SUPPLEMENTARY MATERIAL

1 DEFINING OUTCOMES

Figure S1: **Primary outcome decision diagram**

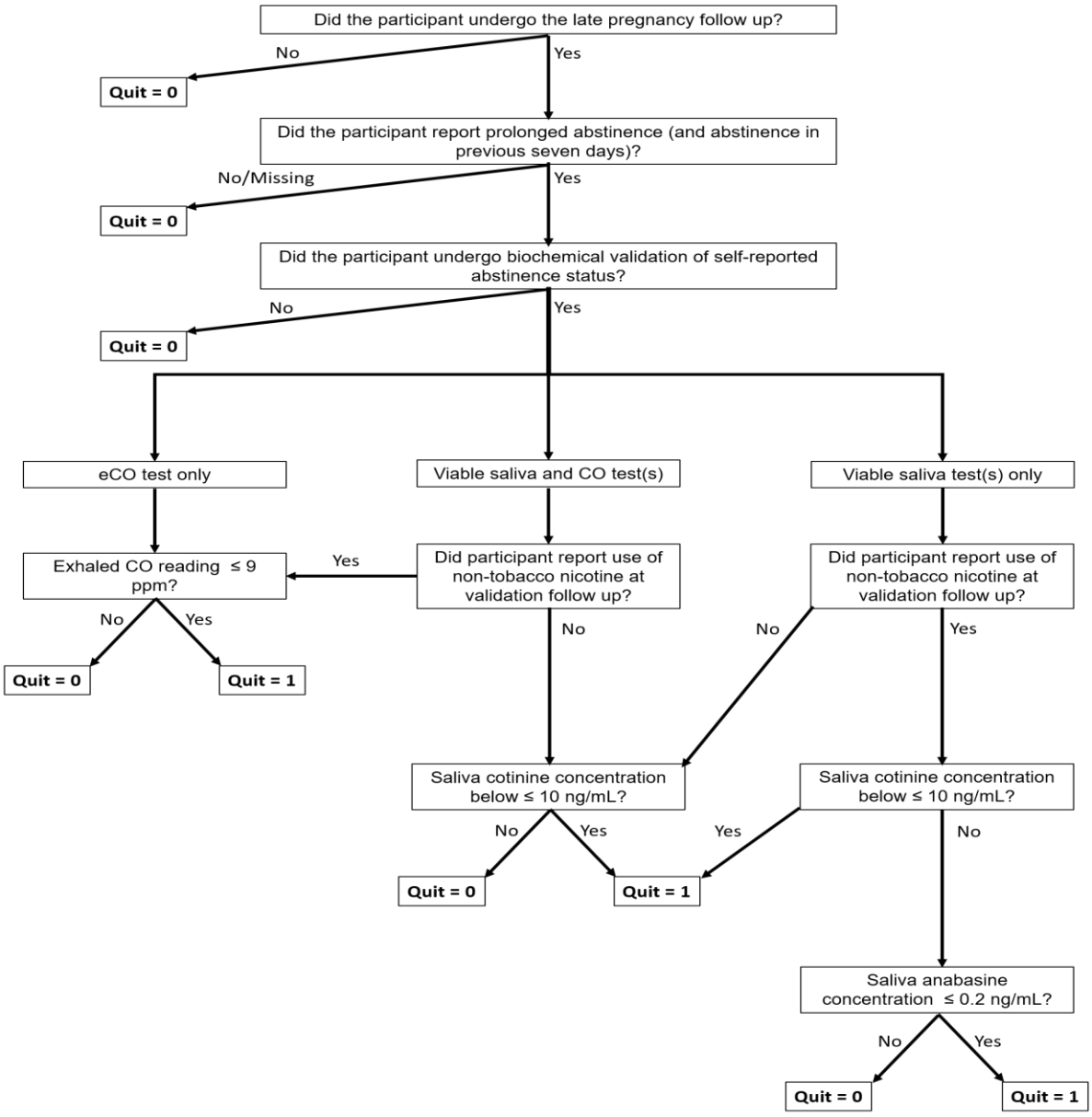
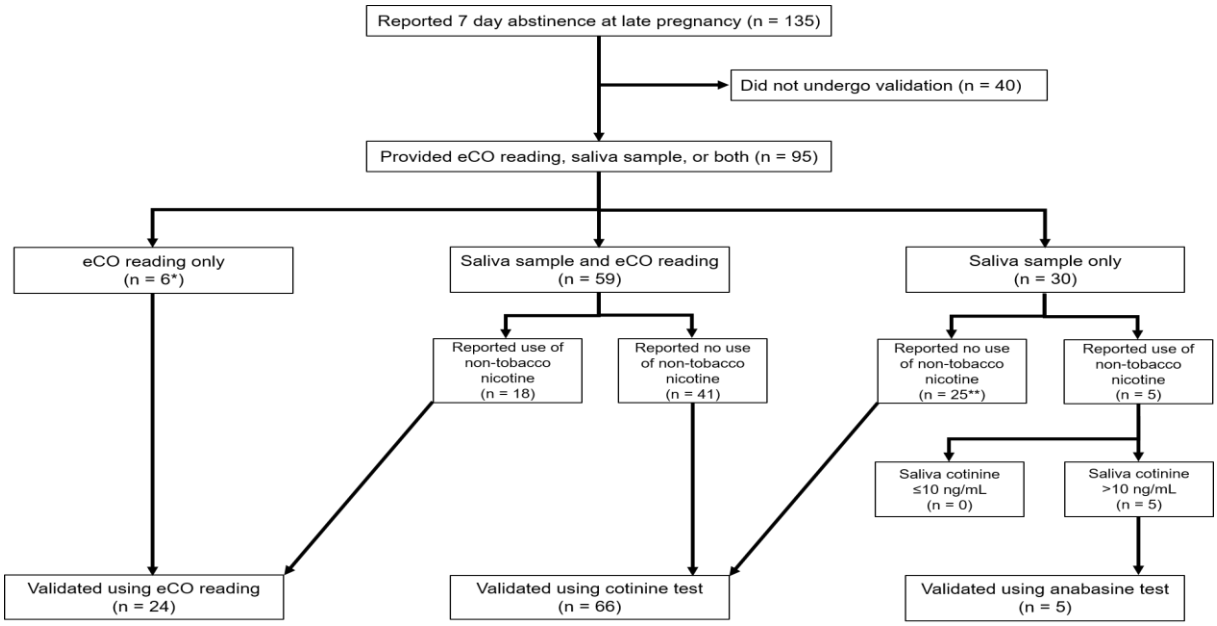


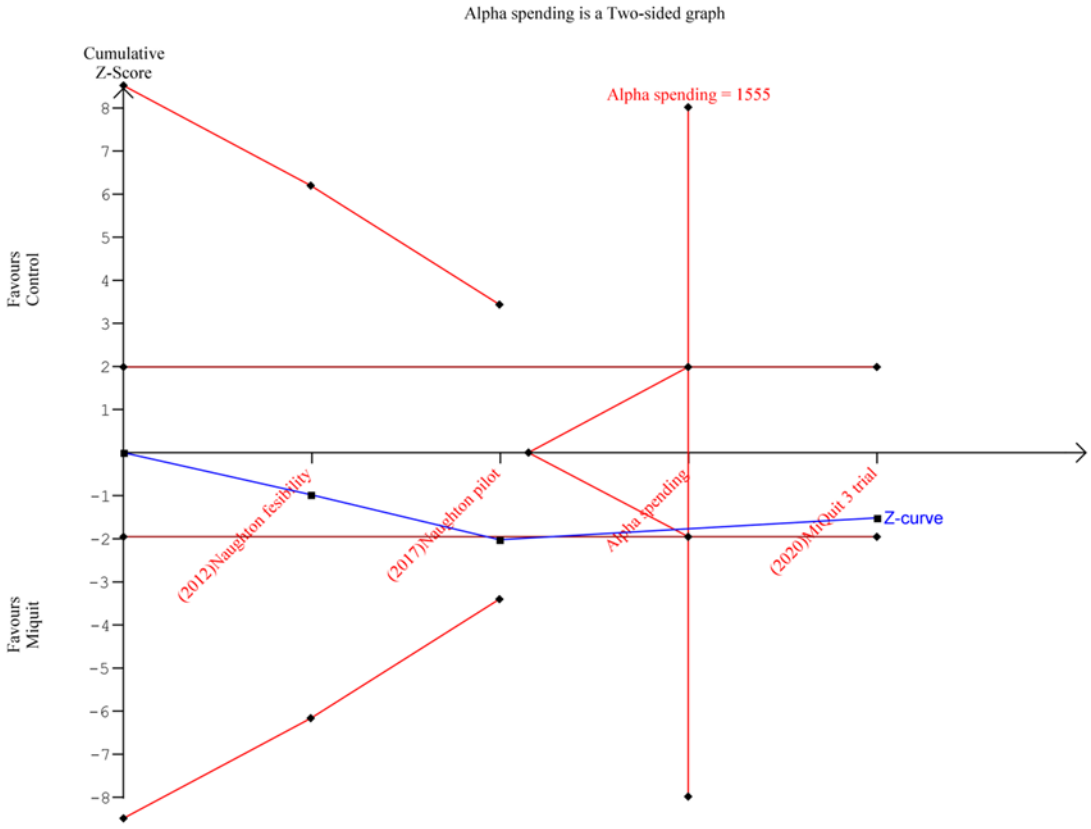
Figure S2: Validation of self-reported 7-day abstinence



*Two participants provided a saliva sample, but there was insufficient sample volume to obtain readings
**One participant was missing information relating to e-cigarette use

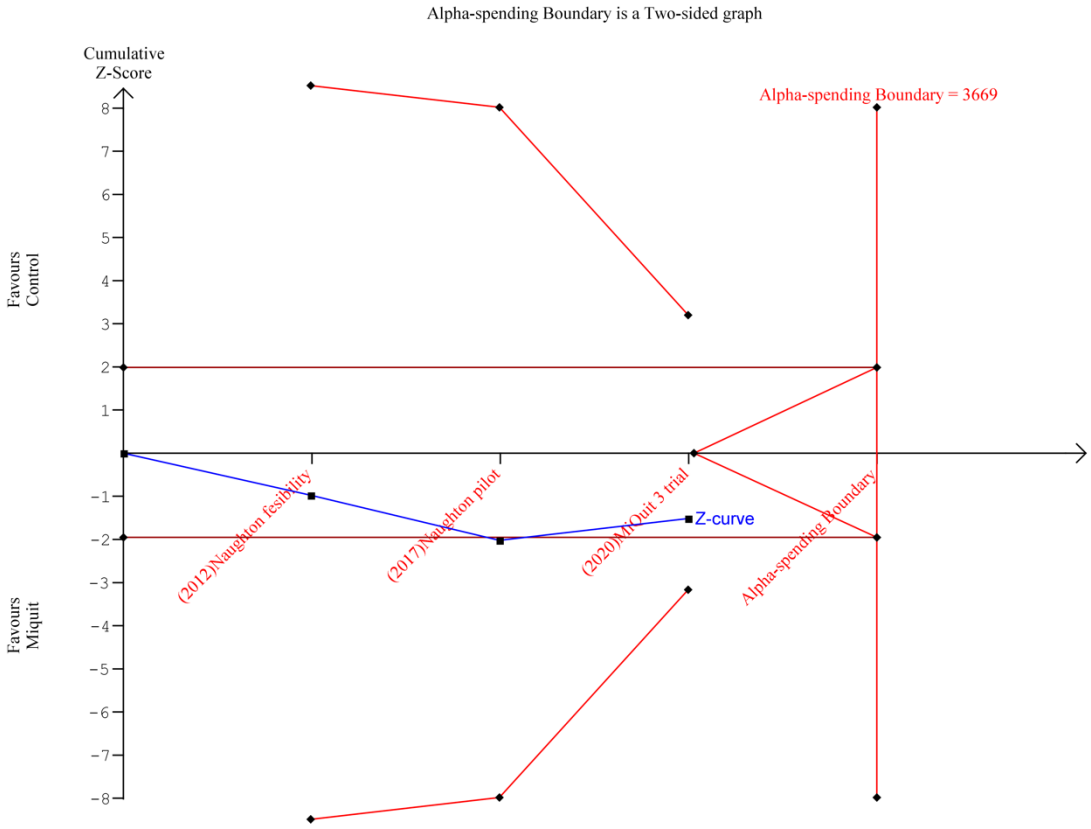
2 SAMPLE SIZE

Figure S3 Trial Sequential Analysis of three trials evaluating MiQuit effectiveness



Compared to control on prolonged abstinence from smoking at 4 weeks after enrolment until 36 weeks' gestation. The vertical red line is the diversity-adjusted optimal information size, i.e., the cumulative sample size required to establish with 90% power and 5% 2-sided significance whether the intervention increases prolonged abstinence from smoking by an absolute difference of 3.4% allowing for repeatedly meta-analysing the accumulating studies. The horizontal red line is a Z score of +1.96, equal to two-sided P = 0.05. The cumulative Z-statistic (blue line) crosses the futility boundary and reaches the optimal information size without crossing ±1.96, indicating evidence of futility such that further trials of this intervention may not be required.

Figure S4 Sensitivity analysis: Trial Sequential Analysis of three trials evaluating MiQuit effectiveness, assuming 2% absolute effect on quit rates



The vertical red line is the diversity-adjusted optimal information size, i.e., the cumulative sample size required to establish with 90% power and 5% 2-sided significance whether the intervention increases prolonged abstinence from smoking by an absolute difference of 2.0% allowing for repeatedly meta-analysing the accumulating studies. The horizontal red line is a Z score of +1.96, equal to two-sided P = 0.05. The cumulative Z-statistic (blue line) does not reach the optimal information size and does not cross the trial sequential monitoring boundary (curved red line), indicating that further trials are required before a firm conclusion regarding the effectiveness of the intervention can be concluded.

3 ECONOMICS

Please note that a publicly-available, latest version of the ESIP model is available at: https://www.nottingham.ac.uk/research/groups/tobaccoandalcohol/smoking-in-pregnancy/esip/index.aspx

This provides further information and a 'deterministic' version of the model (i.e. without probabilistic sensitivity analyses). Anyone can download this version of the model to gain familiarity with it prior to using the full ESIP model which is available on request.

Figure S5 Cost-effectiveness acceptability curve for MiQuit 3.

The curve suggests that approximately 55% of iterations are cost-saving. The curve asymptotes to around 61%, suggesting that there is a 61% chance that the MiQuit 3 intervention is cost-effective compared to the comparator at high willingness to pay per additional QALY.

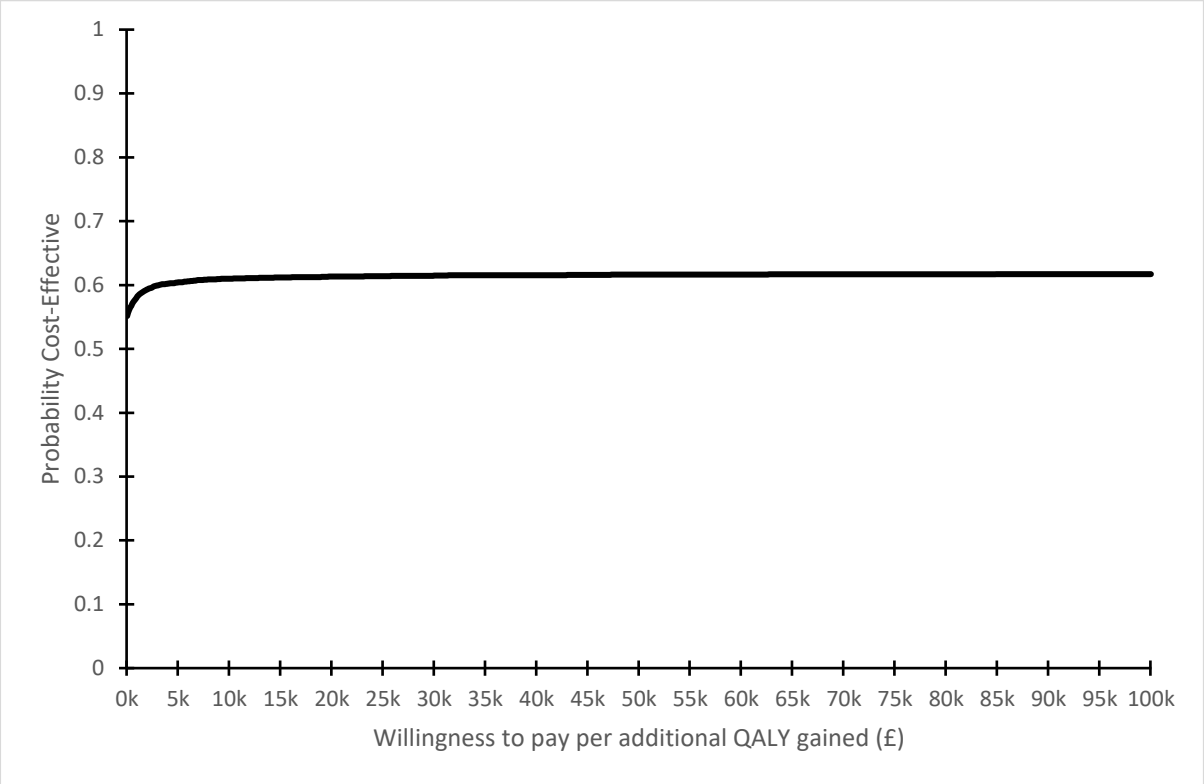
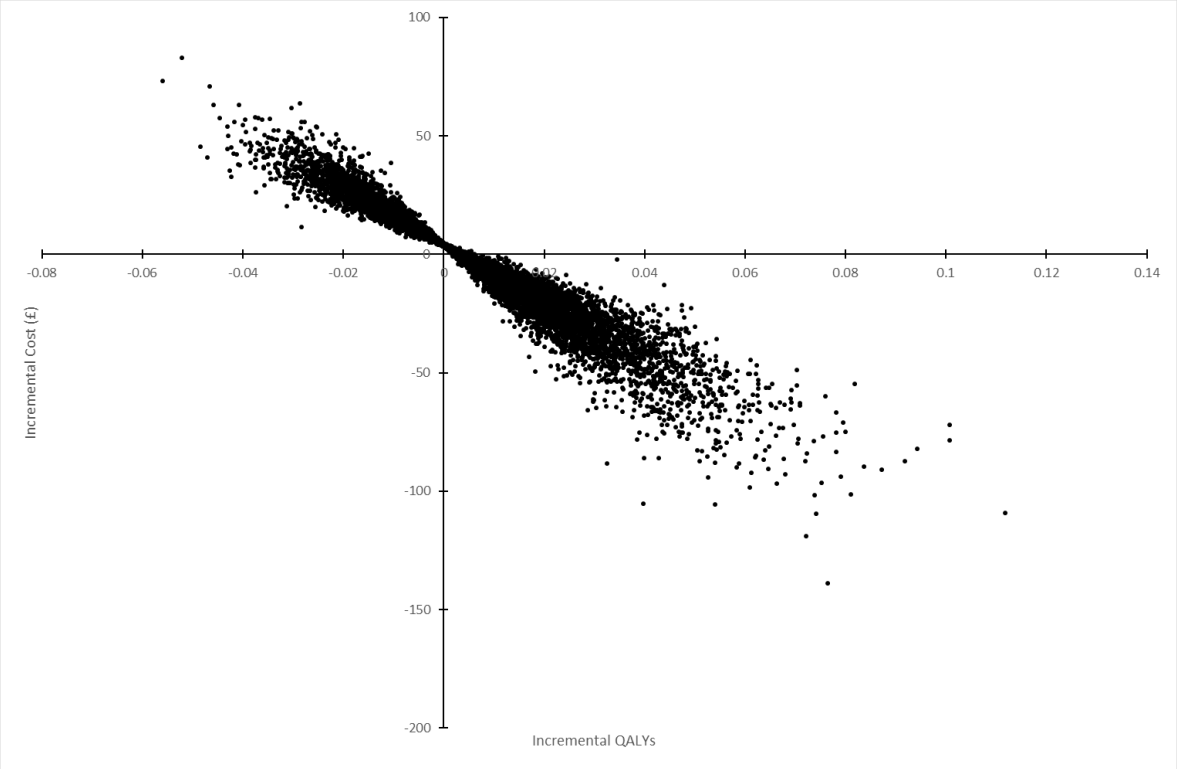


Fig S6 Scatterplot for MiQuit 3 intervention on the cost-effectiveness plane for combined maternal and offspring outcomes over the lifetime.

The probabilistic sensitivity analysis suggested that iterations covered three quadrants of the cost-effectiveness plane. Most iterations lie to the right of the vertical axis, suggesting that there is health gain and, where, these are below the horizontal axis, potential cost savings. However, some iterations are to the left of the vertical axis suggesting that in some iterations, MiQuit 3 led to a health loss.



4 FINDINGS

Table S1: Definitions of abstinence outcomes

Outcome	Definition
Abstinence 1 (primary)	Self-reported prolonged abstinence (fewer than six cigarettes smoked) between four weeks post randomisation and late pregnancy, with biochemical validation of self-reported abstinence during the previous seven days at the later time point. Participants missing self-reported abstinence data at late pregnancy, and participants who reported abstinence, but were missing validation data were assumed non-abstinent.
Abstinence 2	Self-reported prolonged abstinence (fewer than six cigarettes smoked) between four weeks post randomisation and late pregnancy. Participants missing self-reported abstinence data at late pregnancy were assumed non-abstinent.
Abstinence 3	Self-reported seven-day abstinence at both four weeks post randomisation and late pregnancy, with biochemical validation at the later time point. Participants missing self-reported abstinence data at either 4 weeks or late pregnancy, and participants who reported abstinence, but were missing validation data were assumed non-abstinent
Abstinence 4	Self-reported seven-day abstinence at both four weeks post randomisation and late pregnancy. Participants missing self-reported abstinence data at either 4 weeks or late pregnancy were assumed non-abstinent
Abstinence 5	Validated seven-day abstinence at late pregnancy. Participants missing self-reported abstinence data at late pregnancy, and participants who reported abstinence, but were missing validation data were assumed non-abstinent
Abstinence 6	Self-reported seven-day abstinence at late pregnancy. Participants missing self-reported abstinence data at late pregnancy were assumed non-abstinent
Abstinence 7	Self-reported seven-day abstinence at four weeks post randomisation. Participants missing self-reported abstinence data at four weeks were assumed non-abstinent

Key baseline data by follow up status

Table S2 provides summaries of key baseline data split by whether or not participants were followed up in late pregnancy. Participants who did not undergo follow up in late pregnancy were broadly similar to those that did for the majority of the characteristics presented in Table S2. One notable exception is maternal education and/or qualifications, where there is some evidence that those who were not followed up were more likely to be in the lowest education strata (i.e. no formal education/qualifications. However, there was little evidence to suggest the missingness mechanism differed by randomised group.

Table S2: Key baseline data by late pregnancy follow up status

Characteristic	Followed up at Late Pregnancy			Not followed up at Late Pregnancy		
	MiQuit (N = 309)	**Control** (N = 337)	**Total** (N = 646)	**MiQuit** (N = 192)	**Control** (N = 164)	**Total** (N = 356)
Age (years)						
N	309	337	646	192	164	356
Mean (SD)	27.4 (5.6)	27.5 (5.6)	27.5 (5.6)	26.6 (5.6)	27.6 (5.8)	27.0 (5.7)
Median (Q1, Q3)	26.8 (23.3, 31.1)	26.9 (23.2, 31.6)	26.9 (23.3, 31.3)	25.4 (22.4, 30.8)	27.0 (23.1, 31.2)	26.2 (22.6, 30.9)
Min, Max	16.7, 43.4	16.4, 43.1	16.4, 43.4	16.9, 40.6	17.5, 43.2	16.9, 43.2
Education, n (%)						
No formal qualifications	41 (13.3)	41 (12.2)	82 (12.7)	37 (19.3)	35 (21.3)	72 (20.2)
GCSEs (or equivalent)	156 (50.5)	184 (54.6)	340 (52.6)	110 (57.3)	81 (49.4)	191 (53.7)
A Levels (or equivalent)	80 (25.9)	80 (23.7)	160 (24.8)	36 (18.8)	29 (17.7)	65 (18.3)
Degree or higher	30 (9.7)	30 (8.9)	60 (9.3)	7 (3.6)	16 (9.8)	23 (6.5)
Missing	2 (0.6)	2 (0.6)	4 (0.6)	2 (1.0)	3 (1.8)	5 (1.4)
Gestation at baseline (weeks)						
N	309	337	646	192	164	356
Mean (SD)	14.9 (4.0)	15.3 (3.9)	15.1 (4.0)	15.0 (3.9)	14.5 (3.5)	14.8 (3.8)
Median (Q1, Q3)	13.1 (12.1, 19.6)	13.7 (12.4, 19.9)	13.4 (12.3, 19.7)	13.3 (12.4, 19.1)	13.1 (12.1, 15.6)	13.2 (12.3, 17.6)
Min, Max	6.0, 24.7	6.3, 24.9	6.0, 24.9	6.4, 24.1	6.0, 23.9	6.0, 24.1
Gestation stratum, n (%)						
< 16 weeks	210 (68.0)	208 (61.7)	418 (64.7)	121 (63.0)	124 (75.6)	245 (68.8)
≥ 16 weeks	99 (32.0)	129 (38.3)	228 (35.3)	71 (37.0)	40 (24.4)	111 (31.2)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Partner's smoking, n (%)						
Single	56 (18.1)	56 (16.6)	112 (17.3)	29 (15.1)	25 (15.2)	54 (15.2)
Partner a non-smoker	54 (17.5)	69 (20.5)	123 (19.0)	36 (18.8)	34 (20.7)	70 (19.7)
Partner a smoker	199 (64.4)	212 (62.9)	411 (63.6)	127 (66.1)	105 (64.0)	232 (65.2)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Cigarettes/day now						
N	309	337	646	192	164	356
Mean (SD)	8.6 (5.5)	8.6 (5.3)	8.6 (5.4)	8.6 (5.5)	9.6 (6.0)	9.0 (5.7)
Median (Q1, Q3)	8.0 (5.0, 10.0)	8.0 (5.0, 10.0)	8.0 (5.0, 10.0)	8.0 (5.0, 10.5)	10.0 (5.0, 12.0)	8.0 (5.0, 12.0)
Min, Max	1.0, 40.0	1.0, 30.0	1.0, 40.0	1.0, 35.0	1.0, 40.0	1.0, 40.0
Time from waking to first cigarette, n (%)						
Within 5 minutes	95 (30.7)	94 (27.9)	189 (29.3)	54 (28.1)	54 (32.9)	108 (30.3)
6 - 30 minutes	94 (30.4)	121 (35.9)	215 (33.3)	66 (34.4)	53 (32.3)	119 (33.4)
31 - 59 minutes	48 (15.5)	49 (14.5)	97 (15.0)	27 (14.1)	26 (15.9)	53 (14.9)
1 - 2 hours	44 (14.2)	51 (15.1)	95 (14.7)	24 (12.5)	20 (12.2)	44 (12.4)
More than 2 hours	28 (9.1)	22 (6.5)	50 (7.7)	21 (10.9)	11 (6.7)	32 (9.0)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Heaviness of Smoking Index						
N	309	337	646	192	164	356
Mean (SD)	1.9 (1.4)	1.9 (1.3)	1.9 (1.4)	1.9 (1.4)	2.1 (1.4)	2.0 (1.4)

Median (Q1, Q3) Min, Max	2.0 (1.0, 3.0) 0.0, 6.0	2.0 (1.0, 3.0) 0.0, 5.0	2.0 (1.0, 3.0) 0.0, 6.0	2.0 (1.0, 3.0) 0.0, 6.0	2.0 (1.0, 3.0) 0.0, 6.0	2.0 (1.0, 3.0) 0.0, 6.0
Strength of addiction, n (%)						
Low addiction	188 (60.8)	223 (66.2)	411 (63.6)	118 (61.5)	96 (58.5)	214 (60.1)
Moderate addiction	116 (37.5)	110 (32.6)	226 (35.0)	72 (37.5)	66 (40.2)	138 (38.8)
High addiction	5 (1.6)	4 (1.2)	9 (1.4)	2 (1.0)	2 (1.2)	4 (1.1)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)

Analysis of abstinence outcomes

Table S3: Analysis of abstinence outcomes 1 - 7

Outcome	**MiQuit** N = 501	**Control** N = 501	**Unadjusted OR** (95% CI)	**Unadjusted difference** (95% CI)	**Analysis model**[1]	**Adjusted OR** (95% CI)[2]	**Adjusted difference** (95% CI)	**AIC**	**Likelihood ratio test**[3]
Abstinence 1 (primary outcome)	26 (5.19%)	23 (4.59%)	1.14 (0.64 to 2.02)	0.60% (-2.07% to 3.27%)	Model 1 (N = 1002)	1.15 (0.65 to 2.04)	0.76% (-2.38% to 3.89%)	411.40	-
					Model 2 (N = 1002)	1.15 (0.65 to 2.04)	0.77% (-2.39% to 3.94%)	412.35	0.63
					Model 3 (N = 1002)	1.13 (0.64 to 2.01)	0.67% (-2.48% to 3.81%)	405.43	0.07
					Model 4 (N = 993)	1.16 (0.66 to 2.07)	0.83% (-2.36% to 4.02%)	407.44	0.13
					Model 5 (N = 993)	1.15 (0.65 to 2.04)	0.77% (-2.46% to 3.99%)	403.75	0.20
Abstinence 2	54 (10.78%)	47 (9.38%)	1.17 (0.77 to 1.76)	1.40% (-2.33% to 5.12%)	Model 1 (N = 1002)	1.19 (0.78 to 1.80)	1.64% (-2.34% to 5.61%)	654.01	-
					Model 2 (N = 1002)	1.19 (0.79 to 1.80)	1.67% (-2.32% to 5.65%)	653.20	0.45
					Model 3 (N = 1002)	1.17 (0.77 to 1.77)	1.47% (-2.51% to 5.44%)	645.77	0.02
					Model 4 (N = 993)	1.19 (0.78 to 1.80)	1.66% (-2.36% to 5.68%)	649.97	0.36
					Model 5 (N = 993)	1.17 (0.77 to 1.78)	1.54% (-2.50% to 5.57%)	641.95	0.18
Abstinence 3	14 (2.79%)	10 (2.00%)	1.41 (0.62 to 3.21)	0.80% (-1.09% to 2.69%)	Model 1 (N = 1002)	1.43 (0.64 to 3.30)	1.18% (-1.47% to 3.83%)	259.95	-
					Model 2 (N = 1002)	1.46 (0.66 to 3.38)	1.26% (-1.39% to 3.90%)	255.89	0.03
					Model 3 (N = 1002)	1.43 (0.64 to 3.29)	1.18% (-1.50% to 3.87%)	257.86	0.95
					Model (N = 993)	1.45 (0.65 to 3.33)	1.25% (-1.44% to 3.93%)	254.32	0.03
					Model 5 (N = 993)	1.40 (0.63 to 3.21)	1.15% (-1.56% to 3.87%)	249.61	0.02
Abstinence 4	27 (5.39%)	16 (3.19%)	1.73 (0.92 to 3.25)	2.20% (-0.31% to 4.70%)	Model 1 (N = 1002)	1.79 (0.96 to 3.42)	2.86% (-0.18% to 5.91%)	376.14	-
					Model 2 (N = 1002)	1.84 (0.99 to 3.53)	2.98% (-0.05% to 6.02%)	367.94	0.01
					Model 3	1.77	2.83%	373.03	0.43

					(N = 1002)	(0.95 to 3.39)	(-0.23% to 5.89%)		
					Model 4 (N = 993)	1.80 (0.97 to 3.43)	2.93% (-0.16% to 6.02%)	370.52	0.05
					Model 5 (N = 993)	1.79 (0.97 to 3.43)	2.92% (-0.18% to 6.02%)	361.85	0.01
Abstinence 5	38 (7.58%)	29 (5.79%)	1.34 (0.81 to 2.20)	1.80% (-1.29% to 4.89%)	Model 1 (N = 1002)	1.34 (0.81 to 2.23)	2.02% (-1.43% to 5.47%)	498.73	-
					Model 2 (N = 1002)	1.34 (0.81 to 2.23)	2.03% (-1.44% to 5.51%)	499.76	0.82
					Model 3 (N = 1002)	1.32 (0.80 to 2.20)	1.90% (-1.55% to 5.35%)	492.01	0.05
					Model 4 (N = 993)	1.29 (0.78 to 2.16)	1.79% (-1.69% to 5.26%)	488.81	0.12
					Model 5 (N = 993)	1.28 (0.78 to 2.14)	1.73% (-1.78% to 5.24%)	484.23	0.17
Abstinence 6	76 (15.17%)	59 (11.78%)	1.34 (0.93 to 1.93)	3.39% (-0.83% to 7.62%)	Model 1 (N = 1002)	1.37 (0.95 to 1.99)	3.73% (-0.65% to 8.11%)	775.09	-
					Model 2 (N = 1002)	1.37 (0.95 to 1.99)	3.74% (-0.64% to 8.13%)	773.85	0.45
					Model 3 (N = 1002)	1.35 (0.93 to 1.96)	3.55% (-0.82% to 7.93%)	767.95	0.05
					Model 4 N = 993)	1.34 (0.93 to 1.95)	3.49% (-0.92% to 7.89%)	766.46	0.36
					Model 5 N = 993)	1.33 (0.92 to 1.93)	3.36% (-1.05% to 7.77%)	758.81	0.24
Abstinence 7	37 (7.39%)	24 (4.79%)	1.58 (0.93 to 2.69)	2.59% (-0.36% to 5.55%)	Model 1 (N = 1002)	1.62 (0.96 to 2.78)	3.11% (-0.26% to 6.49%)	474.89	-
					Model 2 (N = 1002)	1.65 (0.98 to 2.84)	3.21% (-0.15% to 6.58%)	465.20	0.00
					Model 3 (N = 1002)	1.61 (0.96 to 2.77)	3.09% (-0.30% to 6.49%)	471.81	0.57
					Model 4 (N = 993)	1.64 (0.97 to 2.81)	3.23% (-0.19% to 6.65%)	469.67	0.11
					Model 5 (N = 993)	1.64 (0.97 to 2.82)	3.24% (-0.18% to 6.67%)	459.10	0.02

[1]All models included allocation, weeks gestation at baseline and recruitment site as fixed effects. Models 2, 3 and 4 included fixed effects for partner's smoking status, strength of nicotine dependence and maternal education respectively, and model 5 included all three of these terms

[2]Confidence intervals based on the penalised profile likelihood

[3]p-value for likelihood ratio test against model 1

5 SENSITIVITY ANALYSES

Analysis model with random intercepts for site

Following the concerns raised by a reviewer, we repeated the primary analysis using a mixed effect logistic regression with fixed effects for allocation and weeks gestation at baseline, and random intercepts for study recruitment site. The estimated OR (95% CI) for allocation from this model is 1.13 (0.64 to 2.02), and the estimate of the between site variance from this model was essentially 0. These results agree exactly with the estimates obtained from an ordinary logistic regression model (fitted by maximum likelihood), with just fixed effects for allocation and weeks gestation at baseline (i.e. with no fixed or random site effects included in the linear predictor).

Analysis using weakly informative priors on site effects

It was anticipated that the proportion of participants classified as abstinent for the purposes of the primary outcome would be small, and therefore numerous sites would have either none, or very few cases. This motivated the decision to use Firth logistic regression for the analysis of the primary and secondary abstinence outcomes. However, the penalisation imposed by the Firth model is relatively weak, and it was anticipated that the standard errors of some of the fitted site effects would remain highly unstable despite the shrinkage offered by the chosen model. Table S4 shows the results of refitting the five models used to analyse the primary abstinence outcome using stronger priors for the site effects. Specifically the site effects were assigned identical normal priors, with a mean of zero and a variance of 1.38. This prior translates to a prior median odds ratio of 1, with lower and upper 95% prior limits of 0.1 and 10 respectively.

Table S4: Analysis of primary outcome using stronger priors on the site effects

Model	**Adjusted OR** (95% CI)	**Adjusted difference** (95% CI)
Model 1 (N = 1002)	1.16 (0.65 to 2.08)	0.67% (-2.01% to 3.34%)
Model 2 (N = 1002)	1.16 (0.65 to 2.08)	0.67% (-2.00% to 3.35%)
Model 3 (N = 1002)	1.14 (0.64 to 2.06)	0.60% (-2.07% to 3.28%)
Model 4 (N = 993)	1.17 (0.65 to 2.11)	0.73% (-1.97% to 3.42%)
Model 5 (N = 993)	1.16 (0.65 to 2.10)	0.68% (-2.00% to 3.37%)

Missing data

Participants provided abstinence data at both four weeks post-randomisation and at late pregnancy. These responses were used to derive the seven abstinence outcomes. For some of these outcomes, abstinence status was derived based on several factors including self-reported abstinence, biochemical validation and use of non-tobacco nicotine. Hence the missingness of a given outcome could depend on several factors. Abstinence outcomes were classified as observed/missing based on the rules given in Table S5. The proportion of participants missing each of the seven abstinence outcomes (under the definitions given in Table S5) is detailed by allocation in Table S6.

Table S5: Rules used to classify abstinence data as observed or missing

Outcome	**Process used to determine missingness**
Abstinence 1 (primary)	Classified as missing if self-reported prolonged abstinence was missing, or if the participant reported prolonged abstinence and 7-day abstinence at late pregnancy, but did not undergo biochemical validation.
Abstinence 2	Classified as missing if self-reported prolonged abstinence was missing.

Abstinence 3	Classified as missing if self-reported 7-day abstinence was missing at either week four or late pregnancy, or if the participant reported abstinence at both time points, but did not undergo biochemical validation.
Abstinence 4	Classified as missing if self-reported 7-day abstinence was missing at either week four or late pregnancy.
Abstinence 5	Classified as missing if self-reported 7-day abstinence was missing at late pregnancy, or if participant reported 7-day abstinence, but did not undergo biochemical validation.
Abstinence 6	Classified as missing if self-reported 7-day abstinence was missing at late pregnancy.
Abstinence 7	Classified as missing if self-reported 7-day abstinence was missing at week four.

Table S6: Proportion of participants missing each abstinence outcome by allocation

Outcome	**Randomised treatment group**		
	MiQuit (N = 501)	**Control** (N = 501)	**Total** (N = 1002)
Missing abstinence 1	210 (41.9%)	178 (35.5%)	388 (38.7%)
Missing abstinence 2	192 (38.3%)	164 (32.7%)	356 (35.5%)
Missing abstinence 3	233 (46.5%)	203 (40.5%)	436 (43.5%)
Missing abstinence 4	226 (45.1%)	198 (39.5%)	424 (42.3%)
Missing abstinence 5	214 (42.7%)	182 (36.3%)	396 (39.5%)
Missing abstinence 6	192 (38.3%)	164 (32.7%)	356 (35.5%)
Missing abstinence 7	147 (29.3%)	117 (23.4%)	264 (26.3%)
Missing any abstinence outcomes	246 (49.1%)	213 (42.5%)	459 (45.8%)

From Table S6 we see that the proportion of participants with missing abstinence outcomes data was consistently around 6% higher in the MiQuit group compared with the control group. The proportion of participants who formally withdrew from follow up was similar in each group (4.2% of the MiQuit group vs 3.4% of the control group), hence the specific reasons for the outcome data being missing are generally not available. The primary analysis was conducted under the assumption that participants who were missing abstinence data were still smoking. This assumption may not be entirely plausible; hence several analyses are presented here to investigate the sensitivity of the results of the primary analysis to variation in the assumptions made regarding the missing values.

Multiple imputation (chained equations) was used to allow for variation in the values imputed for the missing abstinence outcomes, while also allowing for uncertainty about these unknown values. In the Statistical Analysis Plan it was stated that this procedure would be performed separately for each randomised group. However due to the rarity of the outcome, and the inclusion of numerous categorical predictors in the imputation model, the decision was taken to impute all participants at the same time in order to mitigate against problems caused by separation in the univariate imputation models fitted as part of the chained equation algorithm. Hence allocation was included in the imputation model along with the following variables; biochemically validated abstinence between four weeks and late pregnancy (abstinence outcome 1), self-reported 7-day abstinence at four weeks (abstinence outcome 7), weeks gestation at baseline (mean centred), recruitment site, partner's smoking status, strength of nicotine dependence at baseline, maternal education and baseline variables identified (using Firth logistic regression models) as being associated with missingness of any of the seven abstinence outcomes. Baseline variables were tested for inclusion in the imputation model using a likelihood ratio test of size 10% (where each model was tested against an equivalent Firth model with all parameters other than the intercept constrained to zero). Table S7 shows the baseline variables assessed. This process identified two additional baseline variables to include in the imputation model, a four level nominal

variable indicating when participants find it most difficult to avoid smoking and the depression and anxiety dimension of the EQ-5D-5L (a five level ordinal variable).

Table S7: Baseline variables tested for inclusion in the imputation model

Baseline predictor	**p-value**	**Included in imputation model?***
Age at baseline	0.155	No
Ethnicity	0.398	No
Previous pregnancies beyond 24 weeks	0.478	No
Cigarettes smoked per day prior to pregnancy	0.740	No
Urge to smoke in previous 24 hours	0.342	No
Seriously planning to quit	0.539	No
Length of longest previous quit attempt	0.271	No
Most difficult to not smoke	0.017	Yes
Main disadvantage of not smoking	0.713	No
Most important reason to stop smoking	0.195	No
Smoking beliefs item 1	0.903	No
Smoking beliefs item 2	0.332	No
Smoking beliefs item 3	0.888	No
Smoking beliefs item 4	0.795	No
Smoking beliefs item 5	0.253	No
Smoking beliefs item 6	0.561	No
Smoking beliefs item 7	0.304	No
Smoking beliefs item 8	0.706	No
Smoking beliefs item 9	0.389	No
EQ-5D item 1	0.184	No
EQ-5D item 2	0.315	No
EQ-5D item 3	0.187	No
EQ-5D item 4	0.558	No
EQ-5D item 5	0.069	Yes
EQ-5D general health	0.336	No

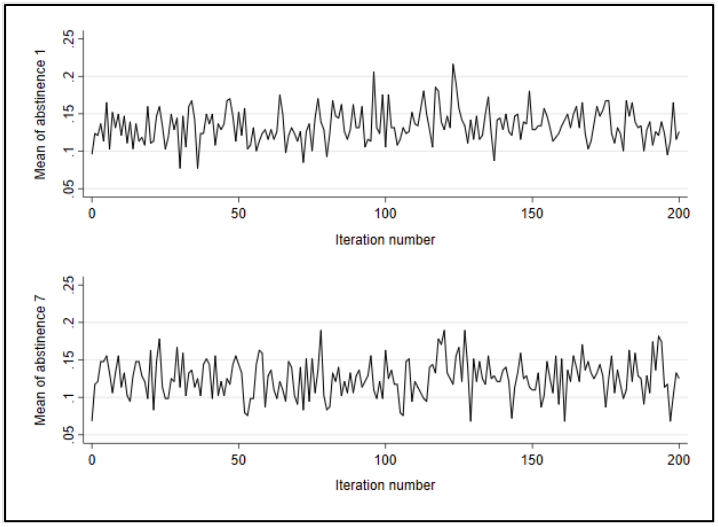
*Baseline variables tested for inclusion in imputation model using a likelihood ratio test of size 10% (model tested against the model with all parameters other than the intercept constrained to zero)

Missing values occurred predominantly in the two abstinence outcomes included in the imputation model, abstinence 1 (38.7% missing) and abstinence 7 (26.3% missing). There was small amounts of missingness in three other variables included in the model, namely maternal education (0.9% missing), the depression/anxiety dimension of the EQ-5D-5L (0.4% missing) and when participants found it most difficult to avoid smoking (0.1% missing). The imputation model was used to impute missing cases in all variables that had missing values. Imputation was performed using Stata's mi impute chained command. The two abstinence outcomes were imputed using binary logistic regression models. Maternal education and the EQ-5D-5L anxiety/depression dimension were imputed using ordinal logistic regression. When participants find it most difficult to avoid smoking was imputed using multinomial logistic regression. As stated, the inclusion of site in the imputation model, as well as several other binary/categorical variables would have led to separation or near separation occurring in at least some of the imputation models. Two steps were taken to mitigate against this issue. Firstly data augmentation was used in all of the univariate models used to impute variables with missing values. Secondly ordinal variables (maternal education and EQ-5D-5L depression/anxiety dimension) were treated as continuous when used to impute other variables.

The chained equation algorithm was run for 200 iterations to check stability and ascertain the length of burn in required. Trace plots of the mean of both the abstinence outcomes over these 200 iterations

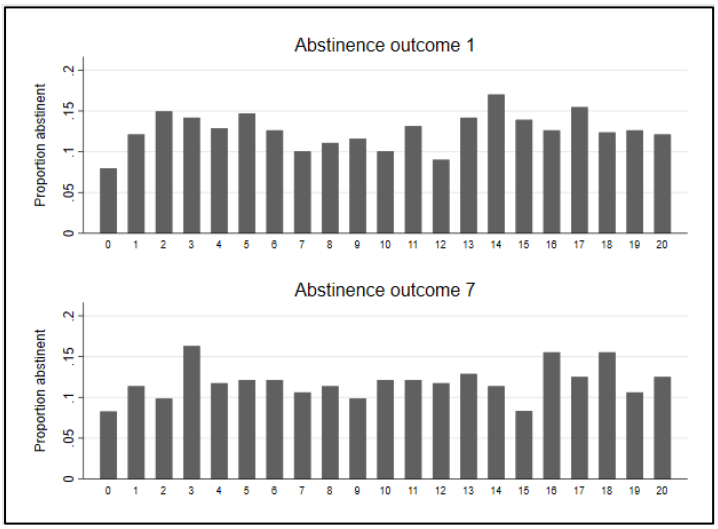
are given in Figure S7. These suggest that the chained equation algorithm reaches a stable state and does so quickly. Hence the data were imputed using 10 burn in iterations per imputation.

Figure S7: Trace plots of the mean of abstinence outcomes 1 and 7 over 200 iterations



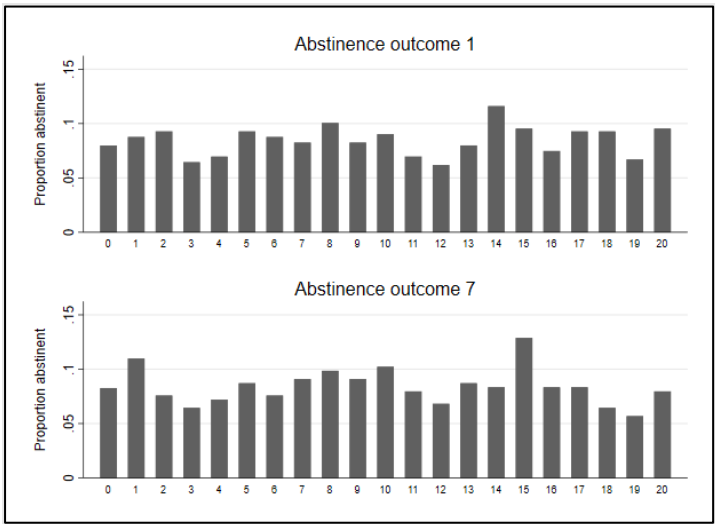
Initially 20 imputed datasets were generated to check the plausibility of the imputations generated by the proposed imputation model. Abstinence rates for both abstinence outcomes were compared between the observed and imputed data. For both these outcomes, the proportion of successes (i.e. participants abstinent) was substantially higher among participants with imputed data than among participants with observed data. This is illustrated in Figure S8, which shows the proportion abstinent in the observed data (labelled 0), and the proportions imputed as abstinent in the 20 imputed datasets (labelled 1 to 20). There is no clear rationale for those with missing data being more likely to have been abstinent. It is perhaps more likely to be an artefact of an inadequate/misspecified imputation model, as opposed to reflecting some actual underlying process. One possibility is that the inclusion of site in the imputation model leads to many near empty strata, despite the steps taken to mitigate against this. This may have led to an overly diffuse approximation of the posterior of the parameters, from which the perturbed parameters used in the univariate imputations are drawn. This in turn may lead to imputations of the missing abstinence outcomes, for which the probability of success (i.e. abstinence) is too high.

Figure S8: Proportion of participants abstinent in the observed data (labelled 0), and imputed data across 20 imputed datasets (labelled 1 to 20) for abstinence outcomes 1 (top) and 7 (bottom)



To investigate whether the inflated proportion of successes in the imputed data could be due to near empty strata resulting from the inclusion of site in the imputation model, an additional 20 imputations were generated omitting site from the imputation model (while keeping all other aspects of the imputation model the same). For this modified imputation model, the average proportion of successes (among participants with imputed outcome data) across 20 imputations was 8.5% and 8.4%, for abstinence outcomes 1 and 7 respectively. Hence the number of successes in the imputed data matches the number of successes in the observed data much more closely when site is omitted from the imputation model. This is illustrated in Figure S9 which shows the proportion abstinent in the observed data (labelled 0), and the proportions imputed as abstinent in the 20 imputed datasets (labelled 1 to 20) generated using the imputation model with site omitted.

Figure S9: Proportion of participants abstinent in the observed data (labelled 0), and imputed data across 20 imputed datasets (labelled 1 to 20) for abstinence outcomes 1 (top) and 7 (bottom), for the imputation model omitting site



The analyses of the primary abstinence outcome (abstinence outcome 1) presented previously all adjusted for site as a fixed effect. This needs to be reflected in our sensitivity analyses using multiply imputed data. The inclusion of site in the analysis models fitted to the multiply imputed datasets necessitates the inclusion of site in the imputation model. However, given the concerns over the imputations generated when site is included in the imputation model, there is a reasonable rationale for also generating imputed data with site omitted from the imputation model, with site also being omitted from the models used to analyse these imputed datasets. We therefore generated two sets of imputed datasets, one using an imputation model which includes site, and another which does not. The first set of imputations was analysed using Firth logistic regression models including site as a fixed effect, and the second set was analysed using the same models, but with the site effects omitted.

Each imputation model was used to generate 250 imputed datasets. These were then analysed using Firth logistic regression models, with the point estimates of the ORs for allocation being combined using Rubin's rules. The 95% CI was obtained by combining the profile penalised likelihoods from the models fitted to the completed datasets as described in [1]. This was implemented using the CLIP.confint function in the most up to date version of the logistf package [2]. Results of analyses including site in the imputation and analysis models are presented in Table S8. Results of analyses excluding site in the imputation and analysis models are presented in Table S9.

Table S8: Estimated ORs for allocation based on multiply imputed data (250 imputations), with study recruitment site included in the imputation model and analysis models

Analysis model	Estimated OR for allocation
Model 1	1.14 (0.66 to 1.98)

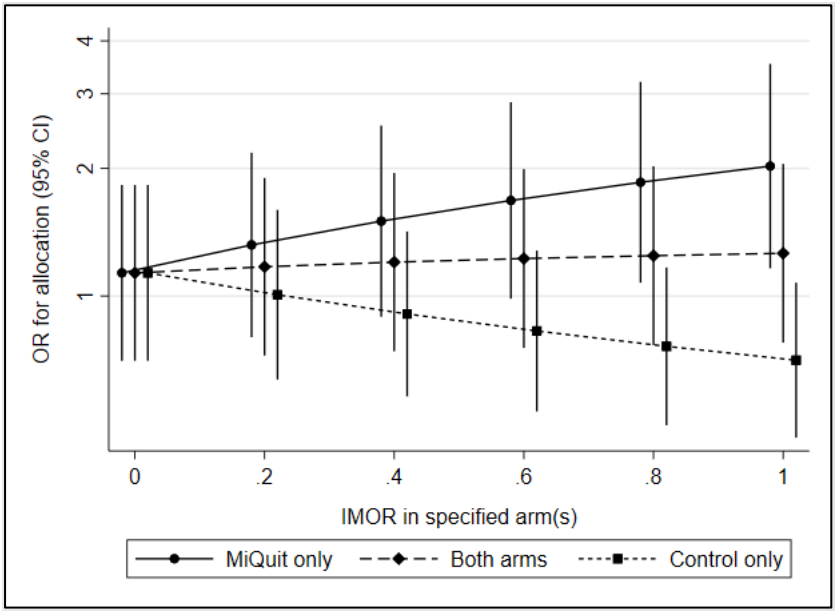
Model 2	1.14 (0.66 to 1.98)
Model 3	1.12 (0.64 to 1.94)
Model 4	1.15 (0.66 to 1.99)
Model 5	1.13 (0.65 to 1.97)

Table S9: Estimated ORs for allocation based on multiply imputed data (250 imputations), with study recruitment site not included in the imputation model or analysis models

Analysis model	Estimated OR for allocation
Model 1	1.18 (0.66 to 2.09)
Model 2	1.19 (0.67 to 2.12)
Model 3	1.16 (0.66 to 2.08)
Model 4	1.19 (0.67 to 2.12)
Model 5	1.20 (0.67 to 2.14)

We explored the sensitivity of the results to departures from MAR less extreme than missing = smoking and allowed the missingness mechanism to vary by randomised group. This was accomplished using the user written Stata command rctmiss [3]. The Firth logistic regression model fitted for the primary analysis, and for the analysis of the multiply imputed data, is not currently implemented in this software, hence a standard (maximum likelihood) logistic regression model was used as the substantive analysis model. To avoid separation, fixed site effects were omitted from this model and clustering by site was accounted for using cluster robust standard errors. Let Δ denote the log odds ratio for missingness conditional on the covariates included in the primary analysis model. The informative missingness odds ratio (IMOR) is defined as exp(Δ), meaning IMOR = 0 is equivalent to assuming missing = smoking (for both groups). The estimated odds ratios for allocation as the IMOR varies between 0 and 1, in each group separately and in both groups are given in Table 3 (main paper). The change in the estimated odds ratio for allocation as the IMOR varies from 0 to 1 is illustrated in Figure S10.

Figure S10: Sensitivity of the primary analysis to variation in the missing not at random assumption used to impute missing primary outcome data



Compliance

Two further analyses of the primary outcome were undertaken to obtain estimates of the average treatment effect among those participants who were able to adhere to the MiQuit programme, or would have been able to adhere to the MiQuit programme had they been offered it (assuming the assumptions of these analyses are met). Since participants in the control group had no way of accessing the MiQuit intervention, the only observable non-adherent behaviour was participants allocated to the MiQuit group not receiving the MiQuit programme as intended. Adherence was defined in a binary manner, with participants in the intervention group being classified as either compliers or non-compliers depending on the extent to which they received the MiQuit programme as intended. It was assumed that any participant in the MiQuit group for whom a STOP text was not received, received the MiQuit intervention as intended and is considered a complier, unless they specifically indicated at the late pregnancy follow up that they did not receive any texts from the study team. Participants for whom a STOP text was received after greater than 4 weeks of texts were also defined as compliers, unless they specifically indicated at the late pregnancy follow up that they did not receive any texts from the study team.

The average causal effect of the MiQuit programme among the compliers will be estimated by two different approaches, both of which assume randomisation to be a valid instrument for treatment received. Upper and lower bounds for complier average causal effect (CACE) (in terms of the causal risk ratio and causal risk difference) were obtained using a non-parametric approach. This was implemented using the most up to date version of the user written bpbounds Stata command, available from the Boston College Statistical Software Components archive [4]. Secondly the adjusted treatment received approach of Nagelkerke at al. [5] was used to obtain a point estimate (and 95% CI) of the CACE. Standard errors were obtained for the estimates from the second stage using the method of Terza [6].

Of the 501 participants randomised to the MiQuit programme, 43 (8.6%) were classified as non-compliers under the definitions given above. None of these participants were classified as abstinent for abstinence outcome 1. Hence all 26 participants classified as abstinent (for abstinence outcome 1) in the MiQuit group were compliers (under the definitions given above). The lower and upper bounds for the causal effect (on the risk ratio scale) of receiving the MiQuit programme as intended are 1.13 and 3.04 respectively. The lower and upper bounds for the causal risk difference are 0.59% and 9.35% respectively. These are the tightest bounds on the causal effect of the MiQuit programme which can be obtained (in the presence of non-compliance), assuming only that randomisation is a valid instrument for treatment received. Clearly these bounds encompass a large range of effects, from little practical significance up to substantial benefit, as might be expected given the minimal assumptions required. These bounds effectively rule out adverse effects of the MiQuit programme among those able to adhere to the programme requirements, but provide little information about a plausible set of values for the average causal effect. The adjusted treatment received estimate of the causal effect (on the odds ratio scale) of MiQuit, and 95% confidence interval is 1.10 (0.57 to 2.13). This is not materially different from the intention-to-treat estimates obtained for the primary analysis, other than the slight loss of precision owing to the use of the two-stage estimator. Hence there is little evidence to suggest that the estimated treatment effect among the compliers (i.e. the CACE) differs meaningfully from the intention-to-treat estimate.

Pregnancy outcomes

Key pregnancy outcomes for the 930 participants with these data available are summarised in Table S10. Maternal hospital admissions and gestational age at birth are summarised in Table S11 for the 919 live births. Finally birthweight and neonatal ICU admissions are summarised in Table S12, for each of the 927 individual live births (i.e. counting twins separately).

Table S10: Key pregnancy outcomes data by allocation

Outcome	Randomised treatment group		
	MiQuit (N = 466)	**Control** (N = 464)	**Total** (N = 930)
Single or multiple birth, n (%)			
Single	462 (99.1)	460 (99.1)	922 (99.1)
Twin	4 (0.9)	4 (0.9)	8 (0.9)
Missing	0 (0.0)	0 (0.0)	0 (0.0)
Type of birth, n (%)			
Live	464 (99.6)	455 (98.1)	919 (98.8)
Stillbirth	0 (0.0)	3 (0.6)	3 (0.3)
Miscarriage	2 (0.4)	6 (1.3)	8 (0.9)
Missing	0 (0.0)	0 (0.0)	0 (0.0)

Table S11: Maternal hospital admission and gestational age at birth (live births only)

Outcome	Randomised treatment group		
	MiQuit (N = 464)	**Control** (N = 455)	**Total** (N = 919)
Maternal hospital admissions, n (%)			
Admitted	10 (2.2)	9 (2.0)	19 (2.1)
Not admitted	454 (97.8)	446 (98.0)	900 (97.9)
Missing	0 (0.0)	0 (0.0)	0 (0.0)
Gestational age at birth (weeks)*			
N	464	455	919
Mean (SD)	38.7 (2.0)	38.5 (2.3)	38.6 (2.2)
Median (Q1, Q3)	39.0 (37.8, 40.0)	39.0 (37.7, 40.0)	39.0 (37.7, 40.0)
Min, Max	27.9, 42.1	26.1, 42.3	26.1, 42.3
Pre-term birth (<37 weeks), n (%)			
Pre-term	54 (11.6)	62 (13.6)	116 (12.6)
Full term	410 (88.4)	393 (86.4)	803 (87.4)
Missing	0 (0.0)	0 (0.0)	0 (0.0)

Table S12: Birthweight and infant ICU admissions (all live infants)

Outcome	Randomised treatment group		
	MiQuit (N = 468)	**Control** (N = 459)	**Total** (N = 927)
Infant ICU admissions, n (%)			
Admitted	45 (9.6)	47 (10.2)	92 (9.9)
Not admitted	423 (90.4)	412 (89.8)	835 (90.1)
Missing	0 (0.0)	0 (0.0)	0 (0.0)
Birthweight (kg)			
N	468	459	927
Mean (SD)	3.1 (0.6)	3.0 (0.6)	3.1 (0.6)
Median (Q1, Q3)	3.1 (2.7, 3.5)	3.1 (2.7, 3.5)	3.1 (2.7, 3.5)
Min, Max	0.6, 4.8	0.6, 4.5	0.6, 4.8

Non-abstinence smoking outcomes

Descriptive summaries of self-reported daily cigarette consumption and number of serious quit attempts for these participants are given in Table S13. These indicate that self-reported daily cigarette consumption was slightly lower in the MiQuit group than the control group, although this difference is small and could quite feasibly be the result of random variation. We also see that participants in the MiQuit group were more likely to report having made at least one serious quit attempt over the course of the study. 77.3% of MiQuit participants followed up at late pregnancy reported making at least one quit attempt, compared with 68.2% in the control group.

Table S13: Descriptive summaries of self-reported daily cigarette consumption and number of serious quit attempts made

Outcome	**Randomised treatment group**		
	MiQuit (N = 309)	**Control** (N = 337)	**Total** (N = 646)
Number of cigarettes smoked daily			
N	304	326	630
Mean (SD)	4.0 (3.9)	4.9 (5.0)	4.5 (4.5)
Median (Q1, Q3)	3.0 (0.0, 5.0)	4.0 (1.0, 7.0)	3.0 (0.0, 6.0)
Min, Max	0.0, 18.0	0.0, 25.0	0.0, 25.0
Quit attempts during study, n (%)			
Zero	64 (20.7)	94 (27.9)	158 (24.5)
One	70 (22.7)	68 (20.2)	138 (21.4)
Two	59 (19.1)	55 (16.3)	114 (17.6)
Three	54 (17.5)	46 (13.6)	100 (15.5)
Four	20 (6.5)	21 (6.2)	41 (6.3)
Five	12 (3.9)	15 (4.5)	27 (4.2)
Six	7 (2.3)	5 (1.5)	12 (1.9)
Seven	4 (1.3)	1 (0.3)	5 (0.8)
Eight	1 (0.3)	1 (0.3)	2 (0.3)
Nine	0 (0.0)	0 (0.0)	0 (0.0)
≥Ten	12 (3.9)	18 (5.3)	30 (4.6)
Missing	6 (1.9)	13 (3.9)	19 (2.9)

References

[1] Heinze G., Ploner M., Beyea j., Confidence intervals after multiple imputation: combining profile likelihood information from logistic regressions, Statistics in. Medicine., 2013; 32:5062-5076

[2] Heinze G., Ploner M., Dunkler D., Southworth H., logistf: Firth's Bias-Reduced Logistic Regression, R package version 1.23

[3] White I., *rctmiss: Stata module to analyse a RCT allowing for informatively missing outcome data* version 0.12.4, Statistical Software Components s458304, Boston College Department of Economics

[4] Palmer T. M., Ramasahai R. R., Didelez V., Sheehan N. A., *Nonparametric bounds for the causal effect in a binary instrumental variable model*, The Stata Journal 11(3): 345-367

[5] Nagelkerke N., Fidler V., Bernsen R., Borgdorff M., *Estimating treatment effects in randomized clinical trials in the presence of non-compliance*, Statistics in Medicine, 2000; 19:1849-1846

[6] Terza J., *Simpler Standard Errors for Two-Stage Optimisation Estimators,* The Stata Journal, 2016; 16(2): 368-385