The Recognition of STEMI by Paramedics and the Effect of Computer inTerpretation (RESPECT): a randomised crossover feasibility study.

Corresponding author
Richard Pilbery
Research Paramedic
Yorkshire Ambulance Service NHS Trust
Springhill, Brindley Way
Wakefield 41 Business Park
Wakefield
WF2 0XQ
Email: r.pilbery@nhs.net
Mobile: 07463 73482

Co-authors
M. Dawn Teare
Design, Trials & Statistics
School of Health and Related Research (ScHARR)
University of Sheffield
Regent Court
30 Regent Street
Sheffield S1 4DA
Email: m.d.teare@sheffield.ac.uk

Steve Goodacre
Health Services Research
School of Health and Related Research (ScHARR)
University of Sheffield
Regent Court
30 Regent Street
Sheffield S1 4DA
Email: s.goodacre@sheffield.ac.uk

Francis Morris
Emergency Department
Northern General Hospital
Sheffield Teaching Hospitals NHS Foundation Trust
Herries Road
Sheffield
S5 7AU
Email: francis.morris@sth.nhs.uk

Word Count: *3023 words*

ABSTRACT

Background The appropriate management of patients with ST-segment elevation myocardial infarction (STEMI) depends on accurate interpretation of the 12-lead ECG by paramedics. Computer interpretation messages on ECGs are often provided, but the effect they exert on paramedics' decision making is not known. The objective of this study was to assess the feasibility of using an online assessment tool, and collect pilot data, for a definitive trial to determine the effect of computer interpretation messages on paramedics' diagnosis of STEMI.

Methods The RESPECT feasibility study was a randomised crossover trial using a bespoke, web-based assessment tool. Participants were randomly allocated 12 of 48 ECGs, with an equal mix of correct and incorrect computer interpretation messages, and STEMI and STEMI-mimics. The nature of the responses required a cross-classified multi-level model.

Results 254 paramedics consented into the study, 205 completing the first phase and 150 completing phase two. The adjusted odds ratio (OR) for a correct paramedic interpretation, when the computer interpretation was correct (true positive for STEMI or true negative for STEMI-mimic), was 1.80 (95% CI 0.84–4.91), and 0.58 (95%CI 0.41–0.81) when the computer interpretation was incorrect (false positive for STEMI or false negative for STEMI-mimic). The intraclass correlation coefficient (ICC) for correct computer interpretations was 0.33 for participants and 0.17 for ECGs, and for incorrect computer interpretations, 0.06 for participants and 0.01 for ECGs

Conclusion Determining the effect of computer interpretation messages using a web-based assessment tool is feasible, but the design needs to take clustered data into account. Pilot data suggest that computer messages influence paramedic interpretation, improving accuracy when correct and worsening accuracy when incorrect.

WHAT THIS PAPER ADDS
Section 1: What is already known on this subject
Timely diagnosis and appropriate management of patients with ST-segment elevation myocardial infarction (STEMI) depends on accurate interpretation of the 12-lead ECG by paramedics.
The design of existing studies do not generally account for clustering of data and/or enable assessment of the effect computer diagnostic messages exert on paramedics' decision making

Section 2: What this study adds
The RESPECT feasibility study has demonstrated that it is possible to conduct a randomised crossover trial to test the accuracy of STEMI recognition by paramedics, using an online assessment tool.
Pilot data suggest that computer messages influence paramedic interpretation, improving accuracy when correct and worsening accuracy when incorrect.

Pilot data show marked clustering, which must be taken into account in study design and analysis

INTRODUCTION
Background
In 2013/14, 31,653 patients in England, Wales and Belfast were diagnosed with STEMI, and over 85% of the 21,602 patients who received either mechanical or chemical clearance of their coronary arteries, were transported to hospital by ambulance.[1] Clinical trials have demonstrated a clear link between patient outcome and time to reperfusion in STEMI.[2] However, timely diagnosis and appropriate management of patients with STEMI depends on accurate interpretation of the 12-lead ECG by paramedics. In the UK, computer aided interpretation of the 12-lead ECG is typically available on ECG monitors carried by ambulance services. Studies have demonstrated that computer interpretation is 58–78% sensitive and 90–100% specific, with false positive rates varying between 19–39%.[3–5] Although there are studies directly examining the effect of computer interpretation on the diagnostic accuracy of doctors, these are not typically limited to STEMI only and none include paramedics.[6–8]

In contrast to early studies examining paramedics' safe administration of thrombolysis, false-positive rates for pPCI referral, have been reported to be 20–31%,[9–12] possibly due to poor ECG acquisition, misinterpretation of the ECG and/or the perception that the consequences of a decision to transfer for pPCI are less severe than administering thrombolytics.[13] However, inappropriate referral to pPCI centres has potential cost implications, may contribute to staff burnout, particularly for hospital staff who are called in from home out-of-hours, and result in longer patient transport times to a regional pPCI centre, rather than the local emergency department (ED). False negatives are equally undesirable, since failure to identify and appropriately manage patients with STEMI, is more likely to result in delayed time to reperfusion, with the subsequent negative impact on mortality and morbidity.[2]

In addition, a weakness of previous studies has been the failure to take account of the clustered nature of their data. In studies of ECG interpretation, data may be clustered by ECG or by clinician, i.e. interpretations of the same ECG are more likely to be similar (correct or incorrect) than interpretations of different ECGs, and interpretations made by the same clinician are more likely to be similar than interpretations made by different clinicians. Standard statistical tests assume that all interpretations are independent of each other. If data are clustered by ECG and clinician then this assumption may be violated, the study may be underpowered and analysis using standard statistical tests may underestimate p-values and confidence intervals resulting in the wrong conclusions being drawn.

The intra-class correlation coefficient (ICC)[14] is a measure of clustering that can be obtained from pilot data and used to estimate the impact of clustering upon study power. We aimed to undertake a feasibility study to determine the feasibility of using an online assessment tool and collect pilot data to assist with sample size estimation for a definitive trial to determine whether computer interpretation messages on a 12-lead ECG have an effect on paramedics' diagnosis of STEMI.

AIMS AND OBJECTIVES

3

The primary aims of the feasibility study were to create and test the web-based assessment tool and collect pilot data to inform sample size calculations for the definitive trial. The feasibility study objectives were to:

- Obtain preliminary estimates of the accuracy of paramedic's interpretation, and determine whether it is appropriate to conduct the main study
- Estimate the intra-class correlation coefficients for participants and ECGs, and the discordant proportions, in order to provide guidance in determining the sample size for an appropriately powered main study
- Construct a conditional logistic regression model to determine the odds ratios relating to paramedics' accuracy in recognising STEMI, taking into account the clustering of participant responses and ECG.

The aim of a definitive study would be to determine the effect of computer interpretation messages printed on ECGs, on the accuracy of paramedics' recognition of STEMI. The main study objectives would be to:

- Obtain precise estimates of the accuracy, sensitivity and specificity of the paramedic's interpretation
- Estimate the effect of computer-generated messages on paramedic interpretation (stratified by correct and incorrect computer interpretations)

METHODS
Study design and setting
The RESPECT feasibility study was a randomised crossover trial with clustering at two crossed levels, utilising a bespoke web-based assessment tool (http://respect.ambulanceresearch.co.uk). Each participant was presented with a range of ECGs to classify. Since there is a limited pool of ECGs to sample from, ECGs were classified multiple times and participants made multiple classifications.

Participants
The study took place between 1st March and 30th April, 2013. Participants were recruited by advertising the study on social media and word of mouth. It was open to Health and Care Professions Council (HCPC) registered paramedics working in the UK, with the exception of those employed by Yorkshire Ambulance Service NHS Trust, since it was anticipated that they would form the sample for a subsequent study. Prior to commencing the study, informed consent was obtained from participants and basic demographic information obtained, including their training route (traditional/vocational or higher education), number of years service as a paramedic, hours spent on continuous professional development (CPD) activities relating to ECG interpretation in the past year, and the number of patients taken for pPCI or thrombolysed in the past 12 months.

Sample size
As this was a feasibility study, no a priori sample size calculations were determined. Instead a pragmatic target of 50 participants was deemed to be sufficient, but no cap was placed on recruitment, given the low risk of the intervention on the participants.

Randomisation

Block randomisation was used for ECG allocation to ensure that each participant was asked to classify the same number of ECGs and the sampling from the full pool of ECGs was balanced. This meant that for every 4 participants, all 48 study ECGs were allocated. The sequences were generated using the random number generator from the website RANDOM.ORG. In addition, the order in which ECGs were presented to participants, and the message visibility, were also randomised. Allocation of ECGs, including ordering and message visibility, was automatically handled by the website assessment tool, ensuring that the researcher and participant were blinded to the ECG allocation sequence (Figure 1).

Interventions

Participants were randomly allocated 12 ECGs to view from a pool of 48, with an equal proportion of true positive, true negative, false positive and false negative ECGs (i.e. 3 from each) based on computer interpretation agreement with the study reference standard:

- The ECG had to be a 12-lead ECG recorded in the out-of-hospital environment
- The ECG had to display a wave morphology consistent with either a STEMI or STEMI-mimic, and have a computer diagnostic message printed on the ECG
- The diagnosis of the ECG (i.e. STEMI or not-STEMI) had to be determined by the independent assessment and agreement of two senior ED doctors with specialist knowledge of ECGs. Any disagreements on diagnosis were resolved by discussion between the doctors. An option for subsequent review by an independent third party, was provided, but not required.

Prior to the study commencing, each ECG was duplicated and the computer diagnostic message removed from one of the pair of ECGs. Once a paramedic had consented into the trial, they were then allocated 12 ECGs. During an attempt by a participating paramedic, each ECG they were allocated to was displayed in a random order and with a randomly allocated message visibility on the participant's computer browser window, for 60 seconds (Figure 2). The participant was asked to identify whether the ECG showed a pattern consistent with STEMI. Once the participant had provided a response, or 60 seconds had elapsed, the ECG was removed from view and the participant invited to view the next ECG.

Once all 12 ECGs had been viewed, participants were given a two-week 'washout' period, during which time the website would not allow participants to attempt the second (crossover) phase of the study. Once two-weeks had elapsed, participants were invited to return to complete the study and the same ECGs that were viewed before, but with the message visibility reversed, were shown in a random order.

Statistical methods

Our design allowed each paramedic to interpret multiple ECGs and each ECG to be used multiple times. This was efficient but required specific statistical analysis. Standard statistical tests assume that all observations are independent of each other. Our study could only have met this assumption if each paramedic only interpreted one ECG and each ECG was only interpreted once. This would be an inefficient design as the pool of

ECGs was limited and simple crossover trials are not efficient as concordant responses (e.g. correct answer with and without message) do not contribute to the analysis.[15] This would clearly not be feasible due to the large number of participants and ECGs required, so we allowed multiple interpretations by each paramedic and multiple interpretations of each ECG. Interpretations by the same paramedic or interpretations of the same ECG were therefore clustered and responses likely to be correlated, so analysis had to take this clustering into account.

The crossover design ensured that paramedics interpreted the same ECG with and without a computer message. A simple crossover trial would be analysed with conditional logistic regression, to produce an odds ratio estimating the effect of a computer-generated message upon paramedic ECG interpretation. However, to address the issue of clustering we added two random effects to the conditional logistic regression model: a random effect for the ECGs and another for the participants. This produced adjusted odds ratios with confidence intervals that take into account clustering by ECG and paramedic.

Statistical data analysis was conducted using the statistics packages R (http://www.r-project.org) and WinBUGS (http://www.mrc-bsu.cam.ac.uk/software/bugs/). The analysis preceded in an incremental fashion, commencing with the calculation of participant accuracy, sensitivity and specificity values and odds ratios, before analysis of conditional logistic regression with random effects. This enabled the estimation of adjusted odds ratio values and intra-class correlation coefficients that took account of the clustered nature of the data around participants and ECGs. Supplementary 1 contains links to the scripts used to prepare and analyse the data.

RESULTS
Participants and electrocardiograms
Figure 3 shows the CONSORT diagram for the RESPECT feasibility study. In total, 254 participants consented into the study, with 205 completing the first stage and 150 completing the second stage, an overall attrition rate of 40.9% (104/254). Demographic information was provided by 218 participants, including 62 participants who did not complete the study (Table 1). There appeared to be little difference in participant characteristics between those who completed and failed to complete the study, with the exception of median hours spent undertaking CPD activities relating to ECG interpretation. Only ECG interpretation attempts from participants who had completed both stages were included in the final analysis. This necessitated the removal of 605 ECG interpretation attempts, leaving 1800 paired ECG attempts for the final analysis. Each ECG in the study was attempted a median of 90 times (interquartile range, IQR 87–93) overall, and attempted a median of 76 times (IQR 74–81) once unpaired ECG attempts had been excluded.

Accuracy
The sum totals of all responses by message visibility and answer accuracy are summarised in Table 2. Participants in the feasibility were correct approximately 80% of the time, irrespective of whether the computer interpretation message was visible. When only ECGs that the computer correctly interpreted were considered, participants were more accurate, particularly when the correct message was visible. Conversely,

participants were less accurate in interpreting ECGs that the computer had also mis-interpreted.

The odds of a correct paramedic interpretation with the computer interpretation visible, compared to the odds of a correct paramedic interpretation with the computer interpretation hidden (i.e. the odds ratio) for all ECGs, was 0.87 (Table 3). In the sub-group of correct computer interpretations, the odds ratio was 1.51, and for incorrect computer interpretations, the odds ratio was 0.60.

Table 2: Summary of participant answer accuracy

	All computer interpretations		
	Participant Interpretation		
Message visibility	*Correct*	*Incorrect*	*Total*
Visible	1424 (79%)	376 (21%)	1800
Hidden	1448 (80%)	352 (20%)	1800
Total	2872(80%)	728(20%)	3600
	Correct computer interpretations		
Visible	785 (87%)	115 (13%)	900
Hidden	758 (84%)	142 (16%)	900
Total	1543(86%)	257 (14%)	1800
	Incorrect computer interpretations		
Visible	639 (71%)	261 (29%)	900
Hidden	690 (77%)	471 (26%)	900
Total	1329 (74%)	471 (26%)	1800

Table 3: Summary of concordant and discordant pairs of interpretation. The number of discordant pairs are highlighted in bold

	Message Hidden		
Message Visible	*Correct*	*Incorrect*	*Total*
	Correct computer interpretations		
Correct	705	**80**	758
Incorrect	**53**	62	142
	785	115	900
	Incorrect computer interpretations		
Correct	562	**77**	690
Incorrect	**128**	133	210
	639	261	900

Intra-class classification

For the subset of computer correct interpretations, the ICC for participants was 0.37 and for ECGs, 0.19. For computer incorrect interpretations, the ICC for participants was 0.05 and for ECGs, 0.01.

Proportion of discordant pairs
Overall, there were 338 discordant pairs (18.8%, Table 3). For the sub-group of computer correct interpretations, there were 133 (14.8%) discordant pairs and for incorrect computer interpretations, 205 (22.8%).

Conditional logistic regression
Table 4 shows the odds ratio of a correct paramedic interpretation when the computer interpretation is displayed, with and without adjustments for clustering. The unadjusted analysis is based on an inappropriate assumption but is included here to show the potential impact of clustering. There is a noticeable difference in the odds ratio and 95% confidence interval for the sub-group of correct computer messages between the adjusted and unadjusted models.

Table 4: Odds ratios of correct participant interpretation with computer interpretation visible

	Correct computer messages		**Incorrect computer messages**	
	Adjustments for clustering		Adjustments for clustering	
	Unadjusted	Adjusted	Unadjusted	Adjusted
OR	1.51	1.80	0.60	0.58
95% CI	1.07–2.14	0.84–4.91	0.45–0.80	0.41–0.81

Treatment, carryover and period effects
Table 5 demonstrates that participants with discordant responses (e.g. correct on 1st attempt and incorrect on 2nd), were more likely to make a correct diagnosis when the message was visible in the correct computer interpretation group, although this was not significant (χ^2 = 0.70, d.f. = 1, p = 0.40). In the incorrect computer interpretation group however, the opposite was true, with significantly more participants making a correct diagnosis when the message was hidden (χ^2 = 6.80, d.f. = 1, p = 0.01). No statistically significant carryover or period effects were found (see Supplementary 2).

Table 5: Correct participant interpretation by computer interpretation and sequence
Note that only discordant pairs are shown

Computer interpretation	**Sequence**	**Correct 1st attempt**	**Correct 2nd attempt**	**Total**
Computer correct	Visible -> Hidden	38	27	65
Computer correct	Hidden -> Visible	26	42	68
Computer incorrect	Visible -> Hidden	29	65	94
Computer	Hidden ->	63	48	111

incorrect	Visible				

DISCUSSION

The RESPECT feasibility study has demonstrated that it is possible to conduct a randomised crossover trial to test the accuracy of STEMI recognition by paramedics, using an online assessment tool. The numbers of paramedics who participated in the feasibility reflect the advantage of using an online, and accessible anywhere, method of delivering the assessment tool to maximise recruitment. However, the level of attrition is concerning and strategies to reduce attrition need to be employed in any future study.

Overall, participants were correct approximately 80% of the time, irrespective of whether the computer message was visible. The sub-group analysis suggests that computer interpretation messages have an effect on participant interpretation, although this must be taken in the context of a non-powered feasibility study. In the sub-group of ECGs where the computer interpretation was correct, the proportion of correct answers by participants when the message was hidden was 84%, increasing to 87% when the message was visible. The adjusted odds ratio in the sub-group of correct computer interpretations was 1.80 (95%CI 0.84–4.91), and the wide confidence interval suggesting little evidence of an effect of the message on paramedics. Conversely, in the sub-group of incorrect computer interpretation, the proportion of correct answers fell to 77% with the message hidden, and to 71% when the incorrect computer interpretation message was visible. In this sub-group, the adjusted odds ratio was 0.58 (95%CI 0.41–0.81), suggesting that there is a significant negative effect of the computer message on paramedics' interpretation of the ECG. Overall, the results suggest that the computer and participant have a tendency to correctly, and incorrectly, interpret similar types of ECGs, which is worth investigating in the main study.

The ICCs reported in this study indicate substantial clustering induced correlation of data by paramedic and ECG. This confirms that interpretations of different ECGs made by the same paramedic are more likely to be similar (i.e. correct or incorrect) than interpretations of different ECGs made by different paramedics. Likewise, interpretations of the same ECG by different paramedics are more likely to be similar (i.e. correct of incorrect) than interpretations of different ECGs by different paramedics. The ICCs will enable and inform the calculation of the design effect of a definitive study. The ICCs estimated from this feasibility mean that the full trial will need to expose more participants to correct message ECGs in order to detect a meaningful result. This is due to a lower discordance rate in this group and stronger ECG and participant effects. Based on the calculated ICCs and the proportion of discordant pairs, the greatest challenge to an adequately powered main study, is the sample size, as this may prove prohibitively large. In addition, the results suggest that more responses to correct computer interpretation ECGs are required, either by increasing the number of participants and/or allocating more computer correct interpretation ECGs to each participant. Increasing the number of ECGs in the pool would also be a possibility.

Implications for practice

The results from the feasibility suggest that incorrect computer interpretations may have a significant effect on paramedic's accuracy at interpreting this group of ECGs.

However, it will be necessary to conduct an adequately powered main study to be sure. In addition, true population proportions of the ECGs that computers correctly and incorrectly interpret will be required prior to the main study commencing. For example, if incorrect computer interpretations do have a significant effect on paramedics' interpretation, but constitute a small minority of overall ECGs in the population, then this may not be clinically significant.

Limitations
Although angiographic confirmation of the diagnosis would have been ideal, in practice, patients with a false-positive 12-lead ECG should never have received angiography. It was more pragmatic, therefore, to use the benchmark of the decision of two senior ED consultants against a paramedic's diagnosis of the 12-lead ECG.

A key drawback with crossover trials is the risk of 'carry over', but the two week wash-out period appears to have been sufficient to ensure participants could not recall their first phase attempt, and the ECGs they had viewed. However, this may have contributed to the high attrition rate, with 24% of participants failing to return to complete the second phase of the study. In addition, a perceived poor performance in the first phase, may have prompted the participants to revise their knowledge on ECG interpretation, or not return at all. However, the overall attrition rate was much higher (40.9%), and due to the anonymised nature of the enrolment, it is not possible to follow up participants who consented into the study, but did not view any of their allocated ECGs.

There was a risk that participants may have utilised textbooks or an expert colleague to assist with their answers, since the study was not supervised by the researcher. However, the time limited nature of the assessment (each ECG was only visible for 60 seconds) and the inability to view the same ECG with a specific message visibility (i.e. visible or hidden) more than once, should have minimised this risk.

CONCLUSION
The RESPECT feasibility study has demonstrated that determining the effect of computer interpretation messages using a web-based assessment tool is feasible but that the design needs to take into account the clustered nature of the data. Feasibility data suggest that computer messages influence paramedic interpretation, improving accuracy when correct and worsening accuracy when incorrect.

COMPETING INTEREST
None

ETHICS APPROVAL

Ethics approval was obtained from the University of Sheffield, School of Health and Related Research's, ethics committee.

REFERENCES

1   Myocardial Ischaemia National Audit Project. *How the NHS care for patients with heart attack. Annual Public Report.* London. 2013. https://www.ucl.ac.uk/nicor/audits/minap/documents/annual_reports/minap-report-2013 (accessed 18 Dec2013).

2   Department of Health. Treatment of heart attack national guidance: final report of the National Infarct Angioplasty Project (NIAP). 2008. http://www.bcis.org.uk/resources/documents/NIAP%20Final%20Report.pdf (accessed 23 Apr2013).

3   Bhalla MC, Mencl F, Gist MA, *et al.* Prehospital Electrocardiographic Computer Identification of ST-segment Elevation Myocardial Infarction. *Prehosp Emerg Care* 2013;**17**:2113

4   Clark EN, Sejersten M, Clemmensen P, *et al.* Automated electrocardiogram interpretation programs versus cardiologistsnsen P, Myocardial Infarction. Angioplasty Project (NIAP). 2008.myocth suspected acute coronary syndrome. 2010.

5   Youngquist ST, Shah AP, Niemann JT, *et al.* A Comparison of Door-to-balloon Times and False-positive Activations between Emergency Department and Out-of-hospital Activation of the Coronary Catheterization Team. *Acad Emerg Med* 2008;**15**:7848

6   Tsai TL, Fridsma DB, Gatti G. Computer Decision Support as a Source of Interpretation Error: The Case of Electrocardiograms. *J Am Med Inform Assoc* 2003;**10**:4783

7   Massel D. Observer variability in ECG interpretation for thrombolysis eligibility: experience and context matter. *J Thromb Thrombolysis* 2003;**15**:1313

8   Goodacre S, Webster A, Morris F. Do computer generated ECG reports improve interpretation by accident and emergency senior house officers? *Postgrad Med J* 2001;**77**:4551

9   Young DR, Murinson M, Wilson C, *et al.* Paramedics as decision makers on the activation of the catheterization laboratory in the presence of acute ST-elevation myocardial infarction. *J Electrocardiol* 2011;**44**:1811

10   Davis DP, Graydon C, Stein R, *et al.* The positive predictive value of paramedic versus emergency physician interpretation of the prehospital 12-lead electrocardiogram. *Prehospital Emerg Care Off J Natl Assoc EMS Physicians Natl Assoc State EMS Dir* 2007;**11**:3997

11   Rokos IC, French WJ, Mattu A, *et al.* Appropriate Cardiac Cath Lab activation: Optimizing electrocardiogram interpretation and clinical decision-making for acute ST-elevation myocardial infarction. *Am Heart J* 2010;**160**:9950

12 Swan PY, Nighswonger B, Boswell GL, *et al.* Factors Associated With False-Positive Emergency Medical Services Triage for Percutaneous Coronary Intervention. *West J Emerg Med* 2009;**10**:2089

13 Smith SW. ST-elevation Acute Myocardial Infarction; A Critical but Difficult Electrocardiographic Diagnosis. *Acad Emerg Med* 2001;**8**:3821

14 Lancaster G, Campbell M, Eldridge S, *et al.* Trials in primary care: statistical issues in the design, conduct and evaluation of complex interventions. *Stat Methods Med Res* 2010;**19**:3490

15 Campbell MJ. *Statistics at square two: understanding modern statistical applications in medicine.* 2nd ed. Malden: Blackwell 2008.
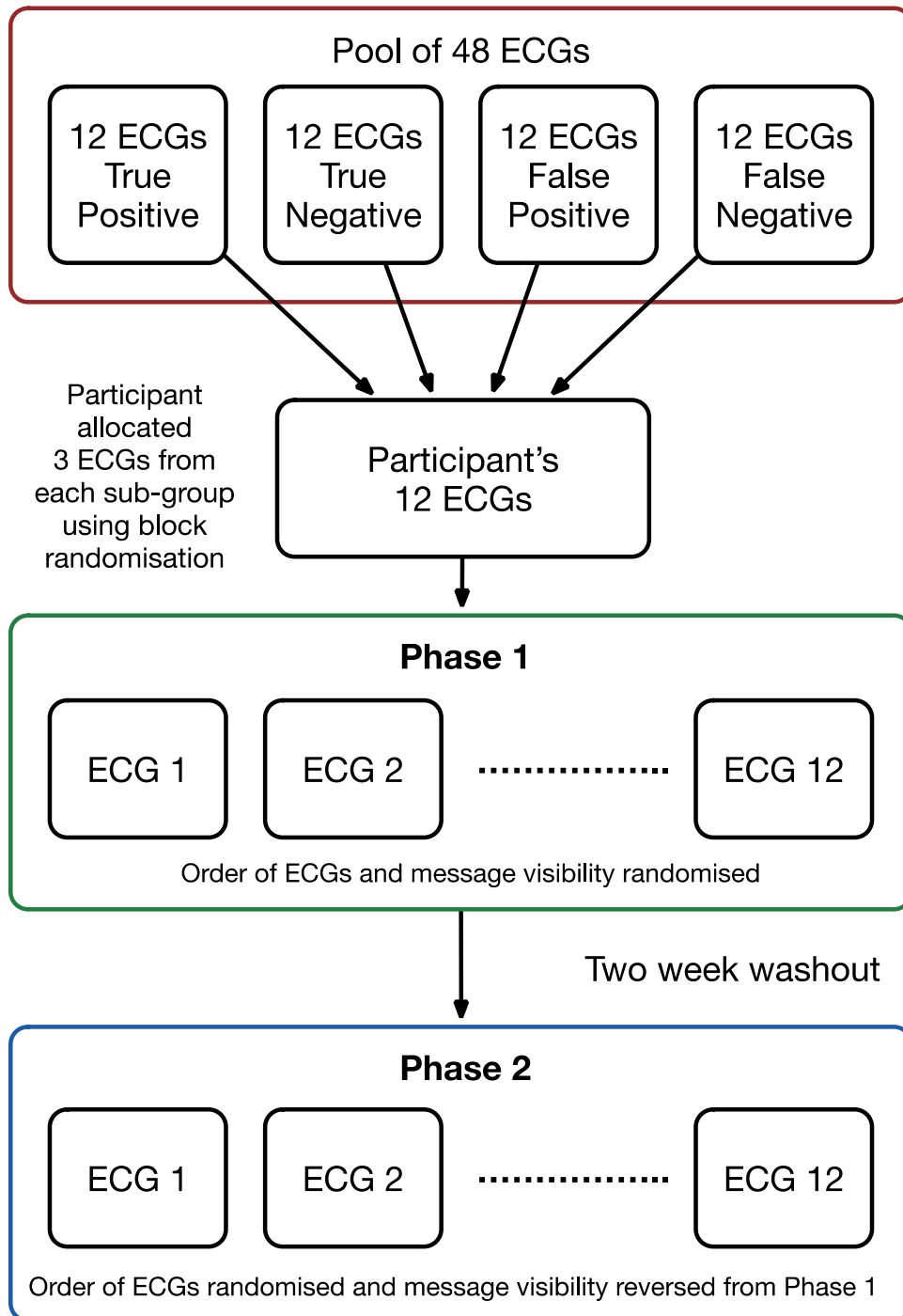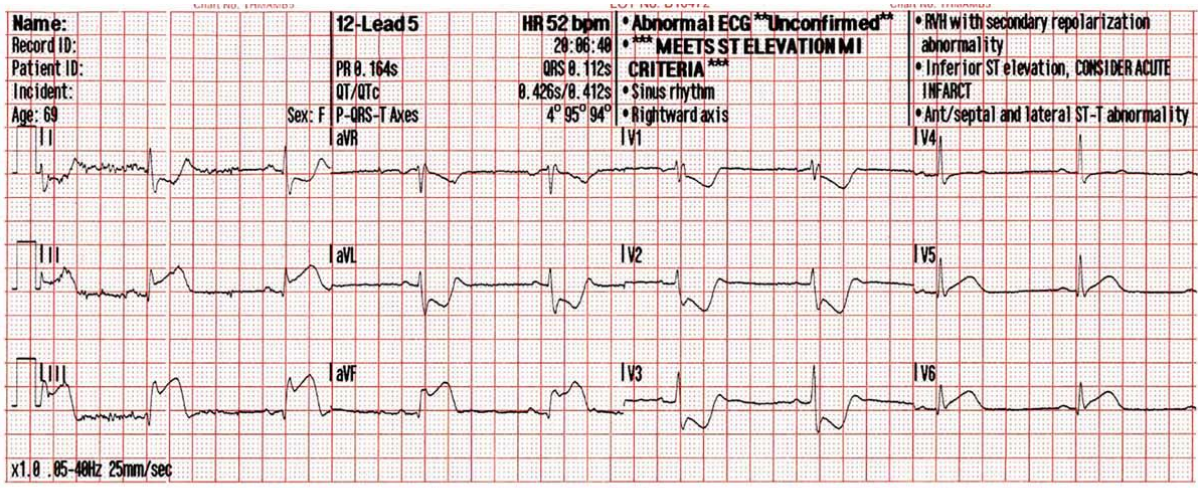
Figure legends

Figure 1: ECG allocation for the RESPECT feasibility study

Figure 2: An example view of the RESPECT feasibility study ECG webpage. Note the form to record the participant's response (bottom left) and the timer, indicating the time remaining before the ECG is removed from view (bottom right)
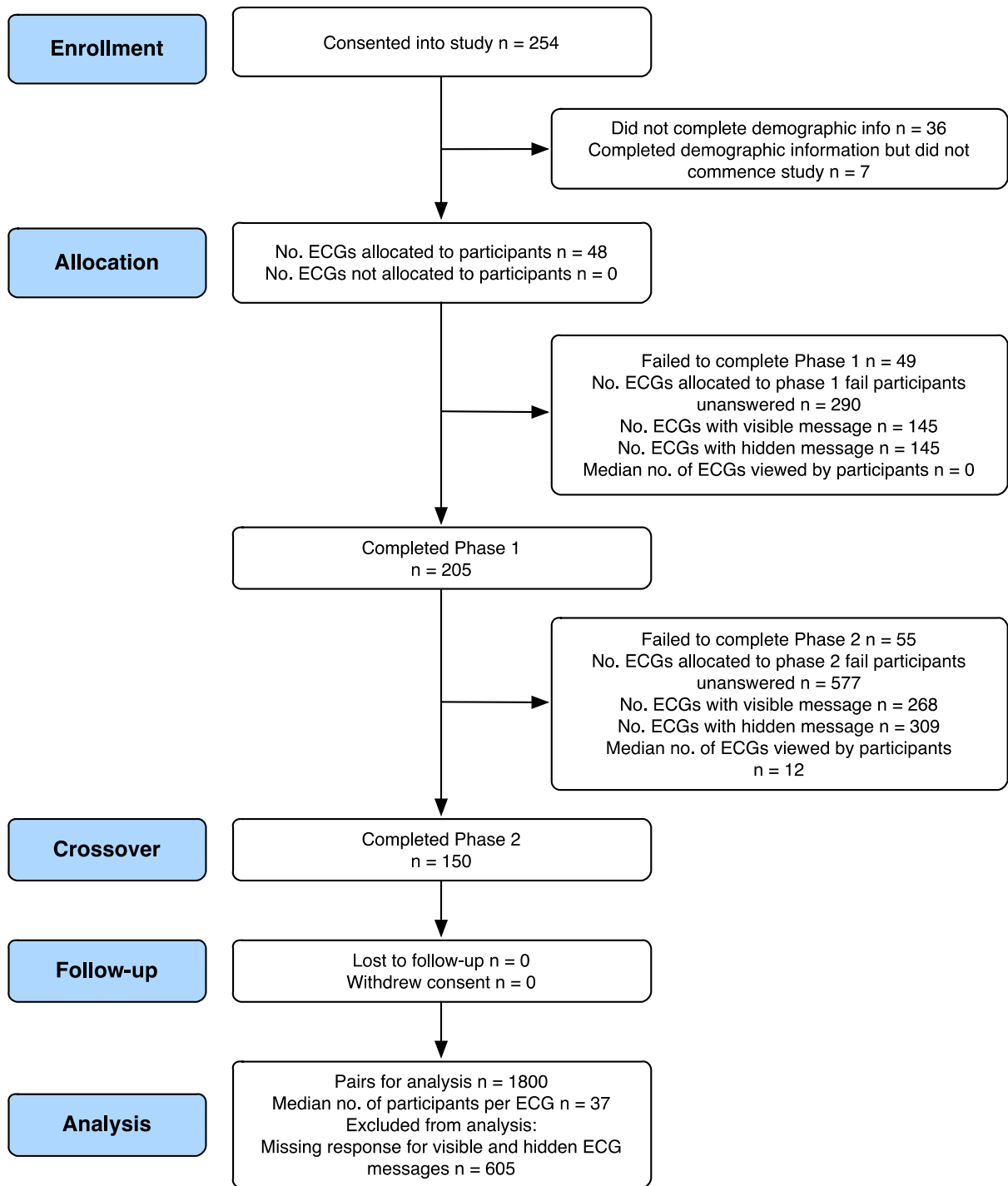
Figure 3: CONSORT for RESPECT feasibility study