Long-term effectiveness and cost-effectiveness of text message and endowment incentives for weight management in men with obesity: A follow-up of the Game of Stones RCT

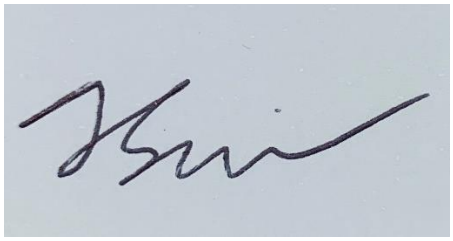
Statistical analysis plan v1.2

19-06-2024

Based on the Protocol Version v7.0 01.11.23

ISRCTN91974895

Signature Page

Trial Statistician: 

(James Swingler)

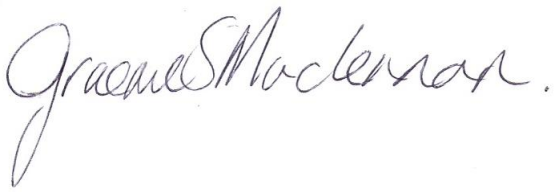
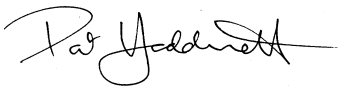


CHaRT Director and Senior Statistician:

(Graeme MacLennan)

Chief Investigator: _____  **_____**

(Pat Hoddinott)

Amendment History

SAP version	Protocol version	Section number changed	Description of and reason for change	Statistician	Date changed
1.0	V7.0	NA	NA – draft 1.0 signed	JS GMAC	23/05/2024
1.1	V7.0	All	Edits to wording and formatting of dummy tables	JS GMAC	06/06/2024
1.2	V7.0	All	Further edits to wording and formatting of dummy tables	JS GMAC	19/06/2024

JS – James Swingler; GMAC – Graeme MacLennan

Contents

Introduction

Obesity increases the risk of type 2 diabetes, heart disease, stroke, mobility problems and some cancers, and its prevalence is rising. Men engage less than women in existing weight loss interventions. A more detailed description of Game of Stones can be found in the protocol.

Game of Stones is a pragmatic RCT; the primary outcome evaluation at 12 months (M) is published.[1] This SAP focuses on the 24M analysis of the trial.

Objectives

Primary objective

To compare the difference in % weight change at 24M from baseline for men with obesity in the following groups: i) Text messages with endowment incentive (Texts + Incentive) vs 12M waiting list for texts delivered for 3M (control); ii) Text messages alone (Texts only) vs control. These groups will hence be referred to as Texts + Incentive, Texts alone and Control.

Secondary objectives

- To assess differences between groups in secondary outcomes
- To assess the cost-effectiveness of Texts + Incentive vs Control and Texts alone vs Control
- To understand men's and service providers' experiences of the intervention
- To compare PHQ-4, Quality of Life (EQ-5D-5L), EQ-5D-5L Anxiety and Depression (AD) dimension measures for the 3 trial groups at 24M

Study methods

Trial design

Game of Stones is a pragmatic, multi-centre, parallel, 3-arm, assessor blind, randomised controlled trial (RCT) comparing weight change at 12M and at 24M for: i) Texts + Incentive vs Control; and ii) Texts alone vs Control.

Sample size

The sample size was fixed by 12M follow-up design stage. There is no update to this calculation and this section is included for information.

The sample size calculation was estimated to give adequate power for the target difference at 12M. The 12M sample size calculations assumptions are included as follows. We will require to follow-up 146 men in each group to detect differences in weight loss between groups of at least 3% at 12M, with 90% power and two-sided alpha equal to 2.5% (to maintain a nominal significance level of 5% with two tests being used, see Framework section below for specification of comparisons). With an expected 25% loss to follow-up as observed in the Feasibility Study at one year, a total of 585 men will need to be randomised: 195 per trial group and 195 (65 per trial group) at each of the three centres. The sample size calculation is based on detecting a mean difference in weight between intervention groups and control of at least 3.3kg, assuming a pooled standard deviation of 8kg. A minimum clinically important weight loss of 3% is recommended by NICE,[2] and the mean difference of 3.3kg is derived from 3% of 109kg (the mean baseline weight in the Feasibility Study).[3] Several trials of SMS-delivered weight management interventions [4] reported an effect size >3.3kg, including the largest study, which was the only trial to include predominantly men, suggesting 3.3kg is an achievable mean weight loss. The standard deviation for absolute weight loss ranged from 4.9kg to 6.3kg in the Feasibility Study (at 12M) and from 2.5kg to 7.3kg in the systematic review.[4] We therefore conservatively assume a standard deviation of 8kg.

Framework

Primary and secondary outcomes will be compared using a superiority framework for the following intervention groups:

1) Texts + Incentive vs Control
2) Texts alone vs Control

Timing of final analysis

There are no planned interim analyses of any outcomes, there is no Data Monitoring Committee. The trial will not be stopped for futility reasons or safety reasons before the collection analysis of 24M outcome data. A single final analysis will be performed, and results shared with co-investigators, after 24M assessment data for the last man has been collected, all data will have been entered and the database is scheduled to be locked on 1 July 2024.

Statistical principles

Confidence intervals and P-values

Primary comparisons will be reported with 97.5% confidence intervals to reflect the two-sided 2.5% alpha level. This is a more stringent level than the usual 95% confidence intervals to adjust for multiple testing due to the multi-arm nature of the trial.

Blinding to allocation

The trial statistician will be not blinded to allocation during the analysis since unblinding occurred during the 12M analysis. Outcome assessors will aim to be blind to allocation when the participant is weighed at 24M, however it may not be practical if they met the participant earlier in the trial. All other investigators will remain blind until after the database has been locked.

Reporting conventions

P-values ≥0.001 will be reported to 3 decimal places; p-values less than 0.001 will be reported as "<0.001". The mean, standard deviation, and any other statistics other than quantiles, will be reported to one decimal place greater than the original data. Quantiles, such as median, or minimum and maximum will use the same number of decimal places as the original data. Estimated parameters, not on the same scale as raw observations (e.g. regression coefficients) will be reported to 3 significant figures.

Adherence

Analysis to account for non-compliance is not necessary, as automated interventions can only be accessed via randomisation, therefore cross-over cannot occur and contamination was minimal in the Feasibility Study.

Protocol deviations

A detailed protocol about weight assessment in Game of Stones participants is provided in the Protocol. In summary, a face-to-face verified assessment should happen; if not possible, every effort will be made to gain a satisfactory verified weight via video assessment, or via an independently confirmed weight by third party. These three approaches of collecting will be considered as following protocol (see Appendix 'SOP 7.5: Remote weight' for more details).

To achieve the gold-standard, the following conditions must be met:

1. Game of Stones research scales for measuring weight
2. Weight measured within a window of +/-23 days from the follow-up due date

If weight is provided but any of the gold-standard criteria not met, then we will use the data to inform secondary analyses.

Analysis population

The analysis population for the 24-month main analysis of Game Stones will be based on an intention-to-treat framework (ITT), given the pragmatic nature of the trial. All randomised participants will be analysed according to the treatment group they were allocated to, and multiple imputation will be used to deal with participants that do not have observed primary outcome. We will include all observed primary outcome data in the ITT analysis. Sensitivity analysis will be done using more stringent requirements for the measurement of the primary outcome, see Appendix: Standard Operating Procedure (SOP) 7.5.

Secondary analysis will include a "per protocol" population (PP) where only the observed data from participants that followed the gold standard will be included. Missing data and weight measurement data that does not meet the gold standard in the **Protocol deviations** above will set to missing and multiple imputation used to analyse imputed weights. For more information pertaining to the gold standard, and weigh assessment protocols, refer to the Appendix (SOP 7.5).

Post-randomisation exclusions

There will be no post-randomisation exclusion of participants in this trial. Any participants randomised then found ineligible at the time of randomisation will be included in all analyses.

Data sources

Data are being collected from the following sources:

- Case Report Form (screening and adverse events)
- Participant questionnaires
- SMS (text) activity data (the numbers of SMS sent, requests to stop SMS and responses sent)
- Activity data from interactive participant web pages
- Qualitative interviews with participants conducted after 24M assessment.

Trial population

Screening of potential participants for eligibility (see Protocol), recruitment processes (which has been reported),[1] the level, and timing of withdrawal with reasons where provided will be presented as descriptive statistics and, where applicable, as part of the CONSORT diagram.

Based on differences in baseline characteristics observed in the Feasibility Study[5] and the 12M findings of this trial[1] we will present descriptive statistics for baseline characteristics of Game of Stones participants by whether participants were approached to take part via GP practice staff (e.g. letter, text, given a flier) or via community i.e. did not first hear about the trial through GP practice staff.

Flow of participants diagram

We will follow the CONSORT guidance for reporting multi-arm trials[6] to report the participant diagram.

Baseline participant characteristics

Demographics and measures collected at baseline are summarised in **Table S1** and are based on the Feasibility study and the 12M findings from this trial,[3,1] FFIT trial[7] socio-economic measures relevant to obesity[8-11] and the recently published STAR-LITE core outcome set for behavioural weight management interventions.[12]

Baseline variable definitions

As already reported for 12M data[1], for socio-demographic variables, we will adhere to the harmonised standards guidance published by the Office for National Statistics,[13] except for employment status where harmonised guidance from the Scottish Government will be used.[14] Perceived wealth[8] and financial strain[10] variables are included to capture additional health inequalities data. Men under-report mental health conditions.[15] A variable called Possible Latent Mental Health Condition is defined as men who do not self-report a mental health condition but have scores on at least one of the following that suggest that participants may have undetected mental health conditions or be at risk of developing one in future:

- PHQ-4 score 3+
- EQ-5D-5L-AD
- WEMWBS
- WSSQ

There is some survey evidence that gains in levels of mental health/wellbeing predict declines in the incidence of mental illness and losses of mental health/wellbeing level predict increases in the incidence of mental illness.[16]

The presence of Multiple Long-Term Conditions (MLTC) is based on existing definitions[17] (i.e., comorbidities, as presented in Table 1). Presence of MLTC will be defined as the co-existence (self-report) of two or more chronic non-communicable disease conditions (comorbidities) where obesity is a recognised risk factor - stroke including mini-stroke, high blood pressure, a heart condition, diabetes, cancer, arthritis, or a mental health condition. Presence of MLTC which includes the presence of self-reported diabetes as one of the comorbidities will also be reported. Self-reported disability (ONS standardised definition) will be reported separately.

Analysis

Outcome measures

Outcomes will be measured at 3 and 6M (two active intervention groups), 12 and 24M (all three groups), unless otherwise indicated. Outcomes are also measured at baseline, where applicable.

Primary outcome

The 24M primary outcome is within-participant change from baseline weight expressed as a percentage of baseline weight at 24M from baseline.

Secondary outcomes at 24 months

- absolute weight change from baseline (kg)
- number of participants achieving any weight loss i.e., >0% (binary yes/no)
- number of participants achieving ≥5% (binary yes/no)
- number of participants achieving ≥ 10% (binary yes/no)
- weight loss categorised (NICE CG189 recommends aiming for 5-10% weight loss, particularly when comorbidities are present, these are the targets set for the Texts alone and Texts + Incentive intervention groups): gaining weight, achieving 0 to <5% weight loss, ≥5 to <10% weight loss & ≥10% weight loss (ordered categories)
- EQ-5D-5L
- EQ-5D-5L-AD
- PHQ-4

Exploratory outcomes/endpoints at 24 months

- weight management strategies used
- satisfaction with weight loss progress
- recommendation of Game of Stones to others

Health economic outcomes

NHS costs, QALYs, incremental cost-per QALY gained and incremental cost per % weight loss over trial follow-up and modelled lifetime. Analysis of these outcomes is covered in the Health Economic Analysis Plan.

Data Collection Time Points

Timing of data collection is summarised in Table 1.

Table 1 – Data collection time points

Data Collection	**Baseline**	**12M**	**24M***
Socio-demographic: deprivation category, comorbidities (physical and mental health), possible latent mental health condition, self-reported disability, ethnicity, age, education, employment, household size, relationship status.	✓		
Perceived wealth, financial strain	✓	✓	✓
Anthropometry - height (for BMI)	✓		
Anthropometry – weight	✓	✓	✓
Participant satisfaction		✓	✓
Participant recommendation of GOS			✓
Weight management strategies used over last 12 months	✓	✓	✓
Patient Health Questionnaire -4	✓	✓	✓
EQ-5D-5L – Anxiety and Depression Dimension (AD)	✓	✓	✓
Health Economic Outcomes: EQ-5D, NHS health care use	✓	✓	✓
Qualitative interview data (experiences, behaviours)		✓	✓

Unintended consequences or adverse events		✓	✓

Scoring of outcomes measured in questionnaires

All questionnaire items will be analysed individually, as set out in the dummy tables (Appendix) and in the 12M results.[1] For mental health, we will select appropriate cut-offs according to the literature (for example, the Patient Health Questionnaire (PHQ-4)[18] will have a cut off scores of 3+ for the GAD2 and PHQ2 scale scores, respectively).

General statistical methods

All continuous variables will be summarised and tabulated using the following descriptive statistics: N (number of valid non-missing responses), number of missing records, mean, standard deviation (SD). Likert-scale variables will be treated as continuous measures. The frequency and percentages (based on the non-missing sample size) of observed levels will be reported for all categorical measures. In general, all data will be listed, sorted by subject and treatment and where appropriate by visit number within subject. The number of missing observations and corresponding completion rates will be reported so that the feasibility of each outcome measure can be assessed.

Primary outcome

In this the 24M analysis plan the primary outcome is considered the change from baseline weight expressed as a percentage at 24M from baseline.

The analysis of the primary outcome will estimate the mean difference in change from baseline weight expressed as a percentage at 24M from baseline between groups (Texts + Incentive vs Control; and Texts alone vs Control), using a linear regression model adjusting recruitment centre and method of recruitment route (GP or community) as a fixed effect. Secondary outcome measures will be analysed similarly, using an appropriate generalised linear model, including binary logit regression for dichotomous outcomes (e.g. number of participants achieving any weight loss) and ordered logit for ordinal outcomes (e.g. categorised weight loss). Statistical significance will be at the 2.5% level, consistent with the assumptions made in the sample size calculation.

Secondary outcomes and exploratory outcomes

All secondary outcome measures will be analysed similarly, using an appropriate generalised linear model, including binary logit regression for dichotomous outcomes (e.g. smoking status) and ordered logit for ordinal outcomes (e.g. alcohol frequency) and presenting two comparisons: Texts + Incentive vs Control; Texts alone vs control.

Table 2 – Analysis strategy for secondary outcomes and exploratory outcomes

Secondary outcome	**Analysis strategy**
Continuous outcome absolute weight change from baseline (kg)	Linear regression model adjusting for recruitment centre and method of recruitment route as fixed effects, and baseline weight.
Binary outcomes number of participants achieving >0%, ≥5%, and ≥10% weight loss	Logistic regression model adjusting for recruitment centre and method of recruitment route as fixed effects

Ordered categories: Achieving 0<5% weight loss, ≥5<10% weight loss & ≥10% weight loss	Ordinal logistic regression adjusting for recruitment centre and method of recruitment route as fixed effects
Continuous outcomes EQ-5D-5L, EQ-5D-5L-AD, PHQ-4	Linear regression adjusting for recruitment centre and method of recruitment route as fixed effects and baseline score.
Exploratory outcomes (not listed individually)	GLM suitable for the outcome distribution, adjusting for recruitment centre and method of recruitment route as fixed effects

Process outcomes

Descriptive summaries will be reported and integrated with the qualitative data analysis, but no formal statistical analysis of these data will be undertaken. More information about the process outcomes is available in the Process Evaluation Analysis Plan.

Pre-planned subgroup analysis

Pre-planned subgroup analysis was conducted for the 12M analysis of the trial. No subgroup analysis had been planned for the 24M analysis. A lack of evidence in the literature in conjunction with the quantitative results observed at 12M, give no reason to carry out an analysis of any subgroup. Any analysis of subgroups will be conducted post hoc if required.

Missing or spurious data

For all baseline characteristics and outcome measures, we will report the level of missing data. When estimating treatment effects, if a relevant baseline score is missing, we will implement best practice by imputing the score (if continuous) or adding an extra category for missing (if categorical). The primary analysis will use multiple imputation of missing outcome data, applying predictive mean matching.[19]

Sensitivity analyses of the primary outcome will examine the data under various assumptions around missingness, including:

- an analysis of all observed cases,
- per protocol analyses (see Table S3 for more details),
- missing weight data being treated as Baseline Observation Carried Forward (BOCF) and Last Observation Carried Forward (LOCF)) (intervention arms only), as recommended in the STAR-LITE core outcome set,[12] for comparability with previous weight loss studies,[20]
- an analysis of all observed cases excluding those individuals who indicated that they took weight loss pills, injections, or meal replacements in the past 24M and a second analysis applying predictive mean matching to impute these values as missing.

Data quality assurance and source data verification will be carried out in accordance with the CTU's standard operating procedures to minimise spurious data. Further data quality checks will be carried out by the trial statistician prior to the analysis and potentially implausible data will be queried with trial office and/or site staff.

Harms

Each initial adverse event (AE) will be considered for severity, causality or expectedness. A serious adverse event (SAE) is any AE that:

- Results in death
- Is life-threatening
- Requires hospitalisation, or prolongation of existing inpatients' hospitalisation
- Results in persistent or significant disability or incapacity is otherwise considered medically significant by the investigator.

Please see the Study Protocol for more details on AEs. The number of AEs and SAEs and the proportion of participants with an event will be presented at each assessment time point. These will be tabulated and summarised by allocated group.

Statistical software

The latest version of Stata available at the time of the analysis will be used.

References

1. Hoddinott P, O'Dolan C, Macaulay L, et al. Text Messages With Financial Incentives for Men With Obesity: A Randomized Clinical Trial. *JAMA.* 2024

2. National Institute for Health and Care Excellence, (NICE). *Obesity: identification, assessment and management.* No. 189 ed. ; 2022

3. Dombrowski SU, McDonald M, van der Pol M, et al. Game of Stones: feasibility randomised controlled trial of how to engage men with obesity in text message and incentive interventions for weight loss. *BMJ Open.* 2020;10(2):e032653. doi:http://dx.doi.org/10.1136/bmjopen-2019-032653

4. Skinner R, Gonet V, Currie S, Hoddinott P, Dombrowski SU. A systematic review with meta-analyses of text message-delivered behaviour change interventions for weight loss and weight loss maintenance. *Obes Rev.* 2020;21(6):e12999. doi:10.1111/obr.12999

5. McDonald MD, Dombrowski SU, Skinner R, et al. Recruiting men from across the socioeconomic spectrum via GP registers and community outreach to a weight management feasibility randomised controlled trial. *BMC Med Res Methodol.* 2020;20(249):249-2. doi:https://doi.org/10.1186/s12874-020-01136-2

6. Juszczak E, Altman DG, Hopewell S, Schulz K. Reporting of Multi-Arm Parallel-Group Randomized Trials: Extension of the CONSORT 2010 Statement. *JAMA.* 2019;321(16):1610-1620. doi:10.1001/jama.2019.3087

7. Hunt K, Wyke S, Gray CM, et al. A gender-sensitised weight loss and healthy living programme for overweight and obese men delivered by Scottish Premier League football clubs (FFIT): a pragmatic randomised controlled trial. *Lancet.* 2014;383(9924):1211-1221. doi:10.1016/S0140-6736(13)62420-4

8. Best M, Papies EK. Lower socioeconomic status is associated with higher intended consumption from oversized portions of unhealthy food. *Appetite.* 2019;140:255-268. doi:10.1016/j.appet.2019.05.009

9. Bridges S, Disney R. Debt and depression. *Journal of health economics.* 2010;29(3):388-403. doi:10.1016/j.jhealeco.2010.02.003

10. French D. Financial strain in the United Kingdom. *Oxford Economic Papers.* 2018;70:163-182. doi:10.1093/oep/gpx030

11. Lillis J, Luoma JB, Levin ME, Hayes SC. Measuring weight self-stigma: the weight self-stigma questionnaire. *Obesity (Silver Spring).* 2010;18(5):971-976. doi:10.1038/oby.2009.353

12. Mackenzie RM, Ells LJ, Simpson SA, Logue J. Core outcome set for behavioural weight management interventions for adults with overweight and obesity: Standardised reporting of lifestyle weight management interventions to aid evaluation (STAR-LITE). *Obes Rev.* 2020;21(2):e12961. doi:10.1111/obr.12961

13. Office for National Statistics. Harmonised standards and guidance. Government Statistical Service. https://analysisfunction.civilservice.gov.uk/government-statistical-service-and-statistician-group/gss-support/gss-harmonisation-support/. Updated 2020. Accessed 8 March, 2021

14. Scottish Government. Scottish Surveys: Core and Harmonised Questions. https://www.gov.scot/publications/scottish-surveys-core-and-harmonised-questions/. Updated 2017. Accessed 8 March, 2021

15. Baker C. Mental health statistics for England: prevalence, services and funding. 2020. https://commonslibrary.parliament.uk/research-briefings/sn06988/

16. Keyes CLM, Dhingra SS, Simoes EJ. Change in level of positive mental health as a predictor of future risk of mental illness. *Am.J.Public Health.* 2010;100(12):2366-2371. doi:10.2105/AJPH.2010.192245

17. The Academy of Medical Sciences. Multimorbidity: a priority for global health research. *The Academy of Medical Sciences.* 2018

18. Löwe B, Wahl I, Rose M, et al. A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. *J.Affect.Disord.* 2010;122(1-2):86-95. doi:10.1016/j.jad.2009.06.019

19. Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. 2014;14(1):75. doi:10.1186/1471-2288-14-75

20. Avenell A, Robertson C, Skea Z, et al. Bariatric surgery, lifestyle interventions and orlistat for severe obesity: the REBALANCE mixed-methods systematic review and economic evaluation. *Health Technol.Assess.* 2018;22(68):1-246. doi:10.3310/hta22680

Supplementary information: Dummy tables

Table S1: Participant baseline characteristics[1]

	Texts + Incentive (N=196)	**Texts alone (N=194)**	**Control (N=195)**
Age [a] **(yrs) - mean (SD); n**			
≥18-<25			
≥25-<45			
≥45-<65			
≥65-<75			
≥75			
Deprivation Category - n (%)			
Most deprived			
More deprived			
Deprived			
Less deprived			
Least deprived			
Ethnic Group [a] **- n (%)**			
Asian/ Asian British			
Black/ African/ Caribbean/ Black British			
Mixed/ multiple ethnic groups			
Other			
Prefer not to say			
White			
Relationship status [a] **- n (%)**			

	Texts + Incentive (N=196)	Texts alone (N=194)	Control (N=195)
Married / civil partnership			
Co-habiting			
Single (never married; never in a civil partnership)			
Divorced			
Separated			
Widowed			
Prefer not to say			
Comorbidities [a] **- n (%)**			
High Blood Pressure			
Mental health condition			
Arthritis			
Possible Latent Mental Health Condition			
Diabetes			
Heart condition such as angina or atrial fibrillation			
Stroke (including mini stroke)			
Cancer			
One or more co-morbidity			
Multiple Long-Term Conditions (MLTC)			
MLTC including self-reported diabetes			
Physical or Mental Disability [a]			
Disability - n (%)			
Perceived wealth [a] **- mean (SD); n**			
Perceives to live in relatively wealthy neighbourhood (0-100, Strongly disagree)			
Feels relatively wealthy compared to others (0-100, Strongly disagree)			

	Texts + Incentive (N=196)	Texts alone (N=194)	Control (N=195)
Feels like they have enough money (0-100, Strongly disagree)			
Financial Strain [a] - n (%)			
Living comfortably			
Doing alright			
Just about getting by			
Finding it quite difficult			
Finding it very difficult			
Prefer not to say			
Household composition [a] - n (%)			
Lives alone			
Lives with partner			
Lives with child/children			
Lives with parents			
Lives with friends			
Other			
Household size - mean (SD); n			
Highest educational qualification [a] - n (%)			
Degree level or above			
Another kind of qualification			
Employment Status [a] - n (%)			
Paid job - Full time (30+ hours per week)			
Paid job - Part time (8-29 hours per week)			
Paid job - Part time (Under 8 hours per week)			
Self-employed			

	Texts + Incentive (N=196)	Texts alone (N=194)	Control (N=195)
Full time student			
Unemployed and seeking work			
Retired			
Not in paid work due to illness or disability			
Not in paid work for other reason			
Other			
Prefer not to say			
Access to self-monitoring equipment [a] - n (%)			
Owns scales for self-weighing			
Scales link to internet/app			
Owns an activity tracker/pedometer			
Highest weight (kg) - mean (SD); n			
Lowest weight (kg) - mean (SD); n			
Intended weight loss in study (kg) - mean (SD); n			
Weight loss attempts - median (P25, P75); n			
Measured weight and height			
Weight (kg) - mean (SD)			
Height (cm) - mean (SD)			
BMI (kg/m^2) - mean (SD)			
≥30-<35; n (%)			
≥35-<40; n (%)			
≥40; n (%)			

a – self-report

Table S2. Weight assessment methods for men retained in the trial at 24 months

	Texts + Incentive (N=)	Texts alone (N=)	Control (N=)	Total (N=)
Weight assessment face to face with a researcher on Game of Stones research scales within 23 days of target date				
Weight assessment face to face with a researcher on Game of Stones research scales outside 23 days of target date				
Researcher blind to group allocation				
Weighed on Game of Stones research scales by video call with a researcher within 23 days of target date				
Self-report weight within 23 days of target date				
Self-report weight outside 23 days of target date				

Table S3: Primary outcome: weight change at 24M from baseline

Weight change (%), mean (SD)	**24 months**			Mean difference (Texts + Incentive vs Control) 97.5% Confidence Interval	Mean difference (Texts alone vs Control) 97.5% Confidence interval
	Texts + Incentive	**Texts alone**	**Control**		
All observed cases					
PP					
EWLM between baseline and 24 months					
EWLM in previous 12months					
BOCF					
LOCF					

PP = per protocol measurements following the gold standard (those in green) shown in the SOP 7.5 appendix, EWLM = excluding participants taking weight loss medications used, BOCF = baseline observation carried forward, LOCF = last observation carried forward, kg = kilogram, M = mean, SD = Standard Deviation.

Table S4: Weight change at 24M from baseline (secondary outcomes)

	24 months			Mean difference or odds ratio/risk difference (Texts + Incentive vs Control) 97.5% Confidence interval	Mean difference or odds ratio/risk difference (Texts alone vs Control) 97.5% Confidence interval
	Texts + Incentive	**Texts alone**	**Control**		
Weight change (kg), mean (SD)					
All Observed cases					
Weight loss dichotomies (all observed cases), n (%)					
Any weight loss					
Weight loss ≥5%					
Weight loss ≥10%					
Weight change categories (all observed cases), n (%)					
Weight gain					
0<5% weight loss					
≥5-<10% weight loss					
≥10% weight loss					

kg = kilogram, M = mean, SD = Standard Deviation.

Table S5: Summarised weight change over time to 24M from baseline

	12 months			24 months		
	Texts + Incentive	**Texts alone**	**Control**	**Texts + Incentive**	**Texts alone**	**Control**
Weight change (kg), mean (SD)						
All Observed cases						
Weight loss dichotomies (all observed cases), n (%)						
Any weight loss						
Weight loss ≥5%						
Weight loss ≥10%						
Weight change categories (all observed cases), n (%)						
Weight gain						
0<5% weight loss						
≥5-<10% weight loss						
≥10% weight loss						

kg = kilogram, M = mean, SD = Standard Deviation.

Table S6: Participant recommendation of the Game of Stones trial and weight loss satisfaction at 24M

Variable - mean (SD); n	24 months			Mean difference (Texts + Incentive vs Control) 97.5% Confidence interval	Mean difference (Texts alone vs Control) 97.5% Confidence interval
	Texts + Incentive	**Texts alone**	**Control**		
Programme recommendation [a]					
Happy with weight loss progress [a]					

All data mean (SD), [a] scored 1-7 (1 = low, 7 = high). Each outcome analysed using linear regression adjusting for baseline and minimization covariates

Table S7: PHQ-4, EQ5D and EQ-5D-5L-AD: Baseline and 24M

	Baseline			24 months			Mean difference (Texts + Incentive vs Control) 97.5% Confidence interval	Mean difference (Texts alone vs Control) 97.5% Confidence interval
	Texts + Incentive	**Texts alone**	**Control**	**Texts + Incentive**	**Texts alone**	**Control**		
PHQ-4								
EQ5D								
EQ5D Visual Analog Score (VAS)								
EQ-5D-5L-AD								

Table S8: Weight management strategies used at baseline and 24M

Variable – n/N (%)	Baseline			24 months			Odds Ratio (Texts + Incentive vs Control) 97.5% Confidence interval	Odds Ratio (Texts alone vs Control) 97.5% Confidence interval
	Texts + Incentive	**Texts alone**	**Control**	**Texts + Incentive**	**Texts alone**	**Control**		
Current weight management strategies								
Kept track of weight by weighing yourself	NA	NA	NA					
Looked up strategies, tips, plans on how to lose weight								
Avoided certain foods								
Had a weight goal to work towards								
Reminded yourself of the reasons you're trying to lose weight								
Swapped one type of food for another								
Swapped one type of drink for another								
Told others about your weight loss goals								
Used a book, website, or app								
Checked the portion size of things you eat								
Kept track of the calorie/nutritional content of the things you eat and drink								
Used a weight loss service to help me manage my weight								
Cut down on alcohol								
Increased the amount of physical activity, sport or exercise that you were doing								
None								
Other*								

** Free text to enable analysis category decision*

Table S9: Use of publicly funded weight loss services, weight loss medications or meal replacements between baseline and 24M

Variable		24 months		
		Texts alone	Texts + Incentive	Control
Local/NHS provided service				
Attended a subsidised (e.g. vouchers) or paid for group-based programme such as Weight Watchers or Slimming World or similar meetings	N, n, %			
Attended a subsidised or paid for gym, leisure centre or local sport facility to swim or take part in other physical activity sessions?	N, n, %			
Attended a subsidised or paid for health trainer programme	N, n, %			
Attended a subsidised or paid for exercise referral scheme	N, n, %			
Attended a subsidised or paid for weight management programme at a Community Pharmacy	N, n, %			
Have taken weight loss pills prescribed by the GP or hospital prescribed	N, n, %			
Have had daily weight loss injections prescribed by the GP or hospital prescribed	N, n, %			
Have had weekly weight loss injections prescribed by the GP or hospital prescribed	N, n, %			
Have taken meal replacement drinks for weight loss prescribed by the GP or hospital prescribed where the NHS has paid for them (e.g. Optifast, Slim-Fast, The Cambridge Diet)	N, n, %			
Attended an appointment with an NHS dietician for weight management	N, n, %			
Any of these	N, n, %			
Attended a subsidised (e.g. vouchers) or paid for group-based programme such as Weight Watchers or Slimming World or similar meetings	N, n, %			
I have paid for the following myself:				
Weight loss pills	N, n, %			
Daily weight loss injections	N, n, %			
Weekly weight loss injections	N, n, %			
Meal replacement drinks for weight loss e.g. Optifast, Slim-Fast, The Cambridge Diet	N, n, %			

Variable		24 months		
		Texts alone	Texts + Incentive	Control
Perceived Wealth				
I feel that I live in a relatively wealthy neighbourhood (0 – 100 Strongly disagree)	Mean (SD), n			
I feel relatively wealthy compared to others (0 – 100 Strongly disagree)	Mean (SD), n			
I feel that I have enough money (0 – 100 Strongly disagree)	Mean (SD), n			
Financial Strain				
Living comfortably	N, n, %			
Doing alright	N, n, %			
Just about getting by	N, n, %			
Finding it quite difficult	N, n, %			
Finding it very difficult	N, n, %			
Prefer not to say	N, n, %			

Appendix: SOP 7.5

24 month primary outcome - assessment per protocol

+/- 3 weeks of target date, weight taken on GoS research scales ***AND:***

- Two fieldworkers present (**one blinded**) face-to-face *OR*
- One fieldworker present, face-to-face and verified in person by one **blinded** independent witness *OR*
- One fieldworker present, face-to-face and verified by blinded fieldworker by video *OR*
- One fieldworker present (**not blinded**), face-to-face but not independently verified *OR*
- No fieldworker present and verified by **blinded** fieldworker by video *OR*
- No fieldworker present and verified by **non-blinded** fieldworker by video *OR*

+/- 3 weeks of target date, weight taken on pharmacy/NHS calibrated scales ***AND:***

- No fieldworker present, verified by **blinded** fieldworker by video
- No fieldworker present, verified by **non-blinded** fieldworker by video
- No fieldworker present, verified by (**blinded**) Pharmacist or health professional face-to-face

+/- 3 weeks of target date, remote weight taken on own scales ***AND:***

- No fieldworker present, verified by **blinded** fieldworker by video
- No fieldworker present, verified by **non-blinded** fieldworker by video

+/- 3 weeks of target date, remote weight taken on own scales ***AND:***

- No fieldworker present, not verified

OR

Any weight provided that does not fit the above categories and is closer to the 12m data collection time point than the 24m data collection timepoint

Missing

Gold standard per protocol sensitivity analysis

Include in intention to treat analysis

Include in intention to treat analysis

Include in intention to treat analysis if closer to 12m than 24m data collection timepoint