ENDO-CARE

Pre-operative intentional weight loss to support post-operative recovery in patients with overweight and endometrial cancer: the ENDO-CARE feasibility randomised controlled trial

(Short title: Could supported weight loss reduce womb cancer surgery complications?)

Statistical Analysis Plan supplementary

Version v1.0 08Apr2025

Based on Protocol version v6.0, 17 Mar 2025

Based on Statistical Analysis Plan (SAP) version v1.0_23Oct2023
Based on Statistical Analysis Plan - Data Definition and Tables (SAP-DD&T or SAP-DDT) version to be determined

Trial registration: IRAS Project ID: 324534

Surgical Intervention Trials Unit (SITU)

CONTENTS

CHANGES FROM PREVIOUS VERSION OF SAP

This is a summary of key changes from earlier versions of the SAP, with particular relevance to protocol changes that have an impact on the design, definition, sample size, data quality/collection and analysis of the outcomes. The protocol version number and date are included.

Version number Issue date	Author of this issue	Protocol Version & Issue date	Significant changes from previous version together with reasons
V1.0_08Apr2025	Martyn Hill	V6.0_17Mar2025	Not applicable as this is the 1st issue

1. INTRODUCTION

The ENDO-CARE SAP includes this text

> *3.7.2 Secondary and other outcomes*
>
> *Further details regarding the secondary outcomes will be provided at a later date, (provisionally in a supplementary document) prior to database lock.*
>
> *Excerpt from ENDO-CARE_SAP_v1.0_14Mar2023.docx*

This is that supplementary document.

2. OVERVIEW

The secondary outcomes will be summarised descriptively by trial group. If appropriate, the effect size and 95% confidence intervals will be estimated with regression models. Both absolute and relative effect sizes will be reported.

Complications will be summarised using one or more of the following:

- count/percentage of participants with any complication
- count/percentage of participants with any complication by grade
- count/percentage of participants with the highest grade of complication reported
- count/percentage of participants with any type of complication
- count of total complications.

3. DEFINITION OF POPULATION FOR ANALYSIS

Main analysis: All randomised participants that underwent surgery will be included in the main analysis on an intention-to-treat basis (ITT) regardless of withdrawal or non-adherence. An intention-to-treat basis assigns all randomised participants to their originally allocated group, whether they received their allocated randomised treatment arm approach or not, or dropped out or not. Participants who had their surgery less than 20 days since randomisation will be excluded from this analysis.

Per-protocol analysis: We will also conduct an analysis on a per-protocol basis. A per-protocol basis includes only those randomised participants who met all eligibility criteria, had no major deviations from the study procedures, and adhered to their assigned interventions. For the intervention group, this includes the participants who achieved ≥5% weight loss from baseline to the day of surgery. For participants in the usual care group, this excludes those who lost ≥5% weight loss from baseline to the day of surgery. Participants who had their surgery less than 20 days since randomisation will be excluded from this analysis.

Adverse events: The adverse event analysis will include the participants in the control group and the participants commencing the intervention in the intervention group.

4. USE OF MODELS

4.1 Input variables

The SAP states that the mean and 95% confidence intervals for the secondary outcome variables will be estimated with adjusted regression models. Those adjustments will be our four input variables, specifically:

- treatment group. There are two treatment groups, the control and the intervention.
- baseline value (where applicable)
- time point (where applicable)
- stratification variables. The two stratification variables are BMI (</≥40 kg/m2) and age at diagnosis ((</≥65 years). These will be combined into one [1][2].

The extension from regression models to mixed models was not explicitly stated but is assumed. Such models will have fixed effects and may be further extended to include a random effect and/or an interaction. Details of those models include:

- The fixed effects are the adjustments
- The random effect is the participant/study ID
- The interaction is multiplicative. For example, an interaction between the treatment group and the time point (both of which are fixed) will be considered.

In models with interaction terms, the treatment effect will be the interaction term. In models without an interaction term, the treatment effect will be the treatment group.

4.2 Output variables

We expect to model all secondary outcome variables, except survival, recurrence, new primary/secondary cancer, re-admission rates, re-operation rates, costs and resource use, and adverse events. The final decision will depend on the available data. For example, if the counts for conversion to open surgery rates are very low, we will not model this outcome.

Therefore, we plan on analysing:

- Morbidity (five variables not including 2iii survival)
- Operative outcomes (seven variables not including re-operation and re-admission rates)
- Hospital stay (two variables)
- Anthropometry (two variables)
- Fitness (one variable)
- Health-related quality of life (HRQoL) (three variables)

That's twenty variables equating to at least 20 models, as some of those variables will be further subdivided (eg HRQoL variable "EQ-5D-5L" has two subdivisions, index and VAS).

4.3 Model selection for a given outcome variable

See Appendix "Use Of The Word "Linear" In Linear Models" for an explanation of the use of the word "linear" even if the underlying curve is a higher polynomial

The output variables may be continuous or categorical and the input variables may be categorical or continuous. This would imply a logistic model (logistic regression or discriminant analysis) or a linear model (linear regression or ANOVA or ANCOVA or mixed) or a generalized model (GLM) approach. As GLMs are the most capable, we will use GLMs for our models where possible.

5. LIST OF SECONDARY OUTCOMES

Table: secondary outcomes

Objectives	Outcome Measures	Timepoint(s)	Type	Distribution
To report between-group differences in				
2. Morbidity	i. Any morbidity	i.-ii. Discharge and 30-days post-op	Binary	Binomial
	ii. Morbidity by grade (I, II, IIIa, IIIb, IVa, IVb)	ditto	Binary	Binomial
	ii. Morbidity by highest grade	ditto	Binary	Binomial
	ii. Number of people with no morbidity	ditto	Binary	Binomial
	ii. Morbidity by type (e.g., arrythmia, wound infection, etc)	ditto	Binary	Binomial
	iii. Survival (Grade V)	iii: Discharge, 30-days post-op, 3 yrs	N/A: not model	
3. Oncological outcomes	iv. Fitness to receive planned adjuvant therapy	iv. 30-days post-op	Binary	Binomial
	v. Recurrence	v-vi: 3 years	N/A: not model	
	vi. New primary/ secondary cancer	ditto	N/A: not model	
4. Operative outcomes	vii. Intraoperative blood loss			
	Count/rate (transfused Y/N)	vii-ix: Discharge	Binary	Binomial
	units transfused (including untransfused)	ditto	Continuous	See lookup
	viii. Operative time	ditto	Continuous	See lookup
	ix. Conversion to open surgery	ditto	Binary	Binomial
	x. Surgical site infection	x-xiii: Discharge and 30-days post-op	Binary	Binomial
	xi. Length of time in intensive care unit	ditto	Continuous	See lookup
	xi. Length of time in high dependency unit	ditto	Continuous	See lookup
	xii. Re-operation rates	xii-xiii: 30-days post-op and 3 years	N/A:not model	
	xiii. Re-admission rates	ditto	N/A:not model	
5. Hospital stay	xiv. Length of hospital stay (fitness to discharge)	xiv: Discharge	Time-to-event	See lookup
	xv. Days alive and out of hospital	xv: 30-days post- operatively	Time-to-event	See lookup
6. Anthropometry	xvi. Weight	xvi-xvii: Baseline, pre-operative assessment 4, and 30 days post-op	Continuous	See lookup
	xvii. Fat-free mass (absolute kg)	ditto	Continuous	See lookup
	xvii. Fat-free mass (relative %)	ditto	Continuous	See lookup
7. Fitness	xviii. Time for sit-to-stand test	xviii: Baseline, 30 days post-op	Continuous	See lookup
8. HRQoL	xix. EQ-5D-5L (Index score)	xix-xx: Baseline, pre-operative assessment 3 and 30 days post-op	Continuous	See lookup
	xix. EQ-5D-5L (VAS)	ditto	Continuous	See lookup
	xx. HADS (anxiety) continuous	ditto	Continuous	See lookup
	xx. HADS (depression) continuous	ditto	Continuous	See lookup
	xxi. EORTC-QLQ-EN24			
	29 individual items	ditto	N/A: not model	
9. Costs & resource use	N/A - dealt with by health economics			
10. Adverse events	xxii. Adverse events (excluding morbidity i-vi)	xxii: Baseline, pre-operative assessment and 30 days post-op	N/A: not model	

Details of each variable are provided below.

6. MORBIDITY

Morbidity

#	Outcome Measures	Control	Intervention	Delta
2a	i. Any morbidity			
	At or prior to discharge	X	X	X
	Between discharge and 30 days	X	X	X
	Other	X	X	X
	Total	X	X	X
	At or prior to discharge OR between discharge and 30 days	X	X	X
2b	ii. Morbidity by grade (I, II, IIIa, IIIb, IVa, IVb)			
	at least one Grade I	X	X	X
	.	X	X	X
	at least one Grade IVb	X	X	X
	Other	X	X	X
	Total	X	X	X
2c	ii. Morbidity by highest grade	X	X	X
2d	ii. Number of people with no morbidity	X	X	X
2e	ii. Morbidity by type (eg arrythmia, wound infection, etc)			
	Type	X	X	X
	.	X	X	X
	Type	X	X	X
	Other	X	X	X
	Total	X	X	X
2f	iii. Survival (grade V)	X	X	X

Note: the "other" categories are included for completeness and will be omitted if absent
Note: this is a list of patients, not complications. Because a patient can have more than one complication of different times/grades, an arm may not add to 100%. See appendices for worked examples

6.1 i. Any morbidity

This secondary outcome is reported regardless of timepoint. The information to calculate it is derived from the morbidity grade as specified on the CRFs Complications Grading 1 and Complications Grading 2. It is a binary variable and collectively a count/rate (0-n/%). The entry is defined as the number of patients with at least one complication divided by the number of patients.

For worked examples of this please see the appendices.

6.2 ii. Morbidity by grade (I, II, IIIa, IIIb, IVa, IVb)

This secondary outcome is reported by grade regardless of timepoint (i.e., a complication can occur at either discharge or 30-days follow-up). The information to calculate it is derived from the morbidity grade as specified on the CRFs Complications Grading 1 and Complications Grading 2. The entry for each grade X is is defined as the number of patients with at least one complication of grade X divided by the number of patients. It is a categorical variable. Each category is a count/rate (0-n/%) and will be modelled as a binary variable.

For worked examples of this please see the appendices.

6.3 ii Morbidity by highest grade

As ii above. This secondary outcome is reported for a single grade regardless of timepoint. If a participant has one coded complication, the grade of that complication will count as the highest graded morbidity. If a participant has at least two complications, the highest grade will be counted for this variable. It is a categorical variable & each category is a count/rate (0-n/%). Each category will be modelled as a binary variable.

For worked examples of this please see the appendices.

6.4 ii. Number of people with no morbidity

As i above.

6.5 ii. Morbidity by type (eg arrythmia, wound infection, etc)

This secondary outcome is reported regardless of timepoint. The information to calculate it is derived from the complication type as specified on the CRFs Complications Grading 1 and Complications Grading 2 (as derived from previous CRFs eg Hospital Stay and Discharge). It is a categorical variable & each category is a count/rate (0-n/%). Each category will be modelled as a binary variable.

6.6 iii. Survival (grade V)

This secondary outcome is reported for a single grade regardless of timepoint. The information to calculate it is derived from the morbidity grade as specified on the CRFs Complications Grading 1 and Complications Grading 2. It is a binary variable and collectively a count/rate (0-n/%). Given the expected low count, it is unlikely that this will be analysed statistically.

7. ONCOLOGICAL OUTCOMES

Oncological Outcomes

#	Outcome Measures	Control	Intervention	Delta
3a	iv.Fitness to receive planned adjuvant therapy	X	X	X
3b	v. Recurrence	X	X	X
3c	vi. New primary/ secondary cancer	X	X	X

7.1 iv. Fitness to receive planned adjuvant therapy

This secondary outcome is reported at 30-days follow-up. The information to calculate it is derived from the CRF Treatment. It is a binary variable and collectively a count/rate (0-n/%).

7.2 v. Recurrence

This secondary outcome is reported at 3 years. The information to calculate it is derived from the recurrence or metastatic type data on the CRF. It is a binary variable and collectively a count/rate (0-n/%). Due to low expected numbers, we do not plan an analysis for this. It will be depicted using survival curves.

7.3 vi. New primary/ secondary cancer

This secondary outcome will be collected at the 3-year follow-up. It is a binary variable and collectively a count/rate (0-n/%). Due to low expected numbers, we do not plan an analysis for this It will be depicted using survival curves.

8. OPERATIVE OUTCOMES

Operative Outcomes

#	Outcome Measures	Control	Intervention	Delta
4a	vii. Intraoperative blood loss			
	Count/rate (transfused Y/N)	X	X	X
	number of units transfused (including untransfused)	X	X	X
4b	viii. Operative time	X	X	X
4c	ix. Conversion to open surgery	X	X	X
4d	x. Surgical site infection	X	X	X
4g	xi. Time in intensive care unit	X	X	X
4g	xi. Time in high dependency unit	X	X	X
4h	xii. Re-operation rates	X	X	X
4i	xiii. Re-admission rates	X	X	X

8.1 vii. Intraoperative blood loss

This outcome is reported using two variables: by the presence of intraoperative blood transfusion and by the number of units transfused (including the untransfused). The information to calculate it is derived from the fact of an intraoperative blood transfusion and the number of blood units used. These variables are recorded on the CRF Hospital Stay and Discharge.

The first variable is a binary variable and collectively a count/rate (0-n/%). The second variables are continuous variables and measurements of amount.

8.2 viii. Operative time

The information to calculate this outcome is derived from the total time taken for the procedure on the CRF Hospital Stay and Discharge. It is a continuous variable and a measurement of time.

8.3 ix. Conversion to open surgery

The information to calculate this outcome is derived from the operation performed as recorded on the CRF Hospital Stay and Discharge. It is a binary variable and collectively a count/rate (0-n/%).

8.4 x. Surgical site infection

The information to calculate this outcome is derived from the complications data on CRF Complications Grading 1 (as derived from previous CRFs eg Hospital Stay and Discharge). It is a binary variable and collectively a count/rate (0-n/%).

8.5 xi. Time in intensive care unit and time in high dependency unit

The information to calculate these two lengths of time is derived from the ward and time details on the CRF Hospital Stay and Discharge. Each one is a continuous variable and a measurement of time.

8.6 xii. Re-operation rates

The information to calculate this outcome is derived from the return to theatre data on the CRFs Hospital Stay and Discharge. It is a binary variable and collectively a count/rate (0-n/%).

8.7 xiii. Re-admission rates

The information to calculate this outcome is derived from the readmission data on the CRF Hospital Stay and Discharge. It is a binary variable and collectively a count/rate (0-n/%).

9. HOSPITAL STAY

#	Outcome Measures	Control	Intervention	Delta
5a	xiv. Length of hospital stay (fitness to discharge)	X	X	X
5b	xv. Days alive and out of hospital	X	X	X

9.1 xiv. Length of hospital stay (fitness to discharge)

We will calculate the numbers of days of hospital stay using the date of admission and the date of discharge. The information to calculate it is derived from the admission and discharge data on the CRFs Admission and Hospital Stay and Discharge. It is a time-to-event variable and a measurement of time. Instructions for modelling time-to-event variables are given in the appendices.

9.2 xv. Days alive and out of hospital

This secondary outcome is displayed by timepoint. We will calculate the number of days based on the dates of admission, discharge, and 30-day follow-up. The information to calculate it is derived from the discharge and other data on the CRFs Hospital Stay and Discharge and Anthropometry. It is a time-to-event variable and a measurement of time. Instructions for modelling time-to-event variables are given in the appendices.

The possibility of censorship must be noted here.

- If somebody has surgery on the 1st, discharged on the 10th and followed-up on the 30th, then on the 30th the state is known: at that date the days alive and out of hospital is known to be 20 days.
- But if somebody has surgery on the 1st, discharged on the 10th and followed-up on the 27th, then on the 30th the state is not known with certainty as they may have died in the meantime.

To prevent this censorship, we will seek confirmation from the local research teams that the person is still alive on the 30th.

10. ANTHROPOMETRY

#	Outcome Measures	Control	Intervention	Delta
6a	xvi. Weight	X	X	X
6b	xvii. Fat-free mass	X	X	X

10.1 xvi. Weight

This secondary outcome is displayed by timepoint. The information to calculate it is derived from the weight data on the CRFs Anthropometry and Screening Visit. It is a continuous variable and a measurement of mass.

10.2 xvii. Fat-free mass

This secondary outcome is displayed by timepoint. The information to calculate it is derived from the weight and body fat data on the CRFs Anthropometry and Screening Visit. It is a continuous variable and a measurement of mass. It is reported both as an absolute value (the fat-free mass in kg) and a relative value (the fat-free mass divided by the weight and depicted as a % of body mass).

11. FITNESS

#	Outcome Measures	Control	Intervention	Delta
7a	xviii. Time for sit-to-stand test	X	X	X

11.1 xviii. Time for sit-to-stand test

This secondary outcome is displayed by timepoint. The information to calculate it is derived from the sit-to-stand data on the CRF Sit To Stand Test. It is a continuous variable and a measurement of time.

12. HRQOL (HEALTH-RELATED QUALITY OF LIFE)

#	Outcome Measures	Control	Intervention	Delta
8a	xix. EQ-5D-5L			
	Index score	X	X	X
	VAS	X	X	X
8b	xx. HADS			
	• Anxiety (continuous)	X	X	X
	• Depression (continuous)	X	X	X
8c	xxi. EORTC-QLQ-EN24			
	individual items	X	X	X

12.1 xix. EQ-5D-5L

This secondary outcome is displayed by timepoint. The information to calculate it is derived from the CRF EQ-5D-5L. It is in two parts: a summary index value derived from the code, and a visual analogue scale score (VAS). Instructions for calculating these values are given in the appendices.

- The EQ-5D-5L summary index value ranges from less than 0 (indicating a health state worse than death) to 1 (indicating full health). Higher scores represent better health states.
- The EQ-5D-5L Visual Analogue Scale (VAS) ranges from 0 to 100. The maximum value is 100 and represents the best health you can imagine. The minimum value is 0 and represented the worst health you can imagine.

We will treat both variables as scores and continuous variables. The scoring rules are in the appendices.

12.2 xx. HADS (anxiety)

This secondary outcome is displayed as both a continuous and categorical variable. The information to calculate it is derived from the CRF HADS and Pre-Surgery Questionnaire. It is a score. The scoring rules are in the appendices. The continuous version is detailed here. The other variants are detailed in the subgroup analysis.

- Continuous. The continuous version of this variable will be a number derived from that score. It will be modelled as a continuous variable

12.3 xx. HADS (depression)

This secondary outcome is displayed as both a continuous and categorical variable . The information to calculate it is derived from the CRF HADS and Pre-Surgery Questionnaire. It is a score. The scoring rules are in the appendices. The continuous version is detailed here. The other variants are detailed in the subgroup analysis.

- Continuous. The continuous version of this variable will be a number derived from that score. It will be modelled as a continuous variable

12.4 xxi. EORTC-QLQ-EN24

This secondary outcome is displayed by timepoint. The information to calculate it is derived from the CRF EORTC QLQ - EN24. It is a continuous variable and a score. The scoring rules are in the appendices. The score will be calculated for the 24 individual items. The 24 individual items will be descriptive only

13. COSTS AND RESOURCE USE

These outcomes will be calculated as part of the health economics section of the trial and are, therefore, not covered in this document.

14. ADVERSE EVENTS

#	Outcome Measures	Control	Intervention	Delta
10a	xxii. Adverse events	X	X	X

14.1 xxii. Adverse events

The information to calculate it is derived from CRF Adverse Events (any time). It is a categorical variable & each category is a count/rate (0-n/%).

#	Outcome Measures	Control	Intervention	Delta
10b	xxii. Adverse events (by severity)			
	N of AEs in category Mild	X	X	X
	N of AEs in category Moderate	X	X	X
	N of AEs in category Serious	X	X	X
10c	xxii. Adverse events (by seriousness)			
	N of AEs in category Serious	X	X	X
	N of AEs in category Non-serious	X	X	X
10d	xxii. Adverse events (misc)			
	N of people with at least one AE	X	X	X
	N of people with at least one moderate AE	X	X	X
	N of people with at least one serious AE	X	X	X
	N of people with at least one SAE	X	X	X
	N of people with at least one moderate AE and/or SAE	X	X	X

15. FURTHER CONSIDERATIONS ON THE ANALYTICAL APPROACH

15.1 Missing data

Missing data will be checked and either corrected if wrong or left if legitimate. Missing data will not be imputed (but see sensitivity analysis). The amount of missing data will be quantified and analyses for variables with missing data will be noted as such.

Mixed effects models usually do not require imputation as they are usually robust to missing data (although this depends if missing data is missing at random or not).

15.2 Outliers

Outliers will be checked and either corrected if wrong or left if legitimate. Outliers will not be removed. Analyses for variables with outliers will be noted as such. There are various methods of identifying an outlier, including the “3SD’s beyond the mean” heuristic. We will use the “1.5IQR rule”: outliers will be those points below Q1 − 1.5IQR, or above Q3 + 1.5IQR [6].

15.3 Subgroup Analyses

As stated in the section named “Statistics and Analysis” in the protocol, no formal subgroup analyses are planned. However, subgroup analyses requested after the protocol was written are permissible on an exploratory and descriptive basis at the discretion of the Trial Statistician. These subgroup

analyses are referred to as "informal". With the exception of the stratification factors (see sensitivity analysis below) those informal analyses are descriptive and without the need to calculate or report p-values.

15.3.1 HADS (anxiety)

#	Outcome Measures	Control	Intervention	Delta
8b	xx. HADS			
	• Anxiety (is +ve change since baseline >= MCID, Y/N)	X*	X*	X†
	• Anxiety (is -ve change since baseline >= MCID, Y/N)	X*	X*	X†

† = difference between arms: ie compare the control change since baseline to the intervention change since baseline.
* = this is a count and percentage of all the patients in that arm whose change since baseline exceeds the MCID.

- Anxiety (is +ve change since baseline >= MCID, Y/N)
  - This is the proportion of participants with a positive change at the pre-op 3 assessment since baseline of at least 1.7 points as the minimum clinically important difference (Y/N). This is a binary variable.

- Anxiety (is -ve change since baseline >= MCID, Y/N)
  - This is the proportion of participants with a negative change at the pre-op 3 assessment since baseline of at least 1.7 points as the minimum clinically important difference (Y/N). This is a binary variable.

The minimum clinically important difference for HADS is 1.7 points, see Lemay KR, et al [8]

15.3.2 HADS (depression)

#	Outcome Measures	Control	Intervention	Delta
8b	xx. HADS			
	• Depression (is +ve change since baseline >= MCID, Y/N)	X*	X*	X†
	• Depression (is -ve change since baseline >= MCID, Y/N)	X*	X*	X†

† = difference between arms: ie compare the control change since baseline to the intervention change since baseline.
* = this is a count and percentage of all the patients in that arm whose change since baseline exceeds the MCID.

As HADS (anxiety) above.

15.4 Sensitivity Analyses

Sensitivity analyses will be done on at least the following:

- **Per-protocol analysis.** The per-protocol analysis will repeat the above models using the per protocol population as defined in section 3.

- **Stratification factors.** The two stratification factors are BMI (</≥40 kg/m2) and age at diagnosis ("< 65 years" and "≥65 years"). For sensitivity analysis we will do the following:
  - Repeat the models with an interaction term between age at diagnosis ("< 65 years" and "≥65 years") and treatment. This will be expanded to "…and treatment and time" if time is an input variable, see below.

- Repeat the models with an interaction term between BMI (</≥40 kg/m2) and treatment. This will be expanded to "…and treatment and time" if time is an input variable, see below.
- Models measuring the state at one timepoint (eg discharge) or the state change between two timepoints will not need a time variable. Models measuring the state at two or more timepoints (eg baseline vs admission vs 30-days post-op) will need a time variable. In such cases, the interaction above will include the time variable. The paper at [7] is an example of this.

- **Other.** Other sensitivity analyses may be conducted if deemed necessary by the Trial Statistician.

16. REFERENCES

- [1] Kahan. 2013. 'Adjusting for multiple prognostic factors in the analysis of randomised trials', BMC Med Res..” This is one of the papers mentioned by CI Koutoukidis, see https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-13-99 and https://doi.org/10.1186/1471-2288-13-99
- [2] Paper 3: Kahan, B. C., and T. P. Morris. 2012. 'Improper analysis of trials randomised using stratified blocks or minimisation', Stat Med, 31: 328-40. This is another of the papers mentioned by CI Koutoukidis, see https://onlinelibrary.wiley.com/doi/epdf/10.1002/sim.4431
- [3] “Intention-To-Treat (ITT) vs. Per-Protocol (PP) analysis: what to choose?”, Clinfo.eu, see https://www.clinfo.eu/itt-vs-pp/
- [4] ICH Topic E9, "Statistical Principles for Clinical Trials, step 5: Note For Guidance On Statistical Principles For Clinical Trials (CPMP/ICH/363/96)", European Medicines Agency, September 1998, see https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf
- [5] Green, J. A. (2021). Too many zeros and/or highly skewed? A tutorial on modelling health behaviour as count data with Poisson and negative binomial regression. Health Psychology and Behavioral Medicine, 9(1), 436–455. https://doi.org/10.1080/21642850.2021.1920416
- [6] Carnegie Mellon University, Statistical Computing, Fall 2013, instructor Prof. Cosma Shalizi, see https://www.stat.cmu.edu/~cshalizi/statcomp/13/labs/05/lab-05.pdf
- [7] Ardoin et al (2014), “Secondary analysis of APPLE study suggests atorvastatin may reduce atherosclerosis progression in pubertal lupus patients with higher C reactive protein”. Ann Rheum Dis. 2014 Mar;73(3):557-66. doi: 10.1136/annrheumdis-2012-202315. Epub 2013 Feb 22. PMID: 23436914; PMCID: PMC4104199. See https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4104199/
- [8] Ref for MCID: Lemay KR, Tulloch HE, Pipe AL, Reed JL. “Establishing the Minimal Clinically Important Difference for the Hospital Anxiety and Depression Scale in Patients With Cardiovascular Disease.” J Cardiopulm Rehabil Prev. 2019 Nov;39(6):E6-E11. doi: 10.1097/HCR.0000000000000379. PMID: 30489438. See the link at https://pubmed.ncbi.nlm.nih.gov/30489438/

17. APPENDIX: SCORING RULES

Where available, scoring rules are given below. To reduce duplication, rules that are explicated in the tables above are not duplicated here.

17.1 EQ-5D-5L: HRQoL using Patient Reported Outcome Measure (PROMs) EQ-5D-5L

- *Sources: https://euroqol.org/wp-content/uploads/2021/01/EQ-5D-5LUserguide-08-0421.pdf*
- *Online calculator: https://euroqol.org/wp-content/uploads/2020/12/ENG_value-set_STATA.txt*

17.1.1 Summary index value

Each variable has a tick box, where each dimension has five response levels: 1=no problems, 2=slight problems, 3=moderate problems, 4=severe problems, 5=unable to /extreme problems.

The answers are concatenated into a code (eg "12213") and converted to a utility code, either manually or via code such as STATA. The scoring rules to translate the code into a utility code can be gotten from the user guide "EQ-5D-5L User Guide" at the EQ-5D website at https://euroqol.org/wp-content/uploads/2021/01/EQ-5D-5LUserguide-08-0421.pdf

Conversely, the scoring rules for translating the code to an utility code can be deduced by looking at the STATA code available https://euroqol.org/wp-content/uploads/2020/12/ENG_value-set_STATA.txt. That dataset gives us this lookup table:

variable	value	weight
mobility	1	0
mobility	2	0.058
mobility	3	0.076
mobility	4	0.207
mobility	5	0.274
selfcare	1	0
selfcare	2	0.05
selfcare	3	0.08
selfcare	4	0.164
selfcare	5	0.203
activity	1	0
activity	2	0.05
activity	3	0.063
activity	4	0.162
activity	5	0.184
pain	1	0
pain	2	0.063
pain	3	0.084
pain	4	0.276

pain	5	0.335
anxiety	1	0
anxiety	2	0.078
anxiety	3	0.104
anxiety	4	0.285
anxiety	5	0.289

So a code of "12213" would translate as 1x0+2x0.05+2x0.05+1x0+3x0.104 = 0+0.1+0.1+0+0.312=0.512

The EQ-5D-5L summary index value ranges from less than 0 (indicating a health state worse than death) to 1 (indicating full health). Higher scores represent better health states.

17.1.2 Visual Analogue Scale (VAS)

The EQ-5D-5L Visual Analogue Scale (VAS) ranges from 0 to 100.

- Maximum Value: 100, representing the best health you can imagine.
- Minimum Value: 0, representing the worst health you can imagine.

17.2 EORTC-QLQ-EN24

- *Sources: https://www.eortc.org/app/uploads/sites/2/2018/08/Specimen-EN24-English.pdf and EN24 manual*

The EORTC QLQ-EN24 is a patient-reported outcome measure to evaluate health-related quality of life among endometrial cancer (EC) patients in research and clinical practice. The questionnaire includes 24 items, numbered 31 to 54. They are divided into two categories: items about functions (sexual interest, sexual activity, sexual enjoyment) and symptoms (the rest)

All the QLQ-EN24 questions have four possible answers. They are

- *Not at all (1)*
- *A little (2)*
- *Quite a bit (3)*
- *Very much (4)*

Those four levels are a Likert scale of four response categories and are linearly converted into a scale from 0 to 100 (ie 0,33,67,100 or 100,67,33,0). Patients are asked to indicate their symptoms during the past week except for the sexuality questions, which ask for the past four weeks.

Table: EORTC-QLQ-EN24 Functional Scales (0 is bad, 100 is good)

#	Scale (code)	Scale (name)	Question
49	SXI	Sexual interest	To what extent were you interested in sex?
50	SXA	Sexual activity	To what extent were you sexually active?
54	SXE	Sexual enjoyment	Was sexual activity enjoyable for you?

Table: EORTC-QLQ-EN24 Symptom scales (0 is good, 100 bad)

#	Scale (code)	Scale (name)	Question
31	LY	Lymphoedema	Have you had swelling in one or both legs?
32	LY	Lymphoedema	Have you felt heaviness in one or both legs?
33	BP	Pain in back and pelvis	Have you had pain in your lower back and / or pelvis?
34	UR	Urological symptoms	When you felt the urge to pass urine, did you have to hurry to get to the toilet?
35	UR	Urological symptoms	Have you passed urine frequently?
36	UR	Urological symptoms	Have you had leaking of urine?
37	UR	Urological symptoms	Have you had pain or a burning feeling when passing urine?
38	GI	Gastrointestinal symptoms	When you felt the urge to move your bowels, did you have to hurry to get to the toilet?
39	GI	Gastrointestinal symptoms	Have you had any leakage of stools?
40	GI	Gastrointestinal symptoms	Have you been troubled by passing wind?
41	GI	Gastrointestinal symptoms	Have you had cramps in your abdomen?
42	GI	Gastrointestinal symptoms	Have you had a bloated feeling in your abdomen?
43	TN	Tingling/numbness	Have you had tingling or numbness in your hands or feet?
44	MP	Muscular pain	Have you had aches or pains in your muscles or joints?
45	HL	Hair loss	Have you lost hair?
46	TC	Taste Change	Has food and drink tasted differently from usual?
47	BI	Body image	Have you felt physically less attractive as a result of your disease or treatment?
48	BI	Body image	Have you felt less feminine as a result of your disease or treatment?
51	SXV	Sexual/vaginal problems	Has your vagina felt dry during sexual activity?
52	SXV	Sexual/vaginal problems	Has your vagina felt short and / or tight?
53	SXV	Sexual/vaginal problems	Have you had pain during sexual intercourse or other sexual activity?

17.3 HADS

- *Sources: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5700594, https://academic.oup.com/occmed/article/64/5/393/1436876, https://score.awellhealth.com/calculations/hads/documentation*
- *Online calculator: https://score.awellhealth.com/calculations/hads*
- *Note: HADS is a generated score in the SITU RedCap for this project.*

The HADS questionnaire measures hospital anxiety or depression level.

It has fourteen questions divided into two subscales. There are seven questions for one subscale (depression) and seven questions for the other (anxiety). Scoring for each item ranges from zero to three, with three denoting highest anxiety or depression level. So depression ranges from 0 (good) to 21 (bad), and anxiety ranges from 0 (good) to 21 (bad). A score of more than 8 means considerable symptoms of anxiety or depression.

Table: items on the questionnaire that relate to anxiety

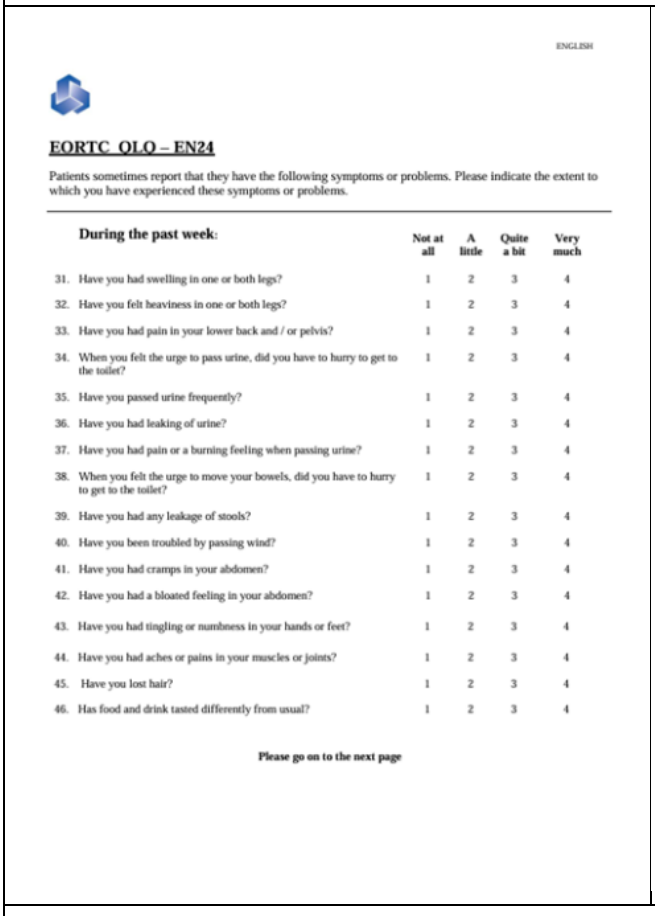
QUESTIONS	ANSWERS	SCORE
I feel tense or wound up	Most of the time; A lot of the time; From time to time, occasionally; Not at all	3-0
I get a sort of frightened feeling as if something awful is about to happen	Very definitely and quite badly; Yes, but not too badly; A little, but it doesn't worry me; Not at all	3-0
Worrying thoughts go through my mind	A great deal of the time; A lot of the time; From time to time, but not too often; Only occasionally	3-0
I can sit at ease and feel relaxed	Definitely; Usually; Not often; Not at all	0-3
I get a sort of frightened feeling like 'butterflies' in the stomach	Not at all; occasionally; Quite often; Very often	0-3
I feel restless as I have to be on the move	Very much indeed; Quite a lot; Not very much; Not at all	3-0
I get sudden feelings of panic	Very often indeed; Quite often; Not very often; Not at all	3-0

Table: items on the questionnaire that relate to depression

QUESTIONS	ANSWERS	SCORE
I still enjoy the things I used to enjoy	Definitely as much; Not quite so much; Only a little; Hardly at all	0-3
I can laugh and see the funny side of things	As much as I always could; Not quite so much now; Definitely not so much now; Not at all	0-3
I feel cheerful	Not at all; Not often; Sometimes; Most of the time	3-0
I feel as if I am slowed down	Nearly all the time; Very often; Sometimes; Not at all	3-0
I have lost interest in my appearance	Definitly; I don't take as much care as I should; I may not take quite as much care ; I take just as much care as ever	3-0
I look forward with enjoyment to things	As much as I ever did; Rather less than I used to; Definitely less than I used to; Hardly at all	0-3
I can enjoy a good book or radio or TV program	Often; Sometimes; Not often; Very seldom	0-3

18. APPENDIX: SCREENSHOT OF EORTC-QLQ-EN24 QUESTIONNNAIRE

Table: Image of questionnaire EORTC-QLQ-EN24



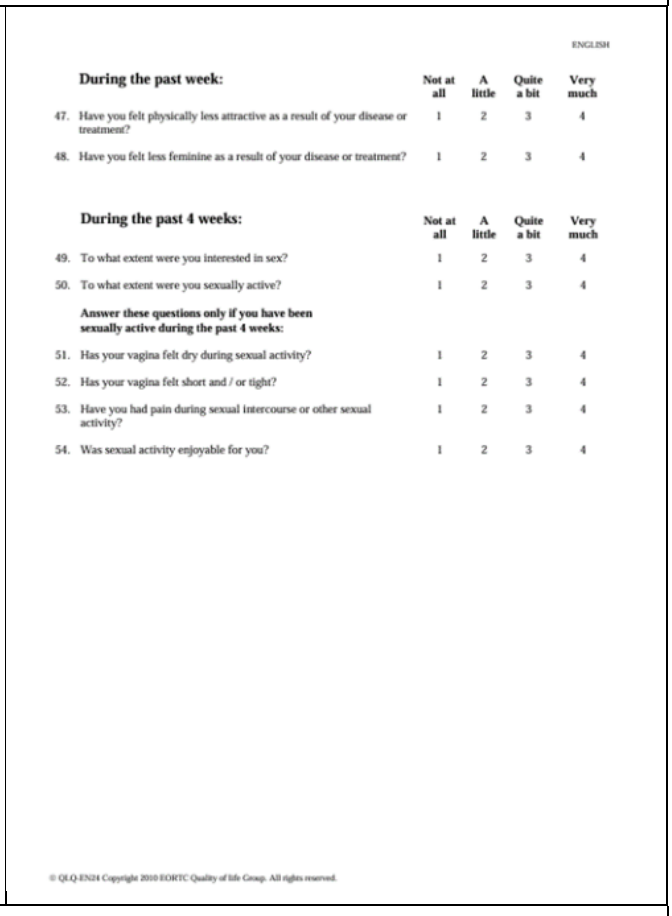
ENGLISH

EORTC QLQ – EN24

Patients sometimes report that they have the following symptoms or problems. Please indicate the extent to which you have experienced these symptoms or problems.

During the past week:	Not at all	A little	Quite a bit	Very much
31. Have you had swelling in one or both legs?	1	2	3	4
32. Have you felt heaviness in one or both legs?	1	2	3	4
33. Have you had pain in your lower back and / or pelvis?	1	2	3	4
34. When you felt the urge to pass urine, did you have to hurry to get to the toilet?	1	2	3	4
35. Have you passed urine frequently?	1	2	3	4
36. Have you had leaking of urine?	1	2	3	4
37. Have you had pain or a burning feeling when passing urine?	1	2	3	4
38. When you felt the urge to move your bowels, did you have to hurry to get to the toilet?	1	2	3	4
39. Have you had any leakage of stools?	1	2	3	4
40. Have you been troubled by passing wind?	1	2	3	4
41. Have you had cramps in your abdomen?	1	2	3	4
42. Have you had a bloated feeling in your abdomen?	1	2	3	4
43. Have you had tingling or numbness in your hands or feet?	1	2	3	4
44. Have you had aches or pains in your muscles or joints?	1	2	3	4
45. Have you lost hair?	1	2	3	4
46. Has food and drink tasted differently from usual?	1	2	3	4

Please go on to the next page



ENGLISH

During the past week:	Not at all	A little	Quite a bit	Very much
47. Have you felt physically less attractive as a result of your disease or treatment?	1	2	3	4
48. Have you felt less feminine as a result of your disease or treatment?	1	2	3	4

During the past 4 weeks:	Not at all	A little	Quite a bit	Very much
49. To what extent were you interested in sex?	1	2	3	4
50. To what extent were you sexually active?	1	2	3	4
Answer these questions only if you have been sexually active during the past 4 weeks:				
51. Has your vagina felt dry during sexual activity?	1	2	3	4
52. Has your vagina felt short and / or tight?	1	2	3	4
53. Have you had pain during sexual intercourse or other sexual activity?	1	2	3	4
54. Was sexual activity enjoyable for you?	1	2	3	4

19. APPENDIX: SCREENSHOT OF HADS QUESTIONNAIRE

Table: Image of HADS questionnaire

Hospital Anxiety and Depression Scale (HADS)

Tick the box beside the reply that is closest to how you have been feeling in the past week.
Don't take too long over you replies: your immediate is best.

D	A		D	A	
		I feel tense or 'wound up':			**I feel as if I am slowed down:**
	3	Most of the time	3		Nearly all the time
	2	A lot of the time	2		Very often
	1	From time to time, occasionally	1		Sometimes
	0	Not at all	0		Not at all
		I still enjoy the things I used to enjoy:			**I get a sort of frightened feeling like 'butterflies' in the stomach:**
0		Definitely as much		0	Not at all
1		Not quite so much		1	Occasionally
2		Only a little		2	Quite Often
3		Hardly at all		3	Very Often
		I get a sort of frightened feeling as if something awful is about to happen:			**I have lost interest in my appearance:**
	3	Very definitely and quite badly	3		Definitely
	2	Yes, but not too badly	2		I don't take as much care as I should
	1	A little, but it doesn't worry me	1		I may not take quite as much care
	0	Not at all	0		I take just as much care as ever
		I can laugh and see the funny side of things:			**I feel restless as I have to be on the move:**
0		As much as I always could		3	Very much indeed
1		Not quite so much now		2	Quite a lot
2		Definitely not so much now		1	Not very much
3		Not at all		0	Not at all
		Worrying thoughts go through my mind:			**I look forward with enjoyment to things:**
	3	A great deal of the time	0		As much as I ever did
	2	A lot of the time	1		Rather less than I used to
	1	From time to time, but not too often	2		Definitely less than I used to
	0	Only occasionally	3		Hardly at all
		I feel cheerful:			**I get sudden feelings of panic:**
3		Not at all		3	Very often indeed
2		Not often		2	Quite often
1		Sometimes		1	Not very often
0		Most of the time		0	Not at all
		I can sit at ease and feel relaxed:			**I can enjoy a good book or radio or TV program:**
	0	Definitely	0		Often
	1	Usually	1		Sometimes
	2	Not Often	2		Not often
	3	Not at all	3		Very seldom

Please check you have answered all the questions

Scoring:
Total score: Depression (D) _____ Anxiety (A) _____
0-7 = Normal
8-10 = Borderline abnormal (borderline case)
11-21 = Abnormal (case)

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5700594/

20. APPENDIX: LINEAR MIXED MODEL THEORY

20.1 Linear mixed models: links

See https://stats.oarc.ucla.edu/other/mult-pkg/introduction-to-linear-mixed-models/

20.2 Linear mixed models: the theory

The formula is:

$$y=xb+zu+e$$

where

Symbol	Description	Type
ny	number of values of the outcome variable	Scalar
np	number of predictor variables	Scalar
nj	number of groups	Scalar
nq	number of random intercepts per group	Scalar
y	outcome variable	ny x 1 vector
x	predictor variables aka fixed-effects variables,	ny x np matrix
b	fixed-effects regression coefficients (parameters)	np x 1 vector
z	random effects variables	ny x nj matrix
u	random-effects regression coefficients	nj x 1 vector
e	residuals (error)	ny x 1 vector

20.3 Linear mixed models: worked example

Consider this scenario:

> *Arm (J=2) indexed by the j subscript each see nj patients. So our grouping variable is the arm. Each arm sees 36 patients. The total number of patients is the sum of the patients seen by each arm. In our example, N=72, so 72 patients were seen by two arms. Our outcome is a continuous variable, output weight. Furthermore we have 3 fixed effects predictors, baseline value, timepoint, and stratification variable, plus a fixed intercept and one random intercept (q=1) for each of the J=2 arms. For simplicity, we are only going to consider random intercepts. We will let every other effect be fixed for now. The reason we want any random effects is because we expect that output weights within arms may be correlated.*

So the formula is

$$y=xb+zu+e$$

That gives us this table below:

Symbol	Description	Type
ny	number of values of the outcome variable	72
np	number of predictor variables	4
nj	number of groups	2
nq	number of random intercepts per group	1
y	outcome variable	72 x 1 vector
x	Four predictor variables aka fixed-effects variables, • Fixed intercept • baseline value, • timepoint, • stratification variable	72 x 4 matrix
b	fixed-effects regression coefficients (parameters)	4 x 1 vector
z	One random effects variable • Random intercept	72 x (1x2) matrix
u	random-effects regression coefficients	2 x 1 vector
e	residuals (error)	72 x 1 vector

21. APPENDIX: USE OF THE WORD "LINEAR" IN LINEAR MODELS

The word "linear" in "linear models" refers to the parameters, not the variables: it can cope with quadratic equations

- https://www.quora.com/Why-is-linear-regression-called-linear-if-we-can-use-quadratic-equations-in-our-model-which-gives-a-curved-line
- https://www.quora.com/Why-is-it-still-called-linear-regression-even-if-we-choose-a-polynomial-of-a-higher-degree-for-the-hypothesis

22. APPENDIX: LOOKUP TABLE

The type of a variable can be used to deduce the variable distribution - for example a variable measuring strength may have a Gamma distribution and a frequency variable may have a Poisson. For the avoidance of doubt we shall graph it as well.

When the variable type and distribution is known, the lookup tables below can be used to deduce which model is needed and which code can be used to do that. Stata code is shown for convenience, although SAS or R code would also suffice. See the Stata help documentation for more details

22.1 GLMs

#	Variable Type	Distribution	Link Fn	Example Stata Code†††
1	Continuous (Normal)	Gaussian	Identity	glm depvar indepvars, family(gaussian) link(identity)
2	Continuous (SPC) †	Inverse Gaussian	Power -2	glm depvar indepvars, family(igaussian) link(power -2)
3	Binary	Binomial (multiple) or Bernoulli (single)	Logit	glm depvar indepvars, family(binomial) link(logit)
4	Count/Rate (Poisson)	Poisson	Log	glm depvar indepvars, family(poisson) link(log)
5	Count/Rate (NB)	Negative binomial	Log	glm depvar indepvars, family(nbinomial) link(log)
6	Continuous (SPC) †	Gamma	Power -1	glm depvar indepvars, family(gamma) link(power -1)
6	Continuous (Time)	Gamma?	Power -1?	glm depvar indepvars, family(gamma) link(power -1)
99	Time-to-event	Exponential††	Log	glm depvar indepvars, family(exponential) link(log)

† SPC = Skewed positive continuous variable.
†† Not available in Stata. Code shown is synthetic. GLMs in this case would require another language
††† For a full list of families, link functions in Stata and their valid combinations, please see Stata help. Note that "inverse" is the same as "power -1", "inverse squared" is the same as "power -2", and so on

22.2 Non-GLMs if GLMs cannot cope or are inferior

Variable Type	Distribution	Link Function	Example Stata Code
Binary (Bernoulli)	Bernoulli	Logit (log-odds)	logit depvar indepvars, family(binomial) link(logit)
Binary (Probit)	Probit	Probit (CDF of standard normal)	probit depvar indepvars, family(binomial) link(probit)
Categorical (Nominal)	Multinomial	Logit	mlogit depvar indepvars, basecategory(1)
Categorical (Ordinal)	Categorical	Probit	ologit depvar indepvars
Categorical (Ordinal)	Cumulative logit	Logit (log-odds)	ologit depvar indepvars, family(ordinal) link(logit)
Categorical (Ordinal)	Proportional odds	Logit (log-odds)	ologit depvar indepvars, family(ordinal) link(logit)
Continuous (Normal)	Normal	Identity (linear)	reg depvar indepvars, family(gaussian) link(identity)
Count/Rate (NB)	Negative binomial	Log	nbreg depvar indepvars, family(nbinomial) link(log)
Count/Rate (Poisson)	Poisson	Log	poisson depvar indepvars, family(poisson) link(log)

22.3 Time-to-event models

Variable Type	Details	Example Stata Code
Time-to-event	Cox Proportional Hazards Model	stset depvar failure(eventvar==1)† stcox indepvars
Time-to-event	Parametric survival models, named distribution	stset depvar failure(eventvar==1)† streg indepvars, distribution(named distn) ††

† depvar would be the time-to-event variable, eventvar=0 if censored, 1 if not
†† named distributions currently supported by Stata are exponential, Weibull, Gompertz, lognormal, loglogistic, and ggmma (short for generalized gamma). "Streg" can handle frailty and accelerated-failure-time models as well as proportional hazards models. See Stata help on "streg" for details

22.4 Green, JA (2021)

The paper Green, JA (2021) (see[5] for more details) also gave a guide for which variable types would need which families. This guide can be used in tandem with the tables above

Variable Type	Family
Count variables	Poisson
Count variables	Negative Binomial
Count variables with excess zeros	Zero-Inflated Poisson
Count variables with excess zeros	Zero-Inflated Negative Binomial
Count variables with excess zeros	Hurdle Models
Overdispersed count variables	Quasi-Poisson
Overdispersed count variables	Quasi-Negative Binomial
Overdispersed count variables	Quasi-Binomial
Continuous skewed variables	Gamma

23. APPENDIX: CATEGORICAL, BINARY, CONTINUOUS

Consider the following synthetic example designed for illustration purposes

Table: synthetic example

ID	Weight	EORTC QLQ-EN24 Category	Not at all	A bit	Quite a bit	Y/N
yyyyyyy1	4.4	Lymphoedema	1	0	0	1
yyyyyyy2	4.4	Lymphoedema	1	0	0	0
yyyyyyy3	4.2	Pain in back and pelvis	0	1	0	0
yyyyyyy4	2.1	Urological symptoms	0	0	1	1
yyyyyyy5	4.2	Urological symptoms	0	0	1	1

23.1 Continuous

This can be considered as a continuous variable that can be modelled with one model

Continuous variable

Continuous variable	Mean Weight
Weight	3.86

23.2 Categorical and binary

Or as a categorical variable with a count/rate that can be modelled with three models

Categorical and binary: a variable with three categories, each with a count/rate

Category	Count/rate
Lymphoedema	2/5
Pain in back and pelvis	1/5
Urological symptoms	2/5

23.3 Categorical and continuous

Or as a categorical variable with three categories, each holding a continuous variable. This can be modelled with one model

Categorical and continuous: a variable with three categories, each with a continuous variable

Category	Mean Weight
Lymphoedema	4.4
Pain in back and pelvis	4.2
Urological symptoms	3.15

23.4 Binary

Or as a binary variable that can be modelled with one model

Binary variable

Binary variable	Count/rate
Y/N	3/5

24. APPENDIX: ABSOLUTE VS RELATIVE

The results of the study are depicted as absolute and relative values, although not all variables will be depicted as both. As for the meaning of "absolute" and "relative", we take our lead from the protocol. It says that "*Continuous variables will be summarised using means, standard deviations, and 95% confidence intervals. Medians with interquartile ranges will be presented where appropriate. Categorical variables will be summarised using counts and percentages.*"

For examples of the absolute vs relative distinction for effect size, please see "Appendix: Worked Examples Of Morbidity"

24.1 Continuous

As per protocol this is a mean and standard deviation (SD). Our synthetic example looks like this.

Continuous variable

Continuous variable	Mean Weight (SD)
Weight	3.86 (1.64)

24.2 Categorical and binary

As per protocol, this is a count and percentage. Our synthetic example looks like this.

Categorical and binary: a variable with three categories, each with a count/rate

Category	Count/rate
Lymphoedema	2 (40%)
Pain in back and pelvis	1 (10%)
Urological symptoms	2 (40%)
Total	5 (100%)

24.3 Categorical and continuous

As per protocol this is a mean and standard deviation (SD). Our synthetic example looks like this.

Categorical and continuous: a variable with three categories, each with a continuous variable

Category	Mean Weight (SD)
Lymphoedema	4.4 (1.6)
Pain in back and pelvis	4.2 (0)
Urological symptoms	3.15 (1.1)

24.4 Binary

As per protocol this is a mean and standard deviation (SD). Our synthetic example looks like this.

Binary variable

Binary variable	Count/rate
Y/N	3 (60%)
Total	5 (100%)

24.5 Discrete

As per protocol, this is a median and interquartile range. We do not have a synthetic example for this.

25. APPENDIX: GENERALIZED LINEAR MODEL THEORY

In statistics, a generalized linear model (GLM) is a flexible generalization of ordinary linear regression (LR).

A ordinary linear regression connects the input variables to the dependent variable via one function. But a GLM uses two functions: the variance function and the link function. This allows the linear model to be related to the response variable via a link function and allows the magnitude of the variance of each measurement to be a function of its predicted value.

Table: distributions and link functions for a GLM

Distribution	Variance Function	Typical Link Function
Normal	$V(\mu) = 1$	Identity: $g(\mu) = \mu$
Binomial	$V(\mu) = \mu(1-\mu)$	Logit: $g(\mu) = \log(\mu/(1-\mu))$
Poisson	$V(\mu) = \mu$	Log: $g(\mu) = \log(\mu)$
Gamma	$V(\mu) = \mu^2$	Inverse: $g(\mu) = 1/\mu$ or Log: $g(\mu) = \log(\mu)$
Inverse Gaussian	$V(\mu) = \mu^3$	Inverse squared: $g(\mu) = 1/\mu^2$
Negative Binomial	$V(\mu) = \mu + \alpha\mu^2$	Log: $g(\mu) = \log(\mu)$

26. APPENDIX: WORKED EXAMPLES OF MORBIDITY

The display of morbidity data can be difficult because problems arise when the morbidity has subdivisions or the data is displayed by patient not complications. To make things easier we provide worked examples below.

Firstly, let us build a synthetic example. This is an example designed for illustration purposes and it looks like this:

Synthetic example part 1: Control arm. three patients, four complications

Patient ID	Complication morbidity grade	Date of complication	Date of discharge	Date difference
xxxxxxx1	I	01Aug2024	15Aug2024	-14
xxxxxxx1	I	01Aug2024	15Aug2024	-14
xxxxxxx2	IIIa	01Aug2024	15Jul2024	17
xxxxxxx3	IV	01Aug2024	15Jun2024	47

Synthetic example part 2: Intervention arm. three patients, four complications

Patient ID	Complication morbidity grade	Date of complication	Date of discharge	Date difference
Xxxxxxx4	I	01Aug2024	15Jul2024	17
Xxxxxxx4	II	01Aug2024	15Jul2024	17
Xxxxxxx5	IIIa	01Aug2024	15Jul2024	17
Xxxxxxx6	IIIb	01Aug2024	15Jun2024	47

26.1 Worked example of "Any morbidity"

The synthetic example gives us this table. It is a list of patients, not complications

#	Outcome Measures	Control	Intervention	Delta
2a	i. Any morbidity			
	At or prior to discharge	1 (33%)	0 (0%)	-1 (-33%)
	Between discharge and 30 days	1 (33%)	2 (66%)	+1 (+33%)
	Other	1 (33%)	1 (33%)	-
	Total	3 (100%)	3 (100%)	-
	At or prior to discharge OR between discharge and 30 days	2 (66%)	2 (66%)	-

26.2 Worked example of "Morbidity by grade (I, II, IIIa, IIIb, IVa, IVb)"

The synthetic example gives us this table. It is a list of patients, not complications

#	Outcome Measures	Control	Intervention	Delta
2b	ii. Morbidity by grade (I, II, IIIa, IIIb, IVa, IVb)			
	At least one Grade I	1 (33%)	1 (33%)	-
	At least one Grade II	0 (0%)	1 (33%)	+1 (+33%)
	At least one Grade III	1 (33%)	2 (66%)	+1 (+33%)
	At least one Grade IV	1 (33%)	0 (0%)	-
	Other	0 (0%)	0 (0%)	-
	Total	3 (100%)	3 (100%)	-

Note: because a patient can have more than one complication, this example does not add to 100%

26.3 Worked example of "Morbidity by highest grade (I, II, IIIa, IIIb, IVa, IVb)"

The synthetic example gives us this table. It is a list of patients, not complications

#	Outcome Measures	Control	Intervention	Delta
2d	iv Morbidity by highest grade			
	Number of patients whose highest grade is Grade I	1 (33%)	0 (0%)	-1 (-33%)
	Number of patients whose highest grade is Grade II	0 (0%)	1 (33%)	+1 (+33%)
	Number of patients whose highest grade is Grade III	1 (33%)	2 (66%)	+1 (+33%)
	Number of patients whose highest grade is Grade IV	1 (33%)	0 (0%)	-
	Other	0 (0%)	0 (0%)	-
	Total	3 (100%)	3 (100%)	-

Note: because a patient can only have one highest grade, this example adds to 100%