Multi-centre double-blind placebo-controlled randomised trial of a combination of methotrexate and gefitinib versus methotrexate alone as a treatment for ectopic pregnancy:

The GEM3 Trial

Statistical Analysis Plan

SAP Version Number
2.0

Author:	Catherine Moakes	Role:	Trial Statistician	Affiliation:	BCTU University of Birmingham
Signature:	CAM	Date:	22/02/2022		
Chief Investigator:	Andrew Horne	Role:	Chief Investigator	Affiliation:	MRC Centre for Reproductive Health The Queen's Medical Research Institute
Signature:	AWH	Date:	23/02/2022		
This Statistical Analysis Plan has been approved by:					
Approver:	Lee Middleton	Role:	Senior Statistician	Affiliation:	BCTU University of Birmingham
Signature:	Lee Middleton	Date:	23/02/22		
This Statistical Analysis Plan has been reviewed by:					
Reviewer:	Jon Bishop	Role:	Senior Statistician	Affiliation:	BCTU University of Birmingham
Signature:	Jon Bishop	Date:	23/02/2022		

Table 1: SAP Amendments

SAP version number	Date Approved	Protocol version number†	Section number changed	Description of and reason for change	Timing of change with respect to interim/final analysis	Blind Reviewer
2.0	22/02/2022	9.0	NA (New version template)	SAP moved to the current BCTU SAP Template.	Prior to database lock for final analysis	Name: Jon Bishop Signature: Jon Bishop Date: 23/02/2022
				Update to definition of adherence to include women who do not receive an initial MTX injection or their allocated intervention as a result of a clinical change in condition (clinical change requiring immediate surgery post-randomisation) to be considered adherent.		Name: Jon Bishop Signature: Jon Bishop Date: 23/02/2022
				Update to analysis population to exclude any women who are randomised and consent was not obtained.		Name: Jon Bishop Signature: Jon Bishop Date: 23/02/2022
				An additional sensitivity analysis added for the primary outcome to exclude women who are found to violate the inclusion/exclusion criteria post-randomisation.		Name: Jon Bishop Signature: Jon Bishop Date: 23/02/2022

SAP version number	Date Approved	Protocol version number†	Section number changed	Description of and reason for change	Timing of change with respect to interim/final analysis	Blind Reviewer
				Update to adjustment covariates. Version 1.0 of the SAP included hCG (log transformed) and BMI as adjustment factors in their continuous form. Log transformation of hCG was considered to account for the fact it is a skewed variable. We propose adjusting for hCG in its categorised form (as per minimisation categories) to account for this. BMI will also be included in its categorised form for continuity in adjustment of analysis models in line with subgroup analyses. Centre will be considered a fixed effect for time to event outcomes.		Name: Jon Bishop Signature: Jon Bishop Date: 23/02/2022
				Sensitivity analysis using imputation methods removed and alternative approach for sensitivity analysis for missing data included (tipping point analysis for each treatment group).		Name: Jon Bishop Signature: Jon Bishop Date: 23/02/2022
				hCG level considered a key subgroup variable, BMI and ectopic size considered exploratory subgroup variables. For all subgroup analyses a ratio of the treatment effect within subgroups will also be provided.		Name: Jon Bishop Signature: Jon Bishop Date: 23/02/2022

SAP version number	Date Approved	Protocol version number†	Section number changed	Description of and reason for change	Timing of change with respect to interim/final analysis	Blind Reviewer
				P-values presented for the primary outcome only (removed from secondary outcomes), in line with current statistical practice.		Name: Jon Bishop Signature: Jon Bishop Date: 23/02/2022
				An additional sensitivity analysis added for the secondary outcome (time to hCG resolution) to change the threshold of resolution to ≤30 IU/L (from ≤15 IU/L) to account for the fact some sites do not follow-up women to a hCG ≤15 IU/L as standard care.		Name: Jon Bishop Signature: Jon Bishop Date: 23/02/2022

†This SAP was written based on information contained in the trial protocol version as listed here.

Abbreviations & Definitions	
Abbreviation / Acronym	**Meaning**
AE	Adverse Event
BCTU	Birmingham Clinical Trials Unit
BMI	Body Mass Index
CI	Confidence Interval
CONSORT	Consolidated Standards of Reporting Trials
CRF	Case Report Form
DMC	Data Monitoring and Ethics Committee
EP	Ectopic Pregnancy
ISRCTN	International Standard Randomised Controlled Trial Number
ITT	Intention to Treat
MNAR	Missing Not at Random
MTX	Methotrexate
NHS	National Health Service
RCT	Randomised Controlled Trial
RD	Risk Difference
RR	Risk Ratio
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SUSAR	Suspected Unexpected Serious Adverse Reaction
TSC	Trial Steering Committee
UK	United Kingdom
Term	**Definition**
International Standard Randomised Controlled Trial Number	A clinical trial registry
Protocol	Document that details the rationale, objectives, design, methodology and statistical considerations of the study
Randomisation	The process of assigning trial subjects to intervention or control groups using an element of chance to determine the assignments in order to reduce bias.
Statistical Analysis Plan	Pre-specified statistical methodology documented for the trial, either in the protocol or in a separate document.

TABLE OF CONTENTS

Introduction

This document is the Statistical Analysis Plan (SAP) for the GEM3 trial and should be read in conjunction with the current trial protocol. This SAP details the proposed analyses and presentation of the data for the main paper(s) reporting the results for the GEM3 trial.

The results reported in these papers will follow the strategy set out here. Subsequent analyses of a more exploratory nature will not be bound by this strategy, though they are expected to follow the broad principles laid down here. The principles are not intended to curtail exploratory analysis (e.g. to decide cut-points for categorisation of continuous variables), nor to prohibit accepted practices (e.g. transformation of data prior to analysis), but they are intended to establish rules that will be followed, as closely as possible, when analysing and reporting data.

Any deviations from this SAP will be described and justified in the final report or publication of the trial (using a table as shown in Appendix A). The analysis will be carried out by an appropriately qualified statistician, who should ensure integrity of the data during their data cleaning processes.

Background and rationale

The background and rationale for the trial are outlined in detail in the protocol. In brief, GEM3 is a multi-centre, double-blind, randomised controlled trial (RCT) comparing the use of a combination of oral gefitinib and intramuscular methotrexate (MTX) against MTX and placebo as a non-surgical treatment to treat tubal ectopic pregnancy (EP).

Trial objectives

The primary objective is to test the hypothesis that administering gefitinib in combination with MTX reduces the rate of surgical intervention ('rescue surgery') for the treatment of tubal EP when compared with MTX alone.

Secondary objectives are as follows:

To assess the effects of a combination of MTX and gefitinib compared with MTX treatment alone, on:

The need for additional treatment with MTX.

The number of days to resolution of the pregnancy (resolution defined as serum human chorionic gonadotropin (hCG) ≤ 15 IU/L).

The number of hospital visits associated with treatment until resolution or scheduled/emergency 'rescue' surgery for EP.

To assess the safety and tolerability of combination MTX and gefitinib compared with MTX treatment alone.

Trial methods

Trial design

GEM3 is a multi-centre, prospective, double-blind, superiority, parallel group, phase III placebo-controlled randomised trial. Participants will be recruited in the United Kingdom (UK). Participants will be randomised to MTX and gefitinib or MTX and placebo. See trial schema in Appendix B.

Trial interventions

The intervention group will be given a single-dose intramuscular MTX (50 mg/m²) injection (standard care) with seven daily doses of oral gefitinib (250 mg). The control group will be given a single-dose intramuscular MTX (50 mg/m²) injection with seven daily doses of placebo (matched to the oral gefitinib tablet).

Primary outcome measure

The primary outcome is surgical intervention for the treatment of EP (salpingectomy or salpingostomy by laparoscopy or laparotomy).

Secondary outcome measures

Secondary outcomes are as follows:

Use of additional MTX (as directed by standardised expected change in serum hCG level).

Time to hCG resolution (days) from randomisation until hCG levels ≤15IU/L.

Number of hospital visits associated with treatment until resolution or scheduled/emergency surgery.

Safety/tolerability (adverse events (AE)).

Acceptability of treatment (assessed 3 months after resolution of EP by Likert score).

Return to menses (assessed 3 months after resolution of EP).

Timing of outcome assessments

The schedule of trial procedures and outcome assessments are given in Appendix C.

Randomisation

Participants will be randomised in a 1:1 ratio to either MTX and gefitinib or MTX and placebo.

Randomisation will be provided by a secure online randomisation system at the Birmingham Clinical Trials Unit (BCTU) using a minimisation algorithm incorporating the following factors:

Baseline serum hCG levels (<1500 IU/L, ≥1500 to <2500 IU/L, ≥2500 IU/L)
Body Mass Index (BMI) (<25 kg/m^2, ≥25 kg/m^2)
Ectopic size on ultrasound (<2cm, ≥2cm)
Centre

A 'random element' will be included in the minimisation algorithm, so that each participant has a probability (unspecified here), of being randomised to the opposite intervention that they would have otherwise received. Full details of the algorithm used will be stored in a confidential document at BCTU.

Sample size

The target sample size of 328 participants is based on data taken from the Phase II study (GEM2), published cohort data [1] and an audit of Edinburgh and Imperial of women undergoing usual care (2012; focusing on those participants with serum hCG >1000 IU/L). The cohort and audit data suggest a 30% rate of surgical intervention in the MTX group, with 15% expected in the MTX and gefitinib group (actual figure from the phase II study was 14% but this has been conservatively rounded up). To detect this absolute difference in proportions of 15% with 90% power and an alpha error rate of 5%, a total of 322 participants would need to be randomised (161 per group). Assuming and adjusting for a 2% loss to follow-up rate, 328 participants need to be recruited.

Framework

The objective of the trial is to test the superiority of one intervention to another. The null hypothesis is that there is no difference in surgery rates between women allocated to MTX and gefitinib vs. women allocated to MTX and placebo. The alternative hypothesis is that there is a difference between the groups.

Interim analyses and stopping guidance

If gefitinib in combination with MTX is overwhelmingly better or worse than MTX alone with respect to reducing the rate of surgical intervention for resolution of EP, then this effect may become apparent before the target recruitment has been reached. Alternatively, new evidence could emerge from other sources to suggest that gefitinib in combination with MTX is definitely more, or less, effective than MTX alone. To protect against any unnecessary continuance of the

trial in this event, interim analyses of major endpoints and safety data will be supplied during the period of recruitment to the study, in strict confidence, to the Data Monitoring Committee (DMC) along with updates on results of other related studies, and any other analyses that the DMC may request.

The DMC will advise the chair of the Trial Steering Committee (TSC) if, in their view, any of the randomised comparisons in the trial have provided both: a) proof beyond reasonable doubt that for all, or for some, types of participant one particular intervention is definitely indicated or definitely contra-indicated in terms of a net difference of a major endpoint, and b) evidence that might reasonably be expected to influence the patient management of many clinicians who are already aware of the other main trial results. Unless this happens, however, the TSC, the collaborators and all of the central Trial staff (except the statisticians who supply the confidential analyses) will remain ignorant of the interim results.

Appropriate criteria of proof beyond reasonable doubt cannot be specified precisely, but a difference of at least $p<0.001$ (similar to a Haybittle-Peto [2] stopping boundary) in an interim analysis of a major endpoint may be required to justify halting, or modifying, the study prematurely. If this criterion were to be adopted, it would have the practical advantage that the exact number of interim analyses would be of little importance, so no fixed schedule is proposed. Given the proposed use of the Haybittle-Peto boundary no adjustment for multiple testing (to control the overall type I error rate) is proposed, i.e. the threshold for statistical significance at final analysis will still be $p=0.05$.

A separate DMC reporting template will be drafted and agreed by the DMC including an agreement on which outcomes will be reported at interim analyses. The statistical methods stated in this SAP will be followed for the agreed outcomes.

Timing of final analysis

The final analysis for the trial will occur after all participants have completed primary and secondary outcomes (where these are expected) and the corresponding outcome data has been entered onto the trial database and validated as being ready for analysis. This is provided that the trial has not been stopped early for any reason (e.g. TSC/DMC advice or funding body request).

Timing of other analyses

Further analysis will occur at 12 months post-randomisation on long-term follow-up data on future pregnancies and fertility. This analysis will subject to a separate SAP.

Trial comparisons

All references in this document to 'group' refer to MTX and gefitinib or MTX and placebo.

Statistical Principles

Confidence intervals and p-values

All estimates of differences between groups will be presented with two-sided 95% confidence intervals, unless otherwise stated. For the primary outcome (ITT, sensitivity/supportive and subgroup analyses), a p-value will be produced, with statistical significance considered at the 5% level.

Adjustments for multiplicity

No correction for multiple testing will be made.

Analysis populations

All primary analyses (primary and secondary outcomes including safety outcomes) will be by intention-to-treat (ITT). Participants will be analysed in the intervention group to which they were randomised, and all participants shall be included whether or not they received the allocated intervention. This is to avoid any potential bias in the analysis [3]. However, in rare instances where a participant is randomised and is later found to violate the inclusion criteria because no consent was collected (as randomised in error), they will be withdrawn and removed from any statistical analyses. Refer to section 5.4 for definition of adherence and definition of the per-protocol population and section 9.10 for further details on such sensitivity/supportive analyses.

Definition of adherence

Adherence to the allocated intervention (gefitinib/placebo) will be monitored by the participant's self-reported account of whether they took their allocated treatment, including total number of tablets taken. This data will be collected and recorded at each study visit. Adherence to MTX will be reported by the clinician on whether the initial methotrexate injection was given. We will define adherence as those participants who receive their initial methotrexate injection and took at least 75% of their allocated treatment prior to resolution (up to a maximum of seven daily doses if resolution not occurred by seven days post-randomisation). Any women who do not receive an initial MTX injection or their allocated intervention as a result of a clinical change in condition (clinical change requiring immediate surgery post-randomisation) will be considered adherent (it is anticipated that this is equally likely to occur in both the MTX and gefitinib and MTX and placebo

groups). Those women who are considered adherent will form the per-protocol (adherent) population.

Handing protocol deviations and violations

A protocol deviation/violation is defined as a failure to adhere to the protocol such as errors in applying the inclusion/exclusion criteria, the incorrect intervention being given, incorrect data being collected or measured or missed follow-up visits. We will include all participants in the primary analysis as per the ITT population described in section 5.3, regardless of deviation from the protocol, unless stated as a planned post-randomisation exclusion in section 5.3 [3]. This includes participants who were randomised but later found to violate the inclusion or exclusion criteria, however the outcomes in these women will be explored in a sensitivity analysis (see section 9.10). This does not include those participants who were randomised (and consented) and went on to withdraw consent post-randomisation for the use of their data. However, these women will be explored as per other missing responses.

Unblinding

For double-blind studies, the unblinding of the Trial Statistician to the intervention code will take place once the database is locked for final analysis unless the DMC request that they review the interim data with knowledge of the intervention groups or the DMC request to be unblinded at an interim analysis.

Trial population

Recruitment

A flow diagram (as recommended by Consolidated Standards of Reporting Trials (CONSORT) [4]) will be produced to describe the participant flow through each stage of the trial. This will include information on the number (with reasons) of losses to follow-up (drop-outs and withdrawals) over the course of the trial. A template for reporting this is provided in section 3 of the supplementary analysis report.

Baseline characteristics

The trial population will be tabulated as per section 6 of the supplementary analysis report. Categorical data will be summarised by number of participants, counts and percentages. Continuous data will be summarised by the number of participants, mean and standard deviation if deemed to be normally distributed or number of participants, median and interquartile range if data appear skewed, and ranges if appropriate. Tests of statistical significance will not be undertaken, nor confidence intervals presented [5].

Intervention
Description of the intervention
A template for reporting information on the intervention is provided in section 7 of the supplementary analysis report.
Adherence to allocated intervention
A cross-tabulation of allocated intervention by the adherence categories stated in section 5.4 will be produced (proportions and percentages). A template for reporting adherence is provided in section 7 of the supplementary analysis report.
Protocol deviations
Frequencies and percentages by group will be tabulated for the protocol deviations and violations as per section 5 of the supplementary analysis report.
Analysis methods
Intervention groups will be compared using regression models to adjust for all covariates as specified in section 9.1 where possible.
Covariate adjustment
In the first instance, intervention effects between groups for all outcomes will be adjusted for the minimisation parameters listed in section 4.6. Centre will be included with a random intercept in the model, and all other factors as fixed effects (with the exception of time to event outcomes where centre will be regarded as a fixed effect). If covariate adjustment is not possible (e.g. the model does not converge), randomising centre will be removed first. If this reduced model still fails to converge, unadjusted estimates will be produced and it will be made clear in the final report why this occurred (e.g. not possible due to low event rate/lack of model convergence). For binary outcomes only, if the (full) adjusted log-binomial model fails to converge (when estimating a risk ratio); a Poisson regression model with robust standard errors will be used to estimate the same parameters [6]. If this also fails to converge, estimates will be produced from the log-binomial model (following rules for removal of variables as outlined above). It will be made clear in the final report why this occurred (e.g. not possible due to low event rate/lack of model convergence).
Distributional checks
There are no continuous outcome measures proposed in the GEM3 trial.

For time to event outcomes (analysed using a cox regression model), the proportional hazard assumption will be assess visually using log cumulative hazard versus log time plots. If these plots identify potential non-proportional hazards, time-varying covariates will be considered.

Handling missing data

In the first instance, analysis will be completed on received data only with every effort made to follow-up participants even after protocol violation to minimise any potential for bias. To examine the possible impact of missing data on the results, and to make sure we are complying with the ITT principle, sensitivity analyses will be performed on the primary outcome measure (surgical intervention) [7]. See section 9.10 for further details.

Data manipulations

See Appendix D.

Analysis methods – primary outcome

A template for reporting the primary outcome is given in section 8.1 of the supplementary analysis report. The primary outcome will be summarised using frequencies and percentages. A mixed effects log-binomial model will be used to generate an adjusted risk ratios (RR) (and 95% CIs). Adjusted risk differences (RD) (and 95% CIs) will also be presented (using an identity link function). Statistical significance of the treatment group parameter will be determined (p-value generated) through examination of the associated chi-squared statistic (this will be obtained from the model which produces the RR). See section 9.1 for covariate adjustment and model convergence.

Analysis methods – secondary outcomes

A template for reporting the secondary outcomes is given in section 8.2 of the supplementary analysis report. See section 9.1 for covariate adjustment and model convergence.

Time to hCG resolution will be considered in a competing risk framework [8]. This is to account for participants who have surgical intervention for their EP, which interferes with the natural path of hCG levels and hence hCG resolution (≤15 IU/L). A cumulative incidence function will be used to estimate the probability of occurrence (hCG resolution) over time. This method is favoured over Kaplan-Meier methods which may overestimate this probability. A Fine-Gray model will be used to estimate subdistribution adjusted hazard ratios (and 95% CIs) directly from the cumulative incidence function. In addition, a further Cox Proportional Hazard model will be fitted and applied to the cause-specific (non-surgical resolution) hazard function and used to generate

adjusted hazard ratios (and 95% CIs) (effectively censoring surgical events at the point of surgery) [9].

Time to event data (return to menses) will be summarised using medians and interquartile ranges. A Cox Proportional Hazard model will be fitted to generate an adjusted hazard ratio (and 95% CI) and a Kaplan Meier plot will be produced to assess the data visually.

Binary outcomes (additional MTX, adverse events) will be analysed as per the primary outcome.

Categorical outcomes of an ordinal nature (acceptability of treatment measured using a Likert scale) will be summarised using frequencies and percentages. An ordinal logistic regression model will be used to generate adjusted odds ratios (and 95% CIs).

For secondary outcome measures which measure counts (number of hospital visits associated with treatment until resolution or scheduled/emergency surgery), this data will be summarised using medians with interquartile ranges. If appropriate, a Poisson model will be used to generate incidence rate ratios (and 95% CIs). An offset term will be included in each model to account for length of exposure time (see data manipulations (Appendix D) for how this will be derived). A negative binomial regression model may be considered if the data are overdispersed.

Analysis methods – exploratory outcomes and analyses

Any data that does not form a pre-specified outcome will be presented using simple summary statistics by intervention group (i.e. numbers and percentages for binary data and means (or medians) and standard deviations (or inter-quartile ranges) for continuous normal (or non-normal) data).

Safety data

AEs are considered a secondary outcome and will be analysed as per section 9.6 of the SAP. The number and percentage of women experiencing any serious adverse events (SAEs) and suspected unexpected serious adverse reactions (SUSARs) will be presented by intervention group alongside the number of events reported. Statistical significance will be determined (p-value generated) through examination of the associated chi-squared statistic. A template for reporting is given in section 9 of the supplementary analysis report.

Planned subgroup analyses
Interpretation of subgroup analyses will be treated with caution (output will be treated as exploratory rather than definitive [10]). Analyses will be limited to the primary outcome only, and the following subgroups: Baseline serum hCG levels (<1500 IU/L, ≥1500 to <2500 IU/L, ≥2500 IU/L) BMI (<25 kg/m^2, ≥25 kg/m^2) Ectopic size on ultrasound (<2cm, ≥2cm) Baseline serum hCG levels is the single key subgroup of interest (i.e. we propose to be able to draw a firm conclusion about any differential effect with respect to this variable only). We hypothesise that there will be a greater differential effect of treatment (reduction in surgical intervention with MTX and gefitinib) for those with lower baseline serum hCG levels. Other subgroup analyses will be considered exploratory. The effects of these subgroups will be examined by including a treatment group by subgroup interaction parameter in the regression model. P-values from tests for statistical heterogeneity will be presented alongside the effect estimate and 95% confidence interval within each subgroup. In addition to this, a ratio(s) (and 95% CI(s)) will be provided to quantify the difference between the treatment effects estimated within each subgroup. For BMI and ectopic size as these subgroup variables contain only two levels, one ratio will be provided. For hCG level, which has three levels, '≥2500 IU/L' will be considered as the reference group and two ratios will be provided for the other levels in comparison to this reference. Given there is only one key subgroup of interest, no adjustments will be made for multiple testing. A template for reporting the subgroup analyses for the primary outcome is given in section 8.1.5 of the supplementary template report.
Sensitivity or supportive analyses
Sensitivity analyses for the primary outcome only will consist of: A per-protocol (adherent) analysis (population described in sections 5.3 and 5.4); A sensitivity analysis which excludes any women who are found to violate the inclusion/criteria. An analysis to assess the effect of missing responses for the primary outcome only. Unadjusted models will be utilised. This analysis will explore the possibility that missing responses are 'missing not at random' (MNAR) using a tipping point approach. In this analysis, for women with missing outcome data, events will be added sequentially in each of the groups in turn to

determine the point where the general conclusion changes (for example from positive to inconclusive in terms of the CI). An assessment can then be made about whether the event rate in the missing responses between groups is likely to be plausible when compared with the event rate in the non-missing data. Two scenarios will be considered as follows:

Scenario A: In the MTX and placebo group, assume all missing responses are non-events (i.e. surgical intervention is no). For missing responses in the MTX and gefitinib group, we will first replace X missing responses with events (i.e. X additional cases of surgical intervention) where X is the number of events such that the event rate in the missing responses is equal to the event rate in the non-missing responses in the MTX and gefitinib group (for example if the event rate in the non-missing data in the MTX and gefitinib group is 30% and we have 10 missing responses we will add 0.3*10 ~ 3 events [number of events rounded to the closest integer >0]). This will be regarded as the base case. All other missing responses in the MTX and gefitinib group will be considered as non-events. An unadjusted model will be fitted. Event rates will be compared between groups. The CI from the treatment estimate (from the RR of the log-binomial model) will be examined and stored. Following the base case, an additional event will be added to the MTX and gefitinib group and the above procedure repeated (unadjusted model fitted and the RR and CI examined). This process will end when the number of events added to the MTX and gefitinib group are equal to the original number of missing responses in this group. The tipping point for the MTX and gefitinib group will occur when enough events have been added such that the upper/lower limit of the CI from the corresponding model differs from that of the primary ITT finding (in regards to whether the CI crosses the null value of one).

Scenario B: In the MTX and gefitinib group, assume all missing responses are non-events (i.e. surgical intervention is no). For missing responses in the MTX and placebo group, we will first replace X missing responses with events (i.e. X additional cases of surgical intervention) where X is the number of events such that the event rate in the missing responses is equal to the event rate in the non-missing responses in the MTX and placebo group (for example if the event rate in the non-missing data in the MTX and placebo group is 30% and we have 10 missing responses we will add 0.3*10 ~ 3 events [number of events rounded to the closest integer >0]). This will be regarded as the base case. All other missing responses in the MTX and placebo group will be considered as non-events. An unadjusted model will be fitted. Event rates will be compared between groups. The CI from the treatment estimate (from the RR of the log-binomial model) will be examined and stored. Following the base case, an additional event will be added to the MTX and placebo group and the above procedure repeated (unadjusted model fitted and the RR and

CI examined). This process will end when the number of events added to the MTX and placebo group are equal to the original number of missing responses in this group. The tipping point for the MTX and placebo group will occur when enough events have been added such that the upper/lower limit of the CI from the corresponding model differs from that of the primary ITT finding (in regards to whether the CI crosses the null value of one).

For both scenario A and scenario B, results will be presented visually. The event rate (surgical intervention rate) in the missing data in the MTX and gefitinib group (scenario A) or MTX and placebo group (scenario B) will be plotted on the x-axis. The RR (MTX and gefitinib vs. MTX and placebo) from the model will be on the plotted on the y-axis. The corresponding CI will be included around each RR. If the CI in the ITT analysis contains one, then the point where the upper CI falls below or above one will be highlighted (tipping point) on the plot. If the CI in the ITT analysis does not contain one (i.e. one treatment is superior to the other), then the point where the CI crosses one will be highlighted (tipping point). The base case will also be highlighted on each plot.

Sensitivity analyses for the secondary outcomes only will consist of:

A sensitivity analysis where a threshold of ≤30 IU/L is used for time to hCG resolution (days).

Analysis of sub-randomisations

Not applicable.

Health economic analysis

No health economic analyses are planned.

Statistical software

Statistical analysis will be undertaken in the following statistical software packages: Stata and SAS.

References

BMC Pregnancy Childbirth. 2013 13:30; Obstet Gynecol Int. 2014 2014:423708

Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV et al. Design and analysis of randomised clinical trials requiring prolonged observation of each patient. I. Introduction and design. Br J Cancer. 1976; 34:585-612.

Gupta SK. Intention-to-treat concept: A review. Perspect Clin Res. 2011;2(3):109-112.

Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c332.

Altman DG, Dore CJ. Randomisation and baseline comparisons in clinical trials. Lancet. 1990;335:149–53.

Zou G. A modified Poisson regression approach to prospective studies with binary data. Am J Epidemiol. 2004;159(7):702-6.

White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. BMJ. 2011;342:d40.

Austin PC, Fine JP. Accounting for competing risks in randomized controlled trials: a review and recommendations for improvement. Statistics in Medicine. 2016; DOI: 10.1002/sim.7215.

A. Latouche et al. A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. 2013; Volume 66, Issue 6, Pages 648:653.

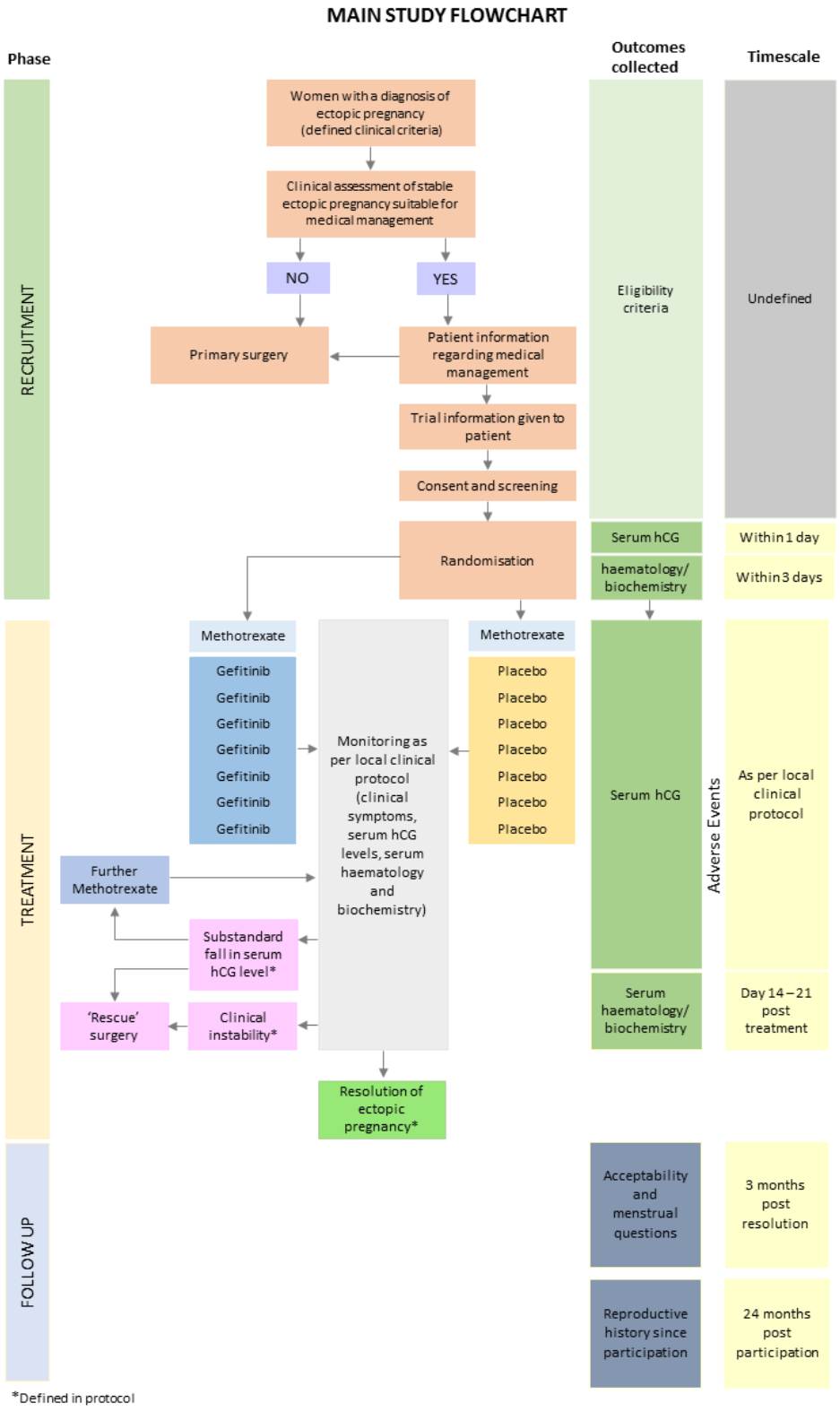
Wand R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Reporting of subgroups analyses in clinical trials. NEJM. 2007;357:2189-94.

Appendix A: Deviations from SAP

This report below follows the statistical analysis plan dated <insert effective date of latest SAP> apart from following:

Section of report not following SAP	Reason
<insert section >	<insert, e.g. exploratory analyses request by TMG>

Appendix B: Trial Schema



Appendix C: Schedule of assessments

	Day of Treatment	Standard Clinical Care Visits	Day 14 – 21^ (during scheduled clinical care visits)	Month 3
Consent	✓	✓*	✓*	✓*
Clinical Assessment	✓	✓	✓	
Serum hCG (within 1 day)	✓	✓	✓	
FBC, U&Es, LFTs (within 3 days of treatment)**	✓		✓	
Adverse Events		✓***	✓	
Compliance		✓		
Telephone consultation				✓

*Ongoing consent will be confirmed at each visit

**Until back to within normal ranges and stop date documented where applicable

***If this clinical appointment falls when research staff are not available, the local research team will contact the participant at the earliest possible date to ascertain any AEs.

^ Sites will define day of MTX treatment as either day 0 or day 1 depending on local policies.

Appendix D: Data Manipulations

The Trial Statistician will derive the following measures as follows:

Outcome measures

Surgical intervention for the treatment of EP (salpingectomy or salpingostomy by laparoscopy or laparotomy)

Reported on the surgical intervention form. If a surgery form has been completed and either 'salpingectomy' or 'salpingostomy/salpingotomy' are yes then this will meet the primary outcome definition (i.e. surgical intervention=yes). On the post-treatment visit form, if a woman has been followed up until her hCG levels are ≤15 IU/L and 'surgical intervention required' is no (or surgical intervention is yes but neither a salpingectomy or salpingostomy/salpingostomy have been performed) then this will not meet the primary outcome (i.e. surgical intervention=no). On the post-treatment visit form, if a woman has not been followed up until her hCG levels are ≤15 IU/L, but we have confirmation from site that no surgery was performed then this will not meet the primary outcome (i.e. surgical intervention=no).

Use of additional MTX

Use of additional MTX is reported on the post-treatment visit form. If 'further dose of MTX required today or since last visit' is yes at any visit (up to resolution) then this will meet the outcome definition (i.e. use of additional MTX=yes). If 'further dose of MTX required today or since last visit' is no at all visits (up to resolution) then this will not meet the outcome definition (i.e. use of additional MTX=no). If a woman does not have a post-treatment visit as a result of immediate surgery following randomisation we will assume additional MTX=no. All other responses will be considered missing.

Time to hCG resolution (days) from randomisation until hCG levels ≤15 IU/L

If a woman is followed up to a hCG ≤15 IU/L (and did not receive surgery or hCG was ≤15 IU/L prior to receiving surgery) the pregnancy will be regarded as having resolved. Date of randomisation will be regarded as time 0 and time to hCG resolution will be derived as follows:

Time to hCG resolution (days) = (Date of hCG measurement*-Date of randomisation)

**Where date of hCG measurement is the earliest date post-randomisation where hCG level is ≤15 IU/L.*

If a woman has not been followed up to a hCG ≤15 IU/L (and did not receive surgery) the pregnancy will be regarded as unresolved and the woman will be censored at the point of her most recent follow-up visit where a hCG measurement was reported. Time to follow-up for censoring will be derived as follows:

Time to follow-up (days) = (Date of hCG measurement**-Date of randomisation)

***Where date of hCG measurement is the most recent date post-randomisation where hCG level is measured.*

If a woman received surgery (salpingectomy or salpingostomy/salpingotomy) prior to hCG resolution (≤15 IU/L), the pregnancy will be regarded as unresolved and the woman will be censored at the point of surgery (as a competing risk). Time to surgery for censoring will be derived as follows:

Time to surgery (days) = (Date of surgery-Date of randomisation)

If time to hCG resolution, time to follow-up or time to surgery are equal to zero, a continuity correction of 0.01 days will be added (Stata unable to process zero time).

Time to hCG resolution (days) from randomisation until hCG levels ≤30 IU/L (Sensitivity analysis)

As per 'time to hCG resolution (days) from randomisation until hCG levels ≤15 IU/L' but hCG threshold changed to ≤30 IU/L.

Number of hospital visits associated with treatment until resolution or scheduled/emergency surgery

Number of hospital visits associated with treatment are reported on the post-treatment visit form.

Remove any post-treatment visits which occur post resolution (where resolution occurs when a woman receives surgery (salpingectomy or salpingostomy/salpingostomy) or her hCG levels fall ≤15 IU/L).

Firstly, we look at the number of extra visits related to the EP:

If 'Have you had any extra visits to hospital related to this EP'=no then number of additional visits=0.
If 'Have you had any extra visits to hospital related to this EP'=yes then number of additional visits should be provided.
If a woman does not have a post-treatment visit as a result of immediate surgery following randomisation we will assume number of extra visits=0.
Where a woman has X post-treatment visits, sum the total number of extra visits (A) as follows:

A=(Number of extra visits at visit 1 + Number of extra visits at visit 2 + … + Number of extra visits at visit X)

Secondly, we look at the number of hospital attendances:

Remove any post-treatment visits where a woman did not have a hospital visit (hCG measurement/USS scan/additional MTX used as a proxy for hospital attendance i.e. if none of these are performed we can assume the visit was conducted remotely). Count the number of post-treatment visits with hospital attendance (X). The number of hospital attendances (B) can be calculated as follows:

If a woman had surgery and this occurred on a different day to a post-treatment visit then

B=X+1

If a woman did not have surgery or surgery occurred on the same day as a post-treatment visit then

B=X

Finally, the total number of hospital visits associated with treatment can then be calculated as follows:

Total number of hospital visits associated with treatment=A+B

Length of follow-up (in days) for the offset term for the Poisson model will be calculated as:

Date of resolution*-Date of randomisation

*Date of resolution is date of surgery (salpingectomy or salpingostomy/salpingostomy) or earliest date post-randomisation where hCG level is ≤15 IU/L. If a date of resolution is missing (no surgery and not followed up to a hCG ≤15 IU/L) and site have confirmed woman's pregnancy has resolved and no surgery performed, date of most recent hCG measurement will be used as date of resolution.

Safety/tolerability (adverse events)
Adverse events are reported on the adverse event log.

Acceptability of treatment (assessed 3 months after resolution of EP by Likert score).
Acceptability of treatment is reported on the acceptability and return of menses questionnaire.

Return to menses (assessed 3 months after resolution of EP)
Return to menses is reported on the acceptability and return of menses questionnaire.

Women will be excluded from this analysis if resolution data unknown (resolution data unknown if a woman is not followed up to a hCG ≤15 IU/L and we do not have confirmation from site if surgery was performed).

If a woman's menses have returned based on the following responses to 'when did your period return after your treatment', then they will considered as having an event and time to menses will be derived as follows:

Time to menses=10 days if 'when did your period return after your treatment'=1-2 weeks.
Time to menses=24 days if 'when did your period return after your treatment'=3-4 weeks.

Time to menses=38 days if 'when did your period return after your treatment'=5-6 weeks.
Time to menses=52 days if 'when did your period return after your treatment'=7-8 weeks.
Time to menses=66 days if 'when did your period return after your treatment'=9-10 weeks.
Time to menses=80 days if 'when did your period return after your treatment'=11-12 weeks.

If 'when did your period return after your treatment'=Not returned then the woman will be regarded as not having an event and will be censored at the point of questionnaire completion. Time to form completion for censoring will be derived as follows:

Time to form completion (days) = (Date of questionnaire completion-Date of resolution*)

*Date of resolution is date of surgery (salpingectomy or salpingostomy/salpingostomy) or earliest date post-randomisation where hCG level is ≤15 IU/L. If a date of resolution is missing (no surgery and not followed up to a hCG ≤15 IU/L) and site have confirmed woman's pregnancy has resolved and no surgery performed, date of most recent hCG measurement will be used as date of resolution.

If no menses data is available, a woman will be censored at time zero (a continuity correction of 0.01 days will be applied to such instances, as Stata is unable to process zero time).

Other measures

Adherence

Reported on baseline (initial data collection) form. If 'Date of MTX' is non-missing then assume initial MTX injection given. Number of tablets of trial medication taken is reported on the post-treatment visit form, surgery form and withdrawal form. The proportion of trial medication taken will be calculated as follows based on two scenarios:

If a woman has resolved* <7 days post-randomisation then:

Proportion of trial medication taken=Number of tablets taken/(Date of resolution-date of randomisation)

If a woman has not resolved* <7 days post-randomisation then:

Proportion of trial medication taken=Number of tablets taken/7

A woman is considered adherent if the initial MTX injection is given and the proportion of trial medication taken is ≥0.75. Any women who do not receive an initial MTX injection or their allocated intervention as a result of a clinical change in condition (clinical change requiring immediate surgery post-randomisation) will be considered adherent.

**Resolution occurs when a woman receives surgery (salpingectomy or salpingostomy/salpingostomy) or her hCG levels fall ≤15 IU/L.*

Ethnicity

White (British/Irish, Other North/West European, East European, South European, Other White)

Asian (Indian, Pakistani, Bangladeshi, Other Asian)

Black (African, Caribbean, Black Other)

Mixed (White/Black Caribbean, White/African, White/Asian, Any other mixed background)

Other (Other, Other Oriental)

Chinese (Chinese)

Woman's age at randomisation (years)

Maternal age=(Date of randomisation-DOB)/365.25