safetxt: A randomised controlled trial of an intervention delivered by mobile phone messaging to reduce sexually transmitted infections (STI) by increasing sexual health precaution behaviours in young people.

Statistical Analysis Plan

Approved by:

Name	Date	Signature
Dr Caroline Free Principal Investigator	10th June 2020	
Dr James Carpenter Trial Statistician	10th June 2020	

Contents

Trial summary

Title: safetxt: a safer sex intervention delivered by mobile phone messaging to reduce sexually transmitted infections among young people in the UK: statistical analysis plan for a randomised controlled trial

Design: A single blind two-arm randomised controlled trial among young people who have recently been diagnosed with chlamydia, gonorrhoea or NSU (non-specific urethritis), recruited from sexual health services in the UK. Participants will be allocated to receive the safetxt intervention (text messages designed to promote safer sexual health behaviours) or to receive the control text messages (monthly messages asking participants about changes in contact details) by a remote online randomisation system.

Outcomes: The primary outcome is cumulative incidence of chlamydia and gonorrhoea infection at one year assessed by nucleic acid amplification tests (NAAT). Secondary outcomes include partner notification, correct treatment of infection, condom use, and STI testing prior to sex with new partners.

Eligibility: Aged 16-24; own a personal mobile phone; able to provide informed consent; and have either been diagnosed with chlamydia, gonorrhoea, or NSU in the last 2 weeks, or have started treatment for Chlamydia, gonorrhoea or NSU in the last two weeks.

Ethical approval: Ethics approval was obtained from NHS Health Research Authority, London – Riverside Research Ethics Committee (REC reference: 15/LO/1665) and the LSHTM Intervention Research Ethics Committee (REC reference: 10464).

Trial Registration: International Standard Randomised Controlled Trials Number: ISRCTN64390461. Registered on 17th March 2016.

1. Introduction

Purpose and scope of the Statistical Analysis Plan (SAP)

The document outlines the planned analysis for the safetxt trial and the results that will be presented in the main paper (shell tables are presented in Appendix 2). The purpose of pre-specifying the analyses is to assure that they are not influenced by the results after unmasking. This SAP will not prevent additional analyses from being conducted, which may become relevant during the analyses outlined in this document. Additional analyses will be interpreted with caution, as they will not have been pre-specified. This SAP will also not prevent the analysis from being adapted if situations arise that necessitate them. This will be done with transparency, will be justified and published. This SAP does not cover the analysis of the qualitative interviews.

Writing of the SAP

This analysis plan was written by Melissa Palmer and Ona McCarthy with input from Phil Edwards, Tim Clayton, James Carpenter and Cari Free. The final version of the analysis plan was approved by CF and JC. A first version was written as the trial was ongoing. The final version was updated once all the trial data had been collected. All researchers were masked to treatment allocation when the final version was written and approved.

Data Checking

Basic checks will be performed to check for abnormal data. These include:

- Checks for missing data
- Values outside expected ranges or impossible values
- Checks for responses to non-applicable questions
- Other inconsistencies between variables (e.g. participants who report not having been diagnosed with an STI since joining the study, but who simultaneously have been identified from clinic data as having received a positive STI diagnosis).

When inconsistencies are found, data will be double-checked and corrected if necessary, or set to missing. All changes will be documented.

Analysis commencement

The analysis program will be prepared using Stata software prior to unmasking. The analysis outlined in this document will be conducted by James Carpenter and Tim Morris (both of whom will be masked) after 1) follow-up data collection is complete, 2) the complete dataset is checked and cleaned and 3) the final SAP was approved and publicly released.

2. Objectives

The primary objective of this trial is to establish the effect of safetxt on the cumulative incidence of chlamydia and gonorrhoea infection at one year. Secondary objectives are to establish the effect of safetxt on partner notification and condom use at four weeks and on condom use and STI testing at one year.

3. Outcomes

Primary outcome

The primary outcome is the cumulative incidence of chlamydia and gonorrhoea infection at one year assessed by NAAT tests: urine for men (with additional pharyngeal and anal swabs for MSM) and self-taken vulvo-vaginal swab for women.

Secondary outcomes

At 4 weeks, the proportion of participants:

- correctly treated for their STI (took the prescribed antibiotic treatment and avoided sex for 7 days after treatment)
- who told the last person they had sex with before the tested positive, that they needed to get treatment
- whose partner attended clinic for treatment (identified from clinic records)
- who report condom use at last sex

At 12 months, the proportion of participants who report:

- condom use at last sex
- 0, 1 or 2+ sexual partners since joining the trial
- sex with someone new since joining the trial
- condom use at first sex with most recent new partner
- sexually transmitted infection testing for self, prior to first sex with most recent new partner (testing confirmed by clinic record)
- that their most recent new partner was tested for sexually transmitted infection prior to sex with them
- car accident in the past year where the participant was the driver
- experience of partner violence in the past year

and

- the proportion of participants who are diagnosed with an STI after joining the trial according to postal test results and clinic records

Process outcomes

At 4 weeks:

- Attitudes towards partner notification
- Self–efficacy in telling a partner about an infection
- Self–efficacy in negotiating condom use
- Correct condom use self-efficacy
- Knowledge related to STIs

Additional data collected

- The proportion of participants in the intervention group who report reading ‘all’, ‘most’, ‘few’, or ‘none’ of the text messages
- The proportion of participants who reported that someone else read the messages (and if yes, how they felt about it)
- Contamination between intervention and control group (the proportion of intervention respondents who shared messages with other trial participants, and the proportion of control respondents who read other trial participants’ messages.)

4. Assessment of outcomes

Self-reported data

Self-reported outcome data is collected at 4 weeks and at 12 months. Hard copy questionnaires collecting outcome data are sent by post to participants. A URL link to a web-based data entry form is also sent to participants via text message and email. Participants can choose their preferred methods to submit outcome data.

Non-responders receive further contact by phone call, email, and text messages. Trial assistants collect outcome data by phone and record this on a hard copy data form. All hard copy data forms are entered into the online trial database. Where possible, discrepant data is verified with participants and corrected.

Objective data

Testing positive for chlamydia or gonorrhoea: at 12 months all participants are sent a self-test NAAT kit (urine for men, with additional pharyngeal and anal swabs for MSM and self-taken vulvo-vaginal swab for women). Additionally, for all participants, data on STI testing and results are collected from all recruiting clinics. Participants who self-report a positive diagnosis of chlamydia or gonorrhoea at 12 month follow-up are also asked to provide information on where they were tested. If a participant reports using a different service (a GP or a sexual health clinic other than that at which they were initially recruited), the service is contacted to verify the diagnosis.

Testing positive for any other STI: the data on STI testing and results collected from all recruiting clinics will include information on diagnoses of STIs other than chlamydia and gonorrhoea.

Sexually transmitted infection testing for self, prior to first sex with most recent new partner: the data on STI testing and results collected from all recruiting clinics will include information on any STI tests conducted during the study period. Only *testing* will be verified by clinic data – we will not be able to verify whether testing occurred *prior* to first sex with most recent new partner.

Clinic attendance by partner for treatment: recruiting clinics will provide data on whether trial participants' sexual partners' have attended the clinic for STI treatment after the participants' initial STI diagnosis. Not all clinics will collect sexual contact testing information.

The proportion of self-reported data that cannot be objectively verified will be reported in the main paper. Methods for handling missing data are described below.

Change in the primary objective during the conduct of the study

There has not been any change in the primary objectives of the study.

5. Study Populations

Participant characteristics

A trial flow diagram of the participants will be presented as recommend by CONSORT guidelines (Appendix 1). A table of the baseline characteristics of participants will be presented (Appendix 2).

Definition of populations for analysis

All analyses will be conducted according to randomised arm, regardless of whether participants received the allocated intervention, i.e. analyses will estimate the intention-to-treat effects.

Major protocol deviations

These will be reported in the results sections of the main trial paper.

If participants are randomised again in error less than 4 weeks after being randomised the first time they will be removed from the trial if allocated to both groups or retained as one participant if allocated to the same group twice.

If participants are randomised again in error more than 4 weeks after the first randomisation then the first randomisation will be retained and any subsequent randomisations will be deleted. The rationale for this is that participants are recruited by clinic staff when they are diagnosed with chlamydia or gonorrhoea or NSU. Participants being recruited and randomised again more than 4 weeks after a first randomisation is likely to be due to them being identified on a subsequent occasion as having chlamydia, gonorrhoea or NSU. i.e. in this circumstance being randomised more than once is contingent on the participant having a subsequent chlamydia/gonorrhoea/NSU diagnosis (the primary outcome) following the first randomisation.

6. Statistical Analysis

General statistical considerations

All statistical tests and confidence intervals will be 2-sided. Significance will be considered at the 0.05 level and confidence intervals will be at the 95% level. Statistical analysis will be performed using the current version of Stata software.

Assumptions about missing data

Data re assumed 'missing at random' (MAR). A MAR assumption assumes that missing data for participants that did not complete follow-up are similar to data from participants who completed follow-up, based on similar baseline covariates (i.e. that missingness is independent of the missing data) (1). We will conduct the primary analysis under a MAR assumption (conditionally on the adjustment variables in the model), then perform sensitivity analysis under different assumptions for the missing data, as explained below. In addition we will conduct a complete case analysis as a supplementary analysis.

Missing baseline covariates

The database requires all items on the baseline questionnaire to be submitted to randomise. Therefore, there will be no missing baseline covariates.

Missing primary outcome data

Missing primary outcome data will occur if:

1. participants do not return their completed STI self-test kit

   AND

2. no testing information is identified from clinic records (either because they did not test at the clinic they were recruited from or they tested at a different health service to the one that they were recruited from/ provided information about and this service did not provide data)

OR

testing information from clinic records showed they received a negative test result for chlamydia and gonorrhoea less than 12 months post-randomisation (i.e. so it is possible that they may have gone on to receive a positive test result during the follow-up period).

Primary Analysis

Analysis of the primary outcome

The primary outcome is binary and we will compare the cumulative incidence of chlamydia or gonorrhoea infection at one year in each group using logistic regression. We will use multiple imputation (MI) by chained equations (MICE). MICE accounts for the uncertainty created by missing data by generating plausible imputed data sets and then combining them. This aims to correct for any potential bias caused by missing data under the assumption that data are MAR. We will adjust the primary analysis regression for the following baseline covariates to improve the efficiency of the analysis and avoid chance imbalances (2):

- Age (continuous)
- Type of infection: Chlamydia/ gonorrhoea/ chlamydia and gonorrhoea/ NSU/ not known
- Sexuality: women who have sex with men (WSW)/ men who have sex with women (MSW)/ women who have sex with women (WSW)/ men who have sex with men (MSM)/ women who have sex with women and men (WSWM)/ men who have sex with men and women (MSMW)/ not stated.
- Ethnicity: White (White British, Other White background)/ Black (Black British – Caribbean, African, other)/ Asian (Asian- British Chinese, Indian, Pakistani, Bangladeshi, other)/ Mixed background/ other background/ not stated.

We will identify the key predictors of the outcome from the complete dataset without allocation groups. We will use forward stepwise regression of outcome on baseline (omitting treatment) with p.enter (0.05), p.exit(0.04) to identify any additional predictors from the available baseline variables:

- condom use at last sex
- condom use at last sex with someone new
- testing for self before last sex with someone new
- new partner tested before last sex with them
- # of partners in last 12 months
- # of partners in last 12 months
- Sex Female/ Male
- Age (continuous)
- Ethnicity:
    - White British/ Other White background
    - Black/Black British – Caribbean/African/other
    - Mixed background
    - Asian/Asian British – Bangladeshi/Chines/Indian/Pakistani/other
    - Other (please state)
- Age left full time education (16 or under,17 or over , still in full time education)
- Infection at baseline: chlamydia, Gonorrhoea, chlamydia and gonorrhoea, NSU, don't know
- Sexual orientation:
    - Women(W) who have sex with (S) men (M)
    - MSW
    - WSW

  - MSM
  - WSWM
  - MSMW
  - Not stated
- Provided a main email (yes/no)
- Provided an alternative mobile number (yes/no)
- Provided an alternative email (yes/no)
- Provided a mobile number of someone we can ask for your current contact details if we cannot contact you (yes/no)
- Provided an email of someone we can ask for your current contact details if we cannot contact you (yes/no)

We will use multiple imputation (MI) by chained equations (MICE) using the predictors of outcome identified in the baseline data (listed above) and in 4 week data (listed below) to impute 12 month outcome data:

- Took treatment (yes/no/unsure)
- Avoided sex for 7 days after treatment (yes/no/unsure)
- Condom use at last sex
- Number of partners since joining the trial (0/1/2+)

- *To what extent do you agree or disagree with the following: (measured on a scale of 1: strongly disagree to 5: strongly agree)*

  - If someone had a sexually transmitted Infection (STI they would know
  - STIs are rare
  - I can tell if someone has an STI
  - Most people who have an STI will tell their partner
  - It's my responsibility to tell a partner if I am diagnosed with a STI
  - If I tell my partner I have an STI, my partner would be glad I let them know
  - If it tell my partner I have an STI my partner would think badly of me

- *How easy or difficult would it be to: (measured on a scales of 1: very easy to 5: very difficult*)
  - Tell the last person you had sex with that you had a STI
  - Tell the last person you had sex with to get treatment
  - Tell a new partner you had an STI
  - Tell a new partner to get treated
  - Put a condom on
  - Keep a condom from drying out during sex
  - Keep a condom from breaking or coming off during sex
  - Keep a condom on while withdrawing the penis after sex
  - Keep a condom on from start to finish

- Imagine that you and your partner have sex but don't use condoms. You want to start using condoms. How easy or difficult would it be for you to tell your partner that you want to use condoms?
- Imagine that you are having sex with someone new. You want to use condoms. How easy or difficult would it be for you to tell them that you want to use condoms?

- Imagine that you are having sex with someone new. You want to use condoms. How easy or difficult would it be for you to tell them that you won't have sex unless you use condom?

We will report the adjusted odds ratios along with the 95% confidence intervals and p-values. Baseline, four week outcome data and randomisation arm treatment will be used as predictors in all imputation model.

Analysis of the secondary outcomes

The analysis of the secondary outcomes will be the similar to the analysis of the primary outcome. We will use MICE and estimate the difference between the groups using logistic regression for binary outcomes and report odds ratios with 95% confidence intervals and p-values. Regressions will be adjusted for the covariates.

Analysis of the process outcomes

The process outcome measures are comprised of multiple ordinal scales. We will refine these measures using latent variable modelling, which will produce a continuous score for each process outcome. We will use linear regression to test for a difference in mean scores between the arms. Regressions will be adjusted for the covariates.

Secondary Analyses

Complete case supplementary analysis

As a comparison to the primary imputation analysis, we will analyse the effect of the intervention on the primary outcome by including only complete primary outcome data in the analysis. We will use logistic regression adjusted for the covariates. We will report the adjusted odds ratios along with the 95% confidence intervals and p-values.

Subgroup analyses

Recognising that the trial is not powered to detect effect differences in subgroups, we will conduct exploratory subgroup analyses for the primary outcome to determine if the intervention effect varies by baseline characteristics. The subgroup analysis will be conducted on the MI dataset. The pre-specified subgroups are presented in Table 1. Across the pre-specified subgroups, we will assess heterogeneity of treatment effect with a test for interaction (3-7). Interaction test p-values will be presented but will be interpreted with caution, due to the exploratory nature, the multiple tests performed and the low power of the interaction test. We will estimate odds ratios along with 95% CIs for each subgroup. Intervention effect estimates by subgroups will be presented in a forest-type plot (Appendix 5). As this is an exploratory analysis of potentially influential characteristics, we will not hypothesise effect directions.

Table 1 Pre-specified subgroups

Variable	Categories
Age	16-19 20-24
Sex	Female Male
Sexual	men who have sex with men OR men who have sex with men AND women

orientation	men who have sex with women ONLY women who have sex with men OR women who have sex with men AND women women who have sex with women ONLY
Ethnic group	White British/other White background Black/Black British (Caribbean/African/Other) All other groups
IMD quintile	1 and 2 (most deprived) 3 4 and 5 (least deprived)

Age and sex are considered the two key subgroups and the analysis of these subgroups we be conducted first.

Analysis of additional data collected

The below data will be presented by arm, but we will not conduct a formal comparison between groups.

Contamination

We will assess the potential for contamination between intervention and control group. At 1 year follow-up participants are asked the following:

Do you know anyone else who took part in the study?

If yes:

Did they read the messages we sent you?

Did you read the messages we sent them?

Based on these questions, we will calculate the proportion of intervention respondents who shared messages with other trial participants, and the proportion of control respondents who read other participants' messages.

Intervention dose

To estimate the intervention dose received, will present the proportion of participants in the intervention group who report reading 'all', 'most', 'few', or 'none' of the intervention messages. We will also report the proportion of messages successfully sent from SMS gateway.

Participants' feelings regarding others reading their messages

For participants in the intervention group who report that someone else read the messages, we will present the proportion who felt 'happy', 'unhappy' and 'unsure' about it.

7. Primary endpoint definitions

Primary outcome

The primary outcome is cumulative incidence of chlamydia and gonorrhoea infection at one year assessed by NAAT tests: urine for men (with additional pharyngeal and anal swabs for MSM) and self-taken vulvo-vaginal swab for women. This data is gathered via postal test and clinic data.

Defining one year

A positive test result for chlamydia or gonorrhoea from any test kit returned while follow-up is active will result in that participant being categorised as primary outcome complete.

One year from the date of randomisation, participants are sent a self-completion STI test through the post. Participants who do not return the completed test kit will receive up to 5 further test kits up until 52 weeks + 10 weeks post-randomisation.

Recruiting clinics will report data on STI tests completed and STI test results for all participants recruited from their clinic. Any positive chlamydia or gonorrhoea test result identified through clinic data while follow-up is active will result in that participant being categorised as primary outcome complete. [N.B. We continue to send postal tests to participants who we identify as testing positive through clinic data.]

Defining infection

Infections will only be identified through laboratory testing.

Coding of the primary outcome

The primary outcome will be coded as follows:

Outcome*	**Postal test / clinic test**	**Notes**
= 1	postal==1 OR clinic==1	Clinic test ordered after randomisation and reported while follow-up is active Postal test up to 52 weeks + 10 weeks
= 0	postal==0 AND clinic==0	
= 0	postal==0 AND self-report==0 AND clinic==.	
= 0	postal==0 AND self-report==1 AND clinic==.	If a participant tests negative on postal test, self-reports that they've tested positive since joining study, but we are unable to find a test result from clinic records, they will be negative for the outcome (postal test trumps self-reported positive not confirmed from clinic data)
= .	postal==. AND clinic==0	Clinic test ordered within 12 months of randomisation (result may have been returned after 12 months)
= 1	(postal==0 OR postal==.) AND clinic==1	Clinic test ordered after randomisation and reported while follow-up is active
= .	postal==. & clinic==.	

* 1 = positive, 0 = negative, . = missing

Verifying infection through clinic data

If a participant reports having received a positive test result for chlamydia or gonorrhoea since joining the trial, they are asked to provide information on where they got tested and where they got treated. If the participant reports having used a different service to where they were initially recruited from, we will contact the service reported to verify if the participant received a positive diagnosis for chlamydia and gonorrhoea between the date of randomisation and the date when follow-up was collected.

We will use the following identifiers to locate participants' health records:

Service	Identifiers	
GUM clinics	Mobile phone number AND Gender	Name AND Date of birth
GP surgeries	Name AND Date of birth	

8. Interpretation of outcomes

If the trial demonstrates a reduction in cumulative incidence of gonorrhoea/chlamydia, we will interpret this as the trial having important public health impact.

9. Adverse events

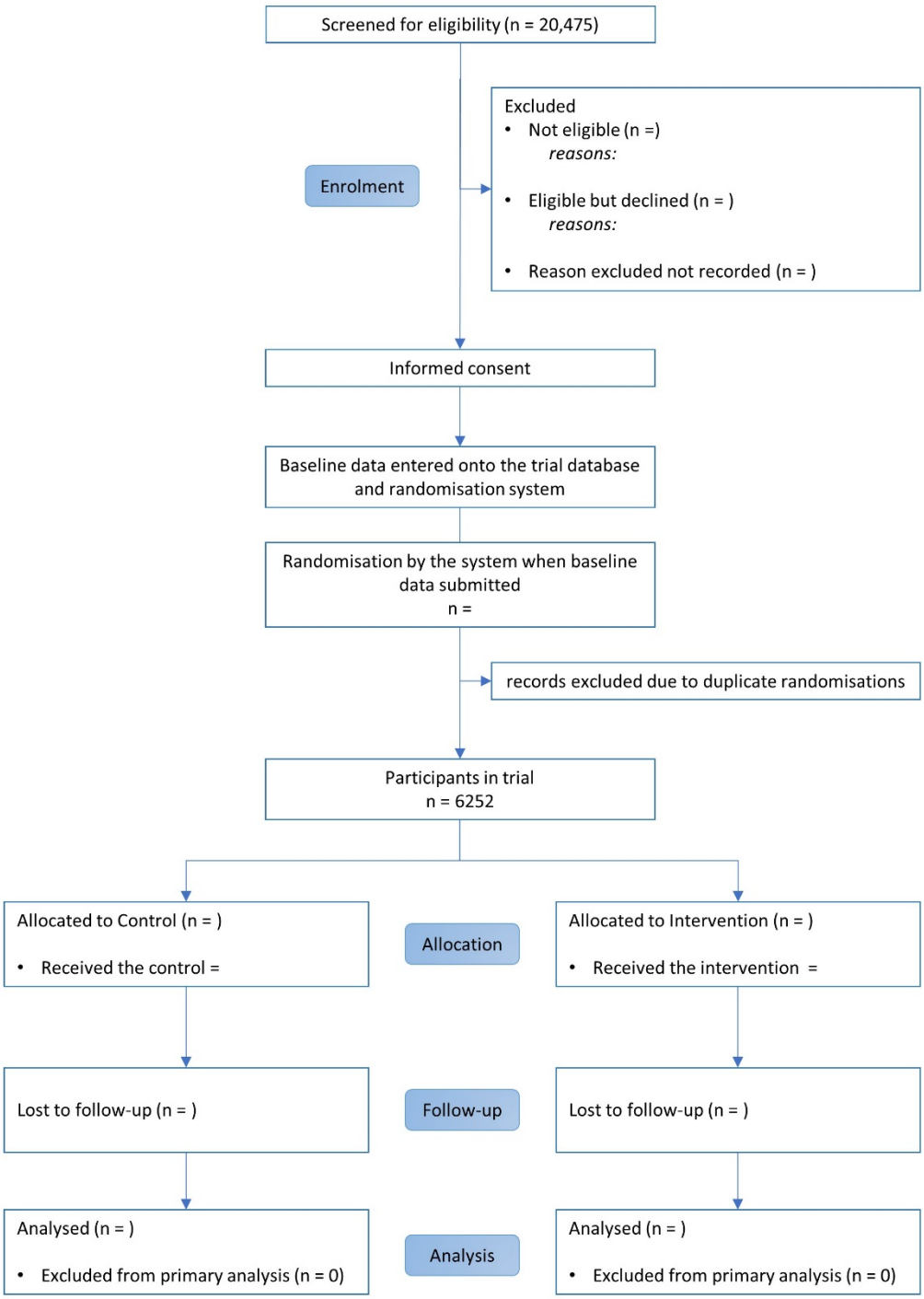
At 12 months we collect data on involvement in car accident where participant was the driver in the last year, and experience of partner violence in the last year. We will present the proportion of participants reporting each adverse outcome by intervention arm and the p-value (calculated by a Chi^2 test or Fisher's exact if less than 5 events).

References

1. O'Kelly M, Ratitch B. Clinical trials with missing data: A guide for practitioners. Chichester, UK: John Wiley & Sons, Ltd; 2014.
2. European Medicines Agency, Committee for Medicinal Products for Human Use. Guideline on adjustment for baseline covariates in clinical trials. London, UK.2015.
3. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Statistics in medicine. 2002;21(19):2917-30.
4. Kasenda B, Schandelmaier S, Sun X, von Elm E, You J, Blumle A, et al. Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications. BMJ (Clinical research ed). 2014;349:g4539.
5. Gabler NB, Duan N, Liao D, Elmore JG, Ganiats TG, Kravitz RL. Dealing with heterogeneity of treatment effects: is the literature up to the challenge? Trials. 2009;10:43.
6. Kent DM, Rothwell PM, Ioannidis JPA, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. Trials. 2010;11:85.
7. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. Journal of clinical epidemiology. 2004;57(3):229-36.

Appendices

Appendix 1 Trial flow diagram



Appendix 2 Baseline characteristics

	Control N = % (n)	Intervention N = % (n)	All participants N = 6,252 % (n)
Sexual behaviour			
condom use at last sex			
condom use at last sex with someone new			
testing for self before last sex with someone new			
new partner tested before last sex with them			
# of partners in last 12 months			
Sex			
Female			
Male			
Age group			
16-19			
20-24			
Ethnicity			
White British			
Other White background			
Black/Black British - Caribbean			
Black/Black British - African			
Mixed background			
Other Black background			
Asian/Asian British - Bangladeshi			
Asian/Asian British - Chinese			
Asian/Asian British - Indian			
Asian/Asian British - Pakistani			
Other Asian background			
Other (please state)			
Age left full time education			
16 or under			
17 or over			
Still in full time education			
Infection at baseline			
Chlamydia			
Gonorrhoea and Chlamydia			
Gonorrhoea			
NSU			
Don't know			
Sexual orientation			
WSM			

MSW			
WSW			
MSM			
WSWM			
MSMW			
Not stated			

Appendix 3 Primary, secondary and process outcomes

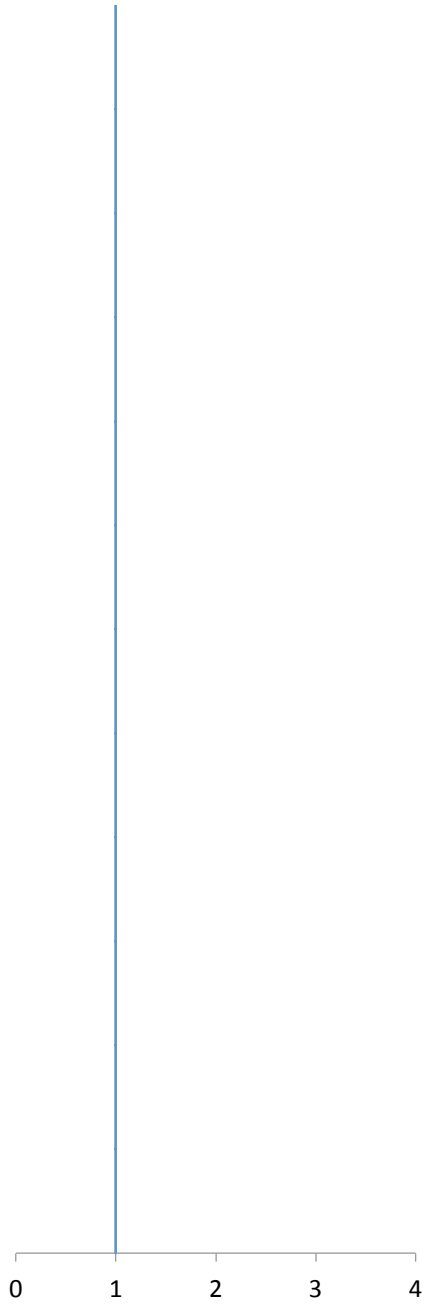
	Control N = % (n)	Intervention N = % (n)	risk diff. (95% CI, p-value)	aOR* (95% CI, p-value)
Primary outcome (at 12 months)				
Cumulative incidence of chlamydia or gonorrhoea				
Secondary outcomes (at 4 weeks)				
Correct treatment				
Notified partner				
Partner known to have attended clinic for treatment				
Condom used at last sex, if had a partner since joining trial				
Process outcomes (at 4 weeks)				
Attitudes towards partner notification				
Self–efficacy in telling a partner about an infection				
Self–efficacy in negotiating condom use				
Correct condom use self-efficacy				
Knowledge related to STIs				
Secondary outcomes (at 12 months)				
STI after joining the trial				
Condom used at last sex				
Number of sexual partners in last 12 months				
0				
1				
2 or more				
Sex with someone new since joining trial				
Condom use at first sex with most recent new partner				
STI testing prior to sex with new partner				

Most recent new partner tested for STIs prior to sex				
Experience of partner violence in the past 12 months				
Car accident where participant was driver in past 12 months				

**adjusted for baseline covariates associated with outcome*

Appendix 4 Primary outcome by pre-specified subgroup

Subgroup	Cont.	Int.	OR (95% CI) Interaction test p-value
Sex			
Female			
Male			
Age group			
16-19			
20-24			
Sexual orientation			
MSM OR MSM+W			
MSW ONLY			
WSM OR WSM+W			
WSW ONLY			
Ethnicity			
White British/Other White background			
Black/Black British (Caribbean/African/Other)			
All other groups			
IMD quintile			
1 and 2 (most deprived)			
3			
4 and 5 (least deprived)			



Appendix 5 Additional data

	Control N = % (n)	Intervention N = % (n)
Number of text messages read		
All		
Most		
Few		
None		
Someone else read messages sent to participant		
Yes		
If yes, how participant felt about it		
Happy		
Unhappy		
Unsure		
Participant knew someone else in the study		
Yes		
If yes, • **someone else read the messages sent to participant**		
Yes		
• **participant read messages send to someone else**		
Yes		