9.2.23

READY

Statistical Analysis Plan

Contents

1. Administrative Information

FULL/LONG TITLE OF THE STUDY A Realist Evaluation of Paramedics Working in General Practice: Work Package 2: An assessment of clinical and cost effectiveness (READY Paramedics)

SHORT STUDY TITLE / ACRONYM READY

1.2 Study registration number IRAS Number: 279490

SPONSORS Number PIMS Project ID: 6424053

1.3 SAP version 2.00, dated 09/02/2023

1.4 Protocol version 5.0, dated 21/11/2022

1.5 SAP revisions

SAP revision history

1.6 Roles and responsibility

Hazel Taylor, Senior Statistician, UHBW

Dr Sarah Voss, Chief Investigator and Associate Professor in Emergency Care, UWE

Dr Matthew Booker, Chief Investigator and GP and NIHR Clinical Lecturer in Primary Care (Urgent Care)

1.7 Signatures of:

	Name	Signature	Date
SAP Authored By:	Study Statistician: Hazel Taylor		
Approved By:	Chief Investigator: Dr Sarah Voss		
	Chief Investigator: Dr Matthew Booker		
	Independent Statistician: Kara Stevens		

2. Introduction

2.1 Study Background and Rationale

General Practice (GP) services are responding to demand by increasing the non-medical workforce supporting front-line service delivery. Paramedics are working in general practice in growing numbers, and their generalist skillset may be well-suited to this setting. Paramedics carry out a range of tasks including home visits, routine and same-day appointments, and telephone triage. There is significant variation in the types of patients that paramedics manage, their models of working and their contractual engagement.

Paramedics have been working in general practice since at least 2002, but there are few studies describing their role and contribution. Existing evidence is largely descriptive and includes many assumptions, such as paramedics reducing GP workload and costs, which have not been tested empirically. To date, no research investigates the variation in paramedic models of working in general practice and the associated impacts on patient safety, clinical or cost effectiveness. Implementation guidance struggles to reflect this variation, making it difficult to make informed decisions around how to successfully implement the paramedic role in differing contexts.

Realist evaluation is a theory-driven approach to understanding complex interventions in complex environments in terms of what works, for whom, in what circumstances, how and why. Our aim is to provide evidence from different models of paramedics in general practice as to how they might: achieve good clinical outcomes for patients; provide safe care; improve patient experience; relieve GP workload; influence the workload of other general practice staff; make efficient use of healthcare resources.

2.2 Aims and Objectives

2.2.1 Aim

To evaluate the role of paramedics in general practice (PGP) and provide evidence about different service delivery models to determine their ability to:

- Achieve good clinical outcomes for patients
- Provide safe patient care
- Improve patient experience
- Relieve GP workload pressure
- Influence the workload of other general practice staff
- Make efficient use of healthcare resources

2.2.2 Objectives

To test the programme theories developed and refined in WP1 (Rapid Realist Review of Paramedics Working in General Practice – separate protocol), using case studies of general practices in England. We will collect qualitative data from patient participants (or their adult carers (individuals) who accompanied the patient participant at their appointment) and general practice health professionals to understand the barriers and facilitators to PGP and the impact it has on access to general practice. We will analyse the implications of differing

models of PGP compared to no PGP on healthcare resource utilisation, costs and patient reported outcomes and safety outcomes to assess clinical and cost effectiveness.

2.2.3 Research Questions

1. How does PGP care impact on patient clinical outcomes (e.g. unplanned hospital admissions, prescriptions, referrals, tests and investigations)?
2. How does PGP care impact on patient reported outcomes (e.g. concern, confidence in health plan, ability to manage symptoms, health related quality of life) compared to non-PGP care?
3. Does PGP result in patient reported safe management?
4. What are the direct costs/savings associated with PGP care and does it provide good value for money?
5. Does PGP lead to improved patient experience; how and for which patients?

3 Study Design

3.1 Study Design

The programme theories developed in WP1 will be tested using a series of case studies with sites (general practices) in England. 24 case study sites will be recruited, to compare 3 different PGP model categories and a control (about 6 sites in each model category and 6 in the control) within a few different pre-specified PGP model configurations. The PGP model configurations and the models within each configuration will be informed from the work in WP1 and will be defined prior to the start of the statistical analysis. 12 sites (with 3 sites in each model category and 3 in the control) will be detailed case study sites where we will collect both retrospective and prospective quantitative data. The remaining 12 sites will be core case study sites where we will collect prospective quantitative data only.

3.2 Sample size

3.2.1 Retrospective Study

Retrospective data between the period June 2021 and June 2022 will be extracted from 12 practices with 3 practices in each PGP model.

3.2.2 Prospective Study

The sample size calculation is based on the outcome; change in PCOQ score between baseline and 30 days. Sample sizes of 138 in each of the PGP models and control groups will be obtained by sampling in 6 practices in each PGP model with an average of 23 subjects with complete data in each practice. This will achieve 90% power to detect a difference between the group means of 0.5 of a standard deviation of the change in PCOQ scores.[36] This assumes an estimated intracluster correlation coefficient (ICC) of 0.02, a coefficient of variation of cluster sizes of 0.65 with a significance level of 0.050 with a two-sided test. To achieve 138 complete datasets per group, assuming a conservative 50% follow-up rate, 276 participants (46 per practice) will be recruited in each of the PGP models.

3.3 Framework

All hypothesis testing will test for superiority.

3.4 Statistical interim analyses and stopping guidance

No formal interim analysis is planned.

3.5 Timing of final analysis

The final statistical analysis will take place only once all data has been collected and entered onto the database.

3.6 Timing of outcome assessments

For the purposes of these analyses, the index appointment is defined as the first eligible appointment during the study period. However, for many patients an individual general practice contact is part of a sustained relationship with a practitioner, and a management plan for long term conditions is built up over time.

3.6.1 Retrospective Analysis

We will extract information on all general practice contacts (including consultation length), tests, medications and referrals during a 30 day (care episode) after the index appointment. A 30-day interval has been selected to provide sufficient time to evaluate outcomes directly related to the care received at the index appointment.

3.6.2 Prospective Analysis

The outcomes will be collected using questionnaires, on, or immediately after, the day of the index appointment and again 30 days later.

4. Statistical Principles

4.1 Confidence intervals and P values

All p values will be 2 sided, and the significance set at 5% level.

All confidence intervals will be 95% confidence intervals.

4.2 Protocol Deviations/ Adherence

N/A for retrospective study.

In the prospective study, if the participant was at paramedic site, then the analysis will assume that they saw a paramedic, regardless of the questionnaire responses.

4.3 Analysis Populations

As both the retrospective and prospective studies are observational studies, data will be analysed from all patients having eligible index consultations. The few ineligible participants recruited to the prospective study will be excluded from the analysis.

5. Study Population

5.1 Screening Data

N/A for retrospective study.

For the prospective study, data will be reported on the number approached, the number eligible, the number recruited and the number followed-up in a Strobe diagram.

5.2 Eligibility

5.2.1 Retrospective Study

The exact eligibility criteria will depend upon the model configurations of PGP identified in WP1.

5.2.2 Prospective Study

Inclusion Criteria

- All adults, 16 and over with capacity to give informed consent. (Carer support if necessary)
- Understanding of English language sufficient to take part in an interview or complete a standardised questionnaire with an interpreter or carer if required. Where necessary, translated versions of questionnaires will be available.
- Registered with a general practice in England.

Exclusion Criteria

- Less than 16 years of age
- Do not have capacity to give informed consent
- Understanding of English language is insufficient to take part in an interview or complete a standardised questionnaire even with an interpreter/ translated version if required.
- Not registered with a general practice in England.

5.3 Recruitment and Consent

5.3.1 Retrospective Study

As this study uses anonymised routine patient data stored on primary care Electronic Health Records, we have HRA permission to run this study without patient consent.

5.3.2 Prospective Study

Eligible patients will be identified and approached with written and verbal information by practice staff at each site at the time of the initial appointment.

5.4 Withdrawal/ follow-up

5.4.1 Retrospective Study

Withdrawals are not possible as the anonymised routine clinic data collected does not require participant consent.

5.4.2 Prospective Study

Participants may withdraw from the study at any time and do not have to provide a reason for doing so. Care will not be affected by a participant's decision to withdraw from the study.

Data collected prior to withdrawal will be included in the study analysis unless a participant specifically requests that their data are removed from the database.

5.5 Baseline characteristics

Baseline patient characteristics, will be tabulated by PGP category within each PGP model configuration to give an overview of the study population. (See example Table 1 for the retrospective data and Table 2 for the prospective data). Data will be summarised using percentages for binary variables and either means/SDs or median/IQR for continuous variables according to their distributional properties.

As this is an observational study, statistical testing, will be used to identify baseline patient differences between the PGP categories in a given model configuration.

GP Practice characteristics will also be tabulated by PGP category within each model configuration and will include: practice size, deprivation, urbanity, and standard mortality weightings.

5.5.1 Retrospective Study

Baseline patient data collected will include age, ethnicity, sex.

5.5.2 Prospective Study

Baseline patient data reported will include age, ethnicity, sex, Primary Care Outcomes Questionnaire (PCOQ[30]), the Patient Reported Experiences and Outcomes of Safety in Primary Care (PREOS-PC[31]) Items 6 -11 from the MODRUM collected at the 30 day follow-up will be summed and used to assess the number of recent attendances and item 12 from the MODRUM on prescribed medications will be used as a surrogate for multimorbidity complexity.

5.6 Potential Confounding Covariates

It is recognised that some patient characteristics could influence whether or not a patient has a paramedic consultation and the outcome. These characteristics could be potential confounding factors in this analysis. Potential confounding factors include those listed in the previous section.

It may be possible to take account of these factors in the statistical analysis, by adjusting for them in a mixed effects model.

6 Analysis

6.1 Outcome definitions

6.1.1 Retrospective Study

Most of the data to be collected from the retrospective study will inform and be reported in the health economic analysis.

The main outcomes to be reported within the statistical analysis will be; the percentage of patients seen by a paramedic, the mean/ median number of paramedic consultations per patient per year, a measure of paramedic dose and the number of patients reconsulting with a GP within 7 days of the index consultation.

The % of patients seen by a paramedic will be calculated from the paramedic data collected in the study and using information on practice list size for the denominator.

Information on practice list size in addition to the retrospective data collected, will be used to calculate the mean/ median number of paramedic consultations per patient per year.

The (Number of WTE Paramedics)/(Number of WTE GP’s) will be used as a measure of paramedic dose.

6.1.2. Prospective Study

We will assess patient (or carer) experience and outcome of the consultation using the Primary Care Outcomes Questionnaire (Murphy et al, 2018) and the Patient Reported Experiences and Outcomes of Safety in Primary Care (Ricci-Cabello et al, 2016), compact version [Oxford University Innovation Limited, 2018). Questionnaires will be administered by post, telephone or email (secure data transfer), depending on patient preference.

PCOQ

The PCOQ measures common outcomes, such as reduction in pain or depression and broader outcomes, such as reduction in concern and a sense of confidence in health plan or an understanding of illnesses/problems and an ability to manage symptoms. It is scored in four domains: Health and wellbeing; health knowledge and self-care; confidence in health provision; confidence in health plan. PCOQ data will be collected on, or immediately after, the day of the index appointment and again 30 days later. It will be analysed as the change in score between the 30 day follow-up and baseline.

The total scores for the four PCOQ domains will be calculated. For the domains; “health and well-being”, “confidence in health provision” and “health knowledge and understanding”, if some items are unavailable, then the PCOQ will be calculated from the mean of the available items, providing at least half of the items in that domain have been completed. All items will be given equal weight in the domain score. The domain score will only be coded as missing, if more than half of the items for that particular domain are incomplete. For “the confidence in health plan”, the score will be calculated from the completed items, if two thirds of the items have been completed. (See https://www.bristol.ac.uk/media-library/sites/primaryhealthcare/documents/PCOQShortUserGuide_Oct2016.pdf)

PREOS-PC

The PREOS-PC has been designed as a tool to comprehensively collect information about patient experience and patient reported safety problems in general practice. It independently assesses five domains of patient safety: practice activation; patient activation; patients’ experiences of safety problems; patient safety outcomes (harm); general perceptions of safety. There is also a VAS score for the patient to rate how safe they felt their health care was in their GP surgery in the past 12 months. It provides discrimination between different levels of patient-reported safety between practices, and is sensitive to change. PREOS-PC data will be collected on, or immediately after, the day of the index appointment and again 30 days later. We have opted to use the compact version after feedback from our public involvement group.

The overall raw scale score for each domain will be computed as the raw mean score for all the items in the scale for which a score is available. All items will be given equal weight in the domain score. A linear transformation will then be applied to rescale the overall raw scale score to a range of 0-100. Responses of “I don’t know” or “Not applicable" will be considered as missing for the purposes of item score computation. For multi-item scales, where responses are missing for more than 50% of the items the whole scale will be scored as missing; otherwise a score will be derived based on the available items without any imputation

A couple of pre-selected individual items on the PREOS-PC will also be reported on, which will be:

Item 4.3. “Thinking about the health care you have received in your GP surgery in the last 12 months, were there any communication problems between GPs, nurses or other health-care staff in your GP surgery and other health-care professionals (such as consultants or hospital

nurses)”

Item 5.1. “Do you think you have experienced harm to your physical health as a result of the health care provided in your GP surgery in the last 12 months?”

6.2 Analysis methods

As the quantitative data analysis is for an observational study using realist methods, the analysis will evolve and be data led.

6.2.1 Retrospective Data

Research question to be addressed, “How does PGP care impact on patient clinical outcomes?”

Descriptive statistics will be reported to characterise the patient sample, the workload and type of PGP care provided within the categories of each PGP model structure. This will include tabulating age, ethnicity, sex, % of patients who saw a paramedic, mean number of paramedic consultations per patient per year, paramedic dose and the % of patients who reconsulted with a GP within 7 days of the index appointment. (See Table 1) Data will be summarised using percentages for binary variables and either means/SDs or median/IQR for continuous variables according to their distributional properties.

As this is an observational study, statistical testing, will be used to identify differences between the PGP categories within each model configuration, using chi-squared tests for categorical variables and 1 way ANOVA or Kruskal Wallis tests for continuous variables. If a significant difference is seen, then post-hoc tests will be used to determine between which categories within a model configuration, the differences lie.

6.2.2 Prospective Data

Research questions to be addressed:

- How does PGP care impact on patient reported outcomes compared to non-PGP care?
- Does PGP result in patient reported safe management?
- Does PGP lead to improved patient experience; how and for which patients?

A STROBE diagram will be used to report the number of patients, screened, recruited and followed-up.

Descriptive statistics will be reported to characterise the patient sample within each PGP model category. This will include tabulating age, gender, ethnicity, number of GP practice appointments in the past month, number of prescriptions (to assess multimorbidity complexity), PCOQ and PREOS-PC. (See Tables 2 and 3). Items 4.3 and 5.1on the PREOS-PC will also be reported on. This will be carried out for the sample who completed the baseline data (Table 2), and for the sample who also completed the 30 day follow-up (Table 3). Data will be summarised using percentages for binary variables and either means/SDs or median/IQR for continuous variables according to their distributional properties. As this is an observational study, statistical testing, will be used to identify patient differences between the PGP categories within each model configuration, using the methods described for the retrospective data.

The PCOQ domains (patient reported outcomes, concern, confidence in plan, symptom management), at baseline and change at day 30 from baseline will be tabulated for each of the PGP categories within each paramedic model configuration (See Tables 2 and 3). The PREOS PC domains (patient reported safety data) will be reported at baseline and day 30. Data will be reported as either means/SDs or median/IQR according to the distributional properties. Note that the PCOQ and the PREOS-PC will be reported in different ways at day 30, as is standard for these scales, with the PCOQ reported as change between day 30 and baseline, and the PREOS-PC being reported as the score at day 30.

As this is an observational study, statistical testing, will be used to identify differences between the PGP categories within each PGP model configuration in the PCOQ and PREOS PC domains and the PREOS PC VAS score, using either a 1 way analysis of variance if the outcome has a normal distribution, or using a Kruskal Wallis test if not. If a statistically significant difference is seen, then post-hoc unpaired t-tests (or Mann Whitney U tests if the data is not normally distributed) will be used to determine between which categories within a model configuration, the differences lie (using a Bonferonni correction to adjust for multiple testing).

It is hoped that it will be possible to fit multilevel models to adjust for confounding factors to explore these differences further. However, this is dependent upon the data fulfilling the necessary assumptions. It is unknown at this stage whether the PCOQ subscales will fulfil these assumptions. There is data available to suggest that the PREOS_PC subscales are highly negatively skewed (skewed to the right), which may impact the validity of any multi-level models fitted. (Ricci-Cabello et al 2017)

For each PGP configuration a separate multilevel model will be fitted for the change in each PCOQ subscale between 30 days and baseline, to compare the outcomes between the different PGP types (including the control) within a PGP model configuration. The multilevel models will adjust for the baseline PCOQ score, fit practice as a random effect and take account of practice level factors (such as practice size, deprivation, urbanity, new registrations, standard mortality weightings) and patient level factors (age, sex, ethnicity, recent attendances, number of prescriptions in the past 30 days). Patient and practice level factors will be included in the model if they are significant at $p<0.05$ by likelihood ratio tests using a backwards stepwise approach. A variable for the type of PGP category within the PGP model configuration will be fitted to the multilevel model. If statistically significant, post-hoc tests (adjusting for multiple testing) will used to compare the PGP categories (including the control category), to see between which PGP categories the differences lie. The coefficient from the multilevel model with 95% confidence interval (equivalent to the adjusted mean difference) for each PGP category will be reported relative to the GP led care (control category), (See Table 4).

If the appropriate distributional assumptions are met, it is hoped that it will be possible to analyse the PREOS-PC subscales at day 30 using similar methods. If there are concerns about the validity of using a multi-level approach with the PREOS-PC subscales fitted as continuous outcomes, the following alternative options may be considered for analysing the data:

1. The significance of the predictor variables in the multilevel model will be validated using a non-parametric boot-strapped percentile based p-value. (Ricci-Cabello et al 2017)
2. If inspection of the data reveals suitable cut-offs, the outcomes will be dichotomised and a multilevel logistic regression analysis will be carried out, using similar methods for variable selection as described above. This method will also be dependent upon

the sample size being large enough with 10 events for every variable offered to the logistic regression model. (Peduzzi et al, 1996) The intended overall sample size is 552, assuming that it is possible to collect a sample of 138 in each PGP model type. So, if the variable was dichotomised so that 10% had the "event", there would be 50 events, and so it would be possible to have 5 variables in the model. It may be that a suitable dichotomisation, will give fewer than 10% with the event, meaning that it will not be possible to include many variables in the model, and so this method of analysis may not be possible. If it is decided to go with this approach, if inspection of the data reveals a choice of suitable cut-off's for dichotomisation, then a sensitivity analysis may be carried out, using the different cut-off's.

3. The analysis will be purely descriptive reporting medians and interquartile ranges. It will not be possible to use a modelling approach, and therefore to do an adjusted analysis.

6.3 Missing Data

6.3.1 Retrospective Analysis

N/A for this component of the study.

6.3.2 Prospective Analysis

For most domains of the PCOQ and PREOS-PC it will still be possible to calculate a score, provided more than 50% of the items for a particular domain have been completed. (See 6.1.2). It is therefore hoped that the amount of missing data for these items will be small.

A sensitivity analysis using multiple imputation may be considered, if the amount of missing data warrants it, where the multilevel models will be rerun on imputed datasets and the results get pooled to give a single result.

6.4 Additional Analyses

No additional analysis is anticipated at this stage. However, as this study is using a realist approach, and the statistical analysis is evolving and informed by the observational data, it is possible that the results from the planned analysis and the qualitative analysis may lead to the need for some additional ad hoc analysis.

6.5 Statistical software

All data will be analysed in Stata Version 14.

6.6 Draft figures and tables

See separate document. (READY SAP Tables Version_2 9 2 23).

7.0 References

Murphy M, Hollinghurst S, Cowlishaw S, Salisbury C. Primary Care Outcomes Questionnaire: psychometric testing of a new instrument. Br J Gen Pract 2018 Jun;68(671):e433-e440

Oxford University Innovation Limited. PREOS-PC Compact (2018). https://innovation.ox.ac.uk/wp-content/uploads/2018/07/PREOS-PC-27-item-version-SAMPLE.pdf.

Peduzzi, Peter; Concato, John; Kemper, Elizabeth; Holford, Theodore R.; Feinstein, Alvan R. (1996). "A simulation study of the number of events per variable in logistic regression analysis". Journal of Clinical Epidemiology. 49 (12): 1373–1379. doi:10.1016/s0895-4356(96)00236-3. PMID 8970487.

Ricci-Cabello I, Avery AJ, Reeves D, Kadam UT, Valderas JM. Measuring Patient Safety in Primary Care: The Development and Validation of the "Patient Reported Experiences and Outcomes of Safety in Primary Care" (PREOS-PC). Ann Fam Med 2016 May;14(3):253-261.

Ricci-Cabello I, Reeves D, Bell BG, et al. Identifying patient and practice characteristics associated with patient-reported experiences of safety problems and harm: a cross-sectional study using a multilevel modelling approach. BMJ Qual Saf 2017;26:899–907.