / High using standard IPAQ thresholds.

Analytical use Descriptive behavioural context.

P9. Economic Derivations

P9.1 Monthly Medication Cost (`med_cost_mo_t`)

$$\text{med\_cost\_mo}_t = \sum_m \frac{\text{weeks supplied}_{m,t} \times \text{weekly unit cost}_m}{4.345}$$

Analytical use Medication cost summaries; cost-offset analyses.

P9.2 Monthly Insulin Cost (`insulin_cost_mo_t`)

Subset of `med_cost_mo_t` including insulin only.

P9.3 Change in Monthly Medication Cost (`med_cost_change_mo_t`)

$$\text{med\_cost\_mo}_t - \text{med\_cost\_mo}_{bl}$$

Negative values indicate cost savings.

P9.4 Programme Cost per Participant (`tmr_cost_pp`)

Total programme delivery cost divided by number initiating TMR.

P9.5 Net Cost Offset (`net_cost_offset_t`)

$$(\text{med\_cost\_mo}_{bl} - \text{med\_cost\_mo}_{t}) \times 12 - \text{tmr\_cost\_pp}$$

Positive values indicate net annual savings.

Closing Note

This appendix defines all derived variables a priori and should be read alongside the Statistical analysis plan (sap), which specifies how these variables are combined, modelled, and interpreted.