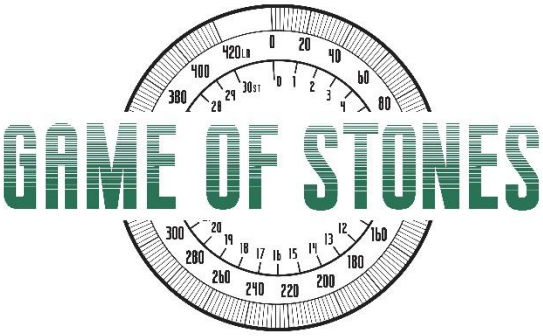

Effectiveness and cost-effectiveness of text message and endowment incentives for weight management in men with obesity: The Game of Stones randomised controlled trial

'The Game of Stones Trial'

Statistical analysis plan v2.0

21-06-2023

Based on the Protocol Version v6.0 11.05.23

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	v4.0	NA	NA – draft 1.0 signed	BG, AE, GMAC	22/06/22
2.0	V6.0	Signature page Dummy tables Tables S1 and S2 and 11a. Trial Statistician References Table 1 Table 2: Table S1/S2	Removal of AE from SAP as no longer associated with Game of Stones Correction to errors in table numbers Edited to reflect the deprivation categories on a country specific 5-point scale Added Trial Statistician JS. Updated citations and reference list. Change in data collected at 24M. Change to correct form of statistical analysis strategy. Changed scale labelling of Deprivation category variable. Formatted and included rows for one or more or no comorbidities section. Amended table to accurately reflect question pertaining to household composition. Perceived wealth corrected to show	JS, GMAC	21/06/23

			it is measured as continuous rather than categorical.		
		Table S10	Added summary rows pertaining to change in vigorous physical activity, moderate physical activity, sedentary behaviour, alcohol consumption and smoking consumption from baseline, respectively.		
		Table S11	Added Change in perceived wealth, financial strain, employment status from baseline, respectively. Removed MLTC including diabetes subgroup analysis.		
		Confirmatory/ Exploratory subgroup analyses	Adjusted layout to match study protocol.		
		MLTC	Removed disability from MLTC definition at baseline.		
		Appendix	Added Standard Operating Procedure (SOP) 7.5 as an appendix.		

BG – Beatriz Goulao; GMAC – Graeme MacLennan; AE; Andy Elders; JS – James Swingler

Contents

Introduction

Obesity increases the risk of type 2 diabetes, heart disease, stroke, mobility problems and some cancers, and its prevalence in rising. Men engage less than women in existing weight loss interventions. This trial builds on our Game of Stones NIHR PHR-funded feasibility study. A more detailed description of Game of Stones can be found in the protocol.

Game of Stones is a pragmatic RCT with primary outcome evaluation at 12M. A secondary outcome evaluation at 24M is planned conditional on the 12m results. This SAP focuses on the 12M analysis of the trial. A subsequent SAP will be prepared for analysis of the 24M data if required.

Objectives

Primary objective

To compare the difference in % weight change at 12M from baseline for men with obesity in the following groups: i) SMS with endowment incentive (SMS+I) vs 12M wait list; ii) SMS only vs 12M wait list.

Secondary objectives

- To assess differences between groups in secondary outcomes
- To assess the cost-effectiveness of SMS+I and SMS only compared to a wait list control
- To understand men's and service providers' experiences of the intervention
- To explore whether there is treatment effect heterogeneity between socio-economic status at baseline
- To compare PHQ-4, Warwick and Edinburgh Mental Wellbeing Scale (WEMWBS), Quality of Life (EQ-5D-5L-AD), Anxiety and Depression (AD) dimension and Weight Self-Stigma Questionnaire (WSSQ) measures for the 3 trial groups at baseline and 12 months
- To undertake exploratory moderator analyses examining interactions between mental health/wellbeing status at baseline and 12-month weight change
- To undertake exploratory moderator analysis for weight change at 12 months by the presence/absence of multiple long-term conditions, presence/absence of obesity related co-morbidities, presence/absence of diabetes at baseline
- To compare secondary Quality of Life, mental health/wellbeing outcomes for men with or without MLTC at baseline

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 for: i) SMS with endowment incentive vs 12-month waiting list control; and ii) SMS only vs 12-month waiting list control.

Sample size

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,[1] and the mean difference of 3.3kg is derived from 3% of 109kg (the mean baseline weight in the Feasibility Study).[2] Several trials of SMS-delivered weight management interventions [3] 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.[3] 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) SMS+ endowment incentive vs 12M waitlist
2) SMS only vs 12M waitlist

Timing of final analysis

There are no planned interim analyses of any outcomes, there will be no Data Monitoring Committee. The trial will not be stopped for futility reasons or safety reasons before the collection analysis of 12-month outcome data. A single final analysis will be performed, and results shared with co-investigators, after 12-month assessment data for the last man has been collected, all data will have been entered and the database will be locked at the end of June 2023. A separate SAP will be written for follow-up data collected at 24 months.

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 blinded to allocation during the analysis. Outcome assessors will be blind to allocation when the participant is weighed at 12m. Fieldworkers undertaking assessments prior to the 12m assessment will not be blind to allocation. 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. This is a self-management intervention, therefore stopping the text messages but continuing to attend weight assessments is compliant. Similarly, endowment incentives for meeting weight targets are offered, but participants can choose whether to accept them. Descriptive data on number of SMS delivered, responses received and engagement in self-reporting progress on the study website are included in the Process Evaluation Analysis Plan.

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 three conditions must be met:

1. Research scales approved by Game of Stones used for measuring weight
2. Blinding of fieldworker to randomised allocation
3. Weight measured within a window of +/-23 days from the follow-up due date (ie 3m/6m/12m after date of randomisation)

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 main analysis of Game Stones will be based on an intention-to-treat framework (ITT), given the pragmatic nature of the trial. All 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)
- Participant questionnaires
- SMS activity data (the numbers of SMS sent, requests to stop SMS and responses sent)
- Activity data from interactive participant web pages

Trial population

Screening of potential participants for eligibility (see Protocol), recruitment processes, 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 [4] 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 [5] 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,[2] FFIT trial [6] socio-economic measures relevant to obesity [7-10] and the recently published STAR-LITE core outcome set for behavioural weight management interventions.[11]

Baseline variable definitions

For socio-demographic variables, we will adhere to the harmonised standards guidance published by the Office for National Statistics, [12] except for employment status where harmonised guidance from the Scottish Government will be used. [13] Perceived wealth [7] and financial strain [9] variables are included to capture additional health inequalities data. Men under-report mental health conditions. [14] 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. [15]

The presence of Multiple Long-Term Conditions (MLTC) is based on existing definitions [16] (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 problem. 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 6 months (two active intervention groups), 12 and 24 months (all three groups), unless otherwise indicated. Details on the 24 months outcome measures and analysis will be available in a separate analysis plan. Outcomes are also measured at baseline, where applicable.

Primary outcome

The primary outcome is within-participant change from baseline weight expressed as a percentage of baseline weight at 12 months from baseline. The trial is powered on a 3% weight loss, which NICE CG189 considers clinically significant, is consistent with STAR-LITE Core Outcome set for obesity trials[11] and with effectiveness of other SMS trials for weight loss.[3]

Secondary outcomes at 12 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 SMS and SMS+I 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
- WEMWBS
- PHQ-4
- Weight Self-Stigma Questionnaire (WSSQ) (See Section 9.12. in Protocol SWAT)

Exploratory outcomes/endpoints at 12 months

- weight management strategies used (including social weight loss)
- self-monitoring activity weight (any activity done to self-monitor one's weight, i.e., self-weighing)
- self-monitoring steps
- physical activity
- alcohol frequency (increased/the same alcohol frequency)
- smoking status (increased/the same smoking frequency)
- satisfaction with Game of Stones
- satisfaction with weight loss progress
- confidence in ability to lose weight
- confidence in ability to maintain weight loss long-term

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 the Health Economic Analysis Plan.

Process outcomes

A separate Process Evaluation Analysis Plan describes descriptive quantitative and qualitative data on recruitment strategies, weight assessments undertaken at 3 and 6 months in intervention groups, assessment characteristics for 12-month weight assessments (Appendix SOP 7.5) and intervention engagement with text messages and study website.

Data collection time points

The outcome assessment schedule has been designed according to the STAR-LITE core outcome set [11] and the Feasibility Study findings. [2] The timing of data collection is summarised in Table 1.

Table 1 – Data collection time points

Data Collection	**Baseline**	**12M**	**24M***
Socio-demographic: derivation category, comorbidities (physical and mental health), possible latent mental health condition, self-reported disability, ethnicity, age, perceived wealth, financial strain, education, employment, household size; relationship status.	✓		
Anthropometry - height (for BMI)	✓		
Anthropometry – weight	✓	✓	✓
Participant satisfaction		✓	
Health behaviours - physical activity, smoking status, alcohol intake	✓	✓	
Weight management strategies used over last 12 months	✓	✓	✓
Patient Health Questionnaire -4	✓	✓	✓
EQ-5D-5L – Anxiety and Depression Dimension (AD)	✓	✓	✓
Weight Self-Stigma Questionnaire (WSSQ) (included in SWAT PhD)	✓	✓	
Confidence in ability to lose weight and maintain weight loss	✓	✓	
Warwick-Edinburgh Mental Well-Being Scale (WEMWBS)	✓	✓	
Health Economic Outcomes: EQ-5D, NHS health care use	✓	✓	✓
Qualitative interview data (experiences, behaviours)		✓	✓
Process outcomes: observed weight meets target and incentives secured/lost at 3M and 6M; number of SMS delivered; SMS responses, web page use over 12M	✓	✓	
Unintended consequences or adverse events		✓	✓

* Analysis of phase 3 of trial not covered in this SAP, a separate SAP will be written for 24-month outcome data if required and will be guided by Trial Steering Committee

Scoring of outcomes measured in questionnaires

All questionnaire items will be analysed individually, as set out in the dummy tables (Appendix), with the exceptions of WEMWBS. In this scale, there are 14 questions (e.g. 'I've been feeling optimistic about the future') each with 5 levels of response. For all items, 'None of the time', 'Rarely', 'Some of the time', 'Often' and 'All of the time' is scored 1, 2, 3, 4 and 5 respectively. The participant-mean value will be imputed if there is a missing value. The overall score is calculated as the sum of the 14 items, with a potential range from 14 to 70. Higher scores indicate greater wellbeing. The overall score should be treated as missing if there are more than three missing items. [17] 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

The analysis of the primary outcome will estimate the mean difference in change from baseline weight expressed as a percentage at 12M between groups (SMS+I vs control; and SMS only 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. smoking status) and ordered logit for ordinal outcomes (e.g. alcohol frequency). 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: SMS+I vs control; SMS only 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, WEMWBS, PHQ-4, Weight Self-Stigma Questionnaire (WSSQ)	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

Subgroup analyses are split confirmatory and exploratory. The confirmatory subgroup analyses are based on hypothesised directions of effect modification of the interventions informed by the weight-loss literature. The exploratory subgroup analyses are based on potential moderators, for which there are gaps or conflicting evidence in the literature.

The NIHR Funding Board asked for subgroup analysis of outcomes by centre, recruitment strategy and deprivation category and funded additional analyses relevant to men living with mental health conditions, obesity related comorbidity and Multiple Long-Term Conditions. Compliant with CHAMP reporting guidelines, [19] we will examine these factors along with other subgroups defined by socio-economic and health status. Subgroup analyses will be performed for the primary outcome and by adding a treatment factor interaction to the model. Confidence intervals will be presented at 99.5%.

Confirmatory subgroup analyses

- comorbidities (at least one comorbidity versus none):
    - obesity related comorbidity [20] present/absent (see protocol for more details)
    - self-report diabetes [21] present/absent

Exploratory subgroup analyses

- Multiple long-term conditions are defined as the presence of two or more obesity related co-morbidities (stroke including mini-stroke, high blood pressure, a heart condition, diabetes, cancer, arthritis, or a mental health problem)
    - MLTC present/absent
    - Self-reported disability (do you have any physical or mental health conditions or illnesses lasting or expected to last 12 months or more AND answers yes to reduce your ability to carry-out day-to-day activities either a little or a lot)
- recruitment route
    - centre (England / Scotland / Northern Ireland)
    - strategy (community / GP)
- socio-economic status

    - deprivation category (based on the postcode address where participants live – UK adjusted index of multiple deprivation (two most deprived categories v three least deprived categories). Data from the separate IMDs in England, Scotland and Northern Ireland will be classified as per the country-specific methodology for the purpose of subgroup classification. [22,23]
    - financial strain (difficult / very difficult vs every other category)
    - perceived wealth
    - highest educational qualification (degree level or above / other qualification / no qualification)
    - work status (in paid work or self-employed / not in paid work)
- relationship status (Co-habiting / living in a couple vs rest of categories)
- mental health and wellbeing status
    - EQ-5D-5L overall utility score (above / below mid-scale at 0.4005)
    - EQ-5D-5L anxiety and depression dimension (not & slightly / moderate, severe / extremely)
    - Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) (less than or equal to 40 versus more than 40 [17]
    - Self-reported mental health condition present/absent (defined via a question in the baseline questionnaire)
    - PHQ-4 (3+)
    - Weight Self-Stigma Questionnaire (WSSQ) (High vs low)
    - Living situation (Lives alone/Lives with others)
    - Alcohol frequency (Every day vs not every day)
    - Possible latent mental health condition (existent/non-existent, definition as specified in section **Baseline variable definitions** of this document)
- social weight loss (lose weight alone / with others)

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. [24]

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, [11] for comparability with previous weight loss studies. [25]

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.

If a data item includes the option “Prefer not to say”, any such responses will be treated as a separate category and not classed as missing data.

Additional analysis

No other additional analyses are planned. A separate Process Evaluation Analysis Plan prospectively describes mixed methods descriptive and qualitative data analyses, and a Health Economics Analysis Plan is also available upon request. There is a separate SWAT statistical analysis plan prepared by the PhD student that does not form part of the analysis of Game of Stones.

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

2. 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

3. 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

4. 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

5. 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

6. 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

7. 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

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

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

10. 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

11. 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

12. 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

13. 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

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

15. 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

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

17. Bianco D. *Performance of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) as a screening tool for depression in UK and Italy.* Thesis for Clinical Psychology ed. University of Bologna; 2012

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. van Hoorn R, Tummers M, Booth A, et al. The development of CHAMP: a checklist for the appraisal of moderators and predictors. *BMC Medical Research Methodology.* 2017;17(1):173. doi:https://doi.org/10.1186/s12874-017-0451-0

20. Dombrowski SU, Avenell A, Sniehott FF. Behavioural Interventions for Obese Adults with Additional Risk Factors for Morbidity: Systematic Review of Effects on Behaviour, Weight and Disease Risk Factors. *Obes Facts.* 2010;3(6):377-396. doi:https://doi.org/10.1159/000323076

21. LeBlanc ES, Patnode CD, Webber EM, Redmond N, Rushkin M, O'Connor EA. Behavioral and Pharmacotherapy Weight Loss Interventions to Prevent Obesity-Related Morbidity and Mortality in Adults: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA.* 2018;320(11):1172-1191. doi:10.1001/jama.2018.7777

22. Abel GA, Barclay ME, Payne RA. Adjusted indices of multiple deprivation to enable comparisons within and between constituent countries of the UK including an illustration using mortality rates. *BMJ Open.* 2016;6(11):e012750. doi:10.1136/bmjopen-2016-012750

23. O'Neill J, Tabish H, Welch V, et al. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. *J Clin Epidemiol.* 2014;67(1):56-64. doi:10.1016/j.jclinepi.2013.08.005

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

25. 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

		SMS Messaging + Incentive N=			SMS Messaging only N=			Control N=			Source
Age (years)	N, Mean, SD										CRF
Weight (kg)	N, Mean, SD										CRF
Height (cm)	N, Mean, SD										CRF
BMI (kg/m^2)	N, Mean, SD										CRF
≥30-<35	N, n, %										
≥35-<40	N, n, %										
≥40-	N, n, %										
Highest weight (kg)	N, Mean, SD										Bas PQ
Lowest weight (kg)	N, Mean, SD										Bas PQ
Intended weight loss in study (kg)	N, Mean, SD										Bas PQ
Weight loss attempts	N, Mean, SD										Bas PQ
Deprivation category											HIC/CRF
1 (most deprived)	N, n, %										
2	N, n, %										
3	N, n, %										
4	N, n, %										
5 (least deprived)	N, n, %										
Perceived wealth											Bas PQ
Perceives lives in relatively wealthy neighbourhood	N, Mean, SD										
Feels relatively wealthy compared to others	N, Mean, SD										
Feel that I have enough money	N, Mean, SD										
Prefer not to say	N, Mean, SD										
Financial strain											Bas PQ
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, %										
Relationship status											Bas PQ
Single (never married; never in a civil partnership)	N, n, %										
Cohabiting	N, n, %										
Married / civil partnership	N, n, %										
Separated	N, n, %										
Divorced	N, n, %										
Widowed	N, n, %										
Prefer not to say	N, n, %										
Comorbidities											Bas PQ
None	N, n, %										
One or more	N, n, %										
Stroke (including TIA)	N, n, %										
High blood pressure	N, n, %										
Heart condition such as angina or atrial fibrillation	N, n, %										
Diabetes	N, n, %										
Cancer	N, n, %										
Arthritis	N, n, %										
Mental health condition	N, n, %										
Possible latent mental health condition	N, n, %										
Disability	N, n, %										Bas PQ
Multiple Long-Term Condition (MLTC)	N, n, %										
MLTC includes self-report diabetes	N, n, %										
Ethnic group											Bas PQ
White	N, n, %										
Mixed/ multiple ethnic groups	N, n, %										
Asian/ Asian British	N, n, %										
Black/ African/ Caribbean/ Black British	N, n, %										
Other	N, n, %										
Prefer not to say	N, n, %										
Household Composition											Bas PQ

Lives alone	N, n, %										
Lives with partner	N, n, %										Bas PQ
Lives with child/children	N, n, %										
Lives with parents	N, n, %										
Lives with friends	N, n, %										
Other	N, n, %										
Household size	N, Mean, SD										
Highest educational qualification											Bas PQ
Degree level or above	N, n, %										
Another kind of qualification	N, n, %										
No qualification	N, n, %										
Prefer not to say	N, n, %										
Employment status											Bas PQ
Paid job - Full time (30+ hours per week)	N, n, %										
Paid job - Part time (8-29 hours per week)	N, n, %										
Paid job - Part time (Under 8 hours per week)	N, n, %										
Paid job – Furloughed due to pandemic	N, n, %										
Self-employed	N, n, %										
Full time student	N, n, %										
Unemployed and seeking work	N, n, %										
Retired	N, n, %										
Not in paid work due to illness or disability	N, n, %										
Not in paid work for other reason*	N, n, %										
Prefer not to say	N, n, %										
Access to self-monitoring equipment											Bas PQ
Owns scales for self-weighing	N, n, %										
Scales link to internet/app	N, n, %										
Owns an activity tracker/pedometer	N, n, %										

***** *Free text – to enable decision for categories*

Table S2: Participant baseline characteristics by recruitment strategy

		Community recruitment N=			GP recruitment N=			Total N=		
Age (years)	N, Mean, SD									
Weight (kg)	N, Mean, SD									
Height (cm)	N, Mean, SD									
BMI (kg/m^2)	N, Mean, SD									
≥30-<35	N, n, %									
≥35-<40	N, n, %									
≥40-	N, n, %									
Highest weight (kg)	N, Mean, SD									
Lowest weight (kg)	N, Mean, SD									
Intended weight loss in study (kg)	N, Mean, SD									
Weight loss attempts	N, Mean, SD									
Deprivation category										
1 (most deprived)	N, n, %									
2	N, n, %									
3	N, n, %									
4	N, n, %									
5 (least deprived)	N, n, %									
Perceived wealth										
Perceives lives in relatively wealthy neighbourhood	N, Mean, SD									
Feels relatively wealthy compared to others	N, Mean, SD									
Feel that I have enough money	N, Mean, SD									
Prefer not to say	N, Mean, SD									
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, %									
Relationship status											
Single (never married; never in a civil partnership)		N, n, %									
Co-habiting		N, n, %									
Married / civil partnership		N, n, %									
Separated		N, n, %									
Divorced		N, n, %									
Widowed		N, n, %									
Prefer not to say		N, n, %									
Comorbidities											
None	N, n, %										
One or more	N, n, %										
Stroke (including TIA)	N, n, %										
High Blood Pressure	N, n, %										
Heart condition such as angina or atrial fibrillation	N, n, %										
Diabetes	N, n, %										
Cancer	N, n, %										
Arthritis	N, n, %										
Mental health condition	N, n, %										
Possible Latent Mental Health Condition	N, n, %										
Disability	N, n, %										
Multiple Long-Term Conditions (MLTC)	N, n, %										
MLTC includes self-report diabetes	N, n, %										
Ethnic group											
White	N, n, %										
Mixed/ multiple ethnic groups	N, n, %										
Asian/ Asian British	N, n, %										
Black/ African/ Caribbean/ Black British	N, n, %										

Other	N, n, %									
Prefer not to say	N, n, %									
Household composition										
Lives alone	N, n, %									
Lives with partner	N, n, %									
Lives with child/children	N, n, %									
Lives with parents	N, n, %									
Lives with friends	N, n, %									
Other	N, n, %									
Household size	N, Mean, SD									
Highest educational qualification										
Degree level or above	N, n, %									
Another kind of qualification	N, n, %									
No qualification	N, n, %									
Prefer not to say	N, n, %									
Employment status										
Paid job - Full time (30+ hours per week)	N, n, %									
Paid job - Part time (8-29 hours per week)	N, n, %									
Paid job - Part time (Under 8 hours per week)	N, n, %									
Paid job - Furloughed	N, n, %									
Self-employed	N, n, %									
Full time student	N, n, %									
Unemployed and seeking work	N, n, %									
Retired	N, n, %									
Not in paid work due to illness or disability	N, n, %									
Not in paid work for other reason	N, n, %									
Prefer not to say	N, n, %									
Access to self-monitoring equipment										

Owns scales for self-weighing	N, n, %									
Scales link to internet/app	N, n, %									
Owns an activity tracker/pedometer	N, n, %									

** Free text to enable analysis category decision*

Table S3: Weight change at 12M from baseline (primary outcome)

Weight change (%), mean (SD)	**12 months**			Mean difference (SMS+I vs control) 97.5% Confidence Interval	Mean difference (SMS only vs control) 97.5% Confidence interval
	SMS + I	**SMS only**	**Control**		
All observed cases					
PP1					
PP2					
BOCF					
LOCF					

PP1 = per protocol measurements following the gold standard (those in green) shown in the SOP 7.5 appendix, PP2 = per protocol measurements following the gold (those in green) and orange standard shown in the SOP 7.5 appendix, BOCF = baseline observation carried forward, LOCF = last observation carried forward, kg = kilogram, M = mean, SD = Standard Deviation.

Table S4: Weight change at 12M from baseline (secondary outcomes)

	12 months			Mean difference or odds ratio (SMS only vs control) 97.5% Confidence interval	Mean difference or odds ratio (SMS only vs control) 97.5% Confidence interval
	SMS + I	**SMS only**	**Control**		
Weight change (kg), mean (SD)					
All Observed cases					
Weight loss dichotomies (all observed cases), n (%)					
>0% weight loss					
≥5%					
≥10%					
Weight change categories (all observed cases), n (%)					
Any weight loss					
Weight gain					
0<5% weight loss					
≥5-<10% weight loss					
≥10% weight loss					

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

Table S5: Game of Stones satisfaction at 12 months

	12 months			Mean difference (SMS+I vs control) 97.5% Confidence Interval	Mean difference (SMS only vs control) 97.5% Confidence interval
	SMS only	**SMS + I**	**Control**		
Programme satisfaction[a] (0-100), mean (SD)					
Happy with weight loss progress[b]					

All data mean (SD), [a] scored 0-100 (higher = more satisfied), [b] scored 1-7 (1 = low, 7 = high)

Table S6: Warwick-Edinburgh Mental Well-being Scale, PHQ-4, WSSQ and EQ-5D-5L-AD: Baseline and 12m

		Baseline			**12 months**			Mean difference (SMS+I vs control) 97.5% Confidence Interval	Mean difference (SMS only vs control) 97.5% Confidence interval
		SMS + I	**SMS only**	**Control**	**SMS + I**	**SMS only**	**Control**		
Warwick-Edinburgh Mental Wellbeing Scale	N, Mean, SD								
PHQ-4	N, Mean, SD								
WSSQ	N, Mean, SD								
EQ5D	N, Mean, SD								
EQ5D Visual Analog Score (VAS)	N, Mean, SD								
EQ-5D-5L-AD	N, Mean, SD								

Table S7: Confidence in weight loss and weight loss maintenance ability Baseline to 12 months

	Baseline			12 months			Mean difference (SMS+I vs control) 97.5% Confidence Interval	Mean difference (SMS only vs control) 97.5% Confidence interval
	SMS + I	SMS only	Control	SMS + I	SMS only	Control		
Confidence in ability to lose weight[a]								
Confidence in ability to maintain weight loss [a]								

All data mean (SD), [a] scored 1-7 (1 = low, 7 = high)

Table S8: Weight management strategies used at baseline and 12m

		Baseline			12 months			Effect size (SMS+I vs control) 97.5% Confidence Interval	Effect size (SMS only vs control) 97.5% Confidence interval
		SMS + I	SMS only	Control	SMS + I	SMS only	Control		
Self-monitoring activity									
Self-weighing									
Never	N, n, %								
Less than once a month	N, n, %								
Once a month	N, n, %								
Once a week	N, n, %								
A few times a week	N, n, %								
Everyday	N, n, %								
Activity Monitor/Pedometer Use									
Never	N, n, %								
Less than once a month	N, n, %								
Once a month	N, n, %								
Once a week	N, n, %								
A few times a week	N, n, %								
Everyday	N, n, %								
Current weight management strategies									
Looked up strategies, tips, plans on how to lose weight	N, n, %								
Avoided certain foods	N, n, %								
Had a weight goal to work towards	N, n, %								
Reminded yourself of the reasons you're trying to lose weight	N, n, %								
Swapped one type of food for another	N, n, %								
Swapped one type of drink for another	N, n, %								

Told others about your weight loss goals	N, n, %								
Used a book, website, or app	N, n, %								
Checked the portion size of things you eat	N, n, %								
Kept track of the calorie/nutritional content of the things you eat and drink	N, n, %								
Used a weight loss service to help me manage my weight	N, n, %								
Cut down on alcohol	N, n, %								
Increased the amount of physical activity, sport or exercise that you were doing	N, n, %								
None	N, n, %								
Other*	N, n, %								

** Free text to enable analysis category decision*

Table S9: Social weight loss at 12 months

		12 months			Mean difference (SMS+I vs control) 97.5% Confidence Interval	Mean difference (SMS+I vs control) 97.5% Confidence Interval
		SMS + I	SMS only	Control		
Who have you told that you are taking part in research about weight loss?						
No-one	N, n, %					
Partner	N, n, %					
Child/children	N, n, %					
Parent/guardian	N, n, %					
Other family member	N, n, %					
Friend	N, n, %					
Flatmate	N, n, %					
Work colleague	N, n, %					
Other*	N, n, %					
Over the past 12 months:						
I have been trying to lose weight	N, n, %					
By myself	N, n, %					
With one other person	N, n, %					
With two or more other people	N, n, %					
I have not been trying to lose weight	N, n, %					

** Free text to enable analysis category decision*

Table S10: Health behaviours at baseline and 12 months

		Baseline			12 months			Mean difference or odds ratio (SMS+I vs control) 97.5% Confidence Interval	Mean difference or odds ratio (SMS only vs control) 97.5% Confidence Interval
		SMS + I	SMS only	Control	SMS + I	SMS only	Control		
Vigorous physical activity in past week (days)	N, M, SD								
Change in vigorous physical activity from baseline (days)								NA	NA
Decrease	N, n, %								
Stayed the same	N, n, %								
Increase	N, n, %								
Moderate physical activity in past week (days)	N, M, SD								
Change in moderate physical activity from baseline								NA	NA
Decrease	N, n, %								
Stayed the same	N, n, %								
Increase	N, n, %								
Sedentary behaviour in past week (hours, min)	N, M, SD								
Change in sedentary behaviour from baseline (days)								NA	NA
Decrease	N, n, %								
Stayed the same	N, n, %								
Increase	N, n, %								
Frequency of alcohol consumption in past month									
Every day	N, n, %								
5 to 6 times a week	N, n, %								
3 to 4 times a week	N, n, %								
Twice a week	N, n, %								

Once a week	N, n, %								
2 to 3 times a month	N, n, %								
Once a month	N, n, %								
Never	N, n, %								
Change in alcohol consumption from baseline									
Decrease	N, n, %								
Stayed the same	N, n, %								
Increase	N, n, %								
Smoking status (current)								NA	NA
Current smoker (regular)	N, n, %								
Current smoker (irregular)	N, n, %								
Ex-smoker	N, n, %								
Never smoked	N, n, %								
Change in smoking status from baseline									
Decrease	N, n, %								
Stayed the same	N, n, %								
Increase	N, n, %								

Table S11: Moderator subgroup analyses: Weight change % at 12 months (complete cases only) by socio-economic status; Multiple Long-Term Conditions (MLTC) and Comorbidity; Mental Health; Recruitment strategy

		12 months			Mean difference (SMS+I vs control) 99.5% Confidence Interval	Mean difference (SMS only vs control) 99.5% Confidence Interval
Weight change (% from baseline)		SMS + I	SMS only	Control		
SOCIO-ECONOMIC STATUS						
Index of Multiple Deprivation						
Less deprived (3, 4, 5)	N, n, %					
More deprived (1, 2)	N, n, %					
Perceived Wealth						
Perceives lives in relatively wealthy neighbourhood	N, Mean, SD					
Feels relatively wealthy compared to others	N, Mean, SD					
Feel that I have enough money	N, Mean, SD					
Prefer not to say	N, Mean, SD					
Change in perceived wealth from baseline						
Decrease	N, n, %					
Stayed the same	N, n, %					
Increase	N, 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, %					
Change in financial strain from baseline						
Decrease	N, n, %					
Stayed the same	N, n, %					
Increase	N, n, %					
Education						
Degree level or above education	N, n, %					
Other qualification	N, n, %					
No qualification	N, n, %					
Prefer not to say	N, n, %					
Employment status						
In paid employment	N, n, %					
Not in paid employment	N, n, %					
Prefer not to say	N, n, %					
Change in employment status from baseline						
Decrease	N, n, %					
Stayed the same	N, n, %					
Increase	N, n, %					
Household status						
Lives alone (household size)	N, n, %					
Lives with others	N, n, %					
Prefer not to say	N, n, %					
Relationship status						
Living in a couple	N, n, %					
Not living in a couple	N, n, %					
Prefer not to say	N, n, %					
HEALTH AND WELLBEING						
Comorbidity						
At least one obesity related comorbidity	N, n, %					

Reports no obesity related comorbidities	N, n, %					
Diabetes						
Yes	N, n, %					
No	N, n, %					
Multiple Long-Term Conditions (MLTC)						
Yes	N, n, %					
No	N, n, %					
Disability						
Yes	N, n, %					
No	N, n, %					
EQ-5D-5L – total utility score						
High	N, n, %					
Low	N, n, %					
EQ-5D-5L Anxiety and depression dimension of EQ-5D-5L						
High	N, n, %					
Low	N, n, %					
Mental Health						
Reports a mental health condition	N, n, %					
Reports no mental health condition	N, n, %					
Possible latent mental health condition	N, n, %					
PHQ-4						
High	N, n, %					
Low	N, n, %					
Warwick-Edinburgh Mental Wellbeing Scale						
High	N, n, %					
Low	N, n, %					
WSSQ						
High	N, n, %					
Low	N, n, %					

Drinks alcohol every day	N, n, %					
RECRUITMENT STRATEGY						
Community	N, n, %					
GP	N, n, %					

Table S12: Exploratory analysis by subgroups for social weight loss reported by participants at 12 months

Table 12A: Weight change (% from baseline)		12 months			Mean difference (SMS+I vs control) 99.5% Confidence Interval	Mean difference (SMS only vs control) 99.5% Confidence Interval
		SMS + I	SMS only	Control		
Told no-one about Game of Stones	N, n, %					
Told others about participating in Game of Stones	N, n, %					
Tried to lose weight by myself	N, n, %					
Tried to lose weight with others	N, n, %					
Did not try to lose weight	N, n, %					

Appendix: SOP 7.5

12 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

Table A1. Weight assessment characteristics for men retained in the trial at 12 months

		12 months			
		SMS + I N=	SMS-only N=	Control N=	Total
Weight assessment in person on Game of Stones research scales within 23 days of target date	n(%)				
Weight assessment in person on Game of Stones research scales outside 23 days of target date	n(%)				
Blind outcome assessment	n(%)				
Weighed on Game of Stones research scales remotely within 23 days of target date	n(%)				
Weighed on Game of Stones research scales remotely outside 23 days of target date	n(%)				
Weighed on NHS/Pharmacy scales within 23 days of target date	n(%)				
Weighed on NHS/Pharmacy scales outside 23 days of target date	n(%)				
Self-report weight within 23 days of target date	n(%)				
Self-report weight outside 23 days of target date	n(%)				