Statistical Analysis Plan (SAP)

Title	A randomized experiment of malaria diagnostic testing and conditional subsidies to target ACTs in the retail sector: the TESTsmART trial AIM 1
CRU/Department/Division/Center	Duke Global Health Institute Clinton Health Access Initiative (CHAI) Moi University
IRB Number	TESTsmART Pro00100425
NCT Number	NCT03810014
Investigators:	
Lead Investigator	Wendy Prudhomme O'Meara
Mentors	
Co-authors (if known)	Theodoor Visser Jeremiah Laktabai
Biostatistician(s)	Ryan Simmons Yunji Zhou
Supervising Biostatistician	Elizabeth Louise Turner
Original Creation Date	08/16/2019
Version Date	11/01/2019
Project Folder Location	[redacted]
Project Goal(s)	Manuscript

Investigator Agreement

☐ All statistical analyses included in an abstract or manuscript should reflect the work of the biostatistician(s) listed on this SAP. No changes or additional analyses should be made to the results or findings without discussing with the project biostatistician(s).

☐ All biostatisticians on this SAP should be given sufficient time to review the full presentation, abstract,

	manuscript, or grant and be included as co-authors on any abstract or manuscript resulting from the analyses. ☐ Publications resulting from this SAP are supported in part by the Duke CTSA and must cite grant number UL1TR002553 and be submitted to PubMed Central. ☐I have reviewed the SAP and understand that any changes must be documented. *Acknowledged by:* Click or tap here to enter text. *Date:* Click or tap to enter a date.
Activity Log	August 21, 2019: Reviewed by Liz Turner August 23, 2019: Updated by Yunji Zhou August 27, 2019: Reviewed by Wendy Prudhomme O'Meara, Liz Turner, Ryan Simmons August 30, 2019: Updated by Yunji Zhou September 04, 2019: Reviewed by Ryan Simmons September 05, 2019: Checklist filled in by Yunji Zhou September 09, 2019: Randomization procedure updated by Yunji Zhou using information from Sarah Laing October 11, 2019: Hypothesis testing procedures updated by Yunji Zhou October 22, 2019: Reviewed by Ryan Simmons October 23, 2019: Reviewed by Liz Turner

1 Study Overview

In response to the high cost of artemisinin-combination therapy (ACT), publicly-funded, retail-sector ACT subsidies were adopted in many malaria-endemic countries. Declining prices of ACTs create a trade-off between access and targeting; lower prices improve uptake of effective therapies by those with malaria but also increase inappropriate use by those without malaria. Curbing inappropriate use and targeting ACTs to malaria cases requires parasitological diagnosis which is virtually absent in the retail sector. It is estimated that at least two thirds of ACTs purchased over-the-counter are consumed by individuals without malaria. Inappropriate use leads to wastage of public funds and prime conditions for the spread of drug resistant parasites, which could dramatically increase global mortality from malaria. Targeting subsidized ACTs to individuals with parasitologically-confirmed malaria would significantly contribute to the sustainability and cost-effectiveness of retail subsidies as well as safeguard the future efficacy of these key drugs. We propose an innovative conditional subsidy approach that links the ACT subsidy to the results of

a rapid diagnostic test (RDT), allowing the subsidy to be targeted only to parasitologically-confirmed malaria cases. This price differential makes information from a test more valuable and could drive appropriate consumption while reducing costs.

This study will be an individually-randomized 2X2 factorial trial and clients will be allocated, in equal proportions, to one of four treatment arms (210 individuals per arm). Each of the four treatment arms combines two factors, each of which have two levels, hence the choice of a 2x2 factorial design. Those two factors are:
1. An offer of a malaria RDT offered at one of two price levels-USD $0.40 (0% subsidy) or USD $0.20 (50% subsidy)
2. An offer of a discounted ACT conditional on a positive test result at one of two price levels- USD $0.40 (67% subsidy) or USD $0.00 (100% subsidy).

The intervention is designed to ensure fixed price point for consumers. Individuals who choose not to be tested for malaria, or who test negative for malaria, may purchase an ACT at the unsubsidized price of approximately USD $1.20. Note that when referring to the two factors in the text below, we will use the terms "RDT subsidy intervention" and "conditional ACT subsidy intervention".

1.1 Study Aims

The goal of this study is to identify the combination of testing subsidies and conditional ACT subsidies that maximizes uptake of testing within specific budget constraints. Several studies, including our own, have shown that uptake of testing and ACT treatment are both sensitive to price. However, very little is known about how these prices should be related in order to maximize appropriate behavior and what effect conditional subsidies may have on treatment decisions. We will use an individually-randomized experiment to determine how different combinations of subsides, allocated between testing and treatment, affect the decision to be tested for malaria before treatment among clients seeking care in the retail sector.

1.2 Study Hypotheses

1.2.1 Primary Hypotheses

RDT uptake (i.e. the proportion of clients who are tested with an RDT) will increase when the price of the RDT is reduced from $0.40 (0% subsidy) to $0.20 (a 50% subsidy), averaged over the ACT price levels.

RDT uptake will increase when the price of the ACT conditional on a positive test is reduced from $0.40 to $0 (67% subsidy vs 100% subsidy), averaged over the RDT price levels.

1.2.2 Descriptive Analysis

Adherence to RDT test results consists of ACT purchasing behaviors among the untested, test negative, and test positive participants. Descriptive statistics of adherence conditional on the ACT price levels (67% subsidy vs 100% subsidy) will be generated to look for possible patterns of ACT purchasing behaviors in different scenarios. We will also assess the effect of seriousness of disease and duration of symptoms on adherence to RDT test results as a sub-analysis.

2 Study Population

The study population will be any individual coming to a participating medicine retail outlet to purchase medicine for an acute, malaria like illness. Ten randomly selected retail medicine outlets in western Kenya will be selected for recruitment, which will happen on random days of the month to avoid individuals seeking to be recruited.

840 participants across 10 drug retail outlets in the study area will be recruited and randomly assigned in equal numbers to each of four arms using secret scratch cards. Children are eligible to participate if they are present with a parent or guardian. Inclusion of children ensures that we get a comprehensive picture of malaria infection and treatment decisions across all age groups. Children are at higher risk of malaria infection and disease which increases their importance to the study. They will also benefit from the opportunity to receive diagnostic testing immediately at the retail outlet, and they, along with their parent or guardian, can use the test results in deciding what drugs to purchase or other treatment actions to take. We will attempt to recruit all eligible participants who attend the outlet on a recruitment day. This will ensure that our sample is representative of the population seeking care in retail outlets in terms of age, gender and ethnicity. We do not have targets for recruitment based on gender, age or ethnicity.

2.1 Inclusion Criteria

- Participants with fever or history of fever or malaria like illness
- Individual with malaria-like illness must be present at recruitment
- Older than one year of age

2.2 Exclusion Criteria

- Any individual with signs of severe illness requiring immediate referral
- Individuals who have taken an antimalarial in the last seven days, including for the current illness
- Individuals who already have a prescription from a facility or medical provider, regardless of whether they have documentation of a test
- Pregnant women will be enrolled and offered an RDT, but will be advised to seek treatment through a health care provider.

2.3 Randomization Procedure

Each client who consents to be part of the study will receive a scratch card that will indicate their intervention arm (i.e. the combination of prices at which they will be offered an RDT and a conditional ACT). We will stratify on retail outlet to ensure that each of the 10 outlets has a similar number of clients who are assigned to the four study arms (approximately 21 per arm, per retail outlet for a total sample of 210 per arm). Due to the nature of the interventions, it is not possible to blind participants and research assistants to the allocation received. Study statisticians will be blinded during the analysis phase. Scratch cards were allocated in batches of 100 or 60 (25 of each or 15 of each arm) and shuffled thoroughly. Batches were assigned to specific outlets based on sales volume (i.e., higher-volume outlets received larger batches of 100 cards, smaller outlets received batches of 60 shuffled cards). Participants were asked to select a card from the shuffled stack. Cards were replenished when ten were remaining.

2.4 Sample Size Justification

Our sample size calculations are based on the expected changes in testing uptake with each of the price changes for the two commodities (RDT and ACT). We expect that reducing the price of the RDT from $0.40 USD to $0.20 will increase the percentage of clients who choose to get tested by 15 percentage points. We also expect that reducing the ACT price, conditional on a positive test result, from $0.40 to $0.00 will increase the uptake of testing by 10 percentage points (see Table 1, which is a copy of Table 2 in the Research Strategy).

As is commonly done for 2X2 factorial trials that seek to separately test for main effects of each of the two factors, our study is powered to detect the effects of the price reductions of each commodity (RDT and ACT) independent of the price of the other commodity. That is, for the purposes of our power calculations we assume no statistical interaction between the two commodities because our previous evidence suggests that any interaction effects between the prices in the two commodities are likely to be small and therefore not large enough in terms of public health importance to choose to power on such interactions [1]. That is, while we expect the uptake of testing to be sensitive to the price of both the RDT and the ACT, we do not expect that price changes in one commodity will have a different effect on testing uptake depending on the price of the other commodity.

ACT Price to client if he/she tests positive for malaria (Price for RDT negative or untested client = $1.20)	RDT Price to client I. $0.20 USD (50% subsidy)	II. $0.40 USD (No Subsidy)	Overall Proportion Tested*
I. $0 USD (100% subsidy)	85%	70%	77.5%
II. $0.10-0.40 USD (67% subsidy)	75%	60%	67.5%
Overall Proportion Tested*	80%	65%	

* Assuming the same sample size in each of the 4 study arms

Table 1: Aim 1 2x2 factorial study design showing the four arms and the assumed level of testing in each arm. The 67% ACT subsidy-level was chosen to match the original AMFm prices in Kenya. We do not test the non-subsidized ACT price because our aim is to identify an ACT discount level that can incentivize individuals to get tested for malaria. Input from policy-makers suggests that fully-subsidized, free RDTs may not be acceptable from a programmatic standpoint (T.Visser, *pers. comm.*). Therefore, we do not include a 100% RDT-subsidy level.

* Assuming the same sample size in each of the 4 study arms

** At current market prices, this subsidy level gives the following prices to the consumer: 6 tablets@$0.10; 12 tablets@$0.15; 18 tablets@$0.20; 24 tablets@$0.40.

Using a standard approach based on a two-sample Z-test for a comparison of proportions for each of the two main effects at a two-tailed 5% Type 1 error rate (alpha), our sample size should provide 90% power to detect each of our two comparisons of interest (i.e. two main effects). To determine the overall sample size of the study, we calculated the sample size needed for each of the two main comparisons of interest, and chose the larger overall sample required (see Table 2). Specifically, because we estimated that we needed a total sample size of 838 participants (i.e. of 418 participants at each of the two conditional ACT price levels) to test for the main effect of the conditional ACT subsidy price levels compared to a total sample size of 374 (i.e. of 187 at each of the two RDT price levels) to test for the main effect of the RDT price levels, we therefore needed to plan the study by using the more restrictive sample size in order to be able to address both study goals. Given that the trial will be conducted within 10 drug retail outlets, there could be clustering of outcomes due to some participants being enrolled from the same shops. Based on our ongoing work in the region, we expect such clustering to be minimal with an intra-class correlation coefficient of at most 0.008. Even with this level of clustering, we would still have at least 80% power to detect our expected effect sizes.

Although a priori we do not anticipate a significant interaction effect, we will perform an additional test to evaluate that assumption. If, as expected, a significant interaction effect is not detected, we will perform our planned tests of the main effects of each of the factors (see the first column in Table 2). Thus, we will have conducted three total significance tests, and will adjust for familywise Type 1 error rate using the Bonferroni correction (alpha=0.05/3=0.0167). As shown in the last column of Table 2, we would still have 80% power to detect the expected effect sizes with the total sample size of 838, given that Bonferroni correction is a conservative method of familywise error rate adjustment. We will use different error rate allocation methods to gain more power in statistical analysis. In the unlikely case that a significant interaction effect exists, we will instead test the conditional effects of each of factor and adjust for Type I error accordingly (as described in section 4.2; see Table 3).

Table 2: Expected effect sizes for the main comparisons of interest and sample size required for 80%,

Description of Comparison	**Expected Increase in Testing Uptake**	**Total sample size required for 90% power (alpha=0.05)**	**Total sample size required for 85% power (alpha=0.0167)**	**Total sample size required for 80% power (alpha=0.0167)**
Effect of reducing conditional ACT price from $0.40 (67% subsidy) to $0.00 (100% subsidy)	10 percentage points (from 67.5% to 77.5%)	838	936	832
Effect of reducing RDT price from $0.40 (no subsidy) to $0.20 (50% subsidy)	15 percentage points (from 65% to 80%)	374	414	370

85%, and 90% power to detect those effect sizes (with two-tailed Type 1 error set at 0.0167 and 0.05, and using standard methods for two-sample comparison of proportions).

2.5 Data Acquisition

Study design	2x2 factorial individually randomized controlled trial
Data source/how the data were collected	Data was be collected by a face-to-face interview at enrolment and follow-up. Data was be entered into ODK on 11 separate tablets and combined into a CSV file on a computer.

Contact information for team member responsible for data collection/acquisition	Stephen Karuru [redacted] Emmah Kimachas [redacted] Josephine Malinga [redacted]
Date or version (if downloaded, provide date)	
Data transfer method and date	Box
Where dataset is stored	[redacted]

3 Outcomes, Exposures, and Additional Variables of Interest

All outcomes of interest are measured at the individual level and will be aggregated by each of the four groups. Importantly, not all outcomes will be defined on all individuals because some are "conditional" outcomes e.g. conditional on having taken a test. As a consequence, although the study design is a randomized experiment, we recognize that the analysis of some outcomes may be more subject to potential confounding than that of outcomes that are defined for all individuals.

3.1 Primary Outcome(s)

Outcome	**Description**	**Variables and Source**	**Specifications**
Testing uptake	Whether client purchases an RDT from the shop before purchasing a drug		Binary: yes/no

3.2 Secondary Outcome(s)

Outcome	**Description**	**Variables and Source**	**Specifications**
ACT purchase among untested	Whether untested client chooses to take an ACT		Binary: yes/no
ACT purchase among test positive	Whether client chooses to take an ACT if test positive		Binary: yes/no

ACT purchase among test negative	Whether client chooses to take another drug or no drug if test negative		Binary: yes/no
Appropriate ACT use	Taking ACT if positive or not taking ACT if negative among the participants who had an RDT test		Binary: yes/no
Targeted ACT use	Taking ACT if positive or not taking ACT if negative among all participants, regardless of RDT test uptake		Binary: yes/no

3.3 Additional Variables of Interest

Variable	**Description**	**Variables and Source**	**Specifications**
Patient gender			Binary: Male/Female
Patient age			Continuous
Household size			Continuous
Highest level of education completed			Categorical: None, Pre-primary, Some Primary, Finished Primary, Some Secondary, Finished Secondary, Some Post-secondary, Finished Post-secondary, Other
Occupation			Categorical: Agricultural, Paid employee, Self-employed (Not in agriculture), Informal employment, Student, Homemaker/Housewife, Not available to work, Unemployed, Others
Wealth category			Binary: under poorest 40th centile or not

Seriousness of disease			Categorical: Not very serious/minor, Moderate, Very serious
Duration of symptoms			Continuous

4 Statistical Analysis Plan

4.1 Demographic and Clinical Characteristics ("Table 1")

We will use principle components analysis to define a wealth index. No significance tests will be performed to test for differences at baseline. Descriptive statistics for continuous variables will include the mean, standard deviation, median, range and the number of observations. Categorical variables will be presented as numbers and percentages.

4.2 Analyses Plan for Aim 1

As per the assumptions used in our sample size calculations (see section 2.4), we anticipate there will be no significant interaction effect between RCT and conditional ACT subsidy levels. However, we will evaluate this assumption to protect our analysis from an unexpected interaction effect. Since the assumption of no interaction effect was in terms of risk differences (see Table 1), the test for interaction will also be performed on the absolute scale, though the primary effect measure of interest is the relative risk ratio (RR) and the test results of interaction effect may differ on different scales. If there is no significant interaction effect as expected, we will report the average effect for each of RDT subsidy and conditional ACT subsidy on the proportion of testing uptake on the relative scale. In the unlikely case that a significant interaction effect is detected, we will report the effect of each factors conditional on the other on the relative scale (see Table 3). In either case, adjustment for familywise Type 1 error rate is necessary for multiple hypothesis tests. In our sample size justification (section 2.4) we used a conservative Bonferroni correction; however, in practice, we will use an alpha allocation strategy that splits a 5% Type 1 error rate across our tests in a manner that reflects the a priori importance of each test and provides more power to detect the expected effect sizes. As mentioned previously, the primary effect measure is on the relative scale. However, due to the nature of the model, interpretable relative risk ratios cannot be derived for the average effects. Instead, approximation of the relative risk ratios on log scale will be reported. Different coding methods will not solve this problem, but we choose to use effect coding over dummy coding because effect coding, with the orthogonality property, allows for interpretable main effects

regardless of the presence of an interaction effect. Details of the statistical analysis are as follows.

In our analysis, we will fit a modified Poisson regression model with log link to estimate risk ratios as the primary effect measure (i.e. an effect measure on the relative scale, see Model 1) [2, 3]. Such an approach assumes a Poisson distribution for the binary outcome and then 'fixes' the estimated standard errors to correct for model misspecification (i.e. using a Poisson rather than binomial model). It is the preferred modelling approach to estimate RRs as it avoids some of the convergence issues that may be encountered when using the alternative log-binomial regression approach whilst maintaining good statistical properties. In addition, we will fit a modified Poisson regression model with identity link to estimate risk differences as the secondary effect measure (see Model 2), since the assumption of no interaction effect is made on the absolute scale. However, the optimization decision will be based the primary model (i.e. on the relative scale). All analyses will be based on the intention-to-treat principle whereby all clients will be included in the analysis irrespective of whether they complied with the intervention.

The models will have the following form:

Model 1: $\log(E(Y_{ij}))= \beta_0 + \beta_1 RDT\_SUBSIDY\_50\%_{ij} + \beta_2 ACT\_SUBSIDY\_100\%_{ij} + \beta_3 RDT\_SUBSIDY\_50\%_{ij} * ACT\_SUBSIDY\_100\%_{ij} + \mathbf{w}_{i'j}\boldsymbol{\alpha}$

Model 2: $E(Y_{ij})= \beta_0 + \beta_1 RDT\_SUBSIDY\_50\%_{ij} + \beta_2 ACT\_SUBSIDY\_100\%_{ij} + \beta_3 RDT\_SUBSIDY\_50\%_{ij} * ACT\_SUBSIDY\_100\%_{ij} + \mathbf{w}_{i'j}\boldsymbol{\alpha}$

where Y_{ij} is an indicator of whether client j in outlet i (j=1,…,84; i = 1…,10), chose to get tested for malaria in the retail outlet (=1 if tested, 0 otherwise) and $E(Y_{ij})$ is its expected value. RDT_SUBSIDY_50%ij is an indicator for whether the client received the lower RDT price of $0.20 (=1 if they received the offer of an RDT at $0.20, =-1 if they received the offer of an RDT at $0.40) and ACT_SUBSIDY_100%ij is an indicator for whether the client received the higher ACT subsidy (=1 if ACT price conditional on a positive test is $0, =-1 if ACT price conditional on positive test is $0.40). Therefore, RDT_SUBSIDY_50%ij * ACT_SUBSIDY_100%ij represents an interaction term for clients who receive both the higher RDT and ACT subsidies or both the lower RDT and ACT subsidies (=1 in two situations: a. RDT is offered at $0.20 and ACT is offered at $0 conditional on a positive test or b. RDT is offered at $0.40 and ACT price conditional on positive test is $0.40, =-1 otherwise). $\mathbf{w}_{ij}$ is a vector of potential confounder variables (e.g., age, gender, wealth, education) to account for possible imbalances between study arms. To account for possible clustering due to multiple individuals being randomized within 10 different (randomly-sampled) outlets we will use a generalized estimating equations (GEE) approach with

exchangeable working covariance matrix, robust standard errors (to correct for model misspecification due to specifying a Poisson distribution) and finite-sample correction (to obtain unbiased standard errors estimates using the Kauermann-Carroll method)[4, 5]. Regression diagnostics, including residual plots, will be used to verify model assumptions. Because all secondary outcomes are binary, we will use the same modelling approach to compare these outcomes.

In the primary model (Model 1), β_0 represents the overall mean of log-mean level (i.e. log-proportion) of testing uptake across all treatment arms. β_1 represents the deviation of the log-proportion of testing uptake from the overall mean when reducing the price of the RDT from \$0.40 to \$0.20, averaged across ACT subsidy levels, while β_2 represents the deviation of the log-proportion of testing uptake from the overall mean when reducing the price of the ACT, conditional on a positive test, from \$0.40 to \$0, averaged across RDT subsidy levels. β_3 represents the interaction effect between the two price changes: it is the additional effect on log-proportion of testing uptake when both commodities are offered at the lower price level or higher price level, and also the negative additional effect if only one of RDT and ACT is offered at a lower price level. Relative risk ratio for the average effect of each of the price reductions cannot be obtained due to the nature of the model. However, an approximation of it can be obtained by exponentiating the parameters and deriving corresponding 95% confidence intervals.

In the secondary model (Model 2), β_0 represents the overall mean of the proportion of testing uptake across all treatment arms. β_1 represents the deviation of the proportion of testing uptake from the overall mean when reducing the price of the RDT from \$0.40 to \$0.20, averaged across ACT subsidy levels, while β_2 represents the deviation of the proportion of testing uptake from the overall mean when reducing the price of the ACT, conditional on a positive test, from \$0.40 to \$0, averaged across RDT subsidy levels. β_3 represents the interaction effect between the two price changes: it is the additional effect on the proportion of testing uptake when both commodities are offered at the lower price level or higher price level, and also the negative additional effect if only one of RDT and ACT is offered at a lower price level. Absolute risk differences and corresponding 95% confidence intervals can be obtained by directly evaluating the model parameters.

Our two primary goals are to determine whether RDT uptake increases significantly when the price of the RDT is reduced from \$0.40 (no subsidy) to \$0.20 (a 50% subsidy), averaged over the ACT price levels (i.e. the main effect), and when the price of the ACT conditional on a positive test is reduced from \$0.40 to \$0 (67% subsidy vs 100% subsidy), averaged over the RDT price levels (i.e. the main effect). That is, we wish to test for the main effects of these two commodities. These can be evaluated by testing the null hypotheses H_0: $2\beta_1=0$ and H_0: $2\beta_2=0$, with 2% Type 1 error rate respectively. Although, a priori, we hypothesized no statistical interaction

on absolute scale in our sample size calculations, a secondary aim is to determine whether the effect of each of the price changes depends on the price of the other commodity. This can be tested by evaluating the null hypothesis H_0: $\beta_3=0$ in model 2 on the absolute scale with 1% Type 1 error rate. If this effect is statistically different from zero, we will report the effect of the price reductions of each commodity conditional on the price of the other commodity, with a Type 1 error rate of 1% respectively. If, however, as we expect, there is no statistically significant or scientifically meaningful interaction (i.e. β_3 is close to zero), we will estimate the average effects from a model with interaction to match the study design with a Type 1 error rate of 2% respectively. The allocation of Type 1 error rate is based on the a priori importance of the hypothesis tests and provides more power than the conservative strategy employed in our sample size calculations (section 2.4). Contrasts of interest are summarized in Table 3.

In the unlikely case that the modified Poisson models fail to converge, we will use a logit-link in order to estimate odds ratios as the measure of effect. As mentioned before, we will additionally provide estimates of absolute effects (risk differences) in order to provide an intuitive measure of the potential public of the interventions. Such an approach is recommended by the CONSORT statement on reporting of cluster randomized trials [6]. These risk differences will be estimated by changing the log-link to an identity link. If convergence is not achieved, we will consider an alternative specification using a normal distribution with identity link and the same approach of using robust standard errors to account for model misspecification. We will fit all models using Stata version 15 [7].

Table 3: Contrasts of interest	**Population**	**Contrast**	**Alpha**
Interaction effect			
(0a) The interaction effect of RDT subsidy and conditional ACT subsidy.	Full cohort	$2\beta_3$ in Model 2	0.01
Primary outcomes (given no significant interaction effect)			
(1a) Effect of RDT subsidy on the uptake of testing, averaged over the ACT price levels.	Full cohort	$2\beta_1$ in Model 1	0.02
(1b) Effect of conditional ACT subsidy on the uptake of testing, averaged over the RDT price levels.	Full cohort	$2\beta_2$ in Model 1	0.02
Primary outcomes (given strong interaction effect)			
(2a) Effect of RDT subsidy on the uptake of testing, given no conditional ACT subsidy.	No conditional ACT subsidy	$2\beta_1$-$2\beta_3$ in Model 1	0.01

(2b) Effect of RDT subsidy on the uptake of testing, given conditional ACT subsidy.	Conditional ACT subsidy	$2\beta_1+2\beta_3$ in Model 1	0.01
(2c) Effect of conditional ACT subsidy on the uptake of testing, given no RDT subsidy.	No RDT subsidy	$2\beta_2-2\beta_3$ in Model 1	0.01
(2d) Effect of conditional ACT subsidy on the uptake of testing, given RDT subsidy.	RDT subsidy	$2\beta_2+2\beta_3$ in Model 1	0.01

5 Addendum for Additional Analyses

5.1 Site for Aim 1

We have chosen to conduct Aim 1 only in Kenya for the following reasons:

- RDTs are not currently used in the retail sector in Kenya. Therefore, individuals presenting to an outlet for their illness would not have a prior expectation about receiving a test or have made a prior decision about whether they would like to receive a test. Customers will not have experience with testing in a retail outlet so they will not have preconceived ideas about how much they should pay. We would like to understand decision-making in the presence of a new tool (RDT) in a 'naïve' system.
- Kenya has maintained stable retail ACT prices for nearly five years. Although these prices are subsidized they are still likely to be a significant expense for households in the area. This offers two advantages – first, individuals know how much they expect to pay for an ACT before coming to the outlet and second, the ACT discounts we will offer (conditional on a positive test) are expected to be attractive to the customer (ie non-negligible).

We expect results from Kenya to be generalizable to other contexts because:

- Most countries have not rolled out RDTs in retail shops which makes them comparable to Kenya (as opposed to Nigeria).
- We have designed our subsidy levels with attention to the ratio of the ACT price to the RDT price, which in the case of a positive test ranges from 0 (positive test, 100% ACT subsidy) up to 2 (positive test, 67% ACT subsidy, 50% RDT subsidy). Studies of RDTs in the retail sector suggest that the relative price of these two commodities has a strong influence on uptake. The subsidy levels can therefore be translated to other contexts, including Nigeria (where Aim 2 will be conducted in addition to Kenya) based on relative price of the RDT and ACT (rather than absolute subsidy amount).

5.2 Intervention Model (Study Design)

We chose a factorial design for Aim 1, because our goal was to identify the combination of ACT and RDT subsidies that maximizes testing within a given budget. We do not need a fully articulated demand curve, we simply need to choose between four realistic subsidy scenarios. A factorial design allows us to maximize power to test the main effects of price changes. This is the case when we hypothesize no statistical interaction of changing the price of the two commodities (the RDT and the ACT) on testing uptake. We anticipate any statistical interaction will, from a public health perspective, be relatively small for two reasons: (1) our previous studies testing combinations of prices for these two commodities found no evidence of a statistical interaction [1] and (2) we expect relatively high levels of testing uptake at all of the price levels we test which minimizes the likelihood of a large statistical interaction effect.

The alternative design for this study would be a parallel design which is powered based on tests for 2-4 combinations of RDT and ACT prices that fit the budget constraint. However, in general, these required much larger sample sizes to have similar levels of power to detect the expected effect sizes between the arms. Moreover, using a parallel design makes it more difficult to disentangle the effects on testing uptake due to

changes in the price of the RDT compared to changes in the price of the ACT. Using a factorial design enables us, within the range of prices that we test, to identify changes in the price of each commodity on testing uptake. It is therefore more generalizable to other contexts.

5.3 Level of Randomization

In Aim 1, we randomize at the individual level because we are testing the effect of RDT and ACT prices on individual clients' decision about whether to get tested for malaria. For a given sample size of clients, randomizing at the individual level gives us greater statistical power for detecting effect sizes than randomizing at a higher level such as the retail outlet.

Since clients generally enter the shop one at a time, we do not expect there to be any spillovers between the treatment arms (i.e we do not expect that offering a client one combination of subsidies will affect the decision-making of a client assigned to another treatment arm within the same shop). Moreover, since the study will be administered by the research assistant (and not the shopkeeper), this should minimize the concern that individual testing decisions might be influenced by the shopkeeper. Randomizing at the individual level also reduces the likelihood that clients will visit specific retail outlets in order to get the highest subsidy levels of the two commodities.

We do not expect there to be significant clustering of the outcome by retail outlet because the 10 retail outlets will be chosen to have similar characteristics in terms of location, size, client load, length of time the shop has been open, inventory, and qualifications of the drug shop owner. We will further minimize clustering by having a research assistant perform the RDT (rather than the shopkeeper), rotating the research assistants across the 10 outlets, and by ensuring that the intervention is delivered in a consistent manner across all outlets. Based on our ongoing work in the region, we expect at most an intra-class correlation coefficient of approximately 0.008, which would still give us 80% power to detect our expected effect sizes. Since we only have 10 shops (i.e. 10 clusters) we will need to use finite-sample correction methods to estimate parameter standard errors. Specifically, the Kauerman-Carroll approach is recommended when the number of individuals per unit is similar across all units [4, 5] .

5.4 Intervention Arms

The intervention arms test 4 different combinations of RDT and conditional ACT prices (see Table 1). These subsidy levels were chosen for two main reasons: (1) the total estimated costs of the subsidies are within the range of total costs that are expected to be feasible from a policy standpoint (from either donor or government funds), and, (2) these are prices at which we expect relatively high levels of uptake of testing, but also prices at which we expect there to be significant price-related variation in demand for testing.

5.5 Analysis Model

We chose to estimate risk ratios (RR) rather than odds ratios (OR) to quantify our intervention effects on the binary outcomes of interest. It is well-known that the OR is commonly interpreted as a RR [8]. This is valid when the reference level of the ratio is low (i.e. < 10%) but it is not true otherwise and instead the OR will overstate the magnitude of effect of the RR (i.e. if RR> 1, then OR>RR>1 and if RR<1, OR<RR<1). In the reference arm (0% RDT subsidy and 67% ACT subsidy), we expect uptake of testing to be approximately 60%.

Therefore, to avoid the potential for misinterpretation of the OR and for overstating of intervention effects, we propose the RR as our measure of relative effect of the intervention. Similarly, as recommendations of the CONSORT statement on reporting of results from trials, we will also present absolute effects quantified by risk differences [6]. As a consequence, we should provide a well-balanced view of the intervention effects for the policy makers who will use the results from our proposed research.

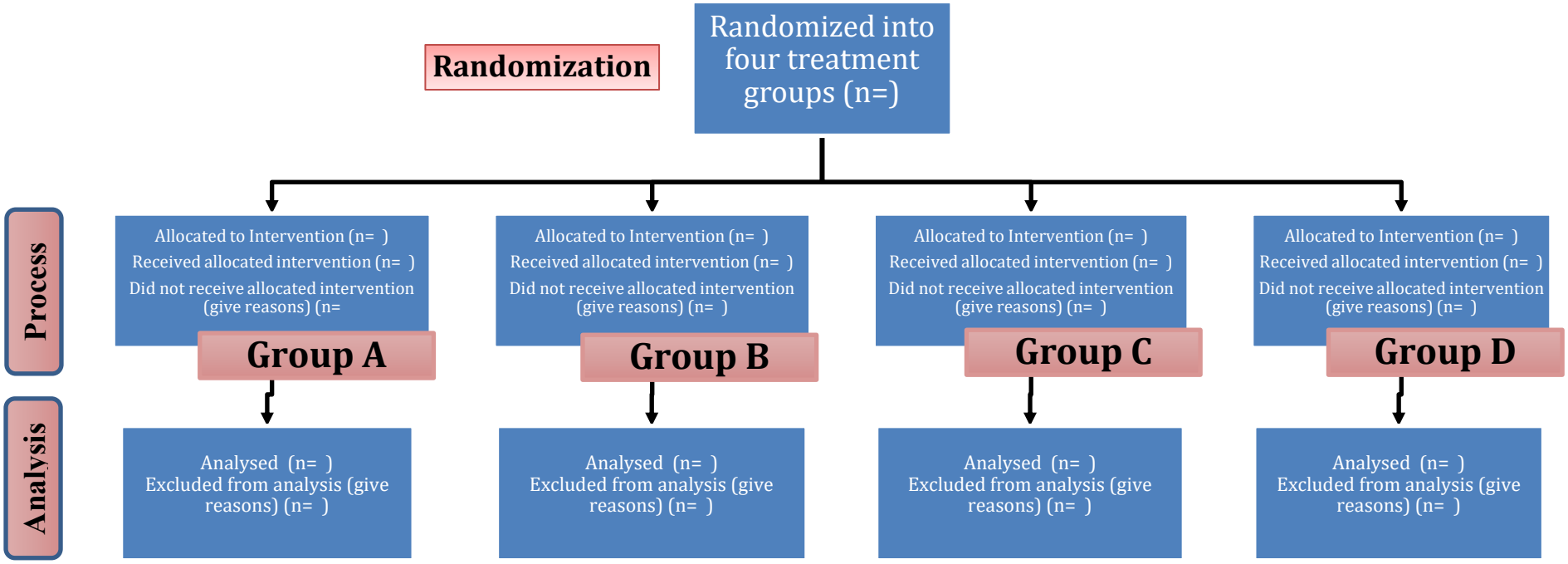
A modified Poisson approach to estimate risk ratios is preferred to a log-binomial approach for clustered binary outcome data because it has fewer computational issues [3]. In addition, the generalized estimating equations (GEE) accounts for clustering by outlet and provide the so-called population-averaged intervention effects (i.e., the average effect of the treatment arms across the population of all individuals represented by our sample) that are commonly of interest for public-health interventions [9].

Since we expect that the impact on uptake of testing for price changes of one commodity will be affected only minimally by the price of the other commodity, we did not power to detect such an interaction (this would have required sample sizes that were not feasible in this context). We will, nonetheless, test for any interaction effect in case it exists and is statistically significant and large in magnitude.

As our interventions are randomly assigned at the individual level, and because we plan to enroll more than 200 clients per treatment arm, we do not expect any confounding related to individual client characteristics. However, in our main model we will include the following covariates that may affect the decision of whether to test for malaria: the age, gender, wealth, and education level of the client (if the client is under the age of 18 then we will include the education level of the parent/legal guardian). Including these covariates will account for any imbalances across the treatment groups that may occur by chance and should also increase the precision of our estimates.

6 Appendix

Appendix A: CONSORT flow-chart for progress of individuals through four treatment groups



Appendix B: Shell tables for main analysis

Table B1. Sample Characteristics by Treatment Group – n (%), unless otherwise noted

	Conditional ACT Subsidy Intervention & RDT Subsidy Intervention (A)	**RDT Subsidy Intervention Only** (B)	**Conditional ACT Subsidy Intervention Only** (C)	**No Subsidy Intervention (Control)** (D)	**All Groups**
Demographic Characteristics					
Febrile individual					
Adult					
Child					
Age (respondent)					
Gender (respondent)					
Relationship (respondent)					
Age (years) – mean (SD)					
Gender					
Socioeconomic Status					
Household size					
Highest level of schooling					
Proxy Respondent for Child					
Adult					
Occupation Category (???)					
Number of animals					
Roof Type					
Owns Land					

Table B2. Sample Means for Testing and Treating Outcomes and Behavior

	Conditional ACT Subsidy Intervention & RDT Subsidy Intervention	RDT Subsidy Intervention Only	Conditional ACT Subsidy Intervention Only	No Subsidy Intervention (Control)	All Groups
	(A)	(B)	(C)	(D)	
Testing Behavior					
Had RDT					
Yes					
Positive					
ACT					
No ACT					
Negative					
ACT					
No ACT					
No					
ACT					
No ACT					

Table B3. Modified Poisson regression results for the outcome of malaria testing (N=?)

Variable	RR (95% CI)
Conditional ACT Subsidy Intervention & RDT Subsidy Intervention	
RDT Subsidy Intervention Only	
Conditional ACT Subsidy Intervention Only	
No Subsidy Intervention	
Grand Mean	
Contrasts (given no significant interaction effect)	
(1a) Effect of RDT subsidy on the uptake of testing, averaged over the ACT price levels.	
(1b) Effect of conditional ACT subsidy on the uptake of testing, averaged over the RDT price levels.	
Contrasts (given strong interaction effect)	
(2a) Effect of RDT subsidy on the uptake of testing, given no conditional ACT subsidy.	
(2b) Effect of RDT subsidy on the uptake of testing, given conditional ACT subsidy.	
(2c) Effect of conditional ACT subsidy on the uptake of testing, given no RDT subsidy.	
(2d) Effect of conditional ACT subsidy on the uptake of testing, given RDT subsidy.	

Table B4. Number of participants in each outlet by treatment group

	Conditional ACT Subsidy Intervention & RDT Subsidy Intervention	**RDT Subsidy Intervention Only**	**Conditional ACT Subsidy Intervention Only**	**No Subsidy Intervention (Control)**	**All Groups**
	(A)	(B)	(C)	(D)	
Number of participants					
Outlet 1					
Outlet 2					
…					

Appendix C: Comparison with Assumptions in Innovative Public-Private Partnership to target Subsidized Antimalarials in the Retail Sector study (IPPP)

Table C1: Estimated proportion of participants opt in for malaria testing in IPPP study

	Conditional ACT Subsidy (Cost to client: $0.65 for adult dose)	**No ACT Subsidy** (Cost to client: $1.25 for adult dose)
RDT Subsidy (Cost to client: $0.00)	75%	50%
RDT No Subsidy (Cost to client: $0.50)	40%	20%

Table C2: Estimated proportion of participants opt in for malaria testing in TESTsmART study

	100% Conditional ACT Subsidy (Cost to client: $0 for adult dose)	**67% Conditional ACT Subsidy** (Cost to client: $0.40 for adult dose)
50% RDT Subsidy (Cost to client: $0.20)	85%	75%
RDT No Subsidy (Cost to client: $0.40)	70%	60%

7 References

1. O'Meara WP, Mohanan M, Laktabai J, Lesser A, Platt A, Maffioli E, Turner EL, Menya D (2016) Assessing the independent and combined effects of subsidies for antimalarials and rapid diagnostic testing on fever management decisions in the retail sector: results from a factorial randomised trial in western Kenya. BMJ Global Health 1:e000101

2. Zou G (2004) A Modified Poisson Regression Approach to Prospective Studies with Binary Data. American Journal of Epidemiology 159:702–706 . doi: 10.1093/aje/kwh090

3. Zou G, Donner A (2013) Extension of the modified Poisson regression model to prospective studies with correlated binary data. Statistical Methods in Medical Research 22:661–670 . doi: 10.1177/0962280211427759

4. Kauermann G, Carroll RJ (2001) A Note on the Efficiency of Sandwich Covariance Matrix Estimation. Journal of the American Statistical Association 96:1387–1396

5. Li P, Redden DT (2015) Small sample performance of bias-corrected sandwich estimators for cluster-randomized trials with binary outcomes: P. LI AND D. T. REDDEN. Statistics in Medicine 34:281–296 . doi: 10.1002/sim.6344

6. Campbell MK, Piaggio G, Elbourne DR, Altman DG, CONSORT Group (2012) Consort 2010 statement: extension to cluster randomised trials. BMJ 345:e5661

7. StataCorp Stata Statistical Software: Release 15. StataCorp LLC, College Station, TX

8. Knol MJ, Duijnhoven RG, Grobbee DE, Moons KGM, Groenwold RHH (2011) Potential Misinterpretation of Treatment Effects Due to Use of Odds Ratios and Logistic Regression in Randomized Controlled Trials. PLoS ONE 6:e21248 . doi: 10.1371/journal.pone.0021248

9. Turner EL, Prague M, Gallis JA, Li F, Murray DM (2017) Review of Recent Methodological Developments in Group-Randomized Trials: Part 2—Analysis. American Journal of Public Health 107:1078–1086 . doi: 2105/AJPH.2017.303707

Statistical Analysis Plan Checklist

Below you will find a checklist of recommended items to include in a statistical analysis plan. Some of these are specific to clinical trials (based on this JAMA paper) and some are other are specific to observational studies (based on STROBE/RECORD guidelines), so every item will not be necessary for every project. The biostatistician should start with the SAP template below (starting on page 6) and add in necessary information from the checklist. Item numbers that are starred (*) are not explicitly included in the SAP template and should be added by the author if relevant to the project.

Section/Topic	Item #	Description	Included (Yes/No/NA)
Administrative Information			
Study Information	1a	Descriptive title that matches the protocol, with SAP either as a forerunner or subtitle	Yes
	1b	Trial registration number, protocol version number, and/or IRB number.	Yes
	1c	CRU/Department/Division/Center/other collaborative unit that the study falls under	Yes
Roles and responsibility	2a	Listing of principal investigators, clinical leads, and co-authors (if known)	Yes
	2b	Name and affiliation of SAP author(s)	Yes
	2c	Names, affiliations, and roles of other SAP contributors (e.g. senior statistician)	Yes
SAP Information	3	SAP version number, with date of current version and original creation date	Yes
Project Information	4a	Project folder location	TBD
	4b	Project goals (e.g. manuscript, abstract, presentation, etc.)	Yes
	4c	Project deadlines (of listed goals)	TBD
	4d	Effort estimate	TBD
Investigator Agreement			
Investigator Agreement	5	Confirmation that BERD Method Core's collaborative process has been reviewed, that all statistical analyses included in an abstract or manuscript should reflect the SAP, no changes should be made to the SAP without discussing with the SAP author, all biostatisticians on the SAP are co-authors on the manuscript, and that publications resulting from the SAP must cite grant number UL1TR002553 and be submitted to PubMed Central	
Signatures	6	Signatures of SAP author, senior statistician, and principal investigator(s)	
Activity Log			
SAP revisions	7a	SAP revision history with dates	Yes
	7b	Justification for each SAP revision	Yes
	7c*	Timing of SAP revision in relation to any interim analyses or submissions	NA
Study Overview			

Background and introduction	8	Synopsis of scientific background and rationale for the study	Yes
Aims and Hypotheses	9a	List of all scientific aims/objectives of the study, with specifications of primary, secondary, etc.	Yes
	9b	List of all statistical hypotheses (corresponding to the scientific aims), with specifications of primary, secondary, etc.	TBD
Variables of Interest	10a	List of all outcome/endpoint variables, with a description of their coding/units, timing, and source, corresponding to the statistical hypotheses. If any variables are defined using ICD or CPT codes, list them out.	Yes
	10b	List of all exposure variables, with a description of their coding/units, timing, and source, corresponding to the statistical hypotheses. If any variables are defined using ICD or CPT codes, list them out.	NA
	10c	List of any additional variables of interest (e.g. covariates, potential confounders, effect modifiers, etc.) in the analysis	Yes
	10d*	Location of data dictionary (or provided as an appendix)	TBD
	10e*	Report category boundaries if continuous variables are collapsed into categories, and describe any other relevant data transformations	NA
Causal Graph	11*	May be helpful to include a DAG or other graph/diagram that describes the way the variables of interest are presumed to relate to each other	No
Study Methods			
Study Plan and Design	12a	Description of the study design (e.g. parallel group randomized trial, case-control study, cohort study, etc.)	Yes
	12b*	Study setting, location, and relevant dates (e.g. periods of enrolment, exposure, follow-up, and collection)	TBD
	12c*	Description of intervention or exposure groups, with allocation ratios, and details of any matching criteria	Yes
	12d*	Details on randomization (e.g. stratification factors) and blinding procedures	Yes
	12e	List of eligibility and/or inclusion/exclusion criteria	Yes
	12f*	Description of screening/enrolment/recruitment processes	Yes
	12g*	Description of patient flow (e.g. CONSORT diagram)	Yes
	12h*	Description of analysis population (e.g. intention to treat, per protocol, etc.)	Yes
	12i*	Definitions of adherence/compliance, protocol deviations, loss-to-follow-up, adverse events, etc.	Yes
	12j*	Time points at which outcomes are measured	Yes

	12k*	Timing of final analyses (are all outcomes analysed collectively, or will short-term outcomes be analysed separately from long-term outcomes, etc.)	NA
Sample Size	13a*	Sample size calculation or justification (either provided in full or summarized, with link to original source)	Yes
	13b*	Description of pre-planned subgroup analyses, power for these analyses, and planned multiple comparison adjustment procedures	NA
Interim Analyses	14a*	Description of what interim analyses will be conducted at which time points, and what methods used to adjust significance levels due to the interim analysis	Yes
	14b*	Details of any guidelines (e.g. safety, futility) for stopping the study early	NA
	14c*	Details of any changes to trial design due to interim analyses (e.g. enrolling more patients)	NA
Data	15a	Description of data collection/acquisition process, with contact information for team member responsible	Yes
	15b	Description of data flow/transfer from primary data collection through to creation of final analysis dataset	TBD
	15c	Data transfer method and date	TBD
	15d	Folder location where datasets are stored	TBD
	15e*	Description of any additional data management, quality control, or processing undertaken	NA
	15f*	If any data are extracted from a database, a description of the database and the query used for the extraction, and whether/how it was merged with any data from outside that database. If the study involved linkage of databases, consider use of a flow diagram to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	NA
	15f*	Description of any other data sources incorporated in the analysis	NA
Missing Data	16a*	Description of sources and magnitudes of missing data	NA
	16b*	Description of how missing data patterns will be presented/summarized (may be helpful to have a table shell or draft CONSORT-style diagram)	NA
	16c*	Description of contingency plans for handling missing data in analysis	NA
Simulations	17a*	If conducting a simulation, a description of the purpose of the simulation and its design (e.g. fully factorial, partially factorial, grid search, etc.)	NA
	17b*	Define the fixed and variable factors or parameters in the simulation, the estimands/targets of the simulation, and the performance measures to be estimated (with justifications of their relevance to the estimands/targets)	NA
	17c*	Description of the tabular and graphical presentations of simulation results and their interpretation	NA

Statistical Analysis Plan

Statistical Significance	18a*	Hypothesis testing framework (e.g. superiority, equivalence, non-inferiority), or description of alternative analytic framework (e.g. evaluation of a posterior in a Bayesian analysis, etc.)	Yes
	18b*	Level of significance for primary hypotheses, including a description and rationale for any multiple comparisons adjustment or Type I error control procedures	Yes
	18c*	Description of any decision-making rules based on confidence intervals, credible intervals, prediction intervals, Bayes' factors, or other alternative inferential methods	No
	18d*	Description of how the results of any hypothesis tests (or alternative inferential methods) will be interpreted with respect to both the statistical hypotheses and scientific aims/objectives of the study	Yes
Descriptive Statistics	19a*	List of characteristics (e.g. demographic, clinical) to be summarized descriptively (e.g. "Table 1")	Yes
	19b*	Description of how these characteristics will be summarized descriptively (e.g. means/medians vs. N (%), tabular displays, graphical displays, etc.)	Yes
	19c*	Summarize follow-up time (e.g. average and total amount) and number of events	NA
Analysis Methods	20a	For each aim/hypothesis (see items 9a/9b), a description of what analysis method will be used and how the results from this method will be reported and interpreted	Yes
	20b*	Description of any transformations, standardizations, covariate or confounder adjustments, weighting, or stratification methods to be used and why.	NA
	20c*	For each analytic method proposed, a description of the assumptions of that method and what processes will be used to evaluate whether or not those assumptions hold	Yes
	20d*	Details of contingency plans/alternative methods to be used if the assumptions are found not to hold	Yes
	20e*	In the case of non-standard test statistics, formulas provided for the test statistic with a description of the mathematical null hypothesis, how significance is determined, and how the test statistic is interpreted	TBD
	20f*	In the case of regression models, formulas provided for the full model with a description of which parameters are to be used, how they will be interpreted, how confidence intervals will be constructed, etc.	Yes
	20g*	In the case of survey, hierarchical/nested, or clustered data, a description of what methods will be used to adjust for the data structure and why (e.g. if using a GEE, describing which correlation structure and why it was chosen, etc.)	Yes
	20h*	For non-continuous outcomes, clearly explain the effect used (e.g. risk difference, risk ratio, odds ratio, etc.), whether it is relative or absolute, and justify why that was chosen as the effect measure of interest	NA
	20i*	Documentation of any non-standard methods used (e.g. using alternative degree of freedom calculation methods, using a non-canonical link function, etc.)	NA

	20j*	Description of any limitations, sources of bias, internal/external validity, and other relevant discussions concerning the interpretation and generalizability of the design or methods used	Yes
Additional Analysis Methods	21a*	Description of any pre-planned sensitivity analyses and how they will be interpreted	TBD
	21b*	Description of pre-planned subgroup analyses, power for these analyses, and planned multiple comparison adjustment procedures	NA
	21c*	Description of any additional post-hoc calculations or analyses (e.g. evaluating interaction/modification effects, calculating mediation or local average treatment effects, evaluation of AUROC curves, etc.)	Yes
	21d*	If conducting any bootstrap analyses, a description of the sampling algorithm and number of iterations used	NA
	21e*	If conducting any cross-validation procedures, a description of how the cross-validation is conducted (e.g. leave-one-out, train/validation/test, etc.)	NA
Exploratory Analyses	22a*	Description and justification for any pre-planned exploratory analyses and what methods will be used to conduct them	TBD
	22b*	Framework for conducting any unplanned exploratory analyses and how they will be integrated into the planned analysis	NA
Software	23*	List of statistical software (along with version numbers) to be used for each phase of the analysis; in the case of R or Stata, additionally list any requisite installed packages and their version numbers	Yes
Other	24*	Description of any additional planned analyses of the data (e.g. a safety analysis looking at adverse event rates for a Data Safety Monitoring Board, etc.)	NA
Tables and Figures			
Table Shells	25*	Example tables related to any of the conducted analyses; if possible including any available preliminary data	Yes
Example Figures	26*	Example figures related to any of the conducted analyses; if possible including any available preliminary data.	TBD
References			
References	27a	References for any non-standard statistical methods used	Yes
	27b	References (and locations) for any relevant protocols, standard operating procedures, or other documents cited in the SAP	Yes
Additional Information			
Appendices	28*	If necessary, appendices may be included (e.g. a full data dictionary, a copy of a Case Report Form, etc.)	Yes
Addendums	29*	Any additional analyses conducted that were not included in the SAP should be documented in an addendum, describing the purpose of the additional analysis, when it was conducted, and by whom	NA