

The clinical effectiveness and cost-effectiveness of telephone triage for managing same-day consultation requests in general practice: a cluster randomised controlled trial comparing general practitioner-led and nurse-led management systems with usual care (the ESTEEM trial)

John L Campbell, Emily Fletcher, Nicky Britten, Colin Green, Tim Holt, Valerie Lattimer, David A Richards, Suzanne H Richards, Chris Salisbury, Rod S Taylor, Raff Calitri, Vicky Bowyer, Katherine Chaplin, Rebecca Kandiyali, Jamie Murdoch, Linnie Price, Julia Roscoe, Anna Varley and Fiona C Warren



**National Institute for
Health Research**

The clinical effectiveness and cost-effectiveness of telephone triage for managing same-day consultation requests in general practice: a cluster randomised controlled trial comparing general practitioner-led and nurse-led management systems with usual care (the ESTEEM trial)

John L Campbell,^{1*} Emily Fletcher,¹ Nicky Britten,² Colin Green,² Tim Holt,³ Valerie Lattimer,⁴ David A Richards,² Suzanne H Richards,¹ Chris Salisbury,⁵ Rod S Taylor,¹ Raff Calitri,¹ Vicky Bowyer,⁶ Katherine Chaplin,² Rebecca Kandiyali,¹ Jamie Murdoch,⁴ Linnie Price,¹ Julia Roscoe,⁶ Anna Varley⁴ and Fiona C Warren¹

¹Primary Care Research Group, University of Exeter Medical School, Exeter, UK

²Institute of Health Service Research, University of Exeter Medical School, Exeter, UK

³Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

⁴School of Nursing Sciences, Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, UK

⁵Centre for Academic Primary Care, School of Social and Community Medicine, University of Bristol, Bristol, UK

⁶Department of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK

*Corresponding author

Declared competing interests of authors: none

Published February 2015

DOI: 10.3310/hta19130

This report should be referenced as follows:

Campbell JL, Fletcher E, Britten N, Green C, Holt T, Lattimer V, *et al.* The clinical effectiveness and cost-effectiveness of telephone triage for managing same-day consultation requests in general practice: a cluster randomised controlled trial comparing general practitioner-led and nurse-led management systems with usual care (the ESTEEM trial). *Health Technol Assess* 2015;**19**(13).

Health Technology Assessment is indexed and abstracted in *Index Medicus/MEDLINE*, *Excerpta Medica/EMBASE*, *Science Citation Index Expanded (SciSearch®)* and *Current Contents®/Clinical Medicine*.

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 5.116

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index and is assessed for inclusion in the Database of Abstracts of Reviews of Effects.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: <http://www.nets.nihr.ac.uk/programmes/hta>

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 08/53/15. The contractual start date was in November 2009. The draft report began editorial review in November 2013 and was accepted for publication in April 2014. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health.

© Queen's Printer and Controller of HMSO 2015. This work was produced by Campbell *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Editor-in-Chief of *Health Technology Assessment* and NIHR Journals Library

Professor Tom Walley Director, NIHR Evaluation, Trials and Studies and Director of the HTA Programme, UK

NIHR Journals Library Editors

Professor Ken Stein Chair of HTA Editorial Board and Professor of Public Health, University of Exeter Medical School, UK

Professor Andree Le May Chair of NIHR Journals Library Editorial Group (EME, HS&DR, PGfAR, PHR journals)

Dr Martin Ashton-Key Consultant in Public Health Medicine/Consultant Advisor, NETSCC, UK

Professor Matthias Beck Chair in Public Sector Management and Subject Leader (Management Group), Queen's University Management School, Queen's University Belfast, UK

Professor Aileen Clarke Professor of Public Health and Health Services Research, Warwick Medical School, University of Warwick, UK

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Peter Davidson Director of NETSCC, HTA, UK

Ms Tara Lamont Scientific Advisor, NETSCC, UK

Professor Elaine McColl Director, Newcastle Clinical Trials Unit, Institute of Health and Society, Newcastle University, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Professor of Health Sciences Research, Faculty of Education, University of Winchester, UK

Professor John Powell Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Institute of Child Health, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Please visit the website for a list of members of the NIHR Journals Library Board:
www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: nihredit@southampton.ac.uk

Abstract

The clinical effectiveness and cost-effectiveness of telephone triage for managing same-day consultation requests in general practice: a cluster randomised controlled trial comparing general practitioner-led and nurse-led management systems with usual care (the ESTEEM trial)

John L Campbell,^{1*} Emily Fletcher,¹ Nicky Britten,² Colin Green,² Tim Holt,³ Valerie Lattimer,⁴ David A Richards,² Suzanne H Richards,¹ Chris Salisbury,⁵ Rod S Taylor,¹ Raff Calitri,¹ Vicky Bowyer,⁶ Katherine Chaplin,² Rebecca Kandiyali,¹ Jamie Murdoch,⁴ Linnie Price,¹ Julia Roscoe,⁶ Anna Varley⁴ and Fiona C Warren¹

¹Primary Care Research Group, University of Exeter Medical School, Exeter, UK

²Institute of Health Service Research, University of Exeter Medical School, Exeter, UK

³Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

⁴School of Nursing Sciences, Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, UK

⁵Centre for Academic Primary Care, School of Social and Community Medicine, University of Bristol, Bristol, UK

⁶Department of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK

*Corresponding author john.campbell@exeter.ac.uk

Background: Telephone triage is proposed as a method of managing increasing demand for primary care. Previous studies have involved small samples in limited settings, and focused on nurse roles. Evidence is limited regarding the impact on primary care workload, costs, and patient safety and experience when triage is used to manage patients requesting same-day consultations in general practice.

Objectives: In comparison with usual care (UC), to assess the impact of GP-led telephone triage (GPT) and nurse-led computer-supported telephone triage (NT) on primary care workload and cost, patient experience of care, and patient safety and health status for patients requesting same-day consultations in general practice.

Design: Pragmatic cluster randomised controlled trial, incorporating economic evaluation and qualitative process evaluation.

Setting: General practices ($n = 42$) in four regions of England, UK (Devon, Bristol/Somerset, Warwickshire/Coventry, Norfolk/Suffolk).

Participants: Patients requesting same-day consultations.

Interventions: Practices were randomised to GPT, NT or UC. Data collection was not blinded; however, analysis was conducted by a statistician blinded to practice allocation.

Main outcome measures: Primary – primary care contacts [general practice, out-of-hours primary care, accident and emergency (A&E) and walk-in centre attendances] in the 28 days following the index consultation request. Secondary – resource use and costs, patient safety (deaths and emergency hospital admissions within 7 days of index request, and A&E attendance within 28 days), health status and experience of care.

Results: Of 20,990 eligible randomised patients (UC $n = 7283$; GPT $n = 6695$; NT $n = 7012$), primary outcome data were analysed for 16,211 patients (UC $n = 5572$; GPT $n = 5171$; NT $n = 5468$). Compared with UC, GPT and NT increased primary outcome contacts (over 28-day follow-up) by 33% [rate ratio (RR) 1.33, 95% confidence interval (CI) 1.30 to 1.36] and 48% (RR 1.48, 95% CI 1.44 to 1.52), respectively. Compared with GPT, NT was associated with a marginal increase in primary outcome contacts by 4% (RR 1.04, 95% CI 1.01 to 1.08). Triage was associated with a redistribution of primary care contacts. Although GPT, compared with UC, increased the rate of overall GP contacts (face to face and telephone) over the 28 days by 38% (RR 1.38, 95% CI 1.28 to 1.50), GP face-to-face contacts were reduced by 39% (RR 0.61, 95% CI 0.54 to 0.69). NT reduced the rate of overall GP contacts by 16% (RR 0.84, 95% CI 0.78 to 0.91) and GP face-to-face contacts by 20% (RR 0.80, 95% CI 0.71 to 0.90), whereas nurse contacts increased. The increased rate of primary care contacts in triage arms is largely attributable to increased telephone contacts. Estimated overall patient–clinician contact time on the index day increased in triage (GPT = 10.3 minutes; NT = 14.8 minutes; UC = 9.6 minutes), although patterns of clinician use varied between arms. Taking account of both the pattern and duration of primary outcome contacts, overall costs over the 28-day follow-up were similar in all three arms (approximately £75 per patient). Triage appeared safe, and no differences in patient health status were observed. NT was somewhat less acceptable to patients than GPT or UC. The process evaluation identified the complexity associated with introducing triage but found no consistency across practices about what works and what does not work when implementing it.

Conclusions: Introducing GPT or NT was associated with a redistribution of primary care workload for patients requesting same-day consultations, and at similar cost to UC. Although triage seemed to be safe, investigation of the circumstances of a larger number of deaths or admissions after triage might be warranted, and monitoring of these events is necessary as triage is implemented.

Trial registration: Current Controlled Trials ISRCTN20687662.

Funding: This project was funded by the NIHR Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 19, No. 13. See the NIHR Journals Library website for further project information.

Contents

List of tables	xiii
List of figures	xvii
Glossary	xix
List of abbreviations	xxi
Plain English summary	xxiii
Scientific summary	xxv
Chapter 1 Introduction	1
Scientific background and review of current literature	1
<i>Feasibility of telephone triage</i>	1
<i>Primary care workload</i>	1
<i>Cost</i>	1
<i>Patient experience of care</i>	2
<i>Patient safety</i>	2
<i>Patient health status</i>	2
Rationale for the research	2
Aims and objectives	3
Chapter 2 Methods	5
Study design	5
<i>Pilot study</i>	5
Ethical and governance arrangements	5
Patient and Public Involvement	5
Health technologies assessed	6
<i>Triage interventions</i>	6
<i>Core triage processes</i>	6
<i>Areas of flexibility</i>	7
<i>Training practices in the two triage systems</i>	7
<i>Computer decision support in nurse triage</i>	8
<i>Usual care comparator</i>	10
Inclusion and exclusion criteria	10
<i>Practices</i>	10
<i>Patients</i>	11
Recruitment and randomisation procedures	11
<i>Practices</i>	11
<i>Patients</i>	13
Outcome measures	16
<i>Describing practices and patients</i>	16
<i>Primary outcome measure</i>	16
<i>Secondary outcome measures</i>	17
Sample size	18
<i>Feasibility of patient recruitment</i>	19
<i>Secondary outcomes</i>	19

Trial delivery in participating practices	20
<i>Run-in period</i>	20
<i>Data collection period</i>	20
Data collection and management	21
<i>Clinician Form</i>	22
<i>Patient questionnaire</i>	22
<i>Case note review</i>	22
<i>Blinding</i>	22
<i>Data entry</i>	22
Monitoring intervention fidelity	23
<i>Integrity checks</i>	23
<i>Use of computerised decision support software in nurse-led telephone triage practices</i>	23
Statistical analysis	24
<i>Practice and patient characteristics</i>	24
<i>Primary outcome</i>	24
<i>Secondary outcomes</i>	25
Economic evaluation	26
<i>Aim</i>	26
<i>Method</i>	26
Process evaluation	29
<i>Aims and objectives</i>	29
<i>Methods</i>	29
<i>Data collection</i>	31
<i>Data analysis</i>	32
Chapter 3 Results	33
Flow of practices and patient participants in the trial	33
<i>Practices</i>	33
<i>Patients</i>	33
<i>Patient loss to follow-up</i>	36
<i>Describing usual care</i>	36
Practice descriptions and patient demographic characteristics	38
<i>Practice descriptors</i>	38
<i>Patient demographics</i>	38
Clinical outcomes	40
<i>Primary outcomes</i>	40
<i>Additional analyses based on primary outcome</i>	42
<i>Safety outcomes</i>	42
<i>Patient management on the index day</i>	47
<i>Resource use</i>	49
<i>Patient self-reported health status (European Quality of Life-5 Dimensions)</i>	52
<i>Patient experience</i>	52
<i>Case complexity</i>	52
Economic evaluation	54
<i>Estimate of the intervention cost (unit cost) for the triage interventions</i>	54
<i>Economic analyses</i>	59
Process evaluation	68
<i>Structure of the results</i>	68
<i>How the trial was experienced</i>	68
<i>How the telephone triage interventions are implemented in different practice settings</i>	76
<i>How triage was experienced</i>	79
<i>Acceptability of triage</i>	91
<i>What influences the extent to which the telephone triage interventions are seen to work or not work?</i>	92

Chapter 4 Discussion and conclusions	95
Summary of findings	95
<i>Primary care contacts in the 28 days following a same-day consultation request</i>	95
<i>Care on the index day</i>	96
<i>Resource use</i>	97
<i>Economic analysis</i>	97
<i>Safety</i>	98
<i>Patient health status</i>	98
<i>Patient experience of care</i>	99
Research findings in context	99
<i>Workload</i>	99
<i>Safety</i>	100
<i>Patient experience</i>	101
Process evaluation	101
Strengths and limitations	102
Implications for future research	105
Implications for health care	107
Acknowledgements	111
References	115
Appendix 1 ESTEEM trial: pilot study report	121
Appendix 2 Summary of changes to original ESTEEM protocol	135
Appendix 3 Audit Log Sheet	137
Appendix 4 Receptionist flow charts	139
Appendix 5 Questionnaire for withdrawn practices	143
Appendix 6 Covering invitation letters	145
Appendix 7 Patient information leaflet	147
Appendix 8 Patient questionnaire	153
Appendix 9 Receptionist Trial Log Sheet	157
Appendix 10 Adverse events reporting procedure	159
Appendix 11 Clinician Form	163
Appendix 12 Case note review	165
Appendix 13 Overview of practice integrity checks	167
Appendix 14 Assessment of nurses' use of computerised decision support software during ESTEEM	169
Appendix 15 Practice Profile Questionnaire	173

Appendix 16 Availability of patient information throughout the trial, consent status and initial patient management	179
Appendix 17 Practice Profile Questionnaire: additional data	183
Appendix 18 Predictors of case note review and questionnaire return	185
Appendix 19 Total primary care and accident and emergency contacts within 28 days of index day (primary outcome) across trial arm: per-protocol analysis	189
Appendix 20 Interactions between intervention and practice-level variables, and between intervention and age category/gender/ethnic group, for primary outcome	191
Appendix 21 Patient-reported health status (European Quality of Life-5 Dimensions): individual question frequencies	193
Appendix 22 Patient experience of care: individual question frequencies	195
Appendix 23 Case complexity	197
Appendix 24 ESTEEM intervention resource use and cost estimates	199
Appendix 25 Data used to estimate staff contact time for triage contact (source: Clinician Form data)	203
Appendix 26 Mean triage contact time by practice	205
Appendix 27 Alternative unit costs for general practitioner and nurse consultations	207
Appendix 28 Sensitivity analyses using alternative unit costs for out-of-hours contacts	209
Appendix 29 Total number of contacts by consultation type (from case note review)	211

List of tables

TABLE 1 Sample size calculation	19
TABLE 2 Outcome data, timing of data collection and source of data	21
TABLE 3 Inter-rater reliability of case note review data extraction	22
TABLE 4 Unit cost data (2012) with data sources for the base case	28
TABLE 5 Process evaluation: participating practices by region	29
TABLE 6 Process evaluation participating practice staff	31
TABLE 7 Process evaluation participating patients	31
TABLE 8 Estimate of the proportion of patients seeking same-day consultations who were ineligible for the trial	36
TABLE 9 Practice Profile Questionnaire	37
TABLE 10 Practice demographics and cluster size coefficient of variation	38
TABLE 11 Baseline patient demographics	39
TABLE 12 Primary outcome measure: total primary care contacts (including A&E) within 28 days of index consultation request across trial arms	41
TABLE 13 Sensitivity analyses based on the primary outcome	43
TABLE 14 Comparison of safety outcomes based on contacts across trial arms	46
TABLE 15 Patient management on index day	47
TABLE 16 Descriptive summary of individual contact types comprising the primary outcome and patient-reported use of NHS Direct	50
TABLE 17 Duration of consultations during trial	51
TABLE 18 Patient-reported health status (EQ-5D): ITT analysis	52
TABLE 19 Patient experience of care: ITT analysis	53
TABLE 20 Training and set up schedule for triage interventions	54
TABLE 21 Organisational training for triage interventions, staff attending (mean count by grade) and mean duration of training	54

TABLE 22 Triage-specific training for triage interventions, staff attending (mean count by grade) and mean duration of training	55
TABLE 23 Clinician contact time for triage intervention, by intervention type	55
TABLE 24 Unit costs used in estimate of triage intervention costs	56
TABLE 25 Mean cost of per-practice training and software required for triage interventions, and estimate of mean cost per same-day appointment request	58
TABLE 26 Estimated unit cost for triage interventions per same-day appointment request	58
TABLE 27 Summary table of sensitivity analyses and impact of different assumptions on estimate of unit cost for triage intervention (for same-day request)	58
TABLE 28 Description of primary cost data (summary view) and primary analyses with comparison by intervention	60
TABLE 29 Description of primary cost data (summary view) primary cost analyses with comparison by intervention type	61
TABLE 30 Secondary analyses and sensitivity analyses against primary cost analyses, total 28-day cost	62
TABLE 31 Same-day contacts: description of data, base case (detailed view) and exploratory economic analyses on the same day as appointment request	64
TABLE 32 Same-day contacts: exploratory analysis and costs for primary outcome contacts on the same day as appointment request	65
TABLE 33 Duration of face-to-face consultations following triage for ESTEEM patients	66
TABLE 34 Cost–consequences matrix, UC costs and consequences compared with triage interventions	67
TABLE 35 Overview of the process evaluation practices	69
TABLE 36 Summarising evidence from ESTEEM considered in reference to UC	109
TABLE 37 Summary table of pilot study outcomes against stopping rules	123
TABLE 38 Staff interview participants in the pilot study	131
TABLE 39 Summary of practices and patients for whom CDSS data were available	169
TABLE 40 Average length of call, number of questions, intensity score and percentage of patients for whom CDSS was used by practice	171

TABLE 41 Patient data availability, intervention status and consent status through the trial	179
TABLE 42 Baseline patient demographics by availability of case note review data and return of completed questionnaire	185
TABLE 43 Demographic associations with consent to case note review and questionnaire response	186

List of figures

FIGURE 1 Practice training structure	9
FIGURE 2 Practice recruitment procedure	12
FIGURE 3 Patient recruitment and consent procedure	14
FIGURE 4 Timeline of interventions	20
FIGURE 5 Consolidated Standards of Reporting Trials diagram	34
FIGURE 6 Patient recruitment by practices	35
FIGURE 7 Primary outcome results by arm	40
FIGURE 8 Redistribution of within-practice workload over (a) the 28-day workload (primary outcome measure) and (b) the index day	45
FIGURE 9 Summary, by arm, of within-practice patient management on the index day	49
FIGURE 10 Main trial practice recruitment at end of pilot (July 2010 to February 2011)	124
FIGURE 11 Flow diagram representing the adverse events reporting procedure for the ESTEEM trial	159
FIGURE 12a Descriptive flow of Clinician Form timings data for health economics analysis for GP triage arm	203
FIGURE 12b Descriptive flow of Clinician Form timings data for health economics analysis for nurse triage arm	204

Glossary

Contact A consultation within primary care of any of the types specified within the primary outcome (e.g. a consultation with a general practitioner or nurse, face to face or over the telephone, within surgery hours or out-of-hours primary care). All contacts are successfully attended by patients (i.e. excludes patients who do not attend) unless stated in the text.

Did not attend Use of the term 'did not attend' applies to both face-to-face and telephone contacts that are unsuccessful (i.e. patient fails to attend face to face or answer the telephone).

General practitioner face-to-face contact A face-to-face consultation with a general practitioner occurring at the practice within working hours.

General practitioner face-to-face out-of-hours primary care contact An out-of-hours, face-to-face consultation with a general practitioner occurring within a clinic setting (i.e. not a home visit).

Index consultation request The patient's request for a same-day consultation which served as the patient's point of entry to the trial and beginning of the 28-day follow-up period.

Index consultation/contact The triage or first management contact following the index consultation request.

Index day The day of the index same-day consultation request.

Per protocol Patients were considered to have been managed per protocol in each trial arm as follows:

- General practitioner triage intervention: patients receiving a general practitioner telephone contact as their first contact on the same day as the index consultation request.
- Nurse triage intervention: patients receiving a nurse telephone contact as their first contact on the same day following the index consultation request.
- Usual care: all patients, as no changes had been made to practices' system for managing same-day consultation requests.

Triage-bookable (appointment) An appointment reserved exclusively for the use of a general practitioner or nurse conducting triage to book for a patient who needs to be seen following the triage consultation.

Triage/first management A triage/first management contact was defined in each trial arm as follows:

- General practitioner triage intervention: a general practitioner telephone contact on the day of the index consultation request.
- Nurse triage intervention: a nurse or general practitioner telephone contact on the day of the index consultation request (recognising that the general practitioners may occasionally assist with the triaging).
- Usual care: the first contact occurring within 7 days of the index consultation request.

List of abbreviations

A&E	accident and emergency	LSOA	lower super output area
APHO	Association of Public Health Observatories	NIHR	National Institute for Health Research
CDSS	computerised decision support software	NT	nurse-led (computer-supported) telephone triage
CONSORT	Consolidated Standards of Reporting Trials	OR	odds ratio
DMC	Data Monitoring Committee	PCRN	Primary Care Research Network
DNA	did not attend	PCT	primary care trust
EMIS	Egton Medical Information Systems	POM	primary outcome measure
EQ-5D	European Quality of Life-5 Dimensions	PPI	Patient and Public Involvement
GLM	generalised linear model	PSSRU	Personal Social Services Research Unit
GP	general practitioner	R&D	research and development
GPT	general practitioner-led telephone triage	RCT	randomised controlled trial
HTA	Health Technology Assessment	REC	Research Ethics Committee
ICC	intracluster correlation coefficient	RM&G	Research Management & Governance
IMD	Index of Multiple Deprivation	RR	rate ratio
ISRCTN	International Standard Randomised Controlled Trial Number	SD	standard deviation
ITT	intention to treat	SF-36	Short Form questionnaire-36 items
		UC	usual care

Plain English summary

Demand for primary care has increased in recent years. 'Telephone triage' involves clinicians assessing patients' needs over the telephone, giving advice or arranging an appointment. We compared general practitioner- and nurse-led triage (GPT and NT, respectively) with 'usual care (UC)' for patients seeking same-day GP appointments, examining patterns of patients' service use, safety and experience of care, and UK NHS costs.

Forty-two practices participated: 15 (7012 patients) were randomly allocated to NT with computer decision-supported software, 13 (6695 patients) to GPT and 14 (7283 patients) to continue 'UC'. We included all patients who were telephoning and seeking a same-day GP appointment. Patients were asked about their opinion of the system via a postal questionnaire 4 weeks later. We interviewed some patients and staff. Information was collected from the records of patients who gave permission; we documented how often patients were seen over the 28 days following their same-day request.

Sufficient practices and patients took part to give us confidence in our results. Both types of triage increased primary care contacts in the 28-day follow-up, but, overall, triage cost almost the same as UC over the 28 days. Across a range of measures, triage appeared to be safe when compared with UC; there were no significant differences between trial arms in the number of deaths, the number of emergency hospital admissions or the number of accident and emergency department attendances. Patients were slightly less happy with NT than with GPT or UC. Interviews identified that the 'success' of triage depended on individual practice 'culture', and highlighted the complexity of introducing a major appointment system change.

We found that GPT or NT for patients seeking same-day GP consultations is potentially a useful approach to support the effective delivery of NHS primary care.

Scientific summary

Background

Demands on UK primary care are increasing, prompting an exploration of alternative ways of managing patients in an attempt to respond to government and societal expectations while continuing to deliver safe, high-quality care.

When combined with telephone consultation, it has been proposed that telephone triage improves the management of demand for primary care by providing more rapid access to health-care advice for patients, reducing non-attendance rates and freeing up opportunities for face-to-face consultation.

The majority of previous research relates to models of triage that involve nurses, with little research addressing the impact of general practitioner (GP) telephone triage. To date, there has been no large-scale randomised controlled trial (RCT) comparing the potential benefits and harms of GP- or nurse-led telephone triage (NT) of patients requesting same-day consultations.

Objectives

The overarching aim of this trial was, in comparison with usual care (UC), to assess the impact of NT- and GP-led telephone triage (GPT) on primary care workload and cost, patient experience of care, and patient safety and health status, for patients requesting same-day consultations in general practice. Specific objectives were to compare the effects on primary care workload and cost, and patient experience of care, patient safety and health status, of (1) NT vs. UC; (2) GPT vs. UC; and (3) NT vs. GPT. We also explored the experiences and views of patients and members of practice staff on the acceptability of telephone triage.

Methods

Design

Pragmatic cluster RCT incorporating economic evaluation and parallel qualitative process evaluation. A preliminary pilot RCT was conducted in six practices to (1) confirm the implementation of GP- and nurse-led triage systems as feasible; (2) confirm the proposed recruitment of practices and refine data collection systems; and (3) confirm the assumed level of clustering of outcomes.

Setting and participants

Forty-two general practices from four regions of England, UK (Devon, Bristol/Somerset, Warwickshire/Coventry and Norfolk/Suffolk). Participants were consecutive patients (aged ≥ 16 years or < 12 years) seeking a same-day face-to-face consultation with a GP. Patients aged 12.0–15.9 years were excluded owing to concerns regarding confidentiality of a mailed questionnaire, as were patients with health-care needs that were deemed too urgent to wait for triage (e.g. difficulties breathing, chest pain) and patients who were unable to communicate in English by telephone.

Randomisation

Individual patient-level randomisation was deemed impractical as it does not reflect the practice-wide reality of triage system implementation and is vulnerable to contamination. Consenting practices (clusters) were randomised in a 1 : 1 : 1 ratio using a secure remote automated allocation system designed by a statistician who was independent of the research team. The allocation sequence was computer generated, and minimised for geographical location, practice deprivation and practice list size.

Interventions

Patients from practices allocated to either of the triage arms who requested a same-day face-to-face consultation with a GP were advised that they would be called by the clinician (GP or nurse using computer decision support software) later that day to discuss their needs and to discuss the most appropriate management option. The options available included (1) self-care advice; (2) a further within-practice consultation on the same day; (3) a further within-practice consultation on a subsequent day; and (4) referral to another NHS service. Triage practices received standardised training on how to implement triage, although some tailoring of the intervention was permitted to suit local needs. Patients in UC practices were managed following the standard protocols for that practice.

Outcome measures

Our primary outcome measure (POM) is the total number of primary care contacts occurring within a 28-day period following a patient's same-day consultation request. We defined primary care as including consultations within general practice (including the triage contact), walk-in centres, out-of-hours primary care services and attendance at accident and emergency (A&E). All POM data were collected by trained researchers conducting case note reviews of electronic practice records.

Other process and outcome data were collected to document NHS resource use, appointment non-attendance and consultation length. Secondary outcomes of patient safety (number of deaths and emergency hospital admissions within 7 days of the index request and A&E attendance at follow-up), NHS resource use and costs, appointment non-attendance and consultation length were collected from case note review and primary care Clinician Forms. Patient-reported health status [European Quality of Life-5 Dimensions (EQ-5D)] and experiences of care (modified English national GP Patient Survey items) were collected by postal questionnaire.

Given the nature of the interventions, it was not possible to blind patients or practitioners to treatment allocation. Although the cluster design of the trial might theoretically allow researchers who are undertaking case note reviews to be blinded, our pilot study showed this was not possible in practice. The data analysis was carried out by a statistician who was blinded to treatment allocation.

Sample size

Based on a previous UK study comparing NT with UC for handling same-day consultation requests, we powered our study to detect a between-arm difference of 1.02 [standard deviation (SD) 0.78] vs. 1.38 (SD 1.79) at follow-up between the triage arms and UC [based on 90% power, intracluster correlation coefficient (ICC) 0.05, two-sided alpha 0.05]. Assuming 47% non-collection of the POM, it was necessary to recruit 7046 patients from 14 practices across each arm to reach our target of 3751 patients per arm for analysis (i.e. a total of 21,138 patients from 42 practices). The pilot study provided confirmation of our assumed ICC of 0.05 (i.e. 0.03, 95% CI 0.00 to 0.08). Furthermore, the pilot study led to a change in the method of patient consent to case note review from written consent (obtained from completed patient questionnaires) to include initial verbal consent obtained from the treating clinician.

Statistical methods

The primary analysis of the POM took the form of a regression analysis using a hierarchical model to take account of the cluster allocation, using a random effect to adjust for potential clustering effect by practice, and allowing for adjustment for practice-level minimisation variables and patient-level covariates shown to differ at baseline. Models were performed twice, initially using the UC arm as reference, and then using the GPT arm as reference, to derive comparison between the two triage arms. Investigation of the effect of missing POM data (owing to lack of availability of a case note review) was undertaken using multiple imputation methods, based on the assumption that missing case note review data were missing at random. Additional analyses were conducted on the POM, and on secondary measures derived from the POM, using the hierarchical generalised linear model methods described above. Some of these analyses were determined a priori; others were determined post hoc following initial inspection of the data.

Economic methods

Primary economic (cost) analyses were undertaken using data collected on the POM contacts taking place over 28 days, and conducted from the perspective of the NHS. Economic analyses were conducted in accordance with the statistical methods above. The primary economic analysis estimated the mean cost of care across each of the trial arms, to include triage (where used) and the items in the primary outcome. Analysis was based on a microlevel costing estimate for the triage intervention, and the use of published unit cost data for other elements of resource use. Estimates of the cost associated with triage interventions were based on incremental costs when compared with UC, with any capital costs and/or training costs depreciated/spread over an appropriate time period in the primary analyses (with other time horizons for these costs explored in sensitivity analyses). Sensitivity analyses were undertaken against the primary analyses to explore the implications of uncertainty in data used and the assumptions made within the primary analyses.

Results

Participating practices ($n = 42$) and patients were well balanced across the three trial arms with respect to key characteristics (practice list size, setting and deprivation; patient age, gender and deprivation). In total, 20,990 patients were eligible for the trial (UC $n = 7283$; GPT $n = 6695$; NT $n = 7012$). POM data were analysed for 16,211/20,990 (77%) participants (UC $n = 5572$; GPT $n = 5171$; NT $n = 5468$). There was some evidence of participation bias, with young adults (aged 16–24 years) being less likely than the reference age group (25–59 years), and women being less likely than men, to have POM data available.

The mean number of POM contacts in the 28-day follow-up period was 1.91 in UC (SD 1.43; total 10,616), 2.65 in GPT (SD 1.74; total 13,720) and 2.81 in NT (SD 1.68; total 15,400). Compared with UC, there was an increase in POM contacts of 33% in GPT [rate ratio (RR) 1.33, 95% CI 1.30 to 1.36] and 48% in NT (RR 1.48, 95% CI 1.44 to 1.52). There was a small increase of 4% in NT (RR 1.04, CI 1.01 to 1.08) compared with GPT.

Triage-arm patients had more diverse patterns of management when compared with UC. For GP face-to-face and telephone contacts combined across the 28-day follow-up period, the RR was 1.38 (95% CI 1.28 to 1.50) in GPT compared with UC, and 0.84 (95% CI 0.78 to 0.91) in NT compared with UC. GP face-to-face contacts decreased by 39% during the 28-day follow-up period in GPT compared with UC (RR 0.61, 95% CI 0.54 to 0.69) and in NT by 20% compared with UC (RR 0.80, 95% CI 0.71 to 0.90). Following the implementation of triage, no impact was observed on contacts with other services outside of the practice (out-of-hours primary care, walk-in centres or A&E) in either intervention arm.

Changes were also observed in the distribution of estimated patient–clinician contact time on the index day following the introduction of triage. Introducing GPT was associated with a small increase in patient–nurse contact time, whereas introducing NT was associated with both a decrease in patient–GP contact time and a substantial increase in overall patient–nurse contact time.

The estimated health-care costs over the 28-day follow-up were similar across all three arms, at a mean cost of approximately £75 (US\$120, €88) per patient.

There was no evidence of differences across the three trial arms with respect to patient safety (patient mortality, emergency hospital admissions and A&E attendance rates) and patient health status did not vary between triage arms and UC, or between GPT compared with NT. NT was somewhat less acceptable to patients than GPT or UC.

Data from 84 qualitative interviews with patients and staff (sampled from 10 practices) found no strong, compelling or consistent narrative about what works and what does not work when implementing telephone triage in primary care. Rather, the qualitative data highlighted the complexity of primary care organisations and the significance of individual practice culture. Both triage models were sometimes experienced positively by staff and patients, whereas others viewed it negatively.

Conclusions

We believe ESTEEM to be the first RCT to assess both nurse- and GP-led telephone triage of patients requesting same-day consultations in primary care. The ESTEEM trial achieved its recruitment target of 22,000 patients across 42 general practices in four regions in England.

Contrary to suggestions that triage can reduce primary care workload, we found that both GPT and NT increased the number of primary care contacts in the 28 days following a patient's same-day consultation request compared with maintaining a UC approach to such requests. Following the introduction of GPT, we observed a redistribution of GP workload from face-to-face to telephone consultations. Following the introduction of NT, we observed a redistribution of workload from GPs to nurses. Triage patients had more diverse patterns of management than those in UC, possibly indicating more flexible approaches to patient management in the triage arms. There was evidence of increased nurse deployment in both triage arms, with this being substantial following the introduction of NT.

When considering the differing patterns and duration of patient care contacts, both forms of triage were cost neutral to the NHS compared with UC. However, we found no important gains in patients' safety, health status or experience of care with triage.

Our process evaluation identified no strong, compelling or consistent narrative about what works and what does not work when implementing telephone triage in primary care. It did, however, identify key issues on which any practice considering implementing triage might want to reflect before adopting a triage system. These included consideration of issues relating to individual practice culture and capacity, and the forward planning, with the whole staff team, of any major changes in access arrangements.

Our results do not support a definitive policy recommendation to roll out across the NHS either nurse or GP triage for the management of same-day appointments. Triage, whether implemented by a GP or by a nurse using decision support software, should be introduced with full awareness of the whole-system implications arising from the decision to implement such a process. Notwithstanding this, clinician triage of patients seeking same-day consultations may offer advantages in supporting the flexible delivery of patient care, and potentially offers a useful approach in the armamentarium of tools facilitating the delivery of effective NHS primary care.

Trial registration

This trial is registered as ISRCTN20687662.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Chapter 1 Introduction

Scientific background and review of current literature

Demands on UK primary care are increasing. The introduction of a new General Medical Services contract in 2004¹ was followed by an estimated 25% increase in primary care workload.² For many practices and staff, addressing this increase in workload has involved an exploration of alternative ways of managing patients, in an attempt to respond to government and societal expectations while continuing to deliver safe, high-quality care.

The introduction of UK NHS walk-in centres, the 24-hour nurse-led telephone advice service NHS Direct and the more recent NHS 111 service, the increased diversity of staff skill mix and the use of remote consultations in primary care all represent organisational responses aimed at increasing the range of services available to patients and improving access to primary health care. When combined with telephone consultation, telephone triage is believed to provide rapid access to health-care advice for patients while freeing up opportunities for face-to-face consultation. Previous research³ has demonstrated the utility of nurse-led telephone triage (NT) of patients requesting same-day appointments in UK general practice. An 'average' practice (7000 patients) might be expected to manage around 20 patients each day requesting a same-day appointment request, representing around 35% of general practitioner (GP) workload.⁴

Some research evidence exists regarding the feasibility, workload implications and cost of telephone triage, and patient experience of care, safety and health status following telephone triage. Most evidence derives from models involving nurse triage; less research has been carried out to address the impact of GP telephone triage to date. There have been no large-scale multipractice studies examining the potential value of general practitioner-led telephone triage (GPT) or NT of patients requesting same-day consultations.

Feasibility of telephone triage

Previous studies suggest that around 50% of nurse triage calls in out-of-hours primary care settings may be handled by telephone advice alone.⁵⁻⁸ However, such studies have been small or focused solely on out-of-hours primary care. Use of telephones (fixed or mobile) is now almost universal in the UK,⁹ and recent years have seen a near quadrupling in the proportion of GP consultations conducted on the telephone.¹⁰

Primary care workload

In the short term, telephone triage, whether by doctor or nurse, appears to reduce GP contacts by around 40%,^{3,11} but it could be that this shifting of GP workload may result in the work undertaken by GPs comprising the more complex cases.¹² The reduction also appears to be associated with an increase in later return consultations of a roughly similar magnitude (30%,³ 50%¹³ between 2 and 4 weeks following a same-day appointment request), in effect smoothing out the peaks and troughs of GP workload that are associated with same-day appointment requests. Although this level of return consultations following triage may raise concerns regarding patient safety, convenience of care and cost-effectiveness, it has been suggested that a proportion of return consultations may be planned routine appointments, resulting from a downgrading of urgency following triage.

Cost

Equivocal results on costs of telephone triage and associated resource service use have been reported across three trials. NHS Direct nurse triage was more expensive than practice nurse triage of patients making same-day consultation requests¹⁴ but similar costs have been reported elsewhere between standard management and practice NT of same-day consultation requests.³ NT in out-of-hours primary care may reduce long-term NHS costs but may not be cost-effective at all times of the day.^{15,16}

Patient experience of care

Equivocal results on patient acceptability and satisfaction have been derived from small studies. One study¹³ reported no difference in satisfaction between telephone and face-to-face consultations. Jiwa *et al.*¹¹ reported that 80% of patients were satisfied with GP telephone management of same-day consultation requests, and Brown and Armstrong¹⁷ have suggested that patients who elect to use GP telephone consultations may do so as an alternative to face-to-face consultations in primary care.

Patient safety

Telephone consultation appears safe and effective.^{18,19} A UK equivalence trial (in which death within 7 days of contact was the primary outcome) established the safety of out-of-hours primary care NT by experienced nurses using computerised decision support software (CDSS) in comparison with usual care (UC).⁵ This is supported by work in the UK by the Richards *et al.*²⁰ study, in which audio tapes of nurse triage consultations to assess decision-making revealed that decision-making was rated by GP or nurse practitioner review to be mostly good, with minimal risk from poor nurse triage decisions. However, one Swedish study^{19,21} noted that nurses often used self-care advice and also over-rode software-determined recommendations on management. A recent Dutch study²² highlighted concerns regarding information-gathering in telephone triage when delivered without the use of CDSS, relying only on clinical protocols. Studies^{8,21,23,24} have highlighted the importance of training in the use of CDSS to address patient safety issues. Nurse telephone triage delivered within a framework of national guidelines (but not with CDSS) was judged to be efficient, although some concerns were raised with respect to patient safety.²⁵ One study,³ adopting a triage system involving 'computerised management protocols developed by the practice' identified a substantial increase in accident and emergency (A&E) attendance, although actual numbers were small. Although computerised, such a system did not provide CDSS (DA Richards, Institute of Health Research, University of Exeter Medical School, 2008, personal communication) such as is now available within a number of NHS primary care computer systems, and which we propose to examine in this study.²⁶ The other trial by Richards *et al.*¹⁴ used CDSS for NHS Direct triage nurses, but not for nurses acting in primary care. A systematic review²⁷ of nine studies of telephone consultation and triage noted that it is unclear if telephone management simply delays visits and thus also the provision of definitive care.

Patient health status

Several randomised studies (but none involving telephone triage) have compared the health status of primary care patients following consultations with a doctor or a nurse by patients with minor problems or after a same-day consultation request. One study²⁸ identified no difference in health status [Short Form questionnaire-36 items (SF-36) scores] between the intervention groups when followed up after 2 weeks. Similar findings have been reported with respect to resolution of patients' symptoms and concerns after 2 weeks,²⁹ or in the proportion of patients reporting themselves as 'cured' or 'improved' 2 weeks after a consultation with either a doctor or a nurse.³⁰

Rationale for the research

The four UK-based trials^{3,5,11,13} of primary care telephone consultation and triage have been conducted in relatively small populations and/or in limited numbers of practice settings (i.e. urban, rural), and most studies examining NT without the use of CDSS, although research undertaken by Lattimer *et al.*⁵ did involve out-of-hours primary care nurse telephone consultations using the CDSS 'Odyssey TeleAssess', provided by Plain Healthcare Ltd. Despite uncertainty about the benefits and costs, many practices operate GPT or NT systems as a way of providing fast access to care for patients and in order to manage practice workload, as demonstrated by the almost fourfold increase in proportion of consultations conducted over the telephone.¹⁰ As an example, in 2008, the NHS Institute for Innovation and Improvement³¹ promoted a model of GPT – the Stour Access system – within which GPs triage all patient requests for care by telephone but without any robust evidence about benefits. We therefore proposed to address this important agenda in a large-scale experimental study of two forms of triage currently being promoted by the NHS for use in UK primary care. Our findings may be generalisable to other health settings, especially those with strong primary care-based health-care systems.

Aims and objectives

The overarching aim of this trial was to assess, in comparison with UC, the impact of NT and GPT mechanisms on primary care workload and cost, patient experience of care, and patient safety and health status for patients requesting same-day consultations in general practice.

The specific research objectives were as follows:

Pilot and feasibility study To undertake an external pilot randomised controlled trial (RCT) in six practices to:

1. confirm the ability of practices to implement the GPT and NT systems
2. confirm the proposed process for recruitment of practices
3. review the assumed level of clustering of outcomes
4. check data collection systems, and
5. identify potential difficulties in implementing the triage systems.

Main trial To undertake a three-arm pragmatic cluster RCT comparing, for patients requesting a same-day consultation in general practice, the effects on primary care workload and cost, and on patient experience of care, safety and health status of:

- i. GPT compared with UC
- ii. NT compared with UC
- iii. GPT compared with NT.

The funding arrangements for this trial through the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) programme involved the delivery of the 1-year pilot study and satisfying a number of key stopping rules (see *Appendix 1*), before progression to the main trial phase.

Chapter 2 Methods

Study design

Consistent with the Medical Research Council framework for evaluating complex interventions,³² a two-stage mixed-methods study was conducted, comprising an external pilot trial, followed by a definitive cluster RCT. The ESTEEM trial was designed as a pragmatic three-arm, multicentre cluster RCT. Practices were randomised 1 : 1 : 1 to receive GPT, NT or UC as comparators. The main trial included a parallel economic evaluation to examine the cost-effectiveness of the two triage interventions and a process evaluation to assess the intervention acceptability from the perspectives of patients and practice staff.

Pilot study

A 12-month, external pilot cluster RCT and parallel process evaluation was undertaken in six practices across Devon and Bristol. This work informed aspects of the main trial protocol published in 2013.³³ We present the methods implemented within the main trial, relating to the triage interventions, practice and patient inclusion and exclusion, practice and patient recruitment, randomisation, and the outcome measures and sample size. A full account of the changes made to the main trial methods can be found in *Appendix 2*.

Ethical and governance arrangements

Multicentre ethical approval for the study was obtained from South West 2 NHS Research Ethics Committee (REC) in October 2009 (reference 09/H0202/53). Local NHS research governance approvals were obtained from the primary care trusts (PCTs) in Devon, Somerset, North Somerset, Bath and North East Somerset, Bristol, South Gloucestershire, Warwickshire, Coventry, Norfolk, Suffolk and Great Yarmouth, and Waveney. At the time this work was undertaken, PCTs were the main unit of administration of primary care, with a total of 152 PCTs in England, each covering an average population of around 330,000 individuals. PCTs were abolished under major changes to the NHS, introduced in 2012, with clinical commissioning groups taking over former PCT functions. The Research Management & Governance Unit, Devon PCT, acted as the study sponsor. A Trial Management Group and an independent Trial Steering Committee and Data Monitoring Committee (DMC) ensured that the study was conducted within appropriate NHS and professional ethical guidelines. The trial was registered with the International Standard Randomised Controlled Trial Number (ISRCTN) Register (reference 20687662).

Patient and Public Involvement

The Patient and Public Involvement (PPI) group supporting ESTEEM initially met to review the study protocol and methods. Practice service users (nurses, GPs, managers and administrative staff) were also consulted to inform the trial design. It was planned that patient service users would be involved at all stages of the pilot study and main trial. Patient representatives were recruited through an existing GP practice patient group local to the Exeter site (not an ESTEEM trial practice). Eleven patient representatives were recruited and were invited to contribute to some or all of the various tasks as they preferred. Although our intention was to recruit representatives on the basis of their demographic characteristics, to reflect variations in the people most likely to be requesting same-day consultations in general practice (e.g. parents of young children, retired people) the majority of the patient representatives were of retirement age.

There were a number of PPI tasks, as listed below, and not all patient representatives wished to contribute to every task over the duration of the study.

- i. Most of the patient representatives contributed to the development of study materials in preparation for ethical review before the pilot study began, and then again at the end of the pilot study for use in the main trial. This involved helping to design the patient information sheets, consent forms, topic guides for patient interviews, and patient questionnaire. Comments were obtained through a combination of face-to-face meetings and via e-mail and post.
- ii. Advice was also sought on ways to improve patient participation in interviews for the process evaluation. This led to the introduction of an option for telephone interviews during the main trial.
- iii. (The same) two patient representatives sat on the Trial Steering Committee for the whole period of the trial and one, if not both, attended every meeting.
- iv. A small number of patient representatives contributed to the process of analysing qualitative data from the process evaluation during the pilot study.
- v. A small number of patient representatives attended meetings and received postal material during the running of the trial in order to be kept informed of study progress.
- vi. All patient representatives were invited to contribute to the dissemination of the study results at the end of the project, by way of providing input to the plain English summary of the trial findings.

Health technologies assessed

Triage interventions

The GPT and NT interventions were complex interventions that involved staff training (clinical and technology based); CDSS to support the delivery of NT; process and organisational change in practices regarding reception activity and appointment system management; and accommodation of patient expectations. Some core elements of triage delivery were common to, and adopted by, all practices in both intervention arms. However, some organisational flexibility was permitted because of the complex nature of the intervention. Full details of the interventions are available from the research team.

Core triage processes

All patients contacting the practice initially spoke to a receptionist. Once the receptionist established that the patient (or a proxy asking on their behalf) was requesting a same-day, face-to-face appointment with a GP, the patient was asked to provide a contact telephone number and was advised that the clinician (GP or nurse, according to the practice's allocation) would call them back within around 1–2 hours. This timescale was suggested as a guide for practices but was not considered mandatory.

On telephoning the patient ('index consultation'), the clinician discussed the presenting condition and had a range of management options at their disposal:

- i. give the patient self-care advice
- ii. book the patient into a 'triage-bookable' face-to-face or telephone appointment with the relevant health professional later the same day
- iii. book the patient into a 'triage-bookable' face-to-face or telephone appointment with the relevant health professional on another day
- iv. book the patient into any routine appointment available
- v. refer the patient to other NHS services where appropriate, including those outside the practice.

Practices kept a number of 'triage-bookable' appointments with GPs and nurses, reserved exclusively for triaging clinicians to book. When patients required face-to-face appointments, they could be triaged to any available, appropriately timed, triage-bookable appointment slot; this may or may not have been with the patient's usual doctor (at the discretion of the practice) or on the same day as the patient's index consultation.

Areas of flexibility

Although the core triage processes were specified, practices had some areas of flexibility in implementing triage interventions, in order to accommodate the complexities of individual practice organisation. These models used were informed by individual practices' audits of appointment requests, GP and nurse capacity information, and guidance and training given to practices prior to starting the trial (see *Training practices in the two triage systems*). Flexibility was allowed in terms of the triaging clinician, triage availability, the triaging of babies, and same-day requests for nurse practitioner consultations.

Triaging clinician

Three alternative staffing models emerged regarding which clinician (GP or nurse) delivered the triage: (1) one clinician (GP or nurse) at a time conducted a triage session (on a rotating basis); (2) a number of clinicians (GPs or nurses) conducted triage sessions on the day (on a rotating basis); or (3) in GPT practices only, all of the GPs conducted triage each day, with GPs triaging their own patients to triage-bookable appointments.

Triage availability

Practices were asked to triage all consecutive eligible patients during opening hours to ensure the least disruption and shortest time of trial data collection possible. In the event that limited staff resources prevented triage of all eligible patients each day, practices were permitted to operate triage within specific agreed sessions ('research window'), amounting to no less than 50% of the week, encompassing all five working days and both mornings and afternoons. These arrangements were agreed with the research team in advance. Practices were advised of the importance of all eligible patients being triaged during an entire session, as opposed to receptionists booking in a set number of triage patients within a session and then 'closing' the triage once the allocated slots were taken. Practices operating triage within 'research windows' had their data collection period extended until the patient recruitment target was reached.

Triaging babies

Practices were advised that babies (i.e. ≤ 2 years of age) were eligible for trial entry according to the inclusion and exclusion criteria, with the parent or guardian identified as the individual who participated in the triage call. As some practices routinely see babies as soon as possible on the same day of the appointment request, it was expected that some practices would opt to book these patients in for a face-to-face appointment rather than add them to a triage list. Although all of the intervention practices were encouraged to offer triage to babies during intervention training, some practices elected to offer face-to-face appointments in line with their existing practice policy. When this happened, the babies were still included in the trial analysis on an intention-to-treat (ITT) basis.

Requests for same-day, face-to-face nurse practitioner consultations

Some practices operated a nurse practitioner-led clinic for acute or minor illnesses, for which the nurse typically managed same-day consultation requests. This request was deemed equivalent to a same-day GP appointment request, as the patient would see a GP if the nurse practitioner was unavailable. Such patients were deemed eligible for trial entry and managed as per the treatment arm allocation.

Training practices in the two triage systems

Role of commercial providers in triage training

To prepare GPT and NT practices for incorporating telephone triage within their appointment systems, we worked with Productive Primary Care Ltd, a Leicestershire-based, NHS-focused commercial organisation that works nationally across the UK with NHS organisations. Plain Healthcare's 'Odyssey PatientAssess' (derived for general practice use from 'Odyssey TeleAssess' or 'TAS') CDSS was used to support the nurses in NT practices. Plain Healthcare, then part of the Avia Health Informatics PLC group, supplies the NHS with CDSS systems (e.g. out-of-hours primary care services). Plain Healthcare liaised directly with NT practices to discuss IT and training requirements for the installation of the software. Following a 'site initiation call' between Plain Healthcare and each practice, installation and testing took place over a

6-week period. The CDSS was fully embedded within the Egton Medical Information System (EMIS, Yeadon, Leeds) 'PCS' and 'LV', and installed as a separate 'stand-alone' window tab to the clinical notes within all other types of practice IT system [e.g. Microtest Evolution (Microtest, Bodmin, Cornwall), Vision (In Practice Systems Ltd, Battersea, London), SystmOne (The Phoenix Partnership, Horsforth)].

Nurses received CDSS training and training in telephone consultation skills. Following this, there was a pretrial period of 1 month during which nurses practised using CDSS in simulated patient scenarios during their daily work. Towards the end of this period, the nurses' use of the system was assessed by Plain Healthcare trainers, with all nurses needing to demonstrate proficiency in its use before being recruited into the trial. It is also important to note that the CDSS is designed to support the nurses' clinical decision-making; there was no requirement in the trial that the recommended advice the CDSS generated had to be followed.

Computer decision support in nurse triage

Odyssey PatientAssess [www.advancedcomputersoftware.com/ahc/products/odyssey-patient-assess.php (accessed 5 November 2014)] is a UK product, developed to support nurses and paramedics to assess the clinical needs of patients. It is already being used by several out-of-hours primary care and NHS walk-in services, and is also the subject of the Department of Health-funded SAFER1 trial focusing on the care of older people who have called the 999 service following a fall, and the HTA-funded SAFER3 trial focusing on the care of adults who have called the 999 services and are not in need of transfer to an emergency department.³⁴

Odyssey PatientAssess has been evaluated in a number of other RCTs, and its clinical effectiveness and cost-effectiveness in settings other than in-hours general practice is already established. For instance, it has been demonstrated to be safe and cost saving in the long term in one study^{5,16} involving a co-investigator compared with GP telephone triage in a trial of out-of-hours primary care consultations. That study remains the largest trial of nurse telephone triage to date.^{5,16} Furthermore, in excess of 60% of PCTs currently commission out-of-hours primary care services that use nurses to triage patients by telephone, supported by Odyssey PatientAssess (Chris Coyne, Plain Healthcare, 2008, personal communication). However, the findings and experience of out-of-hours primary care (providing care for around 10.8 million contacts per year in UK) cannot necessarily be generalised to the very different system providing in-hours primary care and experiencing around 1 million contacts per working day.

Odyssey PatientAssess provides the user with a network of assessment prompts and guided responses relating to over 465 presenting complaints. It allows for multiple symptoms to be evaluated simultaneously, supported by evidence-based frameworks for referral and self-care, with all assessment data remaining visible at all times. It supports the clinician's judgement and expertise through enhancing normal consultation processes. The clinical database comprises several hundred assessment and examination guidelines and protocols, each linked to triage, treatment and advice guidelines, differential diagnoses, patient information and education. These are maintained by an in-house clinical development team, which reviews the entire clinical content at least annually to ensure that it reflects current best practice, including National Institute for Health and Care Excellence (NICE) guidance.

The assessment screens include drop-down menus that provide regularly updated referenced information on differential diagnoses and rationales for lines of enquiry for each type of presentation – so reminding the user about the importance of different lines of enquiry. For the purposes of the ESTEEM trial, Odyssey PatientAssess was either embedded within GP computing systems (EMIS and SystmOne) or installed to run in parallel alongside all other systems (Synergy, Microtest, Vision). Odyssey PatientAssess guides and stores documented records of the assessment, advice and/or referral of each patient producing a fully auditable record. Based on the data elicited during the telephone assessment, Odyssey PatientAssess suggests an appropriate care plan (e.g. patient advice, same-day appointment, home visit, routine appointment, 999 referral).

Overview of practice training

Practices were trained in research procedures and the delivery of triage interventions, following a model (Figure 1) devised during the pilot study. Training was delivered in six stages, and by a combination of Productive Primary Care (GPT and NT only), Plain Healthcare (NT only) and trial researchers (all arms). 'Triage training' workshops were organised for GPT and NT practices (training components 3–5, described below) and delivered as a package to staff from a number of practices at a central location in each of the recruitment areas. If practices were unable to attend a central training event, the various components were delivered on site at the practice.

1. *Inception meeting* An initial whole-practice meeting with the trial researcher was held to outline the trial, staff group involvement and timeline of training. The meeting was convened for the practice manager, at least three receptionists, at least two GPs – to include the two GPs identified as clinical leads for the trial – and up to four nurses (NT only).
2. *Appointment request audit* To assess patient demand for same-day and prebookable appointments, a 1-week audit was conducted by the practice reception team. Capacity information was collated relating to the number of GP and nurse sessions provided per week and the number and typical duration of appointments offered within each session (including the proportion reserved for same-day and prebookable use). The trial researcher provided the reception team with Audit Log Sheets (see Appendix 3) and asked them to record all incoming calls each day for a week (Monday to Friday), indicating whether each call was a same-day or prebookable appointment request. The researcher maintained contact with the practice during the first day or two of the audit to ensure that the procedure was followed appropriately.

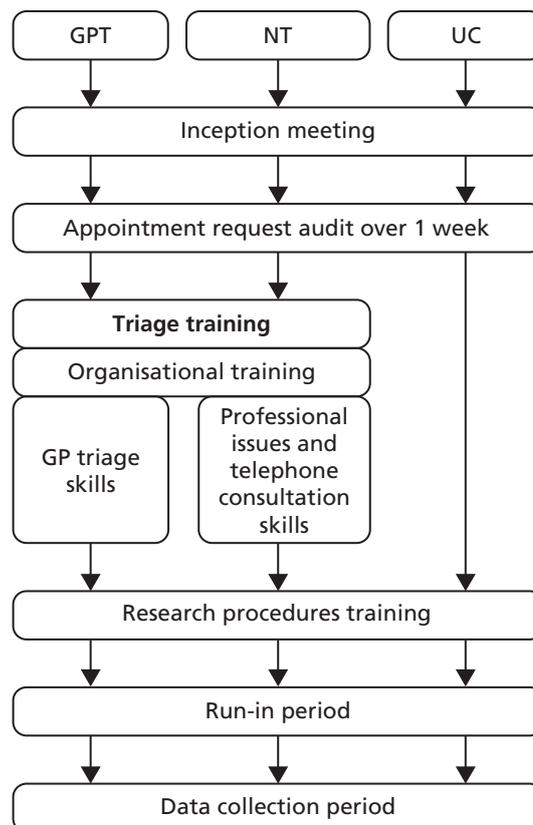


FIGURE 1 Practice training structure.

3. *Organisational training and guidance in organising the triage system* Productive Primary Care trained the practice manager, at least two receptionists, at least two GPs (GPT) and up to four nurses and one GP (NT). The training incorporated advice for the reception teams on how to communicate the introduction of the triage system to patients. Practical guidance on setting up the appointment system to incorporate triage consultation slots and face-to-face appointment slots for the use of the triaging clinician was also given. Using the data collected during the appointment request audit, Productive Primary Care used a spreadsheet (the Appointment Redesign Tool) to estimate the number of appointments that the practice needed to provide and the proportion that should be 'triage bookable' based on their previous experience of working with a range of practices. The details of the calculations for these estimates are the intellectual property of Productive Primary Care.
4. *GP triage skills* In GPT practices, Productive Primary Care delivered training to GPs on how to communicate the operation of the new triage system to patients, and general guidance for consulting over the telephone. At least two GPs at each practice were required to receive this training and were expected to disseminate the information to their other GP colleagues.
5. *Professional issues and telephone consultation skills* All nurses at each NT practice who would undertake triage were required to receive this training. A learning and development adviser from Plain Healthcare presented information and advice on the professional issues that may arise from nurses consulting over the telephone. Guidance on telephone consulting skills, including role-play scenarios, was provided. In addition, nurses received one-to-one remote training from an advisor while they practised using the CDSS on computerised simulated patient scenarios in the weeks leading up to the trial, culminating in a proficiency assessment.
6. *Research procedures training* Trial researchers provided briefing sessions to practice staff in data collection procedures. Staff groups included the reception team, practice manager, key IT and administrative staff and the GPs and nurses to be involved in triage or dealing with patients requesting same-day consultations.

Usual care comparator

Practices were asked to continue with their standard consultation management systems for handling same-day consultation requests. UC practices also (1) received training through the inception meeting; (2) performed the appointment request audit; and (3) received the standard research procedures training, as outlined for the intervention arms.

To describe the range of systems that constitute UC, all 42 practices were asked to complete a Practice Profile Questionnaire before randomisation to collect details on their current staffing and appointment system arrangements to manage same-day consultation requests, including the extent to which they already used triage of any sort.

Inclusion and exclusion criteria

Practices

Practices already implementing a documented triage system for routinely handling same-day GP consultation requests were excluded. We defined a documented triage system as a system involving telephone triage (by a GP or a nurse) managing > 75% of all same-day consultation requests received. This definition recognised that many GP practices already undertake some form of telephone triage. Practices also had to be willing to be randomised to one of the three trial arms.

Patients

Inclusion criteria

All consecutive patients making telephone requests for a same-day, face-to-face consultation with a GP during the data collection period were potentially eligible. When an individual patient had multiple same-day consultation requests during the recruitment period, only the first contact was included in the study to avoid confusion in following up the contact (although subsequent calls would be triaged in the same way, according to trial arm). All patients aged < 12 years and ≥ 16 years requesting eligible appointments were included with respect to the primary and secondary outcome measures. Parents or guardians of children aged < 12 years provided consent on behalf of the child.

Exclusion criteria

Temporary residents or young people aged 12.0–15.9 years were excluded, as the study involved receipt of a postal questionnaire along with written consent to review case notes. We believed that receipt of the questionnaire at the young person's address may have inadvertently led to a breach of confidentiality should third parties have access to the young person's mail. Although excluded from trial participation, both temporary residents and young people received the intervention as per trial allocation. Adults aged ≥ 16 years were included, unless the practice wished to screen out patients for whom they felt it would be inappropriate to send a questionnaire (e.g. patients with recent bereavement, vulnerable adults).

Patients seeking urgent or emergency care on account of the following conditions were excluded:³⁰ too ill to participate (severe chest or abdominal pain or severe difficulty breathing; vomiting blood; altered consciousness; seizures; pregnancy-related problems; or severe psychiatric symptoms); unable to speak English or difficulties with communicating on the telephone (hearing or speech). Reception staff members were trained to refer to a standardised 'Receptionist flow chart' to ascertain patient eligibility (see *Appendix 4*).

Patients ringing on a second occasion were not included in the trial again (i.e. their point of entry into the trial was on the first occasion they rang the practice seeking a same-day consultation). However, in practice, receptionists would have recorded the patient as being eligible and sent their information to the research team. Our database would have flagged them as a duplicate and excluded them from being entered into the trial database again.

Recruitment and randomisation procedures

Practices

We approached all practices within our four participating centres (Devon, Bristol/Somerset, Warwickshire/Coventry and Norfolk/Suffolk). To maximise recruitment we ran the trial in conjunction with the NIHR Primary Care Research Network (PCRN), including the networks in the South West, West Midlands (South) and East of England. In recognition of the challenges of recruitment into large-scale clinical trials of complex interventions,³⁵ a two-stage recruitment process was adopted (*Figure 2*). A written letter inviting participation was sent to all practices in the four geographical areas. This letter was co-signed by the trial Chief Investigator, the local principal investigator and local PCRN clinical lead. Practices expressing preliminary interest were sent an information pack and offered a practice visit by a researcher. At this recruitment meeting (for which the practices were remunerated), the trial design and methods were explained, and staff were given the opportunity to ask questions. We prioritised recruiting practices in workable proximity to recruitment centres (taking account of relevant sampling issues).

Randomisation procedure

Individual patient-level randomisation was deemed impractical as it was unable to reflect the practice-level triage system implementation and was vulnerable to contamination. A cluster RCT design was adopted, with practices that agreed to take part subsequently randomised to one of the three trial intervention

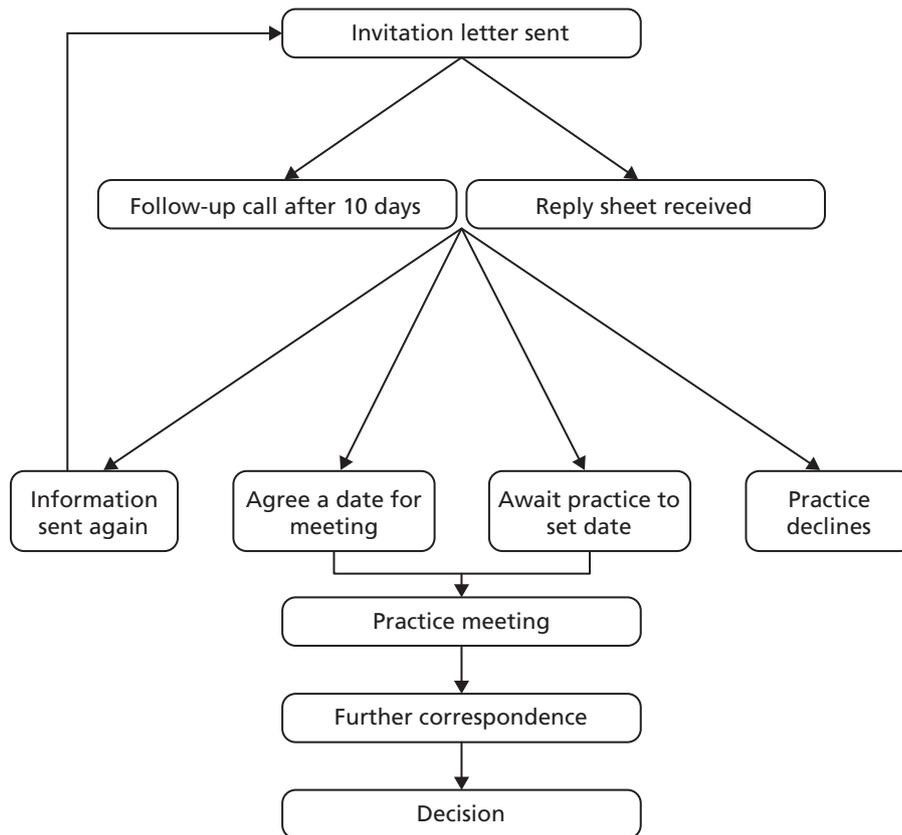


FIGURE 2 Practice recruitment procedure.

arms (GPT, NT, UC).³⁶ To manage the recruitment process, practices were randomised in three distinct 'waves'. Wave 1 took place during spring and summer 2011, wave 2 during winter 2011/12 and wave 3 during spring and summer 2012.

The sequence of practice randomisation was computer generated and minimised for geographical location (Devon, Bristol, Norwich, Warwick), practice deprivation [based on the data provided by the Association of Public Health Observatories (APHO)³⁷ at time of randomisation, practices were classed as 'deprived' (i.e. score is above the average for England) and 'non-deprived' (score is average or below average) and practice list size of 'small' (< 3500 patients), 'medium' (3500–8000 patients) and 'large' (> 8000 patients)]. A stochastic element was included within the minimisation algorithm to maintain concealment. Practice allocation was undertaken using a password-protected web portal (developed and validated by the UK Clinical Research Collaboration (UKCRC)-accredited Peninsula Clinical Trials Unit) by a statistician who was independent from the trial statistical team. Allocation was communicated by e-mail to the trial manager who then informed the local researchers, who, in turn, informed practices of their allocation.

Replacing practices that withdrew post randomisation

To maintain balance between groups, practices that withdrew post randomisation were replaced with a practice from a waiting list in the same geographical area, and of a similar size and deprivation level, where possible. Owing to the limited number of waiting list practices, it was not possible to select a matched practice randomly. Instead, replacement practices were purposively allocated to a trial arm based on locality, practice size and deprivation index. However, other than knowing that the practice was matched for the above characteristics, the researcher making the practice selection was unaware of any other characteristics of that practice. Replacement practices were not aware of their potential allocation when the researcher checked that they were still willing to enter the trial. A questionnaire exploring reasons for withdrawal was sent to practice managers (see *Appendix 5*).

Patients

During the intervention period all practices were provided with a poster to be displayed in the practice to inform participants of the trial and its purpose, and some practices incorporated a voice message onto their telephone systems to explain that the triage system was being trialled. All incoming calls were received by a practice receptionist, who (referring to the 'Receptionist flow chart') ascertained whether the patient was requesting a same-day, face-to-face consultation with a GP at the surgery. Reception staff asked patients to briefly outline the nature of their problem in order to facilitate timely care (but patients did not have to disclose this information). If the patient was not requiring urgent or emergency care (see *Inclusion and exclusion criteria*), the receptionist informed the patient of the current practice consultation arrangements (according to trial arm) and managed the request following the standard operating procedure for that practice.

Patient consent

The patient recruitment and consent procedure is summarised in *Figure 3*.

Consent to trial participation

Consistent with the cluster RCT design, all of the eligible patients were managed according to the practice's trial allocation. Patients were not asked to give individual consent on the basis that all UK practices are free to organise patient care to fit their circumstances. Reception staff simply explained the current consultation arrangements to patients, and informed patients that they would receive a postal questionnaire in 4 weeks' time regarding their experience of care and that their completion of the questionnaire would be greatly appreciated.

Consent to the case note review

At the end of the clinical interaction during the index triage or UC consultation, the clinician reminded the patient that they may receive a questionnaire in 4 weeks' time regarding their experience of care that day. The clinician referred to standard text, asking the patient for their initial verbal consent to the case note review aspect of the study. Twenty-eight days after the index consultation, patients were sent a questionnaire pack by the practice.

Each pack included a covering invitation letter (see *Appendix 6*), an information sheet describing the study (see *Appendix 7*), a questionnaire (see *Appendix 8*) and a prepaid return envelope. The pack also contained a flyer providing details of an incentive to return the questionnaire: entry to a prize draw to receive one of 20 prizes of £25 worth of shopping vouchers. Patients' written consent to a case note review was sought on the final page of the questionnaire. Patients who had given initial verbal consent at the index consultation could opt out of the case note review at this stage if they wished, or opt in if they had initially declined, with written consent always taking priority over the previous verbal response. A name and contact details were provided if patients required any further information about the trial. Patients who did not wish to complete a questionnaire were encouraged to return a blank questionnaire in the prepaid envelope in order to receive no further contact in relation to the study. Non-responders after 2 weeks were sent a reminder questionnaire pack.

Capturing the patient sample

The receptionist flagged a triage or UC appointment as being allocated for a study participant by entering a free-text comment ('ESTEEM') and the date of the appointment request as a note on the appointment slot. Within a week, a dedicated member of the practice staff applied a study-specific Read Code to each 'ESTEEM'-flagged patient, thus defining the eligible patient sample. This Read Code was dated as per the receptionist's free-text comment, allowing the practice to run electronic searches to generate an ESTEEM patient list that was passed to the research team as the denominator for the study database.

In the triage practices, if for any reason an eligible patient was booked a face-to-face or any other type of appointment instead of triage (e.g. a patient being unable to receive a telephone call while at work), such patients were still captured as part of the eligible sample and followed up according to the study protocol.

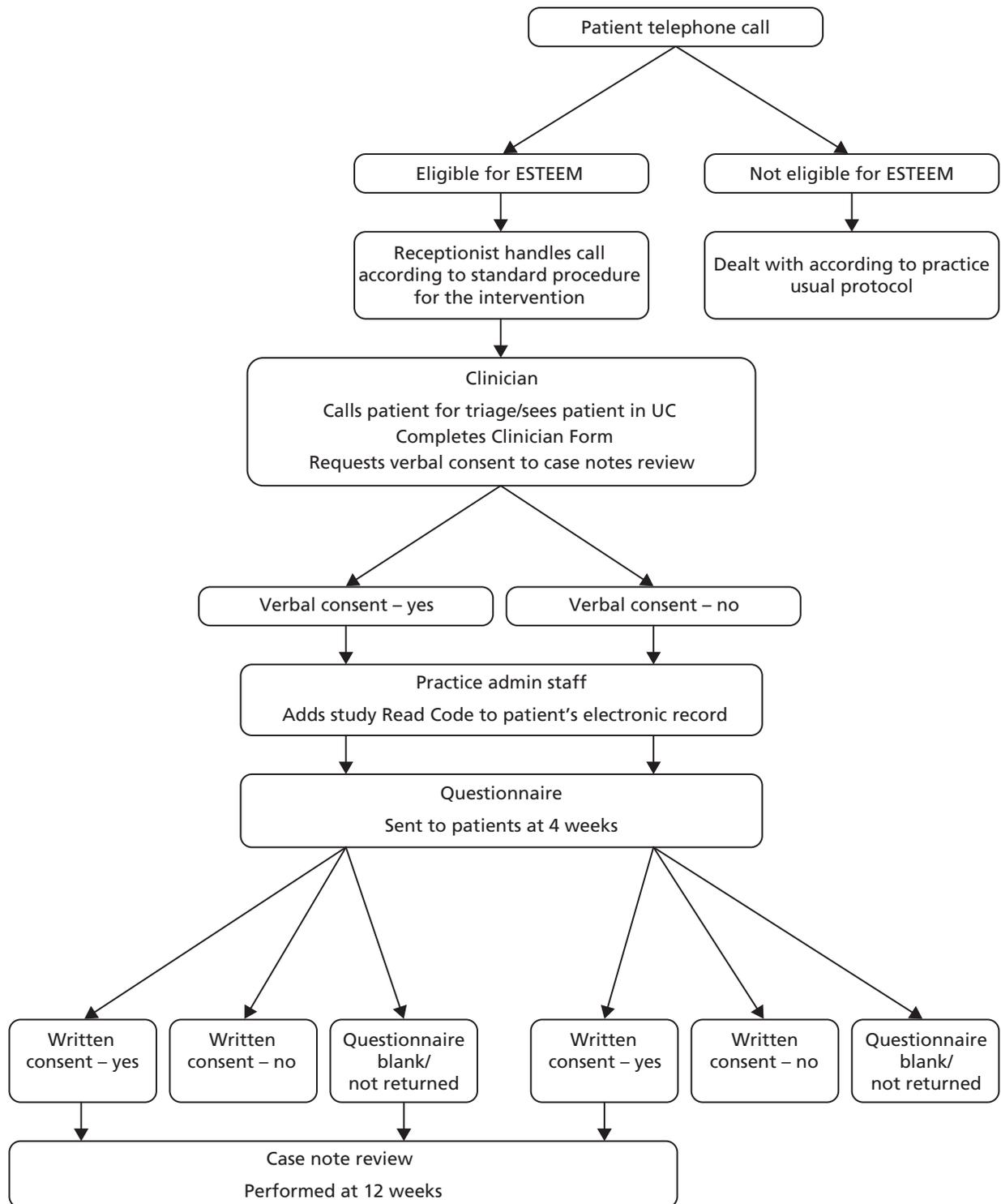


FIGURE 3 Patient recruitment and consent procedure.

Similarly, if a patient was not booked into an appointment (of any type), despite having made an eligible request (e.g. a patient failed to complete the telephone call with the receptionist), the patient's record was still flagged by reception staff as eligible and included within the study denominator.

It was not possible to routinely record eligibility for all patients telephoning the practice. However, on two occasions during run-in and once during data collection (see *Integrity checks*) receptionists logged all incoming calls and recorded patient eligibility, flagging those patients who were requesting same-day appointments and who were too unwell or could not communicate clearly (see *Appendix 9*). We used these data to estimate the proportion of patients making same-day requests who were ineligible for the trial. However, receptionists did not complete the log sheets as fully as we would have liked. For example, in some cases a call had been recorded, but there was no record of whether or not it was a same-day request. Additionally, for other patients recorded as requesting a same-day appointment but who were ineligible, the reason why was sometimes omitted.

Confidentiality

All personal information obtained about patients or staff for the purposes of recruitment or data collection (e.g. names, addresses, contact details, personal information) was kept confidential and held in accordance with the Data Protection Act. Each patient included in the trial was assigned a research number, and all data were encrypted and stored without the subject's name or address. Electronic data were held on a secure database accessible by only members of the study team, and paper-based information was held in a locked filing cabinet in the research team office. Names and participant details were not passed on to any third parties and no named individuals have been included in the reports.

Regarding case note review, the clinical record was viewed only when patients had given verbal or written consent. Patient records were accessed by researchers holding relevant approvals and letters of access granted by PCT research governance units and NHS staff. Case note reviews took place in the practice and no patient records were removed from the premises. Any data extracted were identified by the research number; all other patient identifiers were removed. When data were temporarily stored on laptops, memory sticks or attached to e-mails prior to transfer to the research offices, these media were encrypted as per NHS security requirements. Researchers adhered to confidentiality agreements stipulated by the practices concerned.

Safety of participants

There were not thought to be any significant risks to patients or staff arising from the trial methods. Triage itself may carry risks but the approaches to triage in this trial were already in routine use in the NHS in other settings. Notwithstanding this, we identified two areas of potential minor risk, and developed systems to ensure that these were minimised.

Minimising risk of delays to care for emergency cases or inappropriate triaging

Eligible patients are those requesting same-day consultations at their practice. Consequently, some patients were likely to perceive themselves to have urgent or emergency health-care needs. Where emergencies were identified (see *Inclusion and exclusion criteria*), patients were to be dealt with according to the practice's usual emergency protocol and excluded from trial participation.

Telephone triage means that some patients who might otherwise have been seen in person are managed over the telephone. It was essential that this did not lead to a patient's care being delayed when same-day care was needed. The triage systems were designed to provide safeguards against inappropriate triaging. Both have been tested in previous studies and have been found to be safe (see *Health technologies assessed*), and are already used within some GP practices in the UK. Clinicians and receptionists received training in the system that they were using, and were provided with ongoing support from a clinical lead in the practice and from the system designers. The NT intervention, which relied on CDSS, required nurses to pass a proficiency test (see *Training practices in the two triage systems*). As the GPT intervention draws on clinical skills that are routinely offered by GPs, no proficiency test was required, although guidance in the practicalities of telephone triage were provided.

Minimising risk associated with unexpected software or triage problems

Steps were taken to minimise the risks associated with any unexpected CDSS or triage problems. Each study practice identified two named clinical leads to be the first point of contact, and who were able to deal with any immediate problems and take corrective action. A form was established for use in practices to alert the clinical leads and the trial manager to any such problems arising. A run-in period of 4 weeks was also built in to allow the interventions to be properly set up, and to allow staff to begin operating the interventions 'hands on' before patient recruitment and data collection began. Finally, for NT practices, the CDSS was installed and thoroughly tested by Plain Healthcare before the trial began.

Outcome measures

The selection of relevant outcomes has been a contentious issue in previous evaluations of triage systems in primary care.²⁷ Most studies assessed primary care and hospital service use and workload. However, these outcomes may not fully capture the aim of triage and more broadly the aim of primary care. We proposed that the aim of a primary care consultation management system is to provide an administrative framework for practices to facilitate (1) the safe, timely and definitive ('first pass') management of patients and (2) the timely and efficient management of primary care consulting time resource. We collected descriptive data on practices and patients, and for the following outcome domains: (1) total primary care workload [primary outcome measure (POM), including out-of-hours primary care, A&E and walk-in centre attendance]; (2) patient safety; (3) NHS resource use and cost, and non-attendance rates in primary care; (4) patient health status and experience of care; and (5) case complexity.

Describing practices and patients

Practice-level descriptive data (e.g. list size, location, rural or urban nature, staffing) were collected on trial entry via the Practice Profile Questionnaire completed by the practice manager. This questionnaire also captured information on the typical management of patients before the trial began, providing a description of 'UC'. Practice-level deprivation was collected from the APHO website.³⁷

We collected patient-level descriptive data (age, gender, deprivation) from practice records. Patient self-reported ethnicity, presence of long-standing health conditions and ease of taking time away from work (where relevant) were derived from the postal questionnaire.

Primary outcome measure

The POM is the total number of NHS primary health-care contacts that took place in a 28-day period, commencing with the day of a patient's index telephone request for a same-day consultation. This number includes the index consultation (initial health-care contact resulting from the index telephone call) and all subsequent NHS primary health-care contacts (as defined below) over the following 28 days, which may or may not be related to the nature of the index consultation. All POM data were collected via a case note review (see *Data collection and management*). The primary health-care contacts included in the POM were:

1. GP practice contacts:
 - i. GP face-to-face consultation (in surgery)
 - ii. GP telephone consultation
 - iii. GP home visit (within surgery hours)
 - iv. GP unspecified
 - v. nurse face-to-face consultation (in surgery)
 - vi. nurse telephone consultation
 - vii. nurse home visit (within surgery hours)
 - viii. nurse unspecified
 - ix. general unspecified

2. out-of-hours primary care contacts:
 - i. GP face-to-face out-of-hours consultation
 - ii. GP out-of-hours home visit
 - iii. GP telephone out-of-hours consultation
 - iv. nurse face-to-face out-of-hours consultation
 - v. nurse out-of-hours home visit
 - vi. nurse out-of-hours telephone consultation
 - vii. unspecified
3. walk-in centre contacts:
 - i. walk-in centre attendance (doctor)
 - ii. walk-in centre attendance (nurse)
 - iii. walk-in centre attendance (unspecified)
4. A&E contacts.

The focus of the trial is on primary care workload; however, it is recognised that changes in access to practice appointments may have a knock-on effect on A&E contacts and, therefore, these contacts were also included within the POM.

Secondary outcome measures

Patient safety

The number of deaths and unplanned 'emergency' hospital admissions (and associated number of bed-days) within 7 days, and A&E attendances within 28 days of the index consultation request were collected at case note review. An unplanned emergency hospital admission was defined when there was no evidence in the notes of advance planning, even on the day the admission occurred. The number of 'planned' hospital admissions was also collected. An 'adverse events' reporting procedure (see *Appendix 10*) was developed to monitor patient safety. The clinical lead(s) at each practice were contacted weekly by the trial researcher and asked to return a log sheet completed with any events arising during the intervention period, such as trial patient deaths, emergency hospital admissions or A&E attendances, as well as any problems with the triage system itself or patient complaints.

It became evident that the number of deaths recorded in the weekly log sheets of adverse events from practices did not always correspond to that identified at case note review. On discussion with the DMC, practices were asked to perform a search of their electronic records at the end of the trial to produce a report of all deaths occurring during the data collection period. This became the definitive source of data on patient deaths.

Patient management on the index day

A description of the management of patients on the index day of the same-day consultation request was based on patients' first two primary care contacts collected for our POM.

NHS resource use

The NHS resources that were combined into the POM were also reported separately as secondary outcomes. Non-attendance rates for allocated appointments in the 28 days following the index request were also collected by case note review. It was not possible to capture patients' contacts with NHS Direct from practice records. Patients' use of NHS Direct during the 28-day follow-up period was assessed through self-report by postal questionnaire.

We collected actual consultation length for the index triage and UC (face to face or telephone) consultations on the Clinician Form (see *Data collection and management*). We also estimated the length of subsequent face-to-face consultations, for both 'same-day' and '28-day follow-up' periods. The former was based on the recorded start and end times of a small sample of consultations captured over two randomly selected days in each practice during the intervention period. The latter was based on consultation length used in published unit costs reported within the NHS Personal Social Services Research Unit (PSSRU³⁸).

Patient-reported outcomes

All patient-reported outcomes were collected by the postal questionnaire.

Health status

Health status was assessed using the EuroQol Group European Quality of Life-5 Dimensions (EQ-5D³⁹) questionnaire, and a question on problem resolution^{28,29} (five-point Likert scale). We had originally intended to use the SF-36;⁴⁰ however, piloting emphasised the need to shorten the overall length of the patient questionnaire for the main trial, and thus the shorter EQ-5D measure of health status was adopted (see *Appendix 1*).

Patient experience of care

Patient experience in the context of this trial relates to an episode of care delivered following a same-day consultation request, potentially involving multiple health-care contacts during the index day. Survey instruments suitable for use in a UK setting ask individuals to evaluate a specific consultation^{41,42} or aggregate (practice-based) care. As no validated instruments assessing the patient's experience of an overall episode of care were identified, we used relevant questions from the national GP Patient Survey instrument,⁴³ modifying the questions to focus on the patient's recent experiences of care rather than over the last 6 months. Items selected included the responsiveness of the consultation management system (e.g. how quickly care was provided, overall satisfaction) and patient evaluations (using a five-point Likert scale)^{44,45} of the timeliness and convenience of the response⁴⁶ to the same-day consultation request.

Case complexity

To define patient case mix, the clinician conducting the index consultation captured the complexity of the case (after Howie *et al.*)⁴⁷ using an eight-point scoring schedule incorporated on the Clinician Form (see *Data collection and management*). Each consultation was scored as having either substantial (2 points), attributable (1 point) or no content (0 points) with respect to each of four domains: physical, social, psychological or 'other' (e.g. administrative) components to the consultation.

Sample size

Our sample size estimate was based on our POM, combined with a methodology for gaining patients' consent to case note review that was not eventually operationalised in either pilot or main trials. A UK study comparing NT with standard practice for handling same-day consultation requests³ reported the number of NHS consultations based on general practice (GP and nurse), A&E and out-of-hours primary care contacts. We believed this to be a good proxy for our POM. Over the 28-day follow-up, that trial reported a mean number of NHS consultations of 1.02 [pooled standard deviation (SD) 0.78] in UC compared with 1.38 (pooled SD 1.79) in the NT arm. Using these data, we estimated trial sample size requirements under four scenarios shown in *Table 1*, based on 80% or 90% statistical power, and intracluster correlation coefficients (ICCs) of 0.01 or 0.05 (chosen from a range of ICCs reported from a survey across a basket of outcome measures collected in trials from primary care settings).⁴⁸

TABLE 1 Sample size calculation

<i>n</i> per group ^a	ICC	Design factor	<i>n</i> patients per group	Final <i>n</i> patients per group ^c	<i>n</i> practices per group
305 ^a	0.01	3.26	994	1867	4
305^a	0.05	12.3	3751	7046	14
228 ^b	0.01	3.26	743	1396	3
228 ^b	0.05	12.3	2804	5267	10

a At 90% power and 5% alpha.
b At 80% power and 5% alpha.
c See text below for patient recruitment and attrition assumptions.

Selecting the most conservative sample size scenario (i.e. 90% power, ICC 0.05) we required 3751 patients for analysis and 14 practices per group to detect a difference in means of 1.02 (SD 0.78) compared with 1.38 (SD 1.79) in our POM at follow-up between NT and UC arms. To inform the study power, and prior to undertaking the pilot study that would yield supporting data, we estimated that of all patients requesting a same-day consultation, 6%^{3,28,29} would be deemed ineligible, 10%^{14,29,30} would initially decline participation in the trial and/or follow-up of their notes when talking to the receptionist, and 30%²⁸⁻³⁰ would not respond to the request to return the questionnaire. Of those who did respond in the questionnaire survey, we estimated 10% (conservative) would decline notes review. Thus, a total of 7046 patients seeking same-day consultations across 14 practices over a 5-week period would be required in each of the three trial arms. In the absence of information about a minimum clinically important difference, we used the same sample size as outlined for the NT and UC comparison above for the comparison between GPT and UC. Therefore, we needed 21,138 patients in total (i.e. 7046 per arm) from 42 practices.

It is important to note that our patient recruitment process was altered before conducting the main trial (see *Appendix 1*) in two ways that both impact on our original sample size estimate. First, we did not incorporate a stage at which patients could initially decline participation in the trial and/or follow-up of their notes when speaking to the receptionist (estimated 10%, above). Second, the pilot study also led to a change in method of patient consent to case note review from written consent only (obtained from completed patient questionnaires) to include initial verbal consent obtained from the treating clinician. As a result, we anticipated that, post pilot study, the proportion of patients agreeing to case note review would be around 78%, a much higher estimate than in our original pre-pilot estimate. The pilot study provided confirmation of our assumed ICC of 0.05 [i.e. mean 0.03, 95% confidence interval (CI) 0.00 to 0.08].

Feasibility of patient recruitment

An average practice (7000 patients) accommodates around 714 consultations per week,⁴⁹ approximately 142 consultations per day. Some have estimated⁴ that up to one-third of these are patients seeking same-day consultations; case definition is, however, important and we proposed a more conservative estimate of around 20 patients per day (100 per week, 14% of consultations). This figure was found by the pilot study to be realistic, confirming that to achieve our sample size for the main trial we would need to recruit patients for a 5-week period in each of the 42 practices.

Secondary outcomes

Patient health status and perception of access to care were to be collected by the postal questionnaire with an estimated response rate after one reminder of 70%,⁵⁰⁻⁵³ after piloting this process was altered to include two reminder packs in the main trial. The questionnaire pack also incorporated written consent for case note review, through which our POM is derived. Before the pilot study, we estimated that at 80% power and 5% alpha, our sample size of 7046 per group would allow us to detect an effect size of

0.18 of a SD at an ICC of 0.01, and an effect size of 0.34 of SD at an ICC of 0.05. Thus, the proposed sample size would allow us to be able to detect a small-to-moderate effect size in our patient-reported secondary outcomes.

Trial delivery in participating practices

Following standardised training, practice systems were allowed to stabilise during a run-in period prior to a period of 'live' data collection (Figure 4).

Run-in period

A minimum 'run-in' period of 2 weeks was required before 'live' trial data collection could proceed. During this time all triage and data collection activities were introduced and allowed to bed in. Practices remained in 'run-in' until their performance met predefined trial criteria indicating protocol compliance (see *Integrity checks*). UC practices required only a 2-week run-in period to familiarise themselves with research procedures. As triage practices used this period to adjust the number and distribution of triage slots and protected appointments for exclusive use by the triaging clinicians, most were in run-in for 4 weeks, although some took longer to demonstrate protocol compliance.

Data collection period

Data collection was planned to last approximately 5 weeks in each practice. Medium-sized practices (3500–8000 patients) were expected to recruit 500 unique patients, whereas small (< 3500 patients) and large practices (> 8000 patients) were expected to recruit 350 and 550 patients, respectively. Practices ceased data collection when the target number of unique patient records was entered into the study database; if the target was not reached within 5 weeks then data collection continued.

All GPT and NT practices were asked to revert to their UC arrangements on completing the data collection period for a minimum of 4 weeks. This was for the purpose of preventing differential continuation with a triage management system (should practices wish to continue using triage) by practices during the 28-day period in which we examined patients' use of primary care services.

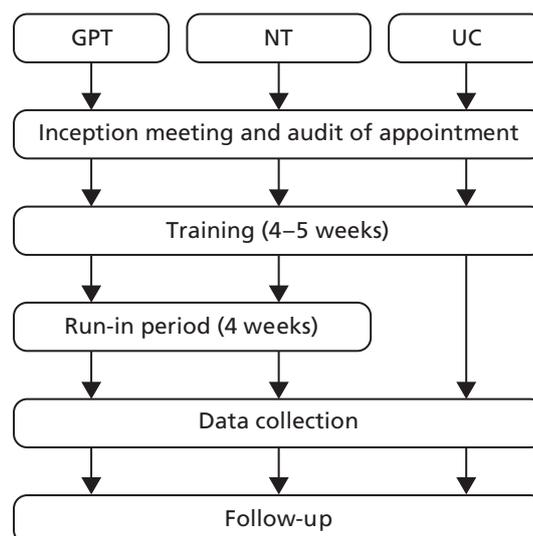


FIGURE 4 Timeline of interventions.

Data collection and management

The sources of outcome data and the timing of data collection are summarised in *Table 2*.

TABLE 2 Outcome data, timing of data collection and source of data

Data	Timing of data collection	Source of data
Practice characteristics (list size, location, deprivation)	Prior to randomisation	List size and location derived from practice; deprivation derived from APHO website ^a
Practice staffing and other clinical information	Prior to randomisation	Practice Profile Questionnaire
Patient age, gender	At entry to trial (i.e. on the day of index same-day request)	Practice records
Duration of first management/triage consultation ^b	Patient's index contact during trial	Clinician Form
Case complexity ^b	Patient's index contact during trial	Clinician Form
Duration of face-to-face contacts subsequent to index consultation request ^c	Two randomly selected days during trial	Recording by GPs and nurses of durations of face-to-face appointments
Adverse events	During the data collection period of the trial	Practice staff weekly log
Patient deprivation data	During the data collection period of the trial	Patients' residential postcode data provided by practices and mapped to ^d IMD 2010 ⁵⁴ via LSOA
Patient satisfaction outcomes	Approximately 28 days after date of index consultation request	Patient questionnaire
EQ-5D	Approximately 28 days after date of index consultation request	Patient questionnaire
Patient self-reported NHS direct use	Approximately 28 days after date of index consultation request	Patient questionnaire
Patient ethnicity, presence of a long-standing health condition, ease of taking time away from work (if relevant)	Approximately 28 days after date of index consultation request	Patient questionnaire
Patient contacts in a 28-day period; includes the date of the index consultation request, deaths and emergency hospital admissions within 7 days of the index consultation request	Approximately 12 weeks after the index consultation request (to allow time for contacts to be recorded at practice)	Case note review by researchers
Patient deaths within 7 days of the index consultation request	End of trial (i.e. following end of data collection period)	Practice records

LSOA, lower super output area.

a Association of Public Health Observatories website: www.apho.org.uk/PRACPROF/ (accessed 5 November 2014).

b The Clinician Form should relate to the first management/triage contact after the index consultation request.

c Face-to-face contact durations were recorded for a sample of ESTEEM patients on 2 days during the trial.

d Index of Multiple Deprivation 2010 source: www.communities.gov.uk/documents/statistics/pdf/1871208.pdf (accessed 5 November 2014).

Clinician Form

A supply of Clinician Forms (see *Appendix 11*) was available in each consulting room for every day of the patient recruitment and data collection period. Clinicians completed this form at the time of the initial 'index' consultation following a patient's same-day request. This index consultation was usually a telephone triage call in the NT or GPT arms, or a face-to-face consultation in the UC arm. The form captured consultation details, including which health professional had been consulted; whether the patient 'did not attend' (DNA); a measure of case mix; treatment and management options chosen (e.g. ordering of tests, recommending subsequent appointment or referral); and the start and end time of the consultation. The clinician also recorded whether the patient had verbally consented to case note review.

Patient questionnaire

Patients were sent a postal questionnaire 28 days after their index request (see *Appendix 8*). The questionnaire measured patient experience and health status, and use of NHS Direct during the 28-day follow-up period (see *Secondary outcome measures*). The questionnaire offered patients the opportunity to confirm their initial verbal consent to the case note review, or to withdrawal if they wished. Patients were asked to send back a blank questionnaire in a prepaid envelope if they did not wish to be contacted again. Patients who did not respond were sent two reminder letters with another copy of the questionnaire, at 2 and 4 weeks after the initial mail-out.

Case note review

Researchers undertook a review of patient's medical records 12 weeks after patients' index request, completing a case note review form (see *Appendix 12*). This delay was to allow time for patients' medical records to be updated with any information on contacts with services outside the practice. The purpose of the review was to extract information regarding our POM (see *Primary outcome measure*) and safety data relating to hospital admissions (see *Patient safety*). A standardised operating procedure was developed to support the four researchers who were conducting these reviews.

The inter-rater reliability of the data extraction process was assessed at three time points during the trial. Any divergence between researchers was documented and the standardised procedure updated as appropriate. The inter-rater reliability was assessed by each researcher reviewing the same random set of 30 case notes. Case notes were selected from different practices, trial arms and practice IT systems (Vision, EMIS LV and SystmOne). The inter-rater reliability results are displayed in *Table 3*.

Blinding

Given the nature of interventions it was not possible to blind patients, clinicians or researchers (conducting case note reviews) to treatment allocation. However, data analysis was carried out by a statistician who was blind to treatment allocation.

Data entry

Electronic records were stored in a bespoke Microsoft SQL Server 2008 database (Microsoft Corporation, Redmond, WA, USA), hosted on a restricted-access secure server maintained by Plymouth University. Data entry was performed via a website encrypted using Secure Sockets Layer (SSL).

TABLE 3 Inter-rater reliability of case note review data extraction

Assessment	Location of practice	Trial arm of practices' notes	Cohen's kappa
Pre-wave 1	Devon	One UC and one NT, practice followed by one GPT practice	0.64 followed by 0.92 ^a
Pre-wave 2	Warwick	One NT practice	0.89
Pre-wave 3	Devon	One NT practice	0.82

^a The initial 'Pre-wave 1' result of 0.64 prompted a re-assessment following more rigorous researcher training.

Monitoring intervention fidelity

Integrity checks

Researchers monitored patient recruitment and adherence to research procedures during both the run-in and data collection periods. Receptionists kept a handwritten log (see *Appendix 9*) of all telephone calls received during three half-day sessions. Integrity checks were preplanned with practices, covering two morning sessions during the latter half of the run-in period and one morning session during the first week of data collection.

For each patient's call that was recorded, the Receptionist Trial Log Sheet captured whether the patient was eligible for the trial, and, if eligible, selected one of the following disposition options:

- triage or telephone appointment booked for today
- face-to-face appointment booked for today
- face-to-face or telephone appointment booked for another day
- patient asked to call again another time (i.e. no appointment booked at all)
- other outcome.

The integrity check documented the following performance indicators:

- at least 75% of patients identified as eligible on the log sheet were subsequently added to that day's triage list
- at least 75% of eligible patients had the study Read Code added to their medical record
- at least 75% of eligible patients had a completed Clinician Form.

Practices satisfying these performance indicators during the run-in period proceeded to data collection. Practices that failed any of the indicators remained in an extended run-in period until another integrity check could be arranged. When practices moved into the data collection period, they were required to undertake one final integrity check during the first week. A practice that failed this final check was permitted to continue with data collection but the trial researcher would have further communication with the relevant members of practice staff regarding the aspect of research procedures on which the integrity check had been failed. An overview of how practices performed at these checks can be found in *Appendix 13*.

In addition, we used the Receptionist Trial Log Sheet completed during the integrity checks to estimate the proportion of all incoming calls to reception which were received from individuals presenting a request for a same-day consultation. We used the same data set to estimate the proportion of individuals presenting same-day consultation requests who met our exclusion criteria (see *Inclusion and exclusion criteria*).

Use of computerised decision support software in nurse-led telephone triage practices

A further approach to monitoring intervention fidelity was put in place for NT practices. We assessed the nurses' use and extent of use of CDSS during each triage consultation. Nurses were advised that the CDSS was intended to act as a support tool, and, although ESTEEM did not aim to evaluate the CDSS itself, as a minimum it should be opened for all triage consultations with ESTEEM patients. The information required for an assessment of use was extracted once the trial was completed by utilising reporting functions within the CDSS. Further details can be found in *Appendix 14*.

Statistical analysis

The methods and reporting of statistical analyses were in accordance with the Consolidated Standards of Reporting Trials (CONSORT) guidelines for cluster randomised trials and pragmatic trials.^{55,56} The combined statistics and health economics plan³³ was reviewed by the Trial Steering Committee and DMC in advance of the analyses.

Practice and patient characteristics

The recruitment of trial practices and disposition of patients during the trial was summarised using a CONSORT flow diagram. Practice characteristics (location, list size and level of deprivation) were compared descriptively across the three trial arms. Practices that withdrew from the study at any point (following initial consent to participate) were scrutinised with regard to their key characteristics, as well as to the trial arm to which they were allocated. The stage of the trial at which they withdrew was recorded, as well as any specific reason for withdrawal (e.g. as the result of staff changes). The results of the baseline Practice Profile Questionnaire (see *Appendix 15*) were also reported descriptively.

Patient demographic characteristics (age, gender, ethnicity and deprivation) were reported descriptively by arm. Age was reported both as a continuous variable and divided into six categories for reporting of frequencies and for use in inferential analyses: 0–4 years; 5–11 years; 16–24 years; 25–59 years (reference category); 60–74 years; and ≥ 75 years. Patients' deprivation status was based on the Index of Multiple Deprivation (IMD) 2010⁵⁴ score and rank of patients' residential lower super output area (LSOA), obtained by mapping the patient's residential postcode to the relevant LSOA.

Availability of patient information throughout the trial was reported descriptively, as was information regarding consent status and initial patient management (specifically, whether a patient in a triage arm was regarded as having been managed per protocol (see *Appendix 16*). Demographic differences between patients who did and did not have case note review performed, and did or did not return a completed questionnaire with at least one completed question (excluding the consent to case note review), were also explored. Potential associations between age, gender and deprivation status, and availability of case note review were described using logistic regression analyses; equivalent analyses were performed for availability of a completed questionnaire. Presence of a long-standing health condition and ease of taking time away from work, if relevant, were also reported descriptively by arm.

Primary outcome

The primary analysis was based on an ITT principle, i.e. analysis of all trial patients in practices according to random allocation. The primary analysis took the form of a regression analysis, using a hierarchical model to take account of the cluster allocation, utilising a random effect to adjust for potential clustering effect by practice, and allowing for adjustment for practice-level minimisation variables (geographical location, deprivation level and list size of practice). For all inferential analyses, patient demographic factors were adjusted for (in conjunction with minimisation variables) if baseline descriptive analyses indicated imbalance of demographic factors across arms. Models were performed twice, initially using the UC arm as reference and then using the GPT arm as reference, to derive comparison between the two triage arms.

A generalised linear model (GLM) was fitted with the appropriate choice of family and link function, according to the type of data and its properties. As the POM (and all secondary outcome measures based on number of patient contacts) was a count, the most appropriate model would be either a Poisson or a negative binomial model, depending on the degree of dispersion in the data. Cluster-level SDs were reported where appropriate, as this parameter approximates the coefficient of variation in underlying cluster rates in certain models.^{57,58} To derive an ICC for the primary outcome, a linear hierarchical model was also fitted.

Additional analyses were conducted on the POM, and on secondary measures derived from the POM, using the hierarchical GLM methods above. Some of these analyses were determined a priori; others were determined post hoc following initial inspection of the data.

A priori analyses

1. A per-protocol analysis including only those patients in the triage arms who received a telephone triage contact by the appropriate clinician type on the index day (all patients in UC were considered to be per protocol).
2. Exclusion of data from two practices in GPT that did not revert to UC as requested following their period of implementing triage and recruiting patients to the trial.
3. Investigation of the effect of missing primary outcome data (due to lack of availability of a case note review) using multiple imputation methods,⁵⁹ based on the assumption that missing case note review data were missing at random.
4. Analyses to investigate interactions between treatment arm and practice characteristics (deprivation, location and list size) on the POM, using interaction terms within the regression models. A series of models were fitted, each model investigating interactions between treatment arm and a specific covariate. Age (as a proxy for case complexity), gender and ethnicity (dichotomised into 'White and other ethnic group', comprising 'Mixed or multiple ethnic groups', 'Asian' or 'Asian British', 'Black/African/Caribbean/black British' and 'Other ethnic group') were also investigated for potential interaction with the treatment arm. A significance threshold of 0.01 was used for hypothesis tests for interaction terms. However, the trial was not powered to investigate subgroup interactions and results should be interpreted with caution.

Results of inferential analyses were presented with 95% CIs (with the exception of inferential analyses including interaction terms, where global p -values for the individual arm by covariate interaction are also stated).

Post hoc analyses

1. Combination of all within-practice contacts on the index day into one overall contact, with contacts on the subsequent days counted individually, as for the primary outcome [e.g. a GP telephone contact and a GP face-to-face contact within the practice on the index day would be counted as one contact only, whereas a GP telephone contact and an A&E contact (or any other contact outside the practice) on the index day would count as two contacts].
2. Analysis of POM contacts on the index day only, and analysis of POM contacts occurring during the 27 days subsequent to the index day only.
3. Analysis of GP face-to-face contacts only, during the full 28-day follow-up period, on the index day only, and on days subsequent to the index day only.
4. Analysis of GP face-to-face and GP telephone contacts only, during the full 28-day follow-up period, on the index day only, and on days subsequent to the index day only.

Secondary outcomes

The following secondary outcomes were analysed inferentially (as well as being reported descriptively) in accordance with the principles of the analyses of the POM, i.e. using a hierarchical regression model appropriate to the nature of the outcome data, adjusting for cluster (practice) effects and with adjustment for minimisation variables and baseline demographic variables. Again, models were fitted twice: once with UC as the reference arm and once with GPT as the reference arm. The ICC was reported for all hierarchical logistic, linear and Tobit regression analyses.

1. Emergency hospital admissions (overall and divided into 'planned' and 'unplanned' emergency admissions) within 7 days of the index day (dichotomised by whether or not a patient had an admission).

2. A&E contacts within 28 days of the index day (dichotomised by whether a patient had at least one contact).
3. EQ-5D status (patients aged ≥ 16 years only).
4. Patient experience aspects derived from the patient questionnaire (for inferential analyses, linearised on a scale of 0–100).⁶⁰

The following secondary outcomes were reported only descriptively:

1. patient deaths within 7 days of the index day
2. number of bed-days resulting from emergency hospital admissions within 7 days of the index day
3. individual components of the primary outcome (i.e. individual contact types)
4. non-attended contacts (DNAs)
5. patient-reported NHS Direct contacts within the 28-day period following, but not including, the index day
6. duration of first management/triage consultations
7. duration of GP and nurse face-to-face consultations
8. case complexity.

Patients who DNA were recorded for each contact type and recorded descriptively. Owing to differences between clinicians in how non-attended telephone contacts were recorded, and in how many attempts may be made to contact a patient, repeated unsuccessful attempts at contacting a patient were recorded as one DNA (if occurring on the same day and if not followed by a successful telephone contact on that day).

Patient management by the practice on the index day was reported descriptively by arm, using the first patient contact recorded on the index day, and the second patient contact on the index day if present.

A composite patient–clinician contact duration on the index day was estimated for each arm. This was calculated using mean contact durations (defined by clinician type and contact type) derived from appropriate sources, including the Clinician Form, the audit of face-to-face consultation durations and standard unit durations,³⁸ and the proportion of patients following specific management pathways on the index day (using pathways that constituted at least 1% of the total patients with available data). This overall patient–clinician contact duration was then subdivided by GP and nurse contact time.

Economic evaluation

Aim

The primary aim of the economic evaluation was to compare the cost incurred over 28 days with respect to primary care contacts plus A&E contacts (using primary outcome data) for GPT, NT and UC, in people requesting same-day appointments in general practice. The perspective of the analysis is that of the UK NHS, i.e. third-party payer.

Method

Intervention cost

Resource use associated with the set-up and delivery of the triage interventions (GPT, NT) comprised initial training costs (set-up), CDSS and licence costs, and staff time spent on delivery of the triage contact. These resource use areas were identified during the pilot study.

Training

Staff time spent at training events, required for set up of the triage intervention, was captured by a within-trial data collection form (completed by the trial researcher). This was a written record of all staff

time (including trainers and trainees) at all relevant training events by staff grade or type, and by practice (see *Training practices in the two triage systems*, above, for an outline of the training schedule in each intervention).

Software and equipment

When comparing triage interventions with UC, pilot study research identified the only additional resource here (i.e. software and equipment) to be the CDSS for the nurse-led, computer-supported triage intervention. The costs of the CDSS and licence fees (required for first year and subsequent years) in each of the NT practices were documented (with information from supplier) within trial.

Triage staff time

Staff time (GP, nurse) used on the delivery of the triage contact (patient contact time) was recorded using the within-trial 'Clinician Form' (see *Data collection and management*). Data collection included staff type by grade, and start time and end time for each triage contact was recorded.

Resource-use data were combined with unit cost data, and market prices (software, licence fees) to estimate the mean cost per triage contact. Costs were reported using 2012 cost data, or with cost data uprated to 2012 costs where required. Costs associated with training and other set-up costs (computer support system) for triage interventions were estimated at a mean cost per practice, and an estimate of the expected number of same-day contacts per practice per year was used to spread the costs across an expected patient group (number of patients requesting same-day appointment per practice). Assumptions on these data and other areas of uncertainty are tested using sensitivity analyses.

Economic outcome: costs of primary care contacts (plus accident and emergency) over 28 days

Primary economic (cost) analyses are undertaken using data collected on the primary outcome, contacts taking place in primary care over 28 days, collected within trial at participant level using a case note review.

Primary care and related contacts by type of contact, as included in the POM, are detailed above (see *Primary outcome measure*). Trial data on service use were combined with unit cost data (*Table 4*) to estimate a mean 28-day cost for primary care service use, for each of the trial arms. Most unit cost data were taken from those reported by Curtis³⁸ (PSSRU unit costs), with other cost data sourced from credible national data sources (see *Table 4*). Contacts recorded as GP or general unspecified ($n = 30$) are treated as GP in surgery consultation, for the purpose of cost analyses. Contacts recorded as 'Nurse contact unspecified' ($n = 15$) were treated as 'Nurse in surgery' consultation for the purpose of cost analyses.

Exploratory analyses report data (primary outcome) on the resource use and costs for 'same-day' care for participants by intervention arm. This analysis includes the data on index contact and other resource use on the same day as the index contact. These exploratory analyses used unit costs for triage contacts as derived from trial data (as above), and published unit cost data on other contacts by type. A sample of data was collected within the trial to provide information on the duration of GP and nurse face-to-face consultations following a triage contact. These data were considered in the context of exploratory analyses of same-day care, and costs of same-day care.

Data analysis: presentation of analysis

Economic analyses were consistent with the methods described for the main statistical analyses (effectiveness outcomes data). The primary economic analyses were based on the ITT trial data (as described above). Data are presented descriptively, and thereafter cost analyses use a random-effects regression model taking account of the hierarchical nature of the study design (i.e. allocation by practice) and allowing for adjustment for practice-level minimisation variables (geographical location, deprivation level and size of practice) and participant-level covariates for age and gender. Data were initially explored using a GLM fitted with the appropriate choice of family and link function according to the type of data

TABLE 4 Unit cost data (2012) with data sources for the base case

Resource use unit	Unit cost (£)	Source
GP telephone triage intervention	14.03	ESTEEM trial (see <i>Table 26</i>)
Nurse telephone triage intervention	7.62	ESTEEM trial (see <i>Table 26</i>)
GP consultation (in surgery)	43.00	Per contact unit cost from Curtis, ³⁸ based on consultation lasting 11.7 minutes, including direct care staff costs with qualifications
GP telephone consultation	26.00	Per contact unit cost from Curtis, ³⁸ based on telephone consultation of 7.1 minutes, including direct care staff costs with qualifications
GP home visit (within surgery hours)	110.00	Per contact unit cost from Curtis, ³⁸ based on home visit lasting 23.4 minutes (including travel time/cost)
Practice nurse consultation (in surgery)	13.64	Derived using cost per minute from Curtis. ³⁸ Assumes 15.5-minute nurse consultation × unit cost per hour/minute (£53/£0.88, including qualifications costs) for face-to-face contact
Practice nurse telephone consultation	5.28	Derived using cost per minute from Curtis. ³⁸ Assumes 6 minutes/telephone consultation cost [duration from advance nurse] × unit cost per hour/minute (£53/£0.88, including qualifications costs) for face-to-face contact
Practice nurse home visit (within surgery hours)	22.00	Derived using cost per minute from Curtis. ³⁸ Assumes 25 minutes/home visit cost [duration from advanced nursing professional] × unit cost per hour/minute (£53/£0.88, including qualifications costs) for face-to-face contact
Walk-in centre attendance (doctor/nurse/unspecified)	41.00	Per contact unit cost from Curtis ³⁸ A&E walk-in service (not admitted)
Out-of-hours contact (doctor/nurse/unspecified) ^a	61.14	Primary Care foundation, 2013 (www.primarycarefoundation.co.uk)
A&E: doctor/nurse/unspecified	112.00	Per A&E attendance, not admitted – national average (weighted averages irrespective of occupation) from Curtis ³⁸

a Unit cost based on average cost per case, including all overheads and oncosts associated with provision of out-of-hours care. Obtained from survey of primary care trusts in England. URL: www.primarycarefoundation.co.uk (accessed 5 November 2014).

and its properties. Based on findings from these analyses (using GLM methods), main analyses are presented using a hierarchical multilevel model, assuming normally distributed total cost data. The ICC is reported for primary cost analyses.

The primary economic analyses present estimates of the mean cost of care across each of the trial arms, as above. Primary analyses report on participants with data on the POM, i.e. a complete case analysis. Regression-based methods (as above) were used to estimate difference in costs for care between trial arms, based on the 28-day data included in the primary outcome. CIs (95%) are estimated using parametric methods. Typically, where a sample has a large number of observations, as in this case (with > 16,000 participants in ESTEEM), incorporating central limit theorem implies parametric tests are appropriate and may be used for analysis of resource use and cost data. Item-level costs are presented descriptively, consistent with the data presentation in the effectiveness analyses.

In secondary analyses, a per-protocol analysis has been performed (for the primary economic outcome only), including only patients who received the triage intervention, this being consistent with the main statistical analysis plan and effectiveness analyses. As no difference is reported on EQ-5D single index values (quality-adjusted life-year weights), by treatment allocation, no exploratory analyses are considered on cost-effectiveness analyses using this outcome.

Sensitivity analyses were undertaken against the primary analyses to explore the implications of uncertainty in data used and the assumptions made within the primary analyses. Sensitivity analyses included an analysis of primary outcome data (total 28-day cost), with missing data imputed via multiple imputation methods (as used in effectiveness analyses).

Results from exploratory analyses are presented descriptively, and with regression-based methods used to provide comparative analyses where appropriate.

Results are presented in tabular format using mean estimates of resource use and cost, with summary measures on the distribution around the mean. A broader presentation of findings are presented in tabular format consistent with the approach described as cost–consequences analyses, presenting estimates of costs alongside the expected impacts associated with interventions, for example safety outcomes, health status and measures of patient experience. Cost–consequences analyses are regarded as a form of full economic evaluation, even though the costs and outcomes are not brought together in a cost-effectiveness ratio.⁶¹

Process evaluation

Aims and objectives

The process evaluation for the ESTEEM trial took place alongside the main trial. Its aims were to:

- describe how the trial was experienced and how the telephone triage interventions were implemented in different practice settings
- describe the experience and acceptability of telephone triage for staff and patients
- elicit patients' and staff views on what influences the telephone triage being seen to work or not to work.

Methods

The process evaluation was composed of semistructured, qualitative interviews conducted in a subsample of trial practices. This was supplemented by a limited amount of observation and interviews with researchers from two of the study sites.

Sampling, recruitment and consent procedures

Process evaluation data were collected from a purposive sample of 10 of the main trial practices (four practices in each triage arm and two in UC) shown in *Table 5*. The practices were sampled from three regions (Devon, Bristol and Warwick) and varied in terms of their list sizes and locations (inner city, urban, suburban and rural).

Patient recruitment

Practices provided the research team with ID numbers for the first 20 consecutive patients eligible for trial entry, within 3 days of a same-day consultation request. Eligible patients were aged < 12 years or ≥ 16 years, with proxy participation by parents or guardians of those patients aged < 12 years. Practices provided the research team with an anonymised list of the patients approached, including patients'

TABLE 5 Process evaluation: participating practices by region

Region	GPT practices	NT practices	UC practices	Total
Devon	2	2	1	5
Bristol	2	1	1	4
Warwick	0	1	0	1

internal practice ID number, patients' initials, age, gender and ethnicity, as well as an indication of their anticipated views on the triage systems if this was known but practices did not always provide this latter information. Clinicians who managed same-day consultations were asked to review the sampling frame and to remove any patient where they believed interviews would be clinically inappropriate and/or likely to cause distress, either to patients themselves and/or to their families (e.g. end-stage terminal illness or death within the immediate family).

We aimed to recruit 15 patients per trial arm, to secure a total of 45 patient interviews. Patients were to be selected purposively to represent as diverse a sample as possible, based on age, gender, ethnicity and anticipated views (where possible) on the triage system. However, sluggish patient recruitment necessitated revising the target sample size downwards, which meant that purposive sampling was compromised.

Practices were provided with letters to send, inviting patients to interview (enclosing patient information leaflets, reply sheets and reply-paid envelopes). Where possible, interviews were secured within 2 weeks of patients' index same-day consultation requests. Patients who were potentially willing to be interviewed returned the study reply sheets containing their preferred contact details directly to the research team. It was intended that five of the first patients in the main trial who replied from each practice would be contacted by the process evaluation researchers to discuss the study further and to agree mutually convenient interview dates and times with those who wished to proceed. This step was to be repeated until four willing patients from each practice, with diverse demographic characteristics, were found. The slow response necessitated revising the recruitment process, so that all patients returning their contact information and consent to be contacted were approached. The recruitment criteria were also revised to provide patients with the option of a telephone interview. This proved critical in improving the response rate, and the majority of the interviews with patients were conducted by telephone. This seems an appropriate method to use in a study of telephone-based interventions.

Interviewees' written consent to participate was obtained by the researchers either at interview, if face to face, or by post if interviews were conducted by telephone. Written consent to telephone interviews was obtained retrospectively in some cases, although all patients were asked to provide verbal consent prior to commencing the telephone interview.

Staff participants

We aimed to recruit five staff members from each practice, to secure a total of 50 interviews. All practice staff within sampled practices were invited to participate in the process evaluation; those who responded to the invitation and went on to participate were self-selected. A purposive sampling strategy was designed, selecting potential staff interviewees for occupational diversity to ensure that the views of GPs, nurses, practice managers and reception staff were represented. Staff members of both sexes and of varied ages were approached. Each staff member received a personally addressed letter signed by the investigators in their respective recruitment sites, inviting them to participate in a study interview. The invitation letters were accompanied by staff participant information sheets, reply sheets and reply-paid envelopes. Staff members who were potentially willing to be interviewed returned the study reply sheets containing their preferred contact details directly to the research team. Potential participants were contacted by the researcher to discuss the study further and to agree a mutually convenient interview date and time with those who wished to proceed. This step was repeated until five staff members, representing each occupational group at each practice, were found. The small number of staff members who expressed an interest, but were not selected for interview, received letters of thanks. Interviewees' written consent to participate was obtained by the researchers at interview. *Tables 6 and 7* show the numbers of patient and staff participants recruited.

TABLE 6 Process evaluation participating practice staff

Trial arm	GPs	Practice nurses or nurse practitioners	Managers	Receptionists	Total
GPT	9	2	4	7	22
NT	7	6	4	6	22
UC	3	1	1	4	9
Total	19	9	9	17	54

TABLE 7 Process evaluation participating patients

Trial arm	Male	Female	Total
GPT	8	12	20
NT	7	12	19
UC	3	3	6
Total	18	27	45

Data collection

One-to-one semistructured interviews of 30–60 minutes' duration were held with selected patients and practice staff. Staff interviews were conducted face to face within practice premises. Interviews with patients were conducted either face to face in patients' homes, or by telephone, according to participants' choices.

Purpose-designed interview schedules were developed by the study team and the PPI group. They guided the interview process throughout and were further developed to reflect emerging interview topics. Patient interviews explored the patient's index consultation (whether on the telephone in the triage conditions or face to face in the UC arm) and specifically their story of that consultation, and its antecedents and consequences. Views on the acceptability, convenience and speed of patients' access to the new telephone triage systems were also sought. Staff interviews explored preparation and training for the trial; expectations, experiences and views held by practice staff of setting up and running the triage systems; the systems' perceived acceptability, problems, solutions (or not) and staff members' hypothetical willingness (or not) to use triage systems post study. In control practices, staff members were asked about the usual management of 'same-day' patient consultation requests at their practice. Interviews were audio-recorded with the permission of the interviewees.

No staff member refused permission to record the interview. However, it was suggested to the researcher that the knowledge that the interview would be recorded had inhibited recruitment in a small number of cases.

Letters were sent to both patient and staff interviewees to thank them for their participation. Both patients and staff were also given the opportunity to confirm and correct descriptive summaries of the preliminary findings (main themes) from the thematic analysis of the study's patient and staff interviews conducted, if they wished to do so.

Data analysis

The 99 interview audio tapes were transcribed verbatim by a professional transcriber and checked and anonymised by the process evaluation researcher. It became clear that UC practices had limited usefulness as a control or comparator in a qualitative study of experience of triage. They were not experiencing triage and UC was not standardised. For various pragmatic and methodological reasons data were analysed from the intervention practices only ($n = 84$). The transcripts were analysed thematically, drawing on the grounded theory techniques of constant comparison⁶² using the qualitative data analysis software package NVivo 8 (QSR International, Warrington, UK). A deductive coding frame based on the process evaluation research questions was agreed by the research team. Within this structure data were coded inductively to allow participants' accounts to inform the analysis to capture the subtleties and subjectivities of a multiplicity of experiences in a range of settings. It also permitted the identification of structural, cultural and organisational factors within practices that framed these experiences, and how they were described and interpreted. A framework approach⁶³ was then adapted to enable both within-case (individual practices) and cross-case (between practices) analyses. The practice emerged as the unit of analysis, as it became evident that data could not be interpreted without considering aspects of individual practices' culture and ways of working. Groups of coded material were summarised and charted onto a matrix to map the range of phenomena emergent from the data, and to enable relationships and interrelationships to be conceptualised in a systematic way.

The analysis of interviews was iterative, with knowledge gained from staff providing insights into patient interviews and vice versa. It also reflected differences in the implementation of the interventions and contextual factors in individual practices, as well as investigating the effect of how the trial itself was experienced and how the intervention was implemented and experienced. Both patient and staff interviews were interrogated in terms of any areas of emerging agreement or disagreement about what worked, or did not work, and any observable conflicts and differences of opinion between and within staff groups. Coding and analysis was undertaken by one researcher (LP) and validated by a second researcher (NB), who reviewed a sample of the transcripts and tested these against the coding frame.

Chapter 3 Results

Flow of practices and patient participants in the trial

The flow of practices and patients through the ESTEEM trial is summarised in *Figure 5*.

Practices

Following an approach to 388 practices, 42 (10.8%) agreed to participate and were randomised to one of the three trial arms (GPT 13, NT 15, UC 14). The 12 post-randomisation withdrawals (UC 2, NT 10) were replaced. Of 42 practices randomised, 11 withdrew and were replaced in a purposeful process in which we drew from a small pool of reserve practices, matching as closely as possible to locality, size and deprivation profile of the original withdrawn practice, and with allocation concealment preserved (see *Replacing practices who withdrew post randomisation*). One of the substitute practices itself subsequently withdrew from the trial and had to be replaced – making a total of 12 replacement practices. Finally, of the 12 practices withdrawing from the trial, 11 withdrew in the very earliest stages post randomisation, usually prior to research procedures training or the run-in period of 4–5 weeks (see *Figure 1*); only one practice withdrew during the trial data collection period.

Reasons for practice dropout

All of the practices that withdrew did so prior to patient recruitment, although one practice in NT did begin the run-in period and had commenced managing patients with a triage system. This practice had two nurses delivering the triage but felt that it could not manage the demand. The practice halted the triage and decided to use the funds made available through the trial to recruit locum nurses to cover the triage. Unfortunately, the practice was unable to find any suitable candidates and felt, therefore, that it could not continue with the trial. The post-withdrawal questionnaire was returned by only one of the two UC practices that withdrew – the practice reporting dissatisfaction with the trial arm to which they had been randomised. We are unable to comment on why the other UC practice withdrew. Nine of the ten practices that withdrew from NT returned a post-withdrawal questionnaire. Although one practice was dissatisfied with the randomisation to NT, the majority of reasons for withdrawal revolved around the practice feeling under-resourced to cope with the demands of the trial: two practices had difficulty recruiting additional staff, four practices had difficulty extending staff hours as needed, and, in one case, nursing staff had left the practice (for reasons other than the trial). Other reasons for practice withdrawal from the trial (each reason cited by only one practice) included retirement of key staff members, key staff on long-term sick leave, insufficient commitment from practice staff, lack of support from the ESTEEM research team, and introduction of a triage system between recruitment and the start of the trial.

Patients

Recruitment of patients by practices as potentially eligible for the trial (i.e. requesting a same-day consultation) over the course of the study, against targets, between May 2011 and December 2012, is shown in *Figure 6*. Data on the numbers of patients excluded from the trial at the point of requesting a same-day consultation as a result of being too unwell or unable to communicate without difficulty were not routinely collected. The Receptionist Log Sheet completed for the integrity checks (twice during the run-in period and once during the data collection period) provided an estimate of the proportion of these patients from the total number of same-day requests (*Table 8*). Missing data from the Receptionist Log Sheet on whether or not the incoming call was a request for a same-day appointment ranged from around 19% in UC to 25% and 29% in GPT and NT, respectively. For a proportion of patients who did make a same-day request we were unable to determine whether they were eligible or ineligible for the trial. The reason for this, as noted in *Chapter 2*, was that receptionists did not reliably record the reason why patients requesting a same-day appointment were ineligible. Missing data from this group of patients ranged from < 1% in UC to 4% in GPT. Available data indicated that between 9% and 19% of patients making same-day requests were ineligible for the trial.

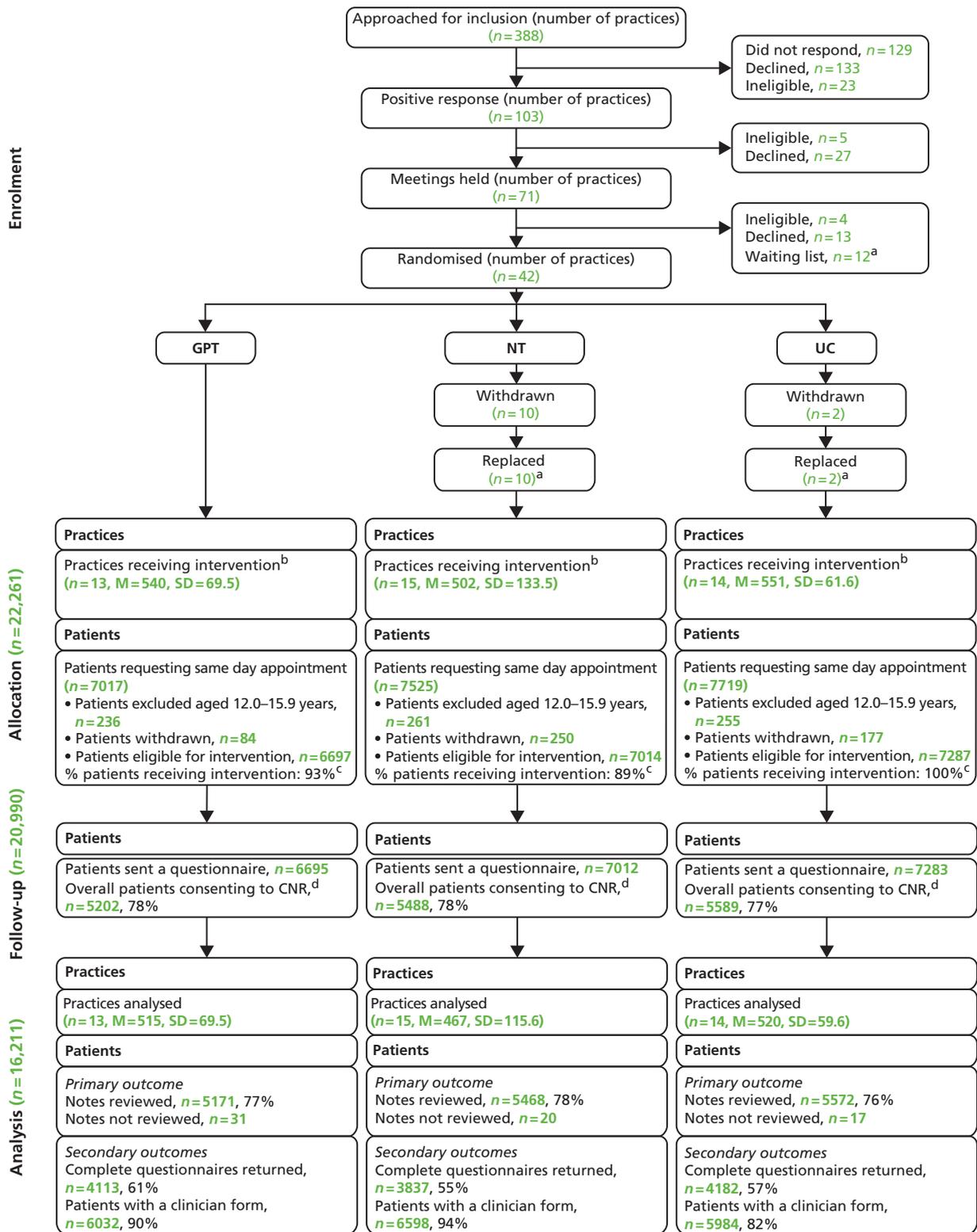


FIGURE 5 Consolidated Standards of Reporting Trials diagram. a, Withdrawn practices were purposively replaced (with practices from the ‘waiting list’) while maintaining allocation concealment; b, n (number of practices), mean number patients (SD) per practice; c, given the nature of the trial it was not possible to determine exactly the number of patients who received the intervention. Assessment of whether patient was treated per protocol was dependent upon the patient having a completed Clinician Form or having their medical notes reviewed. Here, % receiving intervention is based upon case note reviews; % receiving the intervention based upon Clinician Forms; and d, CNR = case notes review. Reproduced with permission from Campbell *et al.*⁶⁴

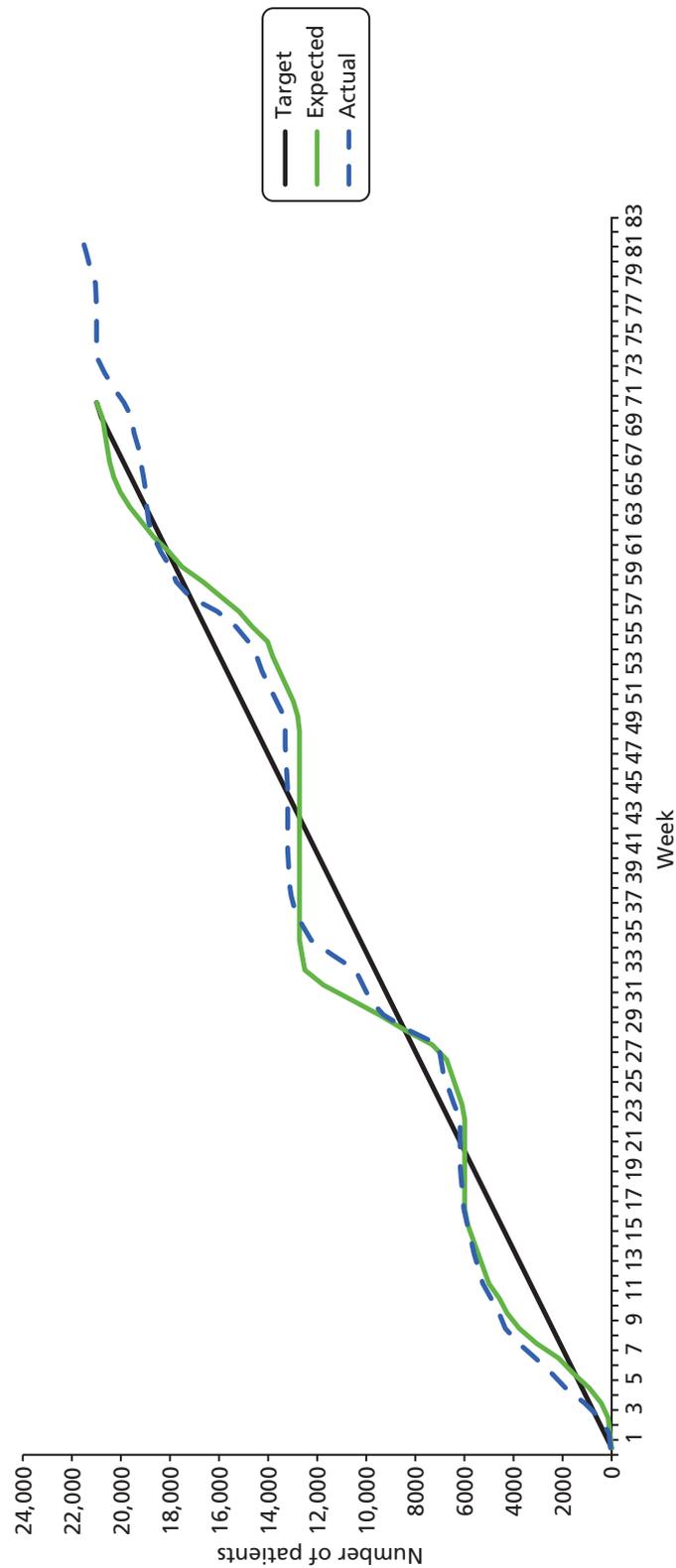


FIGURE 6 Patient recruitment by practices.

TABLE 8 Estimate of the proportion of patients seeking same-day consultations who were ineligible for the trial^a

Receptionist log sheet data	UC	GPT	NT
Total proportion of patients requesting a same-day consultation (from total recorded telephone calls)	48% (1964/4090)	36% (1382/3865)	37% (1896/5115)
Proportion of patients who were ineligible due to being unwell or unable to communicate without difficulty (from total requesting a same-day consultation)	10% (189/1960)	9% (119/1320)	19% (358/1843)

a These data were based on log sheets completed by reception staff. Explanations other than data inaccuracy do exist – there may be heightened awareness of the nature of a consultation request in intervention practices, or there may indeed be a true difference in the proportion of same-day requests incurred by chance in the UC practices. We can only speculate on this matter, however, and to extend beyond what we have said would be unwarranted.

Of the patients identified by receptionists as requesting a same-day consultation, who were both well enough and able to communicate without difficulty, a total of 20,990 patients [6695/7017 (95%) in GPT, 7012/7525 (93%) in NT and 7283/7719 (94%) in UC] were eligible for the trial, following exclusion of patients aged 12.0–15.9 years and withdrawal of a number of patients for various reasons. Reasons included administrative errors in applying Read Codes to patients' records, patients being entered into the trial multiple times and incorrect inclusion of temporary residents, patients making non-same-day requests and patients with hearing impairment or dementia. In the NT arm, 211 patients were withdrawn after being identified for the trial by practices during periods in which no triage system was operating. The proportion of patients who received the intended intervention (thus treated per protocol) was similar across the intervention arms [4796/5171 (93%) in GPT and 4860/5468 (89%) in NT].

Patient loss to follow-up

All eligible patients ($n = 20,990$). Eight patients withdrew (various reasons) prior to being sent a questionnaire.

Of all patients sent a questionnaire (excluding post-questionnaire withdrawals), consent to case note review (providing primary outcome data) was given by 5202/6695 (78%) in GPT, 5488/7012 (78%) in NT and 5589/7283 (77%) in UC. The procedure for obtaining a patient's consent to case note review was complex. For ease of presentation we have reported the final decision following verbal and written consent. The specific flow – the numbers of patients providing verbal consent and then opting out via written consent, and the numbers of patients refusing (or not having any record of) verbal consent but opting in via written consent – can be seen in *Appendix 16*.

In total, 68 patients who provided consent to case note review did not have primary outcome data extracted (< 0.01% of patients providing consent: 68/16,211). This was because these patients returned a questionnaire including written consent to case note review (which was not obtained verbally) after the research team had completed case note reviews for the relevant practice. The case note review follow-up period ended in January 2013.

Describing usual care

Information regarding the practices, their staff numbers and use of telephone triage (prior to the commencement of the trial) is set out in *Table 9*. In general, across all of the arms, a large proportion of practices did not allocate time for telephone triage (26/42). Of those that did, telephone triage was mainly used to manage patients seeking same-day consultations once face-to-face appointment slots had been taken (19/22). The majority of practices tended to manage < 25% of patients seeking same-day consultations via telephone triage (13/22). Staff composition and services provided by the practice can be seen in *Appendix 17*.

TABLE 9 Practice Profile Questionnaire

Practice Profile Questionnaire items	UC (N = 14)	GPT (N = 13)	NT (N = 15)
List size: mean (SD), <i>n</i>	8257 (3792), 14	8023 (3619), 13	8498 (3600), 15
Number of GPs at practice (including full-time and part-time): ^a mean (SD), <i>n</i>	5.9 (2.6), 13	6.4 (3.0), 13	6.5 (2.8), 15
Number of nurses at practice (including full-time and part-time): ^a mean (SD), <i>n</i>	3.6 (1.8), 14	3.5 (1.6), 13	3.3 (1.4), 15
Practice rurality: n (%)			
Rural	7 (53.9)	7 (58.3)	6 (42.9)
Urban	6 (46.2)	5 (41.7)	7 (50.0)
Inner city	0 (0)	0 (0)	1 (7.1)
Total	13 (100.0)	12 (100.0)	14 (100.0)
Does the practice allocate time for doctors or nurses to carry out telephone triage? n (%)			
Yes	3 (21.4)	4 (30.8)	9 (60.0)
No	11 (78.6)	9 (69.2)	6 (40.0)
Total	14 (100.0)	13 (100.0)	15 (100.0)
Which option best describes how telephone triage is used by doctors/nurses at your practice? n (%)			
Patients seeking a same-day consultation are triaged on the telephone only if all of the appointment slots are taken	3 (100)	6 (75.0)	10 (90.9)
Telephone triage is not used across the whole practice but at least one doctor triages his/her own patients	0 (0)	2 (25.0)	1 (9.1)
Total	3 (100.0)	8 (100.0)	11 (100.0)
What proportion (%) of patients seeking a same-day consultation are triaged on the telephone? n (%)			
< 25	2 (66.7)	6 (75.0)	5 (45.5)
25–50	0 (0)	1 (12.5)	4 (36.4)
51–75	0 (0)	0 (0)	1 (9.1)
76–100	0 (0)	0 (0)	0 (0)
Cannot estimate	1 (33.0)	1 (12.5)	1 (9.1)
Total	3 (100.0)	8 (100.0)	11 (100.0)

^a If data were missing for either full- or part-time clinicians, the number was assumed to be 0 and the total calculated using only the available data.

Practice descriptions and patient demographic characteristics

Practice descriptors

Given the relatively small number of practices in each arm, the practice characteristics (location, list size and deprivation) were generally well balanced across the arms (*Table 10*).

Patient demographics

Patient age, gender and deprivation status were well balanced across trial arms (*Table 11*). In all three arms, female patients constituted the majority (approximately 60%): patients in GPT were slightly older than in UC or NT [mean age (years) in GPT was 44.7 (SD 25.0), in UC 41.6 (SD 23.7) and in NT 41.5 (SD 25.2)]. Among questionnaire respondents, across the three arms, 95–97% were white people and 46–50% reported that they had a long-standing health condition. Based on the results of analyses of predictors of availability of case note review (see *Appendix 18*), it was decided to conduct multiple imputations analyses for the primary outcome (and total costings) only, on an ITT basis.

TABLE 10 Practice demographics and cluster size coefficient of variation

Practice demographic variables	UC (N = 14)	GPT (N = 13)	NT (N = 15)	Total included (N = 42)	Withdrawn (N = 12) ^a
Location: n (%)					
Bristol	3 (21.4)	4 (30.8)	3 (20.0)	10 (23.8)	2 (16.7)
Devon	4 (28.6)	3 (23.1)	4 (26.7)	11 (26.2)	3 (25.0)
Norwich	4 (28.6)	3 (23.1)	3 (20.0)	10 (23.8)	4 (33.3)
Warwick	3 (21.4)	3 (23.1)	5 (33.3)	11 (26.2)	3 (25.0)
Deprivation:^b n (%)					
Deprived	5 (35.7)	3 (23.1)	3 (20.0)	11 (26.2)	5 (41.7)
Non-deprived	9 (64.3)	10 (76.9)	12 (80.0)	31 (73.8)	7 (58.3)
List size:^c n (%)					
Small	1 (7.1)	1 (7.7)	2 (13.3)	4 (9.5)	1 (8.3)
Medium	6 (42.9)	5 (38.5)	4 (26.7)	15 (35.7)	4 (33.3)
Large	7 (50.0)	7 (53.9)	9 (60.0)	23 (54.8)	7 (58.3)
Coefficient of variation for cluster size	0.115	0.135	0.247	0.176	

a All withdrawn practices withdrew post randomisation: 2 in UC, 10 in NT.

b Obtained from Public Health England National General Practice Profiles: www.apho.org.uk/PRACPROF/ (accessed 5 November 2014) – ‘Deprived’, above average deprivation for England; ‘Non-deprived’, average/below-average deprivation for England.

c Small, < 3500 patients registered; medium, 3500–8000 patients registered; large, > 8000 patients registered.

TABLE 11 Baseline patient demographics

Individual patient characteristics for total cohort	UC (N = 7283)	GPT (N = 6695)	NT (N = 7012)
Gender, n (%)			
Male	2920 (40.1)	2735 (40.9)	2774 (39.6)
Female	4363 (59.9)	3960 (59.2)	4238 (60.4)
Age (years): mean (SD)	41.6 (23.7)	44.7 (25.0)	41.5 (25.2)
Age by category (years), n (%)			
< 5	690 (9.5)	605 (9.0)	830 (11.8)
5–11	470 (6.5)	379 (5.7)	463 (6.6)
16–24	834 (11.5)	675 (10.1)	726 (10.4)
25–59	3368 (46.2)	2875 (42.9)	3058 (43.6)
60–74	1350 (18.5)	1317 (19.7)	1204 (17.2)
≥ 75	571 (7.8)	844 (12.6)	731 (10.4)
Deprivation (IMD 2010⁵⁴ score^a): mean (SD), n	17.6 (10.3) 7235	17.1 (11.8) 6671	17.7 (11.7) 6930
Deprivation (IMD 2010⁵⁴ quintile, based on rank), n (%)			
Quintile 1 (most deprived)	460 (6.4)	524 (7.9)	653 (9.4)
Quintile 2	1694 (23.4)	995 (14.9)	1348 (19.5)
Quintile 3	1857 (25.7)	1992 (29.9)	1673 (24.1)
Quintile 4	1879 (26.0)	1916 (28.7)	1783 (25.7)
Quintile 5 (least deprived)	1345 (18.6)	1244 (18.7)	1473 (21.3)
Total N	7235	6671	6930
Individual patient characteristics for questionnaire respondents only			
UC (N = 4182) GPT (N = 4113) NT (N = 3837)			
Ethnicity by ethnic group, n (%)			
White	3956 (96.5)	3876 (96.0)	3573 (95.3)
Mixed/multiple ethnic groups	33 (0.8)	36 (0.9)	27 (0.7)
Asian/Asian British	82 (2.0)	79 (2.0)	110 (2.9)
Black/African/Caribbean/black British	15 (0.4)	34 (0.8)	24 (0.6)
Other ethnic group	15 (0.4)	12 (0.3)	17 (0.5)
Total N	4101	4037	3751
Able to attend surgery during work hours, n (% total N;% total N relevant^b)			
Yes, easily	794 (19.6; 38.2)	790 (19.8; 41.2)	736 (19.8; 39.7)
Yes, with difficulty	883 (21.8; 42.5)	830 (20.8; 43.3)	778 (21.0; 42.0)
No	402 (9.9; 19.3)	296 (7.4; 15.4)	340 (9.2; 18.3)
Total N relevant	2079	1916	1854
Not relevant, n (% total N)	1974 (48.7)	2068 (51.9)	1857 (50.0)
Total N	4053	3984	3711
Long-standing health problems n (%)			
Yes	1940 (48.0)	1985 (50.0)	1716 (46.4)
No	2101 (52.0)	1983 (50.0)	1985 (53.6)
Total N	4041	3968	3701

a IMD 2010⁵⁴ score and rank derived from residential postcode data mapped to LSOA.

b Excludes children and non-working adults.

Clinical outcomes

Primary outcomes

The proportion of patients having zero or more primary care contacts is displayed as a function of trial arm in *Figure 7*. The modal number of contacts under UC was one, whereas the modal number of contacts under both GP- and nurse-led telephone triage was two.

Intention-to-treat analyses included 16,211 patients (UC: $n = 5572$; GPT: $n = 5171$; NT: $n = 5468$) and showed an increase in the rate ratio (RR) of contacts comparing GPT with UC (RR 1.33, 95% CI 1.30 to 1.36) with adjustment for practice characteristics (list size, location and deprivation status) and patient demographic characteristics [gender, age (categorised) and deprivation quintile; *Table 12*]. The same model comparing NT with UC yielded a RR of 1.48 (95% CI 1.44 to 1.52); comparing NT with GPT, the RR was 1.04 (95% CI 1.01 to 1.08). The equivalent per-protocol analysis showed an intensification of the observed effects of both GPT and NT (see *Appendix 19*). A sensitivity analysis excluding two GPT practices that did not revert to UC following the end of the trial period yielded results similar to those of the primary ITT analysis (results not shown). The observed ICC derived from a hierarchical linear model was 0.015 (95% CI 0.009 to 0.025), lower than the ICC of 0.05 in the original sample size calculations (see *Sample size*).

Using data derived from multiple imputations combined with observed data where available, the results of ITT analyses were similar to those derived from observed data only (see *Table 12*) and hence the complete case results appeared to be robust compared with analyses including imputed data. There was some evidence of interactions between treatment arm and practice level covariates (location and practice deprivation), and between treatment arm and age category; however, no clear patterns were evident (see *Appendix 20*).

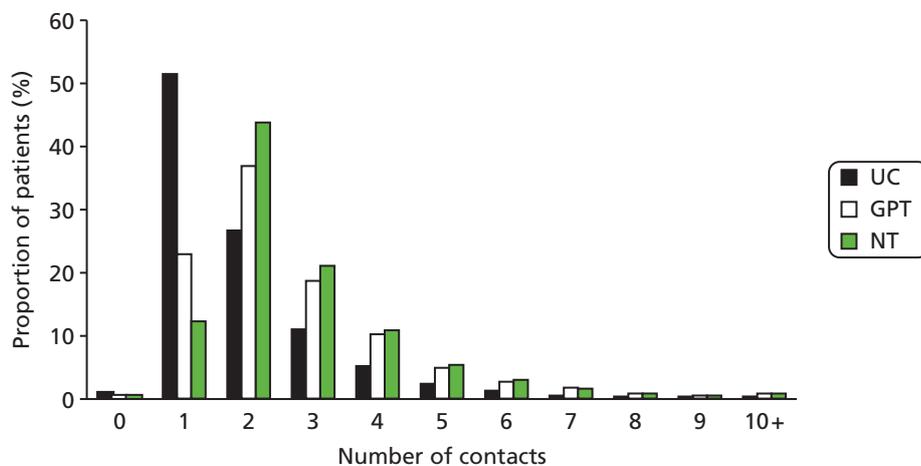


FIGURE 7 Primary outcome results by arm.

TABLE 12 Primary outcome measure: total primary care contacts (including A&E) within 28 days of index consultation request across trial arms

	UC (N = 5572)	GPT (N = 5171)	NT (N = 5468)
Total primary care contacts: ^{a,b} total, mean (SD)	10,616 1.91 (1.43)	13,720 2.65 (1.74)	15,400 2.82 (1.68)
Total primary care contacts:^{a,b} n (%)			
0	45 (0.8)	14 (0.3)	16 (0.3)
1	2863 (51.4)	1179 (22.8)	670 (12.3)
2	1480 (26.6)	1902 (36.8)	2391 (43.7)
3	609 (10.9)	966 (18.7)	1146 (21.0)
4	291 (5.2)	526 (10.2)	591 (10.8)
5	135 (2.4)	255 (4.9)	287 (5.3)
6–10	136 (2.4)	300 (5.8)	342 (6.3)
≥ 11	13 (0.2)	29 (0.6)	25 (0.5)
ITT analyses using complete case data, RR (95% CI)			
	GPT vs. UC	NT vs. UC	NT vs. GPT
Poisson GLLAMM model ^c with robust residuals: adjusted ^d (n = 16,095), RR (95% CI)	1.33 (1.30 to 1.36)	1.48 (1.44 to 1.52)	1.04 (1.01 to 1.08)
Cluster-level SD ^e	0.09		
Mean difference: adjusted^d (n = 16,095)			
	0.69 (0.52 to 0.85)	0.90 (0.73 to 1.06)	0.21 (0.05 to 0.37)
ICC (95% CI)	0.015 (0.009 to 0.025)		
ITT analyses using complete case data and imputed data^{a,b} (UC: N = 7281, GPT: N = 6689, NT: N = 7012), RR (95% CI)			
	GPT vs. UC	NT vs. UC	NT vs. GPT
Poisson regression model ^c with ordinary residuals; adjusted ^d	1.37 (1.28 to 1.46)	1.48 (1.38 to 1.58)	1.08 (1.01 to 1.15)
Cluster-level SD ^e	0.08		
GLLAMM, Generalized Linear and Latent Mixed Models.			
a Patients who died within 28 days of index day excluded.			
b Includes all 20 contact types constituting the primary outcome.			
c All models include random effect on practice.			
d Adjusted models adjust for practice variables: location (Devon, reference; Bristol; Warwick; Norwich), practice list size [large (> 8000 patients), reference; medium (3500–8000 patients); small (< 3500 patients)], practice deprivation [non-deprived (at or below average deprivation for England by APHO ratings), reference; deprived (above average for England by APHO ratings)] and patient variables: age (categorised, 0–4; 5–11; 16–24; 25–59, reference; 60–74; ≥ 75 years), gender (reference: female), IMD 2010 ⁵⁴ deprivation based on residential postcode categorised into quintiles by rank (reference: least deprived, quintile 5).			
e Approximates coefficient of variation for Poisson model with log link.			

Additional analyses based on primary outcome

Adjusting the primary outcome so that all within-practice contacts on the index day were combined into one overall contact, an increased rate of contacts was still observed when comparing GPT with UC (*Table 13*), with an RR of 1.10 (95% CI 1.04 to 1.17), and when comparing NT with UC (RR 1.12, 95% CI 1.00 to 1.26). Considering primary outcome contacts on the index day only, the RRs for both GPT and NT compared with UC were significantly increased (RR 1.51, 95% CI 1.43 to 1.59 and RR 1.72, 95% CI 1.63 to 1.82), respectively; see *Table 13*). Also, the RR of index day contacts was increased in NT compared with GPT (RR 1.14, 95% CI 1.09 to 1.20). Further analyses included the primary outcome contacts on days 2–28 of the follow-up period (i.e. including only those contacts occurring subsequent to the index day); whilst the increased rate of contacts continued to be apparent in the GPT and NT arms vs. UC (RRs 1.46, 95% CI 1.34 to 1.59, and 1.34, 95% CI 1.22 to 1.48, respectively), there was little difference in the rate of contacts comparing NT vs. GPT (RR 0.97, 95% CI 0.93 to 1.00).

Analysis of GP face-to-face contacts only, on the index day only, demonstrated significant reductions in the rate of contacts for both GPT (RR 0.45, 95% CI 0.37 to 0.55) and NT (RR 0.69, 95% CI 0.57 to 0.84) compared with UC (see *Table 13*). A similar analysis using GP face-to-face contacts across the full 28-day follow-up period yielded contact rates that were again reduced in comparison with UC, but to a lesser extent than when considering only the index day (RR 0.61, 95% CI 0.54 to 0.69 in GPT; 0.80, 95% CI 0.71 to 0.90 in NT). Including GP face-to-face and telephone contacts only, across the full 28-day follow-up period, the index day only, and the days subsequent to the index day only, also demonstrated increased contact rates for the GPT arm compared with UC (see *Table 13*).

Triage resulted in a redistribution of primary care contacts. Although GPT, compared with UC, was associated with an increased rate of overall GP contacts (face to face and telephone) over the 28 days of 38% (RR 1.38, 95% CI 1.28 to 1.50), the rate of GP face-to-face contacts was reduced by 39% (RR 0.61, 95% CI 0.54 to 0.69). NT was associated with a reduction in the rate of overall GP contacts of 16% (RR 0.84, 95% CI 0.78 to 0.91), including a reduction in GP face-to-face contacts of 20% (RR 0.80, 95% CI 0.71 to 0.90), which should be viewed in the context of increased numbers of nurse contacts incurred by the NT process (see *Table 16*).

Figure 8a presents the redistribution of workload across the 28 days evident in the three trial arms, highlighting the magnitude of the redistribution of workload from face to face to telephone (in GPT) and from doctor to nurse (in NT). *Figure 8b* presents the redistribution of workload on the index day as a helpful comparison with 28-day workload redistribution.

Safety outcomes

There were eight deaths within 7 days of the index consultation request across all three treatment arms (*Table 14*): one in UC (0.1/1000 patients), five in GPT (0.7/1000 patients) and two in NT (0.3/1000 patients). Owing to the small number of deaths within 7 days during the trial, no formal inferential analysis between groups has been performed. 'Two independent adjudicators deemed that the circumstances of the deaths were not associated with the trial group or procedures' (Campbell *et al.*⁶⁴).

No patient had more than one emergency hospital admission within the 7-day follow-up period. ITT analyses indicated no evidence of increased risk of emergency admission in either of the triage arms compared with UC (see *Table 14*). Durations of emergency admissions were also similar in all arms. Furthermore, there was no evidence of an increased risk of at least one A&E admission within the 28-day follow-up period in GPT or NT compared with UC.

TABLE 13 Sensitivity analyses based on the primary outcome

	UC (N = 5572)	GPT (N = 5171)	NT (N = 5468)
Primary care and A&E contacts within 28 days of index day, with combination of all within-practice contacts on index day into one contact^{a,b}			
Total primary care contacts: ^c total, mean (SD)	10,450, 1.88 (1.39)	11,189, 2.16 (1.65)	11,592, 2.12 (1.60)
	GPT vs. UC	NT vs. UC	NT vs. GPT
Poisson GLLAMM with robust residuals; adjusted ^d (n = 16,095), RR (95% CI)	1.10 (1.04 to 1.17)	1.12 (1.00 to 1.26)	1.01 (0.87 to 1.18)
Cluster-level SD ^e	0.08		
Primary care and A&E contacts on index day only^{a,b}			
Total primary care contacts: ^c total, mean (SD)	5490, 0.99 (0.31)	7705, 1.49 (0.57)	9276, 1.70 (0.55)
	GPT vs. UC	NT vs. UC	NT vs. GPT
Poisson regression model with ordinary residuals; adjusted ^d (n = 16,095), RR (95% CI)	1.51 (1.43 to 1.59)	1.72 (1.63 to 1.82)	1.14 (1.09 to 1.20)
Cluster-level SD ^e	0.05		
Primary care and A&E contacts on days 2–28 following index day only^{a,b}			
Total primary care contacts: ^c total, mean (SD)	5126, 0.92 (1.39)	6015, 1.16 (1.64)	6124, 1.12 (1.60)
	GPT vs. UC	NT vs. UC	NT vs. GPT
Poisson GLLAMM with robust residuals; adjusted ^d (n = 16,095), RR (95% CI)	1.46 (1.34 to 1.59)	1.34 (1.22 to 1.48)	0.97 (0.93 to 1.00)
Cluster-level SD ^e	0.18		
GP face-to-face within-practice contacts on days 1–28			
Total contacts, mean (SD)	8113, 1.46 (0.85)	4766, 0.92 (0.91)	6496, 1.19 (0.89)
	GPT vs. UC	NT vs. UC	NT vs. GPT
Poisson regression model with ordinary residuals; adjusted ^d (n = 16,095), RR (95% CI)	0.61 (0.54 to 0.69)	0.80 (0.71 to 0.90)	1.30 (1.15 to 1.46)
Cluster-level SD ^e	0.15		
GP face-to-face within-practice contacts on index day only			
Total contacts, mean (SD)	5091, 0.91 (0.30)	2195, 0.42 (0.50)	3598, 0.66 (0.48)
	GPT vs. UC	NT vs. UC	NT vs. GPT
Poisson regression model with ordinary residuals; adjusted ^d (n = 16,095), RR (95% CI)	0.45 (0.37 to 0.55)	0.69 (0.57 to 0.84)	1.54 (1.27 to 1.87)
Cluster-level SD ^e	0.24		

continued

TABLE 13 Sensitivity analyses based on the primary outcome (continued)

	UC (N = 5572)	GPT (N = 5171)	NT (N = 5468)
GP face-to-face within-practice contacts on days 2–28 only			
Total contacts, mean (SD)	3022, 0.54 (0.82)	2571, 0.50 (0.77)	2898, 0.53 (0.80)
	GPT vs. UC	NT vs. UC	NT vs. GPT
Poisson regression model with ordinary residuals; adjusted ^d (n = 16,095), RR (95% CI)	0.87 (0.77 to 0.99)	0.95 (0.83 to 1.07)	1.09 (0.96 to 1.23)
Cluster-level SD ^e	0.15		
GP face-to-face and telephone within-practice contacts on days 1–28			
Total contacts, mean (SD)	8718, 1.56 (1.01)	11,324 2.19 (1.29)	7341, 1.34 (1.08)
	GPT vs. UC	NT vs. UC	NT vs. GPT
Poisson regression model with ordinary residuals; adjusted ^d (n = 16,095), RR (95% CI)	1.38 (1.28 to 1.50)	0.84 (0.78 to 0.91)	0.61 (0.56 to 0.66)
Cluster-level SD ^e	0.10		
GP face-to-face and telephone within-practice contacts on index day only			
Total contacts, mean (SD)	5163, 0.93 (0.30)	7076, 1.37 (0.53)	3717, 0.68 (0.49)
	GPT vs. UC	NT vs. UC	NT vs. GPT
Poisson regression model with ordinary residuals; adjusted ^d (n = 16,095), RR (95% CI)	1.49 (1.32 to 1.69)	0.72 (0.63 to 0.82)	0.48 (0.43 to 0.55)
Cluster-level SD ^e	0.15		
GP face-to-face and telephone within-practice contacts on days 2–28 only			
Total contacts, mean (SD)	3555, 0.64 (0.97)	4248, 0.82 (1.18)	3624, 0.66 (0.99)
	GPT vs. UC	NT vs. UC	NT vs. GPT
Poisson GLLAMM with robust residuals; adjusted ^d (n = 16,095), RR (95% CI)	1.19 (1.13 to 1.24)	1.00 (0.94 to 1.07)	0.84 (0.80; 0.89)
Cluster-level SD ^e	0.14		

GLLAMM, Generalized Linear and Latent Mixed Models.

a Patients who died within 28 days of index day excluded.

b All models include random effect on practice.

c Includes all 20 contact types constituting the primary outcome.

d Adjusted models adjust for practice variables: location (Devon, reference; Bristol; Warwick; Norwich), practice list size [large (> 8000 patients), reference; medium (3500–8000 patients); small (< 3500 patients)], practice deprivation [non-deprived (at or below average deprivation for England by APHO ratings), reference; deprived (above average for England by APHO ratings)] and patient variables: age (categorised, 0–4; 5–11; 16–24; 25–59, reference; 60–74; ≥ 75 years), gender (reference: female), IMD 2010⁵⁴ deprivation based on residential postcode categorised into quintiles by rank (reference: least deprived, quintile 5).

e Approximate coefficient of variation for Poisson model with log link.

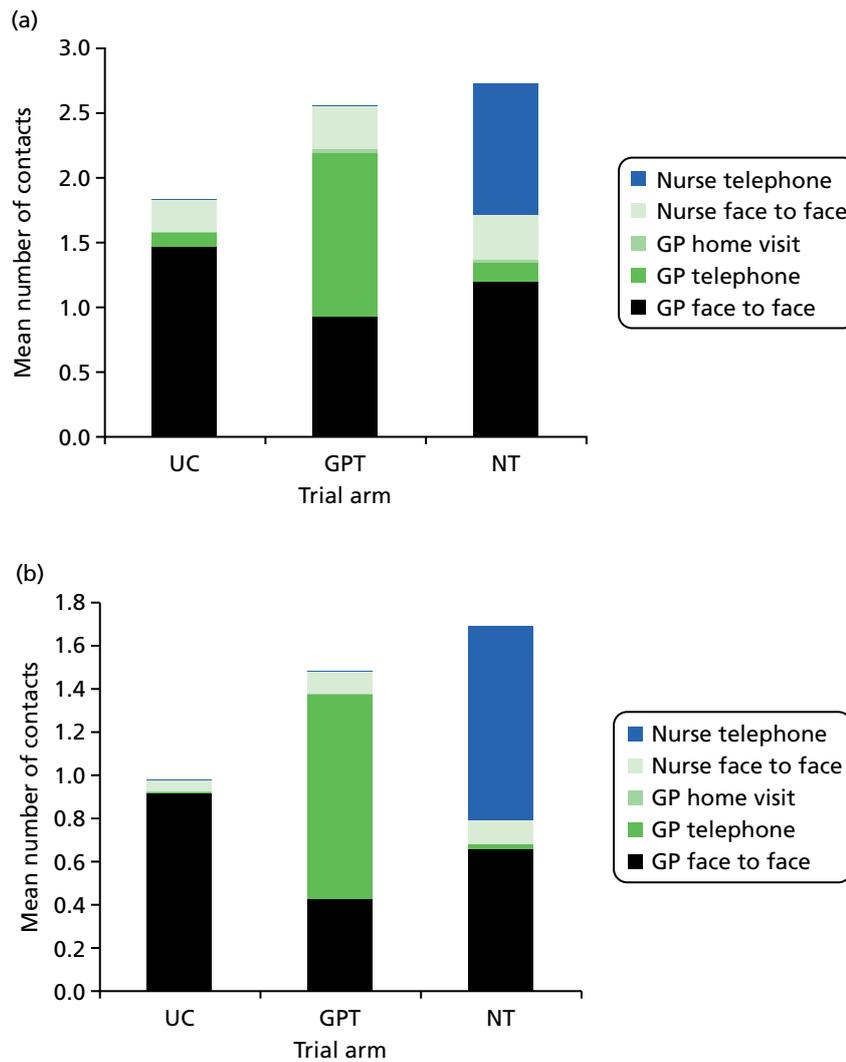


FIGURE 8 Redistribution of within-practice workload over (a) the 28-day workload (primary outcome measure) and (b) the index day.

TABLE 14 Comparison of safety outcomes based on contacts across trial arms

	UC (total <i>N</i> = 7283; case note review <i>N</i> = 5572)	GPT (total <i>N</i> = 6695; case note review <i>N</i> = 5171)	NT (total <i>N</i> = 7012; case note review <i>N</i> = 5468)
Deaths with 7 days of index day (all patients)			
Total deaths (<i>n</i> /1000 patients)	1 (0.1)	5 (0.7)	2 (0.3)
Emergency hospital admissions (patients with case note review only)^a			
Did the patient have at least one emergency hospital admission within 7 days of index day? ^b <i>n</i> (%)	52 (0.93)	59 (1.14)	69 (1.26)
Did the patient have at least one emergency hospital admission, planned in advance by health professional within 7 days of index day? <i>n</i> (%)	38 (0.68)	38 (0.73)	36 (0.66)
Did the patient have at least one emergency hospital admission, unplanned by health professional within 7 days of index day? <i>n</i> (%)	14 (0.25)	21 (0.41)	33 (0.60)
At least one emergency hospital admission within 7 days of index day:^{a-c} OR (95% CI)			
	GPT vs. UC	NT vs. UC	NT vs. GPT
Logistic regression, adjusted ^d (<i>n</i> = 16,095)	1.17 (0.75 to 1.85)	1.31 (0.83 to 2.07)	1.12 (0.73 to 1.72)
ICC ^e (95% CI)	0.023 (0.004 to 0.122)		
Number of bed-days for patients who had an admission; mean (SD), <i>n</i>	3.8 (6.4), 52	3.4 (3.7), 59	4.5 (5.7), 68
A&E contacts within 28 days of index day (patients with case note review only)^a			
Did the patient have at least one A&E attendance within 28 days of index day? <i>n</i> (%)	166 (3.0)	171 (3.3)	156 (2.9)
At least one A&E contact within 28 days of index day:^{a-c} OR (95% CI)			
	GPT vs. UC	NT vs. UC	NT vs. GPT
Logistic regression; adjusted ^d (<i>n</i> = 16,095)	1.18 (0.87 to 1.61)	1.09 (0.80 to 1.49)	0.92 (0.67 to 1.26)
ICC (95% CI)	0.022 (0.007 to 0.063)		
Number of A&E attendances within 28 days of index day, total, mean (SD)	187, 0.03 (0.21)	180, 0.03 (0.19)	183, 0.03 (0.22)
<p>a Patients who died within 28 days of index day excluded.</p> <p>b A total of 180 patients (1.1%) had one admission; no patients had more than one admission.</p> <p>c All models include random effect on practice.</p> <p>d Adjusted models adjust for practice variables: location (Devon, reference; Bristol; Warwick; Norwich), practice list size [large (> 8000 patients), reference; medium (3500–8000 patients); small (< 3500 patients)], practice deprivation [non-deprived (at or below average deprivation for England by APHO ratings), reference; deprived (above average for England by APHO ratings)] and patient variables: age (categorised, 0–4; 5–11; 16–24; 25–59, reference; 60–74; ≥ 75 years), gender (reference: female), IMD 2010⁵⁴ deprivation based on residential postcode categorised into quintiles by rank (reference: least deprived, quintile 5).</p>			

Patient management on the index day

The disposition of all patients on the index day (covering the first two contacts on the index day if the first contact was within practice) is set out in *Table 15*. For patients who had their first contact outside the practice (only three patients in UC, one of whom had an emergency hospital admission on the index day), no further patient disposition was covered. Within each arm, only a small number of patient management pathways covered the overwhelming majority (> 90%) of patients. Within UC, only three management pathways individually accounted for > 1% of the total patients with case note review data; in combination, these pathways represented > 90% of patients. Within GPT and NT there were five patient management pathways that individually accounted for > 1% of patients with case note review data, and in combination represented > 95% of patients within each arm. The proportion of patients who received no within-practice management on the index day was higher in UC (5%; 279/5572) than in the triage arms (approximately 1% in both triage arms). These patient management pathways on the index day are summarised in *Figure 9*. In GPT of 4796 patients who were treated per protocol, 1017 (21.2%) were definitively managed (i.e. had no further recorded contact with primary care or A&E within 28 days or hospital admissions within 7 days). In NT, of 4860 patients treated per protocol, 394 (8.1%) were definitively managed.

TABLE 15 Patient management on index day

Contact ^a	UC (N = 5572): n (%)	GPT (N = 5171): n (%)	NT (N = 5468): n (%)
No contact (primary care or hospital admission) on index day			
	279 (5.0)	51 (1.0)	52 (1.0)
First contact on index day: consultation request outside practice (including emergency hospital admissions)^b			
	3 (0.1)	0 (0)	0 (0)
First contact on index day: GP face to face			
Total patients with first contact on index day: GP face to face	5034 (90.3)	293 (5.7)	480 (8.8)
Second contact on index day			
No second contact on index day	4874 (87.5)	269 (5.2)	461 (8.4)
GP face to face	19 (0.3)	11 (0.2)	2 (< 0.1)
GP telephone	14 (0.3)	4 (0.1)	4 (0.1)
Nurse face to face	84 (1.5)	7 (0.1)	7 (0.1)
Nurse telephone	3 (0.1)	0 (0)	1 (< 0.1)
Other practice	0 (0)	0 (0)	0 (0)
Outside practice ^c	40 (0.7)	2 (< 0.1)	5 (0.1)
First contact on index day: GP telephone			
Total patients with first contact on index day: GP telephone	54 (1.0)	4796 (92.7)	22 (0.4)

continued

TABLE 15 Patient management on index day (continued)

Contact ^a	UC (N = 5572): n (%)	GPT (N = 5171): n (%)	NT (N = 5468): n (%)
Second contact on index day			
No second contact on index day	40 (0.7)	2357 (45.6; 49.1 ^d)	12 (0.2)
GP face to face	14 (0.3)	1850 (35.8; 38.6 ^d)	8 (0.2)
GP telephone	0 (0)	60 (1.2; 1.3 ^d)	1 (<0.1)
Nurse face to face	0 (0)	465 (9.0; 9.7 ^d)	0 (0)
Nurse telephone	0 (0)	1 (<0.1; <0.1 ^d)	0 (0)
Other practice	0 (0)	36 (0.7; 0.8 ^d)	1 (<0.1)
Outside practice ^c	0 (0)	27 (0.5; 0.6 ^d)	0 (0)
First contact on index day: nurse face to face			
Total patients with first contact on day of index consultation request: nurse face to face	197 (3.5)	25 (0.5)	50 (0.9)
Second contact on index day			
No second contact on day of index consultation request	180 (3.2)	23 (0.4)	33 (0.6)
GP face to face	15 (0.3)	1 (<0.1)	13 (0.2)
GP telephone	0 (0)	0 (0)	1 (0.02)
Nurse face to face	0 (0)	1 (<0.1)	3 (0.1)
Nurse telephone	2 (<0.1)	0 (0)	0 (0)
Other practice	0 (0)	0 (0)	0 (0)
Outside practice ^c	0 (0)	0 (0)	0 (0)
First contact on index day: nurse telephone			
Total patients with first contact on day of index consultation request: nurse telephone	3 (0.1)	1 (<0.1)	4860 (88.9)
Second contact on index day			
No second contact on day of index consultation request	1 (<0.1)	0 (0)	1220 (22.3; 25.1 ^e)
GP face to face	1 (<0.1)	0 (0)	3034 (55.5; 62.4 ^e)
GP telephone	0 (0)	0 (0)	73 (1.3; 1.5 ^e)
Nurse face to face	0 (0)	0 (0)	467 (8.5; 9.6 ^e)
Nurse telephone	1 (<0.1)	0 (0)	35 (0.6; 0.7 ^e)
Other practice	0 (0)	0 (0)	15 (0.3; 0.3 ^e)
Outside practice ^c	0 (0)	1 (<0.1)	16 (0.3; 0.3 ^e)
First contact on index day: other practice contact			
	2 (<0.1)	5 (0.1)	4 (0.1)

a Includes first and second contacts only on index day.

b Patients whose first contact was outside the practice are not considered further.

c Includes primary care out-of-hours contacts, walk-in centre contacts, A&E contacts and unrelated/emergency hospital admissions.

d Second percentage is based on denominator of all patients who had a GP telephone contact as their first contact on the index day (n = 4796).

e Second percentage is based on denominator of all patients who had a nurse telephone contact as their first contact on the index day (n = 4860).

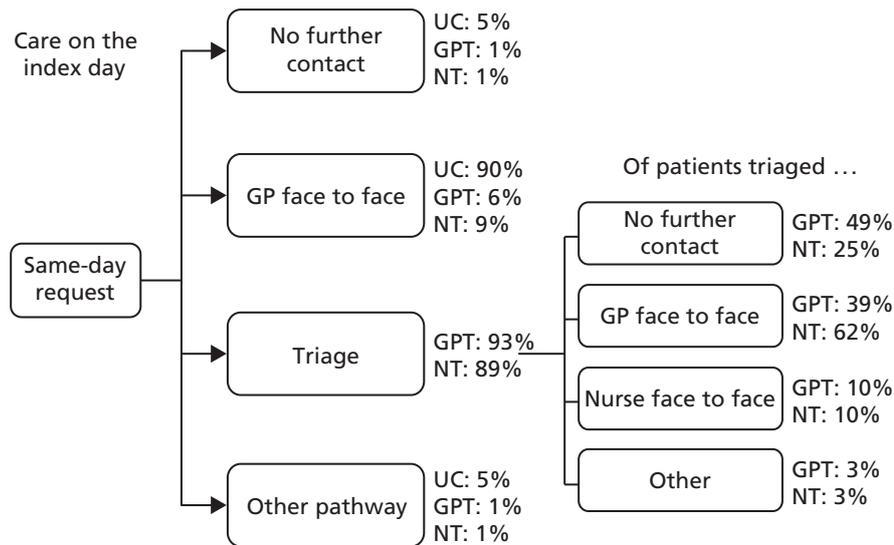


FIGURE 9 Summary, by arm, of within-practice patient management on the index day.

Resource use

Components of the primary outcome

Descriptive data regarding the 20 individual contact types that comprise the primary outcome are shown in *Table 16*; also shown are descriptive data for all GP within-practice contacts, all nurse within-practice contacts, all within-practice contacts (including GP, nurse and unspecified), all walk-in centre contacts (numbers are similar across all arms, indicating that any differences in recording of walk-in centre contacts across the locations has not resulted in a recording bias across the arms) and all out-of-hours primary care contacts.

Did not attend contacts

Total DNAs had similar means across all treatment arms, and were a slightly higher proportion of the total number of contacts (including attended and non-attended contacts) for the UC arm [2.5% (265/10,447); see *Table 16*] compared with the triage arms. The proportions of DNAs for telephone contacts made by the triaging clinician type in the two triage arms were similar [1.7% (113/6671) for GP telephone contacts in the GPT arm, and 1.1% (61/5560) for nurse telephone contacts in the NT arm].

Patient self-reported resource use

The proportions of patients who contacted NHS Direct were similar across the three arms, as were the mean numbers of contacts (see *Table 16*).

Duration of consultations within trial

Based on Clinician Form data, the mean contact duration in UC was longer than for either of the triage arms [9.5 (SD 5.0) minutes compared with 4.0 (SD 2.8) minutes in GPT and 6.6 (SD 3.8) minutes in NT; *Table 17*, including first management/triage contacts only]. The duration of GP and nurse face-to-face consultations for ESTEEM patients (taken for a sample of patients on 2 days) indicated that the mean duration of GP face-to-face consultations in the GPT arm was slightly longer than in NT [12.4 (SD 7.1) minutes compared with 11.5 (SD 6.4) minutes; see *Table 17*] and longer than in UC [9.8 (SD 5.1) minutes] when including consultations on the index day or subsequent day only. Nurse face-to-face consultations had a mean duration of 13.9 (SD 8.8) minutes in GPT, 11.0 (SD 6.6) minutes in UC and 11.0 (SD 8.1) minutes in NT, again including only patients being seen on the index day or the following day.

Estimated patient–clinician contact times indicated little difference in overall contact time comparing GPT with UC, although a slightly increased amount of the total clinician time was fulfilled by nurses in the GPT arm. The overall patient–clinician contact time was noticeably longer in the NT arm, with a much

TABLE 16 Descriptive summary of individual contact types comprising the primary outcome and patient-reported use of NHS Direct

Contact type	UC (N = 5572)	GPT (N = 5171)	NT (N = 5468)
Within-hours surgery contacts: total contacts, mean (SD)			
GP face to face ^{a,b}	8113, 1.46 (0.85)	4766, 0.92 (0.91)	6496, 1.19 (0.89)
GP telephone ^{a,b}	605, 0.11 (0.39)	6558, 1.27 (0.77)	845, 0.15 (0.50)
GP home visit ^{a,b}	48, 0.01 (0.12)	143, 0.03 (0.21)	107, 0.02 (0.19)
GP unspecified ^{a,b}	2, <0.01 (0.02)	12, <0.01 (0.05)	10, <0.01 (0.04)
Nurse face to face ^{a,b}	1361, 0.24 (0.71)	1673, 0.32 (0.78)	1928, 0.35 (0.82)
Nurse telephone ^{a,b}	50, 0.01 (0.11)	77, 0.01 (0.15)	5499, 1.01 (0.49)
Nurse home visit ^{a,b}	1, <0.01 (0.01)	0, 0 (0)	4, <0.01 (0.03)
Nurse unspecified ^{a,b}	2, <0.01 (0.02)	6, <0.01 (0.03)	7, <0.01 (0.04)
General unspecified ^{a,b}	0, 0 (0)	3, <0.01 (0.02)	3, <0.01 (0.02)
Walk-in centre contacts: total contacts, mean (SD)			
^a Walk-in centre: doctor	3, <0.01 (0.02)	10, <0.01 (0.05)	5, <0.01 (0.03)
^a Walk-in centre: nurse	24, <0.01 (0.07)	13, <0.01 (0.05)	21, <0.01 (0.07)
^a Walk-in centre: unspecified	5, <0.01 (0.03)	9, <0.01 (0.04)	5, <0.01 (0.03)
Primary care out-of-hours contacts: total contacts, mean (SD)			
Out-of-hours GP face to face ^a	61, 0.01 (0.11)	63, 0.01 (0.12)	52, 0.01 (0.10)
Out-of-hours GP telephone ^a	66, 0.01 (0.12)	111, 0.02 (0.18)	94, 0.02 (0.16)
Out-of-hours GP home visit ^a	3, <0.01 (0.03)	31, 0.01 (0.10)	16, <0.01 (0.06)
Out-of-hours nurse face to face ^a	3, <0.01 (0.02)	4, <0.01 (0.03)	9, <0.01 (0.04)
Out-of-hours nurse telephone ^a	12, <0.01 (0.05)	28, 0.01 (0.09)	12, <0.01 (0.05)
Out-of-hours nurse home visit ^a	1, <0.01 (0.01)	0, 0 (0)	0, 0 (0)
Out-of-hours unspecified ^a	69, 0.01 (0.14)	33, 0.01 (0.08)	104, 0.02 (0.18)
A&E attendances: total contacts, mean (SD)			
A&E attendances ^a	187, 0.03 (0.21)	180, 0.03 (0.19)	183, 0.03 (0.22)
Subtotals: total contacts, mean (SD)			
Total within-practice GP contacts ^{a,b}	8768, 1.57 (1.02)	11,479, 2.22 (1.33)	7458, 1.36 (1.13)
Total within-practice nurse contacts ^{a,b}	1414, 0.25 (0.73)	1756, 0.34 (0.82)	7438, 1.36 (1.00)
Total within-practice contacts ^{a,b}	10,182, 1.83 (1.31)	13,238, 2.56 (1.60)	14,899, 2.72 (1.55)
Total walk-in centre contacts ^a	32, 0.01 (0.08)	32, 0.01 (0.08)	31, 0.01 (0.09)
Total primary care out-of-hours contacts ^a	215, 0.04 (0.27)	270, 0.05 (0.35)	287, 0.05 (0.33)
DNAs: total DNAs, mean (SD); % n total contacts (attended and DNA) by type			
Failed consultations (patient DNA/was not contacted by telephone – DNA) ^a within-hours practice consultations only; GP face-to-face consultation	110, 0.02 (0.14); 1.3	66, 0.01 (0.12); 1.4	69, 0.01 (0.12); 1.1
GP telephone consultation	82, 0.01 (0.13); 11.9	113, 0.02 (0.15); 1.7	73, 0.01 (0.12); 8.0
Nurse face-to-face consultation	62, 0.01 (0.11); 4.4	46, 0.01 (0.10); 2.7	55, 0.01 (0.10); 2.8
Nurse telephone consultation	9, <0.01 (0.04); 15.3	16, <0.01 (0.07); 17.2	61, 0.01 (0.11); 1.1
Total DNAs (all within-hours practice contact types)	265, 0.05 (0.24); 2.5	241, 0.05 (0.23); 1.8	267, 0.05 (0.23); 1.8

TABLE 16 Descriptive summary of individual contact types comprising the primary outcome and patient-reported use of NHS Direct (*continued*)

Contact type	UC (N = 5572)	GPT (N = 5171)	NT (N = 5468)
Patient-reported health-care contacts^c			
<i>NHS direct contacts within 4 weeks of index day, excluding index day: n (%)</i>			
0	3633 (95.7)	3474 (94.5)	3194 (94.1)
1	116 (3.1)	141 (3.8)	150 (4.4)
2	27 (0.7)	44 (1.2)	28 (0.8)
3	13 (0.3)	11 (0.3)	10 (0.3)
4	3 (0.1)	1 (0.03)	4 (0.1)
5	0 (0)	1 (0.03)	4 (0.1)
≥6	5 (0.1)	5 (0.1)	3 (0.1)
Total n	3797	3677	3393
Mean (SD), n (excluding ≥6)	0.06 (0.31), 3792	0.07 (0.35), 3672	0.08 (0.38), 3390

a Derived from case note review; patients who died within 28 days of index day excluded.

b Excludes non-attended contacts or failure to contact by telephone (DNAs).

c Derived from questionnaire.

TABLE 17 Duration of consultations during trial

	UC (N = 7283)	GPT (N = 6695)	NT (N = 7012)
Duration of consultation (minutes) based on Clinician Form^{a,b} data: mean (SD), n			
First management/triage contacts only ^c	9.5 (5.0), 5693	4.0 (2.8), 5508	6.6 (3.8), 5510
Duration of face-to-face consultations (minutes) for ESTEEM patients:^d mean (SD), n			
GP face-to-face consultations on day of index consultation request or day after index consultation request	9.8 (5.1), 631	12.4 (7.1), 244	11.5 (6.4), 415
Nurse/nurse practitioner face-to-face consultations on day of index consultation request or day after index consultation request	11.0 (6.6), 61	13.9 (8.8), 58	11.0 (8.1), 57
Estimated composite patient–clinician contact time (minutes) on the index day^e			
Overall estimated patient–clinician contact time	9.6	10.3	14.8
Estimated patient–GP contact time	9.1	9.0	7.7
Estimated patient–nurse contact time	0.6	1.3	7.1

a Clinician Form data included only if Clinician Form is dated within 7 days of index consultation request (UC) or dated same date as index consultation request (GPT/NT).

b Includes 98 Clinician Forms that were recorded as patient DNA, but did include duration data.

c A triage contact is defined as a GP telephone contact on the date of the index consultation request in GPT, or a nurse telephone or GP telephone contact on the date of the index consultation request in NT. A first management contact in UC is defined as any contact within 7 days of the index consultation request.

d Durations of face-to-face consultations recorded by form completion or electronically from practice computer system; practices were requested to record data on two separate days – one Monday/Friday and one Tuesday/Wednesday/Thursday; one practice collected data only on a Tuesday and Wednesday, two practices collected data on one day only.

e Patients who DNA any within-practice contacts on the index day, or who were first managed outside the practice, were excluded. Patient management pathways that represented < 1% of patients in each arm were excluded (overall 92% of patients in UC were included, 97% in GPT and 96% in NT). Average durations for each contact type were derived from the Clinician Forms, audit of face-to-face GP/nurse contacts and standard unit timings (Curtis³⁸) as appropriate.

greater proportion of the time being covered by nurses, largely due to the nurse telephone triage aspect of patient management in this arm (see *Table 17*).

Patient self-reported health status (European Quality of Life-5 Dimensions)

Intention-to-treat analyses using Tobit regression models (*Table 18*) showed no evidence of a significant difference in EQ-5D score across the three trial arms (frequencies for the component questions of EQ-5D are shown in *Appendix 21*).

Patient experience

Patients evaluated that it was easier to get through to the practice on the phone in the GPT vs. UC arm (mean difference -8.70 , 95% CI -16.50 to -0.89 ; *Table 19*). There was no evidence to indicate any difference between the UC and GPT arms in terms of ease of accessing prompt care, whereas the NT patients reported that it was significantly more difficult to access prompt care (mean difference 7.02 , 95% CI 3.60 to 10.45 , vs. UC; mean difference 6.63 , 95% CI 3.23 to 10.03 , vs. GPT). NT patients also reported increased difficulty in seeing a GP or nurse compared with the UC arm, and increased difficulty in getting medical help or advice compared with both the UC and GPT arms. Patients in the NT arm also found their care to be less convenient and reported lower overall satisfaction with their care compared with the other two arms; the mean difference in patient satisfaction was 3.94 (95% CI 1.88 to 5.99) compared with UC, and 2.60 (95% CI 0.58 to 4.63) compared with GPT. However, there was no evidence across the arms to indicate any differences in patient self-assessment of the problem (the reason for the index consultation request) at the time of responding to the questionnaire. Frequencies for the individual responses to the patient experience questions are shown in *Appendix 22*.

Case complexity

One or more of the case complexity items was missing for a high proportion of patients and this proportion varied across the arms (see *Appendix 23*); 25% of the GPT patients did not have full complexity data available compared with 18% of patients in UC and 16% in NT (excluding patients with a Clinician Form that indicated non-attendance). In all three arms it was evident that the degree of missingness for the social, psychological and administrative complexity parameters was much greater than that for physical complexity. For those patients who had full case complexity data available, there appeared to be some evidence for systematic variation across the arms; the distribution of the overall complexity score varied by arm. The highest mean case complexity score was seen in the UC arm [2.32 (SD 0.88) compared with 1.92 (SD 0.73) in GPT and 2.20 (SD 1.04) in NT]; although the modal score was '2' in each arm, the distributions around the mode were different in each arm.

TABLE 18 Patient-reported health status (EQ-5D): ITT analysis

	UC (N = 7283)	GPT (N = 6695)	NT (N = 7012)
Total EQ-5D: mean (SD), n	0.769 (0.271), 3281	0.764 (0.274), 3246	0.775 (0.270), 2925
	GPT vs. UC	GPT vs. UC	NT vs. GPT
Tobit model; ^{a,b} coefficient (95% CI); adjusted (n = 9404)	0.01 (-0.02 to 0.04)	0.02 (-0.01 to 0.06)	0.01 (-0.02 to 0.05)
ICC (95% CI)	0.006 (0.003 to 0.014)		

a All models include random effect on practice.

b Adjusted models adjust for practice variables: location (Devon, reference; Bristol; Warwick; Norwich), practice list size [large (> 8000 patients), reference; medium (3500–8000 patients); small (< 3500 patients)], practice deprivation [non-deprived (at or below average deprivation for England by APHO ratings), reference; deprived (above average for England by APHO ratings)] and patient variables: age (categorised, 0–4; 5–11; 16–24; 25–59, reference; 60–74; ≥ 75 years), gender (reference: female), IMD 2010⁵⁴ deprivation based on residential postcode categorised into quintiles by rank (reference: least deprived, quintile 5).

TABLE 19 Patient experience of care: ITT analysis

	UC (N = 7283)	GPT (N = 6695)	NT (N = 7012)
	GPT vs. UC	NT vs. UC	NT vs. GPT
How easy or difficult was it to get through to the practice on the phone? [Mean difference (95% CI)]			
Linearised on scale of 0–100 (ascending difficulty); ^{a,b} adjusted (n = 11,719)	–8.70 (–16.50 to –0.89)	–2.21 (–10.00 to 5.59)	6.49 (–1.26 to 14.25)
ICC (95% CI)	0.133 (0.090 to 0.192)		
How easy or difficult was it to receive prompt care? [Mean difference (95% CI)]			
Linearised on scale of 0–100 (ascending difficulty); ^{a,b} adjusted (n = 10,945)	0.39 (–3.01 to 3.80)	7.02 (3.60 to 10.45)	6.63 (3.23 to 10.03)
ICC (95% CI)	0.033 (0.021 to 0.053)		
How easy or difficult was it to see a GP or nurse if you wanted to? [Mean difference (95% CI)]			
Linearised on scale of 0–100 (ascending difficulty); ^{a,b} adjusted (n = 10,945)	3.63 (–0.42 to 7.68)	7.30 (3.23 to 11.37)	3.67 (–0.37 to 7.71)
ICC (95% CI)	0.036 (0.022 to 0.057)		
How easy or difficult was it to get medical help or advice for the problem you presented? [Mean difference (95% CI)]			
Linearised on scale of 0–100 (ascending difficulty); ^{a,b} adjusted (n = 11,062)	–0.28 (–2.68 to 2.12)	4.82 (2.38 to 7.25)	5.09 (2.69 to 7.50)
ICC (95% CI)	0.016 (0.009 to 0.027)		
How convenient was the care provided by your GP surgery on that day? [Mean difference (95% CI)]			
Linearised on scale of 0–100 (ascending inconvenience); ^{a,b} adjusted (n = 11,721)	1.86 (–0.70 to 4.42)	5.54 (2.96 to 8.13)	3.68 (1.13 to 6.24)
ICC (95% CI)	0.019 (0.011 to 0.032)		
Thinking about the reason why you contacted the GP surgery or health centre that day, is the problem now ...? [Mean difference (95% CI)]			
Linearised on scale of 0–100 (ascending worsening; excluding 'Don't know'); ^{a,b} adjusted (n = 11,630)	–2.15 (–4.41 to 0.10)	–1.74 (–4.04 to 0.55)	0.41 (–1.86 to 2.67)
ICC (95% CI)	0.013 (0.008 to 0.023)		
Overall, how satisfied or dissatisfied were you with the care received on that day? [Mean difference (95% CI)]			
Linearised on scale of 0–100 (ascending dissatisfaction); ^{a,b} adjusted (n = 11,767)	1.33 (–0.69 to 3.35)	3.94 (1.88 to 5.99)	2.60 (0.58 to 4.63)
ICC (95% CI)	0.013 (0.007 to 0.023)		

a All models include random effect on practice.

b Adjusted models adjust for practice variables: location (Devon, reference; Bristol; Warwick; Norwich), practice list size [large (> 8000 patients), reference; medium (3500–8000 patients); small (< 3500 patients)], practice deprivation [non-deprived (at or below average deprivation for England by APHO ratings), reference; deprived (above average for England by APHO ratings)] and patient variables: age (categorised, 0–4; 5–11; 16–24; 25–59, reference; 60–74; ≥ 75 years), gender (reference: female), IMD 2010⁵⁴ deprivation based on residential postcode categorised into quintiles by rank (reference: least deprived, quintile 5).

In the light of these observations, we felt that this measure of case complexity was not sufficiently robust to be used as a potential case-mix adjuster. In its place we elected to adopt patient age as a rough proxy for case complexity, acknowledging and accepting the limitations of that approach.

Economic evaluation

Estimate of the intervention cost (unit cost) for the triage interventions

Resources required for triage interventions

The GPT and NT interventions involve resource use and cost for (1) training and set-up and (2) the GP or nurse contact time for the triage event. In addition, for the NT intervention there is a requirement for CDSS to support the intervention. Resource-use estimates presented here are the incremental (additional) resources required for delivery of triage interventions, in addition to the UC (control).

Training and set-up

The training schedule required to establish and set up the triage interventions consists of two components, first training on organisational plans and arrangements for triage, regardless of type, and, second, specific skills training for either GP or nurse triage skills (*Table 20*). This table also highlights the need for CDSS to support the NT intervention.

Tables 21 and 22 report data collected within trial on the staff time, by staff grade, associated with training events for triage interventions. All active intervention practices in the trial ($n = 28$) had training on an organisational approach to triage; this training was provided by Productive Primary Care. Organisational training sessions were group training events, open to practice staff involved in management of same-day requests by patients. These sessions were relatively short, with a mean (SD) duration of 84 (27) minutes for GPT, and 107 (31) minutes for NT.

TABLE 20 Training and set up schedule for triage interventions

Components (format)	GPT	NT
Organisational training (group)	✓	✓
GP triage skills training (group)	✓	N/A
Nurse triage skills training (group)	N/A	✓
Nurse triage skills (individual/remote)	N/A	✓
CDSS requirement (licence/set-up, implementation)	N/A	✓
N/A, not applicable.		

TABLE 21 Organisational training for triage interventions, staff attending (mean count by grade) and mean duration of training

Triage (n)	Staff input (number) by staff grade: mean (SD) [range]					Training duration (minutes): mean (SD) [range]
	GP	Nurse	Nurse practitioner	Practice manager	Clerical/administrative	
GPT (13)	2.5 (1.1) [1–4]	0.08 (N/A) [0–1]	0.08 (N/A) [0–1]	1.2 (0.4) [1–2]	2.7 (1.6) [1–7]	84.2 (27.1) [60–120]
NT (15)	0.7 (0.6) [0–2]	1.4 (1.0) [0–3]	0.13 (0.5) [0–2]	1.1 (0.6) [0–2]	2.1 (1.7) [0–7]	107.0 (30.5) [60–180]
N/A, not applicable.						

TABLE 22 Triage-specific training for triage interventions, staff attending (mean count by grade) and mean duration of training

Triage (n)	Staff input (number) by staff grade: mean (SD) [range]					Training duration (minutes): mean (SD) [range]
	GP	Nurse	Nurse practitioner	Practice manager	Clerical/administrative	
GPT (13)	2.5 (1.0) [1–4]	0.15 (0.6) [0–2]	0.08 (N/A) [0–1]	1.1 (0.6) [0–2]	2.1 (1.8) [0–7]	76.2 (27.0) [45–120]
NT (15)	0.1 (0.3) [0–1]	2.7 (1.0) [1–5]	0.2 (0.6) [0–2]	0.3 (0.6) [0–2]	0.1 (N/A) [0–1]	335.3 (110.7) [90–450]

N/A, not applicable.

General practitioner triage skills training, provided by Productive Primary Care, was delivered in one session with a mean time of 76 (SD 27) minutes. Nurse triage skills training, provided by Plain Healthcare, was delivered using two sessions spread over one full day, with a combined mean duration of 335 (SD 111) minutes. Nurse triage training also included individual training sessions, delivered remotely as one-to-one training with individual staff (nurses only) delivering the intervention; training was delivered over 3 hours (180 minutes) per person (1 × 2-hour session, followed by 1 × 1-hour session). The number of nurses per practice attending these individual training sessions ranged from one to seven, with the mean number being 3.0 (SD 1.4).

Data collected within the trial indicated that the provider of training for GPT practices (Productive Primary Care) delivered both the organisational training and the GP triage skills training, with the time for each component being 84 minutes and 76 minutes, respectively (as above), i.e. a total training time at a mean of 160 (SD 49, range 120–240) minutes.

Further information on data collected to inform the estimate of training costs is presented in *Appendix 24*.

Software (nurse triage)

Software is used to support the NT intervention. The CDSS provided by Plain Healthcare is renewed annually (see cost below in *Table 24*), with annual licence renewal, and is inclusive of 24-hour support. A minimum of one software licence is required per NT practice, on a single-user basis; if a practice has multiple users then each user requires a separate licence or software purchase. Implementation (set-up) of the software requires support from Plain Healthcare, with a 2-day implementation process (and associated fee, see *Table 23*).

Clinician contact time (general practitioner and nurse time) for triage contact

Data on clinician contact time for the triage contact (index contact) are from within-trial data collection, and have been reported earlier (see *Table 17*) by triage intervention. Summary data are presented in *Table 23*. To estimate clinician contact time for triage contacts, in cost analyses, participant-level data were used when a GP delivered the GPT intervention ($n = 5567$) and when a nurse delivered the NT intervention ($n = 5535$); see *Appendix 25* for detail. Mean contact time for GP and nurse triage contacts are 4.00 (SD 2.83) minutes and 6.56 (SD 3.83) minutes, respectively. *Appendix 26* reports mean triage contact time by practice.

TABLE 23 Clinician contact time for triage intervention, by intervention type^a

Intervention	n	Mean (SD), minutes	Percentile range, 5–95 observed (minutes)	Range (minutes)
GPT contact	5567	4.00 (2.83)	1–9	1–58
NT contact	5535	6.56 (3.83)	2–13	1–60

^a See *Appendix 26* for further detail.

Estimate of costs for delivery of triage interventions

Resource use data reported above are combined with unit cost data and market prices to estimate the mean patient-level cost per triage contact for GPT and NT. *Table 24* presents the unit cost data used to estimate intervention costs. Unit costs for GP and nurse time include cost elements associated with staff costs, practice expenses (associated staff costs, office and general business expenses – including telephone, overheads and capital costs) and an allowance for non-contact time (ratio of direct contact to non-contact time with patients); see the method set out by Curtis.³⁸ In the 'base-case' unit cost estimate the cost per minute for GPs includes cost allowance for qualifications,³⁸ and excludes costs associated with other direct staff costs (reference to direct staff costs applicable only to GP costs). This latter assumption is based on the nature of the telephone triage contact, as a single telephone consultation contact. These unit costs are subject to other assumptions on cost components in sensitivity analyses.

TABLE 24 Unit costs used in estimate of triage intervention costs

Resource use unit	Unit cost (£)	Source
GP time (minutes)	3.40	Per-minute unit cost from Curtis, ³⁸ table 10.8b, p. 183 (includes qualifications, excludes direct care costs)
Nurse time (minutes)	0.88	Calculated from the cost per hour of face-to-face contact time reported in Curtis, ³⁸ table 10.6, p. 180 (includes qualifications)
Nurse practitioner time (minutes)	1.52	Calculated from the cost per hour of client contact reported in Curtis, ³⁸ table 10.7, p. 181 (includes qualifications)
GP management staff, e.g. practice manager, IT manager (minutes)	0.59	Salary based on median FTE band 6 Agenda for Change health staff 2011, working 37.5 hours week/42 weeks per annum [source: www.nhs.uk/workers/pay-and-benefits/agenda-for-change-pay-rates/ (accessed 10 November 2014)] Calculation of unit cost per minute based on cost structure reported in Curtis ³⁸ for similar salary grade worker (see <i>Appendix 24</i> for detail)
Receptionist/clerical worker/ secretary (minutes)	0.41	Salary based on median FTE band 4 Agenda for Change health staff 2011 working 37.5 hours week/42 weeks per annum (source: www.nhs.uk/workers/pay-and-benefits/agenda-for-change-pay-rates/) Calculation of unit cost per minute based on cost structure reported in Curtis ³⁸ for similar salary grade worker (see <i>Appendix 24</i> for estimates for detail)
Training fee: organisational training for triage interventions (group)	1350	GP consultant, fee charged. Productive Primary Care, flat fee for 1-day training
Training fee: organisational training for triage, and/or GP triage skills (group)	1350	GP Consultant, fee charged. Productive Primary Care, flat fee for 1-day training
Nurse triage skills training (group)	1129	Plain Healthcare, flat fee for 1-day training event
Nurse triage skills training; individual one-to-one training (remote), per nurse	292.50	Plain Healthcare, 3 hours per nurse at £97.50 per hour
Software (annual licence)	5709	Plain Healthcare, 2012 price for CDSS. Price is annual cost, including 24-hour support (24/7), for single-user licence
Software (project management and implementation fee)	1760	Plain Healthcare, 2012 price for project management and implementation of software Price is standard 2-day cost of implementation per practice

FTE, full-time equivalent.

Table 25 presents the estimated cost associated with the training and CDSS required to set up and support the triage interventions. These costs represent upfront training and CDSS costs, in the first instance, but also reflect an expected ongoing cost to support triage intervention in a practice setting over time (i.e. annual software cost, recurrent training need). In order to estimate a mean patient level unit cost for triage interventions, estimated training, set-up and software costs are distributed over an estimate of the number of patients per practice expected to request a same-day appointment (with a GP), and therefore reflect the expected estimated workload against triage interventions (where available).

Estimates of triage workload (per practice per year) are based on data reported indicating a rate of 3.8 same-day requests per 1000 patients (on GP list) per day;⁴ this rate is applied using an average assumed practice size of 7000 patients to estimate the mean patient level unit costs, resulting in a projected number of same-day requests per annum at approximately 7000 (estimate 6916 requests). In sensitivity analyses we apply data collected across 1 week in each of the ESTEEM practices, where data indicate a rate of 5.53 (SD = 2.16) same-day appointment requests per 1000 patients on the practice list; this leads to an estimate of 10,000 same-day requests per year with a practice size of 7000.

Triage costs associated with training and software are estimated assuming that costs are distributed over a 12-month period (for base-case cost estimates). In sensitivity analyses the cost is estimated when training and software costs are distributed over a 24-month period (expected 2-year workload) and over a 60-month time frame (although in this latter scenario there are likely to be additional costs associated with the retraining associated with staff turnover).

Using the above assumptions, the estimated mean cost for the training component in the GPT intervention is £0.43 per triage contact, and £1.84 per nurse triage contact for training and software cost components. Sensitivity analyses explore the impact of assumptions on number of same-day requests per year per practice (7000 patients, base case), assuming that the workload for same-day requests per practice is 10,000 requests per year (based on data from ESTEEM), the training component in the GPT intervention costs £0.30 per triage contact, and training and software components cost £1.29 per nurse triage contact.

The higher cost on NT reflects the costs associated with software requirements, although training costs for NT are also twice those estimated in the GPT intervention. These cost estimates involve a number of assumptions; however, this component of cost is a relatively small part of the total mean estimated cost for the triage interventions (see Table 26), and adjustments to these assumptions make a negligible difference to the estimated 12-, 24- or 60-month triage intervention costs [see *Sensitivity analyses (primary cost analyses)*].

Table 26 presents estimates of the cost for the triage interventions. GPT is estimated to cost £14.03 per triage contact, and NT has an estimated mean cost of £7.62 per contact. The difference in cost estimates reflects the relatively low cost per minute for nurse time compared with GP time. The training cost component for GPT is a relatively small proportion of triage cost (3.2%) compared with NT, for which the training and software costs represent 24% of the estimated cost for the NT intervention.

Table 27 presents summary results for sensitivity analyses. Sensitivity analyses indicate mean triage costs per contact for GPT and NT interventions of £13.82 and £7.11, respectively, for which cost estimates are based on calculation of costs over 24 months and distribution of costs over a 24-month workload (same-day requests). The corresponding figures when costs are distributed over a 60-month (5-year) workload are £13.69 and £6.81, respectively. Sensitivity analyses indicate that the estimate of the mean NT cost is £8.44 per contact when the practice needs to purchase two licences for the software used.

Where an alternative estimate is used on expected number of same-day requests per practice (assuming a practice size of 7000), applying data from ESTEEM indicating 5.53 requests per day per 1000 patients, we estimate unit costs of £13.90 and £7.07, respectively, for GP and nurse triage contacts.

RESULTS

TABLE 25 Mean cost of per-practice training and software required for triage interventions, and estimate of mean cost per same-day appointment request

Training component	GPT (mean, £)	NT (mean, £)
Trainer's fees: ^a		
Trainer's fee for organisational training session	1350	1350
GP triage skills training		N/A
Nurse triage skills	N/A	1129
Nurse triage skills 1 : 1 training (£292.50 per nurse attending)	N/A	878
Practice staff training cost: based on mean count and duration data reported (see Tables 21 and 22, plus detail on 1 : 1 training) and unit costs (see Table 24, detail provided below)	1666	2056
Training subtotal	3016	5413
Software ^b	N/A	7499
Training and set-up subtotal	3016	12,912
Estimated training/software cost per triage contact (assuming 7000 requests per year per practice for a same-day appointment ^d)	0.43	1.84
N/A, not applicable.		
a In the ESTEEM trial, the same provider was used for the organisational training and GP triage skills; this provider had a flat rate fee of £1350 per day. This rate was not reduced if organisational training was provided alone, as in the NT intervention training sessions.		
b Includes software cost of £5709 plus implementation fee at £1760. Assumes that the average practice can maintain triage with a licence on a single PC/single-user system/basis. For additional users additional licences cost £5709.		

TABLE 26 Estimated unit cost for triage interventions per same-day appointment request

Cost component for triage per same-day appointment request	GPT (mean, £)	NT (mean, £)
Training and software cost	0.43	1.84
Clinician time for delivery of triage intervention	13.60 (SD 9.61)	5.78 (SD 3.37)
Estimated mean triage cost (per same-day appointment request) [assuming 7000 requests for a same-day appointment per year per practice]	14.03	7.62

TABLE 27 Summary table of sensitivity analyses and impact of different assumptions on estimate of unit cost for triage intervention (for same-day request)

'Scenario' for unit cost estimates; sensitivity analyses reflect base-case methods with stated difference	GPT (mean, £)	NT (mean, £)
Base case	14.03	7.62
Intervention costs for set-up/training/equipment spread over 24 months	13.82	7.11
Intervention costs for set-up/training/equipment spread over 60 months	13.69	6.80
Where multiple licences (two licences) needed for NT	14.03	8.44
Within-trial estimation of same-day appointment requests (n = 10,000 p.a. vs. 7000 pain base case)	13.90	7.07
Staff unit costs without qualifications (GP cost at £2.80 minutes vs. £3.40 minutes; nurse cost at £0.75 minutes vs. £0.88 minutes)	11.63	6.90
Staff unit costs including GP cost per minute where direct care costs are included in estimation of GP cost (£3.70 minutes vs. £3.40 minutes)	15.25	7.63

In base-case estimates we apply unit costs of £3.40 and £0.88 per minute for GP and nurse time; however, when making different assumptions (see *Appendix 27*) and applying the published unit cost (per minute or hour) excluding cost allowance for qualifications, these being £2.80 and £0.72, respectively, for GP and nurse time, we estimate unit costs of triage contact at £11.63 and £6.90 for GP and nurse triage, respectively. When using staff unit costs where there is an allowance for costs associated with other direct care staff costs, aligned to the GP role, we apply unit costs of £3.70 and £0.88 (same as base case for nurse) and estimate unit costs of triage at £15.25 and £7.62 for GP and nurse triage contacts, respectively.

Economic analyses

In this section we present the analyses of trial outcome data, in combination with unit costs to estimate the costs associated with the primary outcome, and related secondary and exploratory cost analyses.

Estimate of resource use and cost: primary analyses

Table 28 presents data on resource use by type of contact and the associated estimate of unit cost by type, for each arm of the trial, and the mean total 28-day cost by intervention arm. Data are presented for the participants with complete data collection in the primary outcome ($n = 16,211$). The resource-use (contact) data are consistent with the effectiveness data presented earlier, and are presented here for consideration alongside cost estimates. The mean 28-day estimates of cost are similar for all three interventions, with all slightly above a mean of £75.

Table 29 presents cost estimates by summary category, with cost estimates for GP contacts (in practice, via telephone and home visits) combined, and the same for nurse contacts. Data presented show that GP costs account for 88% of the mean 28-day cost in UC, and for 69% and 68%, respectively, in GPT and NT. When the triage cost is combined with the GP cost in the active triage intervention arms, the combined triage and GP costs account for 85% of the mean total 28-day cost. When nurse contacts are added to triage and GP contacts this subtotal of primary care 'in-practice' contacts account for 92–93% of the mean total 28-day costs, with remaining contacts in the categories for out-of-hours primary care, walk-in centres and A&E, representing a small proportion of the mean total 28-day cost (7.5–8%). The mean costs associated with triage consultations are £13.01 for GPT and £6.83 for NT.

The results are presented in *Table 29* for unadjusted cost differences across the three interventions, for mean total 28-day costs, and also for regression-based statistical analyses, using a multilevel model (see *Data analysis*), with adjustment for practice-level clusters and other prespecified covariates. Primary cost analyses, using these base-case estimates (regression estimates), show no statistically significant differences in 28-day mean total cost by intervention arm.

Estimate of resource use and cost: secondary analyses

Secondary analyses, using base-case assumptions on a 'per-protocol' participant basis (see *Glossary*), shows the 28-day mean costs for GPT and NT slightly reduced compared with ITT analyses, but with no difference reported across all comparisons of cost data in base-case adjusted analyses (see *Table 30*).

Sensitivity analyses (primary cost analyses)

Table 30 presents sensitivity analyses, against the primary analyses on total 28-day cost estimates, using different assumptions on analyses. Sensitivity analysis presents results using base-case assumptions but with imputation methods used to replace missing data (imputation methods are consistent with those used in multiple imputation analyses for the primary outcome; see *Table 12*). Here we see no difference in cost estimates or adjusted differences across the three trial arms. *Table 30* reports sensitivity analyses. Sensitivity analysis considers use of different regression methods, using a GLM regression approach with two different combinations of family and link (using Gaussian and gamma distributions; the former being equivalent to assuming normally distributed data). These differing specifications of GLM regression models show no difference in results, and indicate that it is appropriate to assume normally distributed total cost data; therefore, base-case results are from a multilevel model, with random effects, assuming normally distributed data and allowing for cluster data. Consistent with the base-case analyses, GLM models show no difference in 28-day total cost across comparisons by intervention type.

TABLE 28 Description of primary cost data (summary view) and primary analyses with comparison by intervention

Contacts by type	Number of contacts over 28 days: mean (SD) [95th percentile range]			Costs for care (£) (primary outcome contacts) over 28 days: mean (SD) [95th percentile range]		
	UC (n = 5572)	GPT (n = 5171)	NT (n = 5468)	UC (n = 5572)	GPT (n = 5171)	NT (n = 5468)
Triage (GP/nurse)	0.00 (0.00) [0 to 0]	0.93 (0.26) [0 to 1]	0.89 (0.31) [0 to 1]	0.00 (0.00) [0 to 0]	13.01 (3.64) [0 to 14.03]	6.83 (2.40) [0 to 7.62]
GP in surgery/practice	1.46 (0.85) [1 to 3]	0.92 (0.91) [0 to 3]	1.19 (0.89) [0 to 3]	62.62 (36.62) [43 to 129]	39.76 (39.15) [0 to 129]	51.19 (38.41) [0 to 129]
GP telephone (excludes triage)	0.11 (0.39) [0 to 1]	0.34 (0.72) [0 to 2]	0.15 (0.50) [0 to 1]	2.82 (10.10) [0 to 26]	8.86 (18.59) [0 to 52]	3.91 (12.94) [0 to 26]
GP home visit	0.01 (0.12) [0 to 0]	0.03 (0.21) [0 to 0]	0.02 (0.19) [0 to 0]	0.95 (12.98) [0 to 0]	3.04 (22.94) [0 to 0]	2.15 (20.77) [0 to 0]
Nurse in surgery/practice	0.24 (0.71) [0 to 1]	0.32 (0.78) [0 to 2]	0.35 (0.83) [0 to 2]	3.34 (9.67) [0 to 13.64]	4.43 (10.70) [0 to 27.28]	4.83 (11.26) [0 to 27.28]
Nurse telephone (excludes triage)	0.01 (0.11) [0 to 0]	0.01 (0.15) [0 to 0]	0.12 (0.38) [0 to 1]	0.05 (0.56) [0 to 0]	0.08 (0.82) [0 to 0]	0.62 (1.99) [0 to 5.28]
Nurse home visit	0.00 (0.01) [0 to 0]	0.00 (0) [0 to 0]	0.00 (0.03) [0 to 0]	0.00 (0.29) [0 to 0]	0.00 (0) [0 to 0]	0.02 (0.59) [0 to 0]
Out of hours (total)	0.04 (0.27) [0 to 0]	0.05 (0.35) [0 to 0]	0.05 (0.33) [0 to 0]	1.63 (12.40) [0 to 0]	1.88 (13.66) [0 to 0]	2.16 (14.70) [0 to 0]
Walk-in centre	0.01 (0.08) [0 to 0]	0.01 (0.08) [0 to 0]	0.01 (0.86) [0 to 0]	0.24 (3.10) [0 to 0]	0.25 (3.32) [0 to 0]	0.23 (3.54) [0 to 0]
A&E	0.03 (0.21) [0 to 0]	0.03 (0.19) [0 to 0]	0.03 (0.22) [0 to 0]	3.76 (23.95) [0 to 0]	3.90 (21.57) [0 to 0]	3.75 (24.28) [0 to 0]
Total	1.91 (1.43) [1 to 5]	2.65 (1.74) [1 to 6]	2.82 (1.68) [1 to 6]	75.41 (57.19) [43 to 172]	75.21 (65.45) [14.03 to 205.31]	75.68 (63.09) [7.62 to 184.9]

TABLE 29 Description of primary cost data (summary view) primary cost analyses with comparison by intervention type

Contact type	Number of contacts over 28 days: mean (SD) [95th percentile range]		Costs for care (£), (primary outcome contacts) over 28 days: mean (SD) [95th percentile range]	
	UC (n = 5572)	GPT (n = 5171)	UC (n = 5572)	GPT (n = 5171)
Triage (GP/nurse)	0 (0) [0 to 0]	0.93 (0.26) [0 to 1]	0 (0) [0 to 0]	13.01 (3.64) [0 to 14.03]
GP all (excludes triage)	1.57 (1.02) [1 to 4]	1.29 (1.33) [0 to 4]	66.39 (43.02) [43 to 138]	51.66 (53.75) [0 to 155]
Nurse all	0.25 (0.73) [0 to 1]	0.34 (0.82) [0 to 2]	3.39 (9.73) [0 to 13.64]	4.51 (10.84) [0 to 27.28]
Subtotal: primary care (items above)	1.83 (1.31) [1 to 4]	2.56 (1.60) [1 to 6]	69.78 (44.97) [43 to 155]	69.18 (55.02) [14.03 to 169]
Out of hours (total)	0.04 (0.27) [0 to 0]	0.05 (0.35) [0 to 0]	1.63 (12.40) [0 to 0]	1.88 (13.66) [0 to 0]
Walk-in centre	0.01 (0.08) [0 to 0]	0.01 (0.08) [0 to 0]	0.24 (3.10) [0 to 0]	0.25 (3.31) [0 to 0]
A&E	0.03 (0.21) [0 to 0]	0.03 (0.19) [0 to 0]	3.76 (23.95) [0 to 0]	3.90 (21.57) [0 to 0]
Total (28-day care/cost)	1.91 (1.43) [1 to 5]	2.65 (1.74) [1 to 6]	75.41 (57.19) [43 to 172]	75.21 (65.45) [14.03 to 205.31]
Comparison by intervention type				
Comparison of total 28-day cost by type, no adjustment for covariates			GPT vs. UC (n = 10,743)	NT vs. UC (n = 11,040)
Primary analyses, total 28-day cost; base-case estimates with adjustment for covariates ^{a-c}			-0.20 (2.75) [5.59 to 5.18]	0.27 (2.33) [-4.29 to 4.83]
			-1.76 (2.52) [-6.70 to 3.18]	-0.19 (2.53) [-5.15 to 4.78]
				GPT vs. NT (n = 10,639)
				-0.47 (3.00) [-6.34 to 5.41]
				-1.57 (2.51) [-6.49 to 3.36]

a Regression covariates: age (by category), sex, site, practice size, deprivation (IMD 2010⁵⁴) by quintile, practice deprivation, cluster by practice ID.
b Regression methods for base-case analyses use multilevel random effects model (xtmixed), using Stata 12.0 (StataCorp LP, College Station, TX, USA) (see sensitivity analyses for GLIM approach).
c ICC = 0.009.

TABLE 30 Secondary analyses and sensitivity analyses against primary cost analyses, total 28-day cost

	Costs for care (£), (primary outcome contacts) over 28 days: mean (SD)		Comparison of mean costs (£), by type of care (for same-day request): mean (standard error ^a) [95% CI] [regression methods ^b]			
	UC	GPT	NT	GPT vs. UC (n = 10,743)	NT vs. UC (n = 11,040)	GPT vs. NT (n = 10,639)
Base case: ITT, adjustment for covariates; xtmixed (multilevel model)	75.41 (57.19) (n = 5574)	75.21 (65.45) (n = 5177)	75.68 (63.09) (n = 5468)	-1.76 (2.52) [-6.70 to 3.18]	-0.19 (2.53) [-5.15 to 4.78]	-1.57 (2.51) [-6.49 to 3.36]
Secondary analyses: per protocol – adjustment for covariates; as per base case (xtmixed)	75.41 (57.19) (n = 5574)	74.82 (65.22) (n = 4796)	75.39 (63.44) (n = 4860)	-2.08 (2.62) [-7.22 to 3.06]	-0.02 (2.64) [-5.19 to 5.15]	-2.06 (2.64) [-7.23 to 3.11]
Sensitivity analyses						
As base case, with imputation of missing data, using predictive mean matching, with covariate adjustment, as used in main analyses for effectiveness data	74.93 (56.65) (n = 7281)	75.05 (65.04) (n = 6689)	76.06 (63.35) (n = 7012)	-1.36 (2.57) [-6.40 to 3.68]	0.22 (2.60) [-4.88 to 5.33]	1.58 (2.57) [-3.47 to 6.63]
Using GLM regression model; with adjustment for covariates; regression model = GLM, Gaussian/identity (family, link)	As per base case			-2.03 (2.75) [-7.41 to 3.36]	-0.79 (2.22) [-5.14 to 3.56]	-1.23 (2.93) [-6.97 to 4.50]
Using GLM regression model; with adjustment for covariates; regression model = GLM, gamma/identity (family, link)	As per base case			-2.14 (2.83) [-7.69 to 3.41]	-1.13 (2.12) [-5.27 to 3.02]	-1.00 (3.01) [-7.00 to 4.99]
As per base case except for change to unit costs for out-of-hours contacts	74.47 (54.71)	74.41 (64.24)	74.56 (60.89)	-1.98 (2.61) [-7.09 to 3.14]	-0.20 (2.63) [-5.35 to 4.95]	-1.78 (2.63) [-6.92 to 3.37]
Low unit cost scenario for GP and nurse time (per-minute costs) from PSSRU (Curtis ⁶⁵), excluding costs for qualifications and excluding costs for direct care	59.47 (48.52)	60.28 (54.08)	60.89 (53.23)	-0.034 (2.00) [-4.27 to 3.59]	1.21 (2.01) [-2.74 to 5.16]	-1.55 (2.00) [-5.47 to 2.37]
Assuming GP consultation following 'same-day' request has shorter duration than that estimated for a broad category of GP consultations: assuming unit cost of £35.15 (vs. £43 base case) for the same-day GP consultation in UC arm only, based on mean reported contact time of 9.5 minutes (vs. 11.7 minutes in base case)	68.23 (54.04)	75.21 (65.45)	75.68 (63.09)	5.43 (2.50) ^c [0.52 to 10.33]	7.00 (2.51) ^c [2.07 to 11.92]	-1.57 (2.51) [-6.49 to 3.36]

a Robust standard error in the case of GLM adjusted covariate.

b Regression covariates: age (by category), sex, site, practice size, deprivation (IMD 2010⁶⁴) by quintile, practice deprivation, cluster by practice ID.

c Statistically significant (95% CI).

Sensitivity analysis (see *Table 30*) addresses uncertainty in choice of unit cost data for out-of-hours primary care contacts: in the base case, one standard unit cost is applied for all out-of-hours primary care contacts compared with sensitivity analysis, for which a different cost structure is used for unit cost data on out-of-hours primary care contacts (see *Appendix 28*). Sensitivity analysis shows no difference in results when alternative data are used for out-of-hours unit costs. Sensitivity analysis presents results for which the unit cost data related to GP time cost, and nurse time cost, are based on the lower estimate of cost presented by Curtis.³⁸

In this scenario, analyses are based on a unit cost of £2.80 per minute for GP time (vs. £3.70 in base case), with no allowance for costs associated with GP qualifications or for costs associated with direct costs for related staff. In this analysis, unit cost data for nurse time, per minute, is £0.75 (vs. £0.88 in base case), with no allowance for costs associated with nurse qualification. The majority of the mean 28-day total costs are related to GP and nurse contacts, and, in this low-cost sensitivity analysis, the mean 28-day cost is lower for each of the intervention types, at £59.47, £60.28 and £60.89 for UC, GPT and NT, respectively, but we see no difference in costs in comparisons across intervention type.

In the trial results (see *Table 17*) we see a mean contact time for the UC index contact (contact resulting from same-day request), which is a lower duration of contact than that seen in the standard costing used here for a GP consultation; in the unit cost estimate applied in base-case cost analyses an allowance is made for 11.7 minutes per GP contact (in surgery). This expected consultation duration is based on analyses of data collected in a 2006–7 UK General Practice Workload Survey (Leeds Information Centre, 2007), which involved a representative sample of 329 practices across the UK, with staff completing diary sheets over a 1-week period (as cited by Curtis³⁸). The estimates of contact duration, and subsequent cost, are based on an allocation of GP time across a recorded number of consultations, and not on a duration recorded per contact. The data collected in ESTEEM, in the UC arm, suggests that GP consultations resulting from requests for appointments on the 'same-day' have a mean (SD) duration of 9.5 (5.0) minutes (see *Table 17*), which is shorter than the estimated 11.7-minute duration reported by Curtis³⁸ in relation to the broader category of GP consultations (in practice), of which same-day requests are one component. Data from ESTEEM, although based on over 5600 observations, are from 14 GP practices, and are collected in a trial setting, and are estimates of mean 'contact time', therefore some caution is needed on the generalisability of these data, and on any direct comparison with data estimated using a 'top-down' approach. Although we succeeded in recruiting large numbers of practices and patients from diverse geographical locations, enhancing the generalisability of our results, our findings may be less applicable to practices serving populations with greater ethnic diversity or those located in inner-city areas with very high levels of deprivation. However, if this difference were to be supported by future research, a difference of 2.2 minutes in consultation duration would reflect a reduction in cost for a GP consultation (in practice) from £43 to £35.15 for a consultation resulting from a 'same-day' request.

Therefore, in sensitivity analysis, we use a unit cost of £35.15 (9.5 minutes × £3.70/minute) for the UC index contact (GP consultation). In this scenario we see that the 28-day cost for UC is less than those estimated for both GPT and NT intervention arms. This difference in 28-day cost is statistically significant, with regression analyses predicting a mean cost difference of £5.43 (lower) for UC compared with GPT, and £7.00 (lower) for UC compared with NT.

Exploratory cost analyses: same-day contacts and cost

Table 31 presents cost estimates for contacts that are reported on the same-day as the same-day request, comprising the index contact and subsequent contacts that constitute the POM, on the same day. The total number of recorded contacts over 28 days, aggregating over all participants ($n = 16,211$ in complete case analyses) is 39,736, and 56% of these ($n = 22,471$) are recorded as contacts on the same day as the initial request (see *Appendix 29*). The mean total number of contacts on the same day are 0.99, 1.49 and

TABLE 31 Same-day contacts: description of data, base case (detailed view) and exploratory economic analyses on the same day as appointment request

Contacts by type	Number of contacts on 'same-day': mean (SD)			Costs for care (£) (primary outcome contacts) on 'same day': mean (SD)		
	UC (n = 5572)	GPT (n = 5171)	NT (n = 5468)	UC (n = 5572)	GPT (n = 5171)	NT (n = 5468)
Triage (GP/nurse)	N/A	0.93 (0.26)	0.89 (0.31)	N/A	13.01 (3.64)	6.83 (2.40)
GP in surgery/practice	0.91 (0.30)	0.43 (0.50)	0.66 (0.48)	39.29 (12.75)	18.28 (21.66)	28.31 (20.71)
GP telephone	0.13 (0.11)	0.02 (0.13)	0.18 (0.13)	0.34 (2.98)	0.43 (3.38)	0.46 (3.47)
GP home visit	0.00 (0.02)	0.01 (0.09)	0.00 (0.06)	0.04 (2.08)	0.83 (9.52)	0.36 (6.30)
Nurse in surgery/practice	0.05 (0.22)	0.10 (0.31)	0.11 (0.31)	0.69 (3.01)	1.40 (3.01)	1.46 (4.28)
Nurse telephone	0.00 (0.04)	0.00 (0.02)	0.01 (0.09)	0.01 (0.23)	0.00 (0.13)	0.04 (0.46)
Nurse home visit	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.30)
Out of hours (total)	0.00 (0.04)	0.00 (0.06)	0.00 (0.07)	0.04 (2.01)	0.12 (2.94)	0.11 (3.09)
Walk-in centre	0.00 (0.01)	0.00 (0.02)	0.00 (0.02)	0.01 (0.55)	0.02 (0.81)	0.01 (0.78)
A&E	0.00 (0.07) 0.01	(0.09) 0.01	(0.08) 0.54 (7.78)	0.82 (9.57)	0.66 (8.54)	
Total	0.99 (0.31)	1.49 (0.57)	1.70 (0.55)	40.96 (14.49)	34.91 (24.50)	38.25 (21.93)

N/A, not applicable.

1.70, for UC, GPT and NT, respectively. Estimates of mean cost for these same-day contacts are £40.96, £34.91 and £38.25, for UC, GPT and NT, respectively. Comparison of these cost data indicates a statistically significant difference in cost when comparing GPT and UC, and NT compared with UC, but no difference between GPT and NT (although here the reported difference in cost of –£3.17 is close to being statistically significant, with a 95% CI of –£6.31 to £0.03). However, cost estimates are presented on same-day contacts for completeness, and it is difficult to interpret the same-day costs in the context of an intervention (triage) that is expected to have an immediate impact on number of contacts, with the effect of the intervention expected to be seen over the longer term (in this instance 28 days).

As discussed above (primary cost analyses) the estimated duration (time allowance) for a GP consultation, in the practice, is 11.7 minutes, using the data reported alongside the unit cost estimated by the PSSRU (Curtis³⁸); a unit cost of £43 is used in base-case analyses (based on 11.7-minute duration). The data in *Table 32* with an adjusted cost for a same-day GP consultation (at £35.15), in the UC intervention arm only, would suggest a UC same-day cost of £33.79 compared with the £40.96 estimated using base-case unit costs, a mean difference (reduction) of £7.17 in the estimated mean same-day costs, with the other triage intervention estimates remaining unchanged. In this scenario (sensitivity analysis; see *Table 32*) we see a difference in costs, same-day cost estimates, across the three intervention arms, with UC having a lesser estimate of same-day costs, with this being statistically significant in comparisons of UC with NT (UC £4.57 less than NT). However, as discussed above, in sensitivity analyses, it is important to be cautious when comparing data on mean duration of a GP contact from prior workload analyses³⁸ and the participant-level mean 'contact time' reported in ESTEEM.

TABLE 32 Same-day contacts: exploratory analysis and costs for primary outcome contacts on the same day as appointment request

	Costs for care (£), (primary outcome contacts) on same day: mean (SD)			Comparison of mean costs (£), by type of care: mean (robust se) [95% CI]		
	UC (n = 5572)	GPT (n = 5171)	NT (n = 5468)	GPT vs. UC (n = 10,743)	NT vs. UC (n = 11,040)	GPT vs. NT (n = 10,639)
Same-day cost; no adjustment for covariates	40.96 (14.49)	34.91 (24.50)	38.25 (21.93)	–6.05 (1.73) [–9.43 to –2.67]	–2.70 (1.72) [–6.08 to 0.67]	–3.3 (2.23) [–7.71 to 1.02]
Same-day cost; total 28-day cost; base-case estimates [with adjustment for covariates ^{a,b,c}				–5.75 (1.61) [–8.91 to –2.59]	–2.58 (1.61) [–5.73 to 0.58]	–3.17 (1.61) [–6.31 to 0.03]
<i>Sensitivity analyses:</i> Assuming GP consultation in UC is shorter (9.5 minutes) than estimated in the base case (11.7 minutes), with unit cost at £35.15, compared with base case of £43	33.79 (12.79)	34.91 (24.50)	38.25 (21.93)	1.38 (1.59) [–1.73 to 4.49]	4.57 (1.59) [1.46 to 7.68]	–3.17 (1.61) [–6.31 to 0.03]

a Regression covariates: age (by category), sex, site, practice size, deprivation (IMD 2010⁵⁴) by quintile, practice deprivation, cluster by practice.

b Regression methods include analyses using GLM models (see sensitivity analysis). GLM model with/without assumption of normally distributed data are similar and similar to those in base-case multilevel model.

c ICC (base-case multilevel model) = 0.033.

Exploratory analyses: contact duration for general practitioner contact (in practice) following triage intervention

Table 33 presents data from a sample of participants from whom data on duration of GP contact was collected following a triage intervention. These data were collected as part of a sample of data gathered in each of the participating GP practices, over approximately 2 days in each practice. The research question of interest was whether there is a difference in GP contact duration (in surgery), following a triage contact, compared with the estimated duration of 11.7 minutes, as used in the base-case cost data.³⁸ Sample data, collected over 2 days in each of the GP practices, indicate that consultation duration for GP contacts (in practice), on the same day as either GP- or nurse-led triage contacts, and are not significantly different from the estimated duration of 11.7 minutes. Data on these contacts (following triage), presented in Table 33, are from a relatively small sample size ($n = 244$; $n = 415$), therefore data are illustrative and are not able to support any conclusions.

Estimated total contact time for 'practice staff' over 28 days

Based on the mean time estimated for the delivery of the triage interventions, and the assumptions over duration of primary care contacts (see Table 33), it is estimated, based on mean number of contacts (reported in Table 12), that total contact time for primary care (practice delivered) contacts, over 28 days, is 21.86 minutes, 22.72 minutes and 27.5 minutes, for UC, GPT and NT, respectively. GP contact time over 28 days is 18.01 minutes, 17.6 minutes and 15.45 minutes, for UC, GPT and NT, respectively. If in UC contacts (only), the same-day request contact (GP in practice) is assumed to have a duration of 9.5 minutes (vs. base-case assumption of 11.7 minutes), the estimate for total contact time (minutes) for practice delivered contacts reduces to 18.86 minutes (vs. 21.86 minutes) and the estimate for the GP contact time, over 28 days, reduces to 16.01 minutes (vs. 18.01 minutes).

Economic analyses: cost–consequence analyses

The cost estimates indicate that there are no differences between interventions by type in the mean 28-day cost estimates (base case). These findings are robust to sensitivity analyses presented, with the exception of the assumption on the unit cost for same-day GP consults. Table 34 presents cost estimates alongside the other potentially relevant outcomes reported in this trial, in the form of a cost–consequence analysis, in which disaggregated cost data are presented alongside outcome data.

Although triage interventions include a higher number of contacts, including the index contact, the 28-day costs, and those costs more directly falling on practice-level budgets, are the same across all three interventions. There is no difference in safety outcomes, DNAs or self-report health status at 28-day follow-up. There is evidence of differences in self-reported patient satisfaction, with NT indicated as not performing as well as UC or GPT in self-reported patient satisfaction outcomes.

TABLE 33 Duration of face-to-face consultations following triage for ESTEEM patients

Duration of face-to-face consultations for ESTEEM patients		UC	GPT	NT
First management (index) GP consultation 'in practice': ^a mean (SD), n		9.5 (5.00), 5693	N/A	N/A
Duration of GP consultations 'after triage contact' on day of after index consultation request: ^b mean (SD), n		N/A	12.4 (7.12), 244	11.5 (6.43), 415
Estimate of mean total contact time (minutes), over 28 days, ^c for:	Practice-delivered contacts	21.86	22.72	27.5
	GP contacts	18.01	17.6	15.45

N/A, not applicable.

a Data collected via 'Clinician Form' within trial, all contacts/all practices.

b Data collected as part of sample data collection, over 2 days in each practice.

c Based on data for mean contacts presented in Table 12, and assumptions on duration of contacts reported in Table 17.

TABLE 34 Cost-consequences matrix, UC costs and consequences compared with triage interventions

Costs and consequences	UC	GPT	NT	Summary comparison
Number of contacts over 28 days (primary outcome)	1.91 (1.43)	2.66 (1.75)	2.82 (1.68)	UC fewer contacts (see Table 12)
Practice-level costs (£) (triage, GP, nurse), over 28 days	69.78 (44.97)	69.18 (55.02)	69.54 (50.33)	NS
Total 28-day costs (£) (primary outcome)	75.41 (57.19)	75.21 (65.45)	75.68 (63.09)	NS
<i>Safety outcomes:</i>				
Emergency 'unplanned' hospital admission, % (n) with at least one event, within 7 days of index contact	0.25% (14)	0.41% (21)	0.60% (33)	Few events; NS (see Table 14)
Safety outcomes: A&E attendance, % with at least one attendance within 28 days of index contact	3% (166)	3.3% (171)	2.9% (156)	Few events; NS (see Table 14)
EQ-5D health-state value (at follow-up)	0.77 (0.27)	0.77 (0.27)	0.78 (0.27)	NS (see Table 19)
Non-attended contacts (DINAs), % of total contacts (n)	n = 3281 2.5 (265)	n = 3246 1.8 (241)	n = 2925 1.8 (267)	–
<i>Patient experience outcomes:</i>				
Reporting difficulty (fairly or very difficult) in receiving prompt care, %	5.1 (n = 3816)	5.8 (n = 3727)	10.2 (n = 3461)	NT reporting greater level of difficulty vs. UC and vs. GPT
Reporting difficulty (fairly or very difficult) in seeing a GP or nurse, %	9.5% (n = 3883)	11.7 (n = 3620)	15 (n = 3501)	NT reporting greater level of difficulty vs. UC and vs. GPT
Reporting difficulty (fairly or very difficult) in getting medical help or advice for the problem they presented, %	4.6 (n = 3819)	4.7 (n = 3798)	7.4 (n = 3502)	NT reporting greater level of difficulty vs. UC and vs. GPT
Reporting care provided by GP surgery on that day was not very or not at all convenient, %	4.7 (n = 4072)	7 (n = 4022)	9.5 (n = 3689)	NT reporting greater level of difficulty vs. UC and vs. GPT
Reporting fairly or very dissatisfied with care received on that day, %	3.2 (n = 4093)	4.4 (n = 4034)	5.5 (n = 3704)	NT reporting greater level of difficulty vs. UC and vs. GPT

NS, not significant.

Process evaluation

The participating practices were diverse in size, location, staffing and their experience of triage. *Table 35* provides a matrix for the process evaluation analysis. Some of the data in this table are based on the subjective accounts of the small number of participants in each practice and should be viewed accordingly. The category 'communication' is based on whether staff participants perceived that they had been included in decision-making about entering the trial and other aspects of practice organisation that affected them. This also included whether they reported the presence or absence of effective formal and informal structures for communication, such as inclusive practice meetings, communal coffee breaks or other occasions in which practice issues could be discussed freely. 'Poor' clinical staff stability indicates recent loss of core medical staff members who had not been replaced and a greater than normal reliance on locum doctors. 'Previous appointment system' summarises staff members' accounts of the appointment system before the trial. 'Overview of triage experience' is based on the summary of participant responses for each practice produced as part of the framework analysis by the process evaluation researchers.

Structure of the results

The results are based on participants' accounts of their perceptions and experiences, and no attempt was made to validate what people said. This was because the aims of the process evaluation were to describe people's experiences and views. Other sections of this report describe quantitative data about the implementation of triage. The results section is structured as follows:

- how the trial was experienced
- how telephone triage interventions were implemented
- how triage was experienced
- acceptability of triage
- the extent to which the telephone triage interventions are seen to work or not work.

How the trial was experienced

One of the challenges of the process evaluation was disentangling experiences and opinions relating to taking part in the trial from those relating to the intervention itself. The former will influence the latter and vice versa in an iterative cycle. This became increasingly apparent during the data analysis. For this reason, experience of the trial will be discussed in some detail. The key themes explored are motivation to participate in the trial and communicative practices in relation to participation, attitudes towards the arm into which practices were randomised, implementation of the study protocol, and perceptions of trial-related workload.

Motivation to take part in the trial

Motivations for entering the trial, and whether these were consensual and effectively communicated within practices, affected how the trial intervention was experienced by practice staff. There was a range of motivations reported for entering the trial.

Research-related motives cited by informants included the fact that theirs was a research practice or a practice that valued research, wanting to know the answer to the research question, and the fact that this was seen to be 'a good trial'.

Challenges in current practice were also catalysts; practices were aware when their appointments system was not working well; they were not coping with increased demand, and increasing workloads and triage was seen as a way of addressing this.

I'm not sure really, just a sort of sense that it would be something that would involve everybody and there's been a lot of talk about, you know, some people finding it difficult to offer enough appointments and you know, that sort of workload issues and the staff having difficulty with the phones and it just seemed like it was an opportunity to try something out, which sort of shows willing to make some changes if changes proved you know, proved to work I suppose.

GP 05, practice 4 (NT)

TABLE 35 Overview of the process evaluation practices

Practice	Triage type	Size of practice and location	Communication	Personal lists	Clinical staff stability	Previous appointment system effectiveness	Triage practised before the trial	Perceived appointment reduction from triage (conversion)	Overview of reported triage experience
Practice 1	Nurse	Large practice, suburban	Very poor	Yes	Poor	Poor	No	No	Very negative
Practice 2	Nurse	Large practice, suburban	Good	No	Good	Good	Yes	No	Quite positive
Practice 3	Nurse	Small practice, rural	Good	No	Good	Good	Yes	Yes	Positive
Practice 4	Nurse	Large practice, suburban	Good	Yes	Good	Reasonable	Yes	Yes	Positive
Practice 5	GP	Small practice, rural	Mixed	Yes	Good	Poor	No	Mixed, variable	Mixed
Practice 6	GP	Large practice, urban	Poor	No	Good	Good	Yes	No	Mixed
Practice 7	GP	Large practice, suburban	Good	Yes	Good	Reasonable	Yes	Divided opinions	Mainly positive
Practice 8	GP	Small practice, urban	Very poor	Yes	Poor	Good	Yes	Yes	Positive

Introducing triage had often been mooted in practices but abandoned because of lack of consensus on how to establish and organise it, and the daunting enormity of the task. The ESTEEM trial was seen as a good way of trying out triage without internal disagreements on how to do it, and with a clear and immutable model and support from outside the practice. Part of the training for the ESTEEM trial was a presentation by the organisation 'Productive Primary Care', taking place after randomisation, and offering a clear model of efficacy of triage, which some practices found inspirational, although this presentation created some problems, which will be discussed below.

Three of the practices were excited by the fact that the trial was something that the whole practice could be engaged in, and its cohesive effect for practice staff was highly valued. However, staff in four of the practices reported lack of consensus about taking part, with some staff feeling that they had not been consulted or that their reservations had been glossed over. In two of the practices, both in which there was no consensus, the motivation to take part was reported to be financial. Often it was a combination of several of these factors, as well as the persuasive powers of one or two highly motivated individuals which drove the decision to take part. Half of the practices reported that the decision had been made with full consultation and consensus from all of the staff.

The range of pathways to entering the trial was echoed in a variety of communication styles about doing so, from 'top-down' authoritarian to democratic and consultative. In practice 1, a NT practice, as an example of the former, one of the partners was reported to have been motivated by the remuneration attached to the trial and himself referred to the 'status and resources' associated with being a research practice. Doctors here had been discussing the possibility of triage for some time without reaching agreement, and felt that this could be a way of introducing it without dissent because the protocol could not be argued with or changed. They discussed it among themselves but not with other practice staff, whose opinion was not a concern to them.

Yes and also people understand that we are a research practice and we are obliged to do so much research. And if we don't do so much research then our status and our resources that come with that will get chopped and if they don't play the game then that will get chopped so even if they don't necessarily think this is best thing since sliced bread, that's going to happen and if they are not on board then . . .

GP 04, practice 1 (NT)

Consequently, other staff, including the nurses who ended up performing the triaging, reported that they had not been included in the decision-making and felt that the decision had been imposed on them by the people least affected by its consequences. The consequent lack of 'ownership' of the trial impacted negatively on how triage was experienced and on its acceptability, and appeared to have a divisive effect on the practice as a whole.

No, I think the people who decided they wanted to do it they just thought the money was an attraction and they didn't think of the consequences. They didn't ask me who would do the appointments and how it would be, and it was just 'this is what we're doing'. And they don't listen to the people that know. It is a shame really.

Manager, practice 1 (NT)

In contrast, inclusive and democratic decision-making appeared to have a cohesive effect on practices and impacted positively on how the trial was experienced. Participating in the trial also provided an opportunity for good communication and developing good staff relationships. In practice 2, also a NT practice, the practice manager drove the decision to take part but the decision was taken after discussion with clinical and all other staff, and everyone in the practice was involved and motivated. Staff members were pleased to have a protocol to guide the introduction of triage and anticipated that the effect of the trial would be to bring the practice together. The nurses were 'very pleased' to be randomised to NT and the learning opportunity this presented.

And the other point I'd make is, I think it brought the practice together in a number of facets, in a number of different ways because we engaged almost the whole practice. Everybody knew it was going on; everybody was involved in it in some shape or form. And with two new nurses, again, I think the rigours of constraints of the ESTEEM trial helped to bring the process into place, that they wouldn't have learnt in any other way. And I think massaging the information and collecting it, was a lot of fun . . .

Practice manager, practice 2 (NT)

Staff at another NT practice, practice 3, reported mixed reactions to entering the trial. Doctors reported the motivation as responding to a perceived need for a better way of handling same-day appointment requests, and were also attracted by the remuneration. Administrative staff, however, felt that they had been 'sold' the idea rather than fully consulted. It is worth noting that neither of the two NT practices that reported being financially motivated to enter the trial provided any additional nurse capacity to absorb the increased workload caused to nurses, as recommended by the research team. Triage was added to nurses' already busy working day. Despite not feeling fully included in the decision-making, staff at practice 3 also reported a cohesive effect on the practice, and excitement about trying something new, perhaps reflecting the excellent relationships in this practice, as observed by the researcher. The practice manager described a mood of excitement at taking part and pleasure that patients might benefit:

But you know, also we were pleased that we were doing it in one sense because it's interesting to do something new and to have a different challenge. And potentially if it was going to have a good effect on the patient and doctor appointments then yeah, we were quite excited I suppose from that point of view.

Practice manager, practice 3 (NT)

In some of the practices there was no consensus among medical staff about entering the trial. Practice 6, a GPT practice, was one of these. Some doctors reported that pressure had been applied by one of the partners to agree to it. Communication in this practice did not appear to be optimal and informants' accounts of a number of issues, including trial participation, were at variance with each other.

I think that we were pressurised by one partner, yes . . . And when I raised concerns it was read as objections and I was pushed down.

GP 03, practice 6 (GPT)

Motivation to enter the trial: key themes

- Reasons for entering the trial, how the decision was made and by whom, and how the decision was communicated all affected how it was experienced.
- Motivations reported were an interest in research, the high quality of the trial, a perceived need to change the appointment system, an interest in trying triage to manage demand, the fact that a clear model of how to introduce triage was provided, and the financial incentive.
- There was a range of communication styles in relation to the decision to take part, from 'top-down' to democratic or consultative.
- There was not always consensus within practices about taking part, even among doctors.
- A number of practices regarded participating in the trial as an exciting challenge.
- Several practices described trial participation as having a cohesive effect on the practice and improving communication.

Attitudes to randomisation

How the trial was experienced by staff was related, at least in part, to their attitudes to the arm into which they were randomised. In two of the NT practices – practices 2 and 3 – staff said that they were pleased to have been randomised to NT and were excited about the learning opportunities NT presented.

Well we were very, you know pleased to be part of research and we liked getting the nurse-led triage because we felt it'd be a good learning curve.

Nurse, practice 2 (NT)

In practice 3, one of the GPs reported being relieved that they had not been randomised to GPT, which was perceived as being more effort for doctors. In practice 1, on the other hand, none of the clinical staff, nurses or doctors was pleased with being allocated to NT. Both of the practice 4 doctors interviewed also had reservations about NT, believing that triage should be done by the most highly trained clinicians, i.e. doctors, especially if they could be triaging their own patients.

I think most people recognise that triage is probably the most difficult thing to do and that it often works best when the most highly trained people do it. And that probably is why personally I feel as though it might be better if I did it for my own patients, which is what I effectively thought I was doing before. And I think that also, you know, that algorithmic medicine is pretty awful and not really as clever as it looks.

GP 05, practice 4 (NT)

In two of the GPT practices (practices 7 and 8) a few staff, but not all, expressed disappointment that they had not been randomised to NT, but most were pleased with GPT, as were those in practice 5.

Signing up for the ESTEEM trial was a gamble really. Because we may have been randomised into one of the other two levels and I tell you what, if we'd been randomised into NT, we'd have been up the creek without a paddle [laughter].

GP 04, practice 5 (GPT)

Reasons given included nurses not being keen and the perception that GPs could do more on the phone. In contrast, practice 6 had a very experienced nurse practitioner who for the last 10 years had triaged, face to face, all patients requesting a same-day doctor's appointment when the appointments had run out. Randomisation to GPT represented a lack of fit with existing experience and expertise within the practice, wasting the resource of an experienced triaging nurse practitioner. Their duty doctor system also caused additional workload for GPs through lost consultation time.

Attitudes to randomisation: key themes

- Of NT practices, two were pleased with their randomisation arm and two would have preferred GPT.
- There was a perception that GPT was a more appropriate model than NT.
- There was a distrust of CDSS.
- Of GPT practices, three were pleased with their randomisation arm and one was not.
- The GPT practice that was displeased had had a highly effective nurse face-to-face triage system in place prior to the trial and no problems managing same-day appointment requests.

Implementation of the study protocol

Although most of the practices had no problem with understanding and adhering to the trial protocol, the process evaluation did reveal some areas of confusion and misinterpretation and infidelities to the protocol. Practice 2, for example, disclosed that they forgot to record all of the phone calls during the run-in audit and remained unsure of the suitability of some calls for the trial, for example children with breathing difficulties.

The few glitches are basically maybe not understanding possibly how it ran 100% from our point of view and being told different things, like if we booked in for tomorrow it wasn't ESTEEM and then it was suddenly ESTEEM and why did it change? . . . The other problem I have – and I'm still not quite sure how we get round that – is that if people . . . people are more inclined now to come down to

the desk at nine o'clock in the morning. Now the triage appointments don't start until . . . I can't remember, at least ten. Now I have a problem saying 'Can you go home, doctor will 'phone you and then you might be coming back' so it just seems a little bit . . .

Receptionist, practice 2 (NT)

Doctors reported forgetting to ask for consent for notes review.

The main problem is remembering to ask their permission, so the last question at the end. And no matter how long you do it for, it does help if it's the second day in a row but we all work part time, and you often find that you've forgotten to ask them the question. Calling them back a second time to say 'by the way I forgot to ask' that is quite difficult. I did that a few times but it gets to the point that we're so busy that all of us stopped doing it.

GP 05, practice 5 (GPT)

Similarly, a patient interviewee from practice 6 reported that she had presented at the desk and been told to go home and wait for a phone call; this same patient denied having requested a same-day appointment in the first place.

Practice 8 also sent home patients who presented at the desk requesting same-day appointments, mainly because receptionists had no appointments to give.

I: So, okay, can I just make sure I've got this straight? So if somebody turns up and wants an appointment that day they can't have it, they have to go home and make the phone call?

R: Yeah, they do really. And we don't like doing that, but it's not a drop-in centre.

Practice manager, practice 8 (GPT)

Practice 5 had previously attempted and abandoned the 'Advanced Access' model promoted by this organisation. There was similar confusion in practice 7.

Because we were doing a triage system and as clinicians we're all trained in triage and using our skills effectively to triage patients. But then the trial was saying no, do today's work today and if they need to be seen book them in. So are we just a booking service then?

GP 01, practice 7 (GPT)

Implementation of the study protocol: key themes

- Most practices had no problem understanding and implementing the study protocol.
- One practice was mistakenly including patients who presented in person to the reception desk to request a same-day appointment. They would have been informed during training that such patients were ineligible for ESTEEM, and it was likely that this had subsequently been forgotten and then not communicated among practice staff.
- Practices at which receptionists had no appointments to give were also sending home people wanting a same-day appointment, telling them to phone in.
- One practice forgot to record all of the phone calls during run-in.
- There was some uncertainty in two practices about patient inclusion and exclusion criteria.
- Doctors forgot to ask for consent for notes review.
- One practice abandoned the trial on busy days.
- One practice believed the ESTEEM protocol required practices do 'today's work on the day', i.e. confused it with the Advanced Access principles promoted by Productive Primary Care.

Perceptions of trial-related workload

Staff in some practices reported that the trial added considerably to their workload, others that workload had decreased, with disagreement about this both within practices and within occupational groups, especially doctors. In practice 1, a practice characterised by poor communication and no prior triage experience, perceptions of the trial were almost universally negative. A receptionist described completing the trial log sheets for the integrity checks as 'just more work' and 'time-consuming', and reported that receptionists hated doing it. One of the nurses complained that the Clinician Forms were badly designed, 'obviously not designed by a nurse'.

... and there's not much space between that bit and that bit, between the patient computer ID and that blue line at the top there, yeah. Yeah it's trying to get quarts into pint pots and get everything on to the sheet.

Nurse, practice 1 (NT)

A GP found the forms 'straightforward and simple' but confessed to often forgetting to complete them. Doing so added to the consultation time.

Obviously it adds length to the consultation which is a bit of a nuisance. And we haven't made any provision for that, so the days that I've had a lot of patients going through the study, I've run late then I would redo them [laughter].

GP 02, practice 1 (NT)

Collating the data was also described by the deputy practice manager as time-consuming; the system of their practice did not allow for bulk Read Coding of patient records to flag that they were part of the trial. Even the visits by the ESTEEM researcher were perceived as something of a time burden. Doctors were in disagreement about whether the trial had increased or decreased their workload, perhaps reflecting real workload disparities and different levels of triaging effectiveness among doctors.

Practice 2 had extensive prior telephone triage experience but had underestimated the requirements of the trial, which had greatly increased their workload. They reported a great deal of tension while it was taking place, particularly in reception. The paperwork was hard to keep up with, although the practice manager found collecting and collating the data 'a lot of fun'. However, doctors reported feeling that they had more control over their workload. They felt that the training had been good and the run-in excellent, and appreciated the approachability of the ESTEEM researcher.

Practice 3 also had prior triage experience and reported no problems with the protocol, which they found clear, or the paperwork, and reported that the training had made everything easy. Nurses' workload was reported to have increased enormously, however.

One informant felt strongly that the initial training was much too complicated and had fostered anxiety and confusion:

Yeah, I just thought, 'Oh I really can't be bothered with this', you know, 'we've got enough to do without this'. And I'll tell you why, it's because the chap who came to show us it, he went into too much detail for us. We didn't need to know all that stuff that he went on about and it made us all as receptionists think, 'Oh my crikey'. Whereas all they really had to say to us was 'We're trying out a new system, this is what you've got to say to patients' ... and it would have been fine. But he made it so complicated. He went through stuff that we didn't need to know about, about what the doctors have got to do, what ... so to me that first talk that we had, we shouldn't have sat through that, because we all felt the same.

Receptionist, practice 4 (NT)

The practice manager in practice 5, which had no prior triage experience, found that the Productive Primary Care session was pivotal in making the trial a success and in getting more sceptical staff enthused about it. However, the GPs in this practice reported extremely high stress levels as triage calls were added to their ever-increasing workload, although two of the doctors conceded that they did feel more in control of their workload despite perceptions of its increased volume.

I think it made life quite difficult, and it is still quite difficult at the moment. It has made it more stressful, it's made it busier, it's made the work more intense, erm, so yes, I mean looking at the benefits from it, there must be some and I'm sure we'll get to that later. But yes, being honest it has been quite difficult.

GP 05, practice 5 (GPT)

Female doctors were under particular pressure, with more patients wanting to consult them. Practice 5 was characterised by poor communication between management and doctors and between management and the reception team.

Doctors' workload in practice 6, a practice with prior triage experience, decreased under the trial or it increased, depending on doctors' accounts. One GP reported having to work extra sessions because of the surgery time lost to the duty triage system, whereas two other doctors claimed a decreased workload as patients were dealt with on the phone.

On the mornings when we're doing the triage I think we are working less intensely than we used to.

GP 02, practice 6 (GPT)

Paperwork was checked and completed every evening by the reception manager, which increased her workload but took the pressure off other staff. Training and support from the ESTEEM researcher was reported to have been good.

In practice 7, which had prior triage experience, response to the trial was mainly positive, although there had been dissent among doctors about it, and some GPs felt that the quality of triage was poor because of the volume of calls. The trial paperwork had not been experienced as a problem. The ESTEEM researcher was regarded as approachable and easy to contact, and the run-in period was considered invaluable. However, a perverse effect of the trial protocol was noted. Patients who would previously have been given an appointment to see a nurse were now put on the triage list if they had asked to be seen that day, thus increasing triaging doctors' workload.

Because we would allocate, you know, if somebody said they had an earache or a sore throat or a chesty cough we would have just automatically allocated them to the practice nurse, nurse practitioner but because of the trial everything had to go on the doctor, you know, the doctor's triage. So people were phoning up who would normally have spoken to, say, T [nurse] or L [nurse] and we were saying to them, no, you've got to go on to the doctor's list so they didn't quite understand that because I think they felt that they were wasting the doctor's time when they could have been seen that, you know, the nurse because it was something that they've seen in the past.

Receptionist, practice 7 (GPT)

Finally, practice 8 had a mixed experience of the trial, with considerable disparity of opinion on how it was experienced and quite a lot of negativity. This practice had previous triage experience but poor communication, with an absence of both formal and informal opportunities for sharing knowledge, information and concerns, and there were major staff continuity problems.

The extra paperwork was described as an additional burden that hindered the ability to manage the workload. This practice did not consider it had received enough support from the ESTEEM team, and had not received the promised feedback from the audit of demand. Setting up had been problematic owing to

the practice starting the trial using a software solution developed during the trial and later abandoned (for details, see *Strengths and limitations*). The software was designed to automatically apply a study-specific Read Code to the electronic records of ESTEEM patients by way of a short keyboard command typed by the receptionist receiving the patient's telephone call. The software required a macro to be set up on all of the computers used by the doctors, which, at this practice, was problematic due to the number of doctors working in different rooms. Staff members were consequently not confident that everything had been set up properly.

Trial-related workload: key themes

- Trial paperwork was perceived as burdensome in most of the practices.
- Trial paperwork and data collation was time-consuming and hard to keep up with.
- The paperwork added to consultation times, sometimes causing GPs to run late.
- Staff sometimes forgot to complete the paperwork.
- Deciding how patients were to be managed under the trial needed to be discussed.
- Visits from the ESTEEM researcher were sometimes considered time-consuming.
- IT issues had caused additional work.

How the telephone triage interventions are implemented in different practice settings

Practices were advised by the study team to triage in set blocks of time (i.e. whole sessions) each day, and all process evaluation practices did so. Most practices triaged in the mornings, although some had both morning and afternoon triage sessions. The variation in arrangements for triage is notable, although this was mostly within the limits permitted by the manual.

Nurse-led telephone triage practices

Two of the NT practices triaged only in the mornings, whereas the other two also included afternoon triage sessions. In practice 1, a large practice in a small town with a personal list system but poor staff continuity, two nurses performed triage in the mornings, although usually only one at a time. There were about 30 triage calls per day. No extra nursing capacity was provided, and nurses had to work extra hours unpaid.

Practice 2, a large practice in a small town with no personal list system had three nurses triaging each morning, a number that had needed to be increased from two due to demand. Two extra nurses were appointed and existing nurse hours were also increased.

Practice 3, a small rural practice with no personal list system, did not provide additional nursing capacity and nurses worked extra hours unpaid to cope with the additional workload. Nurses triaged alone for an hour and a half in the morning, and the same in the afternoon. Because one of the doctors finished at 4 pm the nurse would come in early in her own time on that doctor's duty day to fit triaged patients into his surgery. Timing was an issue in this practice, as in others, because doctors' clinics coincided with triage times, which sometimes did not leave patients enough time to get in to see the doctor, especially those dependent on the bus. This could leave appointment slots unfilled and doctors having to work late. The system was reported as not working well on Friday afternoons when people had to be seen before the weekend.

The fourth NT practice, a large urban practice with a personal list system, appointed two additional nurses for the trial. Initially, triaging took place all day, although demand was found to be lower in the afternoons, and some days triaging took place only in the mornings. GPs changed their surgery times to later in the day to allow triaged patients who needed an appointment time to get to the surgery.

Nurse-led telephone triage was a challenge for practices with a personal list system at which patients needing an appointment were triaged into the next available slot.

I think the things which we've found awkward are that we are a practice where we encourage in favour of personal lists and we agreed to this, we understand it's part of the trial, so it's not a surprise, but nevertheless, you know, we've had days when say four of us are here and each of us are seeing each other's patients. Now I'm seeing someone of [GP name] and he's seen some of mine, which is not very helpful . . . And that's annoying. I think we wouldn't want that to continue. That would tend to grate and would break down a good doctor–patient relationship after a while.

GP 04, practice 4 (NT)

Implementation of nurse-led triage: key themes

- Two practices triaged only in the morning, whereas the other two also did so in the afternoons.
- The number of nurses triaging at the same time varied from one to three.
- Two of the practices provided extra nursing capacity for the trial. Two did not.
- Timing of triage in relation to surgery time could be a difficulty, and doctors and nurses had to adjust their hours to accommodate this.
- Triaging into 'next available slot' is a challenge for personal lists.

General practitioner-led triage practices

Although the ESTEEM intervention manual prescribed many aspects of how GPT was to be implemented, it was essential that practices had some flexibility to design the intervention based on their local context. Consequently, there was marked variation in the ways in which GPT was implemented between the practices. There were two main organisational structures for GPT: a duty triage doctor system, for which GPs took it in turns to provide a dedicated triage service for patients, and an integrated system, for which triage (usually of doctors' own patients) was combined with normal surgery. However, practices seemed to have evolved systems that suited them, often involving a combination of the two organisational structures, especially in practices operating personal lists. Two practices (practices 5 and 7) combined a duty doctor system with presurgery triage, in which non-duty doctors picked their own patients off the triage list and called them back before the duty doctor took over triaging. When an appointment was necessary, patients were triaged to an appointment with the triaging duty doctor, to their 'own' doctor or to the first available appointment, depending on the practice. This is arguably a third organisational pathway for GPT, being neither purely a duty system (because all doctors triaged their own patients rather than just one doctor triaging all patients on the list) nor a combined system (being performed outside surgery time).

Both systems presented issues of workload disparity:

- The duty triage system meant that doctors on triage duty on busy mornings, especially Mondays, struggled to get through the triage list.
- Disparities in the integrated system resulted from extra demand falling on certain doctors. For example, some female doctors faced extra demand from female patients with certain conditions; some doctors were 'more popular' than others, whereas other GPs had a 'more challenging' or 'demanding' patient list.

Our difference in terms of numbers of phone calls is immense. If you work it out per session per year, some doctors are . . . the number of phone calls you do per session per year may be in the region of 250 phone calls per year. Some of us do 2000 phone calls per session per year . . . I was taken on as an extra partner here so naturally anybody who wasn't happy with their own GP immediately came to me. So by definition I had a motley crew.

GP 06, practice 7 (GPT)

In practice 5, a small rural practice with a personal list system, all the doctors triaged their own patients for an hour and a half in the mornings followed by their surgery, at which point a duty triage doctor system came into operation. Although this worked well, the four minutes per patient scheduled for triage in this practice caused some doctors to get behind with the triage list and, consequently, their surgery.

Receptionists in this practice also had extra appointments to offer, so that walk-in patients and those unsuitable for triage could be given appointments; this was not the case in all practices. Advance appointments were also available (following a failed attempt to follow the influential Productive Primary Care 'Advanced Access'-based model of 'do today's work today', which had impressed many of the practices). Surgery times were adjusted to accommodate triaged patients. There was still a flaw in the system, in that people phoning in the late end of the morning session could not be phoned back until 3 pm because triage sessions were immediately followed by surgeries, then lunch, then home visits. In addition to the long wait for ring-back that this created, if people were phoned at 3 pm there was sometimes no time for them to get to the surgery for their appointment, especially if they travelled by bus. Doctors did stay later in the evening to accommodate this but still found that their appointment slots were all gone before afternoon triage was finished.

Practice 6, a large urban practice with no personal list system, operated a duty triage system with one doctor on duty at a time on a rota system. The duty doctor later saw, where possible, the patients he/she had triaged, which operationalised one of the main perceived advantages of GPT – that the consultation has been started on the phone, which shortens the appointment time, and the patient's history does not have to be taken more than once. It also addressed the continuity problems that can be inherent in triage. However, practice 6 had used a uniquely effective system for same-day appointment requests prior to the trial, in which patients were triaged face to face by the nurse practitioner. GPT, according to the ESTEEM protocol, did not allow for patients wanting urgent appointments to be seen by the nurse practitioner. This increased the doctors' workload at this practice so that extra hours had to be worked and locums employed.

Practice 8 is a small urban practice with personal lists but very poor staff continuity. Their triage system was a duty doctor one, although this practice chose to include other things on the 'ring-back' list, such as queries and requests for results, resulting in 'pages and pages' of ring-backs to do, described as a particular challenge for the doctor on duty on Mondays.

Implementation of general practitioner-led triage: key themes

- There were two main systems: a duty triage doctor system, in which doctors triaged patients but did not see them, and an integrated system, in which triage was combined with surgery.
- Two personal list practices adapted the duty doctor system so that all doctors triaged their own patients from the list before surgery and the duty system took over.
- Both duty doctor and integrated systems produced workload disparities between GPs.
- The duty doctor system meant that surgery appointments were lost while doctors were triaging, increasing workload for the other doctors.
- The integrated system in which triage was combined with surgery produced work overload and the risk of running late.
- Timing of triage in relation to appointments required adjustment of working hours.
- Some practices triaged patients needing appointments to the triage duty doctor, some to the patient's own doctor and some to the next available appointment slot.
- The system where triaged patients were not given appointments with their own doctor was a challenge for some doctors in practices with personal lists.

How triage was experienced

It is evident from the previous sections that how triage was experienced in the trial practices was mediated by how the trial itself was experienced, and those factors combined to influence how acceptable telephone triage was felt to be. Introducing a major organisational change under trial conditions is not the same as introducing it in real life, when a multiplicity of factors such as the organisational structure and culture, staff skills and experience, perceived need, staffing issues, geographical location, patient demographics and so on will influence the model of triage chosen, how it will be organised, who will perform it and under what circumstances and so on. Being randomised into an intervention arm, which may not necessarily be viewed as the most appropriate one for that particular practice, and working to an immutable trial protocol under externally given time constraints, are not necessarily the optimal conditions for effective change management to take place. The following report of how triage was experienced should be read in light of this.

No clear patterns emerged from the data on staff experience of triage, either within practices, within occupational groups or in relation to whether GPT or NT was being discussed. There were many positive and negative experiences reported and a complexity of opinions on what worked, why and for whom, sometimes from the same informant. However, there were several key themes to emerge of factors that affected the experience of triage and whether it was found to be acceptable. Many of these echoed the experience of the trial itself and were underpinned by the same aspects of practice organisation, interrelationships and culture. These included:

- the quality of relationships, consultation and communication within the practice
- whether all stakeholders felt fully consulted and informed about the introduction and organisation of triage
- how it was decided to enter the trial and by whom
- how triage was organised
- whether there was adequate staff capacity to perform triage effectively
- whether triage was perceived to adversely affect continuity and clinician–patient relationships
- how busy the practice was
- doctors' workload volume
- whether staff felt in control of their workload
- how staff perceived their role, ideas about what constitutes good clinical practice and good care, safety and risk
- the organisation and effectiveness of the appointment system before the trial
- perceptions of the intervention arm into which they were randomised.

In addition, all staff members interviewed were individuals, whose personalities, relationships, attitude to change, perception of their role and a host of other factors are all different. For triage itself, how it was organised, the 'model of the patient' within the practice culture, patient expectations and behaviour, perceived need for change, and many other 'soft' factors that were difficult to evaluate underpinned attitudes and experiences.

What worked well for whom? The benefits and advantages of triage

The positive aspects of triage reported by practice staff have been organised into the four main themes: (1) effects on work and workload, (2) benefits for individuals, (3) advantages for the practice as a whole and (4) benefits for patients as perceived by staff. A further theme is patients' positive experiences and views (5).

Effects on work and workloads

Practice staff often reported that they had previously felt that they were battling against a tidal wave of increased and increasing workload demand, and beneficial effects of triage on workload featured strongly in their accounts. GPs in both intervention arms reported experiencing less pressure and stress, and having more flexibility and control over their work. When NT was working well, appointments were more appropriate, as fewer minor conditions were referred and work pressure was dramatically reduced. A GP informant in practice 3, for example, reported that appointments had been reduced by half.

In GPT, face-to-face consultation time was felt to have been reduced as doctors had been able to begin the consultation on the phone.

This was particularly felt to be the case in practices where doctors combined triage and surgery time, usually those with a personal list system such as practices 5 and 7, where triaging doctors booked patients into appointments with themselves. Some doctors found telephone work less taxing than face-to-face consultation.

Effects on workloads: key themes

- Appointments were reduced by up to half (only one practice reported this).
- NT referred appointments were appropriate.
- GPT was felt to reduce face-to-face consultation times, especially in practices with a combined GPT system with personal lists.
- There were fewer interruptions to GPs from administrative staff.
- There was no more 'morning rush' of phone calls. Receptionists were more in control of their workload.

Benefits for members of staff

Doctors reported the benefits of triage as relating to aspects of their workload, as described above, and as partners, those for the practice as a whole, which are described below. For triaging nurses there was the bonus of learning new skills, having new responsibilities and the opportunity to form relationships with a wider range of patients, all of which increased their job satisfaction. The CDSS used by nurses for triaging was described positively by a number of them. It was found to be excellent for triage training and for enhancing learning and skills in general, even for those with prior triage experience. It was useful for triggering what questions to ask, and to help understand how patients present, thus increasing knowledge and confidence.

I would say it's good . . . it increases . . . the person who's triaging, it increases their knowledge base definitely I would say and increases their overall nursing experience really, you know, it gives you knowledge. So it gives you . . . you're a bit more empowered when you see patients then face to face. It gives you confidence as well chatting to the patients because after you've done it for a bit it gives you that little bit more confidence to talk to people and . . . yeah, and the layout was very good. It was very quick and, you know, you could get on to things very quickly.

Nurse, practice 3 (NT)

Some nurses commented that the layout of the CDSS was good; relevant sections could be accessed quickly, it was comprehensive, and nothing that they were presented with was not covered. It was particularly good on self care, an invaluable tool for teaching patients the self-management skills and the judgement to know if their condition really was urgent, that was held to be key to managing demand.

I personally think it's been a lot easier, so I think they're getting, they're getting more used to it and trying home care a bit more now than perhaps they were. Because I'm sure all surgeries are the same, you have people whose name crops up and up and up. And at the beginning of the triage system the same names were coming, always being booked in. And I'm guessing that perhaps they weren't getting the doctor's appointments they were after, and have learnt now actually is this really urgent, can it wait, can I do this and this first, and book an appointment in two days' time?

Receptionist, practice 3 (NT)

Receptionists were delighted with the triage system, which made their job easier and less stressful, as they no longer had to attempt to find rare appointments, make judgements about urgency or deal with irate or distraught patients; they just had to put them on the triage list.

It is brilliant, it's absolutely brilliant. Because I think the worst part of my job and any of the receptionists that you ask, has always been that there's never enough appointments to go around for the people that want them.

Receptionist, practice 5 (GPT)

Benefits for members of staff: key themes

- Nurses learned new skills and had more responsibility.
- Nurses generally enjoyed triaging.
- There were enough appointments.
- Patient demand for appointments was under control.
- CDSS was a good training tool for nurses.
- Receptionists' job was much easier.
- Doctors were more in control of their workload.
- Doctors' workload was reduced.

Benefits for the practice

Triage was seen as benefiting practices by providing a more rational and efficient way of allocating appointments, in contrast with the 'informal triage by receptionists' that had been happening before, and appointments were more appropriate, with fewer patients with trivial conditions taking appointments they did not need.

I think it's just been really good for the surgery to see a different way of maybe booking appointments. A much more efficient way of doing the appointments instead of leaving it to us. Because you know, we have no medical knowledge. We're not able to say to someone that's urgent and that's not, whereas they are. So I think from that point of view the patient will get what they need much quicker than they would have done.

Receptionist, practice 7 (GPT)

Triage was appreciated as being an equitable and fair system, with appointments allocated according to need, in which there were always appointments available. It was regarded as the best possible use of resources and was cost-effective because the number of appointments appeared to have been reduced. In practices in both intervention arms staff expressed surprise at how few patients actually needed to be seen. Access was improved in practices where it had previously been poor, and, in those where accessibility had been good before, appointments were reduced by the ability of triaging nurses to provide reassurance and teach self-management skills and the ability of triaging doctors to deal with conditions over the phone. There was a suggestion that good access can perpetuate rising demand by creating its own expectations, and that teaching self-management was a sustainable way of reducing demand as patients began to learn to manage their conditions themselves and to know when an appointment was not necessary, and so would become less likely to request one in the future. Practices therefore had more control over demand. Using CDSS was considered excellent for promoting self-management.

Yeah, it was interesting actually. Some were . . . yeah, some . . . after talking at length to them on the 'phone they actually then realised actually they're not quite as bad as what they think they are and that they are . . . a few measures they could do themselves and then you make an appointment for later in the week and then say, 'Actually if you just try these different measures,' like steam inhalation and things like that, 'you could actually make yourself feel a lot better.' And then a lot of people don't, they don't take any painkillers. So they'll say, 'Oh I've gone over on my foot today, something's

cracked.’ ‘Have you tried . . . have you taken any paracetamol? Rested it?’ ‘No, just want to see a doctor.’ You know, ‘Well try these measures first’.

Nurse, practice 3 (NT)

Staff liked the triage system, against managers' expectations in some cases, which made for a happier atmosphere in surgeries. Several practices reported how much they had enjoyed the sense of cohesion and common purpose the triage trial had produced, and the improved communication and better relationships that came with it. It was an opportunity to be innovative and dynamic, which had a positive impact on practice cultures.

Relationships with patients were also widely reported to have been improved, not only by triaging nurses, who were given the opportunity to get to know patients better and interact with a wider range of patients than before, but also by reception teams, who reported that a lot of the interactional tension and stress had vanished from their encounters with patients.

Benefits for practices: key themes

- Practices could provide a more rational and equitable appointment system. There was more control over demand.
- It appeared to be cost-effective and a good use of resources.
- Appointments were felt to have been reduced.
- There is potential for a sustainable reduction in demand as patients are taught self-care skills.
- Introducing triage gave practices a sense of cohesion and common purpose.
- An opportunity to be dynamic and introduce change.
- Relationships with patients were improved.

Staff perceptions of benefits for patients

Staff reported that patients liked the triage system once they, and staff, had had the chance to get used to it. Like all change, it took time to become embedded.

There were a number of perceived benefits for patients, not least of which was the perception that they were receiving a better quality of care and faster and more equitable access to clinical expertise when the sickest people get seen soonest and everybody gets 'the right help at the right time'.

I think most people were reassured by the, that thing about whoever gets, this is really about making sure that the people are sickest get seen soonest.

Receptionist, practice 7 (GPT)

Patients no longer had to exaggerate symptoms or 'pitch' in order to get an appointment because there were more appointments available and they were therefore more relaxed and less stressed and anxious, which improved their relationship with the practice.

Well, appointment basis, there are, we can always give something now, whereas before we were having to say, even all our emergencies have gone. Because they got the patients, sorry, the patients got used to, if they couldn't get an appointment they'd say, it was an emergency, because they knew we had to give them one.

Receptionist, practice 3 (NT)

In practices whose appointment system had not worked well for patients before the trial, the very fact of being able to speak to a clinician was a huge benefit of the triage system.

Patients could obtain immediate reassurance, and it was felt that the opportunity to learn self-management skills was empowering for patients, enabling them to take control of their own health.

Teaching self-management seemed to be a particular bonus of NT, supported by the CDSS; none of the GPT practices mentioned it. Such knowledge was confidence building for patients, who also got to learn what was and was not an emergency, which could pre-empt anxiety as well as unnecessary appointment requests. Patients were reported to be happy to talk to a nurse and some preferred it, being aware that doctors' time was valuable and being reluctant to 'waste' it.

And I think they're just grateful. And they often say I think they, sometimes they quite like the nurse ringing them back. I think because they think it's perhaps a bit silly speaking to the doctor about it but actually they don't feel quite so uncomfortable speaking to the nurse, you know I think that's been a positive side of it.

Nurse, practice 4 (NT)

Some people sometimes were reported to have found it easier and less embarrassing to explain things to nurses. The triage call with a nurse could itself be therapeutic. Patients were frequently grateful not to have to come to the surgery; it saved them time and effort. For some people, being called back by a doctor or nurse represented individualised care, which made them feel cherished.

Staff perceptions of benefits for patients: key themes

- Patients liked the system after a period of adjustment.
- Fairer. The sickest patients got seen soonest.
- Patients no longer had to claim it was an emergency to get an appointment. Less stressful for patients so they were more relaxed with staff.
- Patients sometimes actually preferred talking to a nurse, which could be less embarrassing than talking to a doctor.
- The ring-back from the nurse could itself be therapeutic.
- Being called back made patients feel cared for.
- Relationships with patients were improved.

Patients' accounts of their experience of triage

Patients at NT practices reported being 'pleased and surprised' at getting a same-day appointment. Others were delighted not to have to go to the surgery and to have their problems resolved with a phone call. The wait to be called back was of an acceptable length; from 10 minutes to 1 hour were the times cited. Patients were aware that other people were getting appointments that were not really necessary, and were glad that a system had been introduced to 'weed out the time-wasters', making appointments available for those who needed them most. However, they also felt that triage could ensure that they themselves were not going to see the doctor for 'something silly'. Patients appreciated getting to speak to someone with medical knowledge straight away.

So being able to get medical trained advice straight away, I think is a really good idea and plus if anybody's got something personal they may not want to discuss with a receptionist, then, you know.

Patient 03, practice 3 (NT)

There were references to relief at not having to disclose personal problems to receptionists, particularly from patients of practice 3, whose receptionists were not permitted to elicit the reason for the appointment request. Patients felt valued, supported and reassured by the fact that someone would call them back. Nurses were felt to offer the benefits of a combination of common sense, sympathy and medical expertise, enabling patients to talk about how they were feeling. Patients at GPT practices also felt that the system was good for weeding out 'time-wasters', and that it made good sense for the doctor, rather than the patient, to make the decision on whether an appointment was necessary. Many were glad not to have to come in to see the doctor in person if they did not have to, feeling that it prevented them from wasting the doctor's time and also saved them time themselves, whereas others described the system as good as it enabled them to get an appointment. One patient said that the system worked but felt he would rather have 'showed' his condition to the doctor. Another patient described triage as 'a responsive

system', which was efficient and convenient. Patients had not had to wait too long to be rung back. They said they were happy to discuss their problem with a doctor on the phone, and that it was reassuring and gave them confidence.

Patients' accounts of their experience of triage: key themes

- Patients were pleased and surprised to get a doctor's appointment.
- Patients were glad not to have to go into the surgery.
- The system was fairer and eliminated 'time-wasters'.
- Triage ensured that they would not see the doctor for something trivial.
- Patients liked the fast access to clinical advice.
- Being rung back made patients feel valued.
- The wait to be rung back was of an acceptable length.
- Patients appreciated nurses' combination of 'common sense' and clinical knowledge.
- Patients were glad that doctors, rather than receptionists, made the decision about whether they needed an appointment.
- Patients were glad not to have to tell receptionists their symptoms.

What worked less well and for whom: challenges of triage

Effects on work and workloads

GPT In contrast with the previous section, staff whose experiences are reported here found that the triage systems had an adverse effect on their work and workloads. GPT, as we have discussed, was broadly organised in three different ways: (1) the duty triage doctor system, in which triage is performed by one or more doctors on a rota system for a period during which this is their sole duty; (2) the integrated system, in which doctors triaged in some combination with their surgeries; and the hybrid of the two, adopted by two of the GPT practices, in which doctors triaged their own patients before surgery, after which a duty triage doctor took over. Each system had an impact on workload. In a duty GPT system, there are particularly busy days, especially Mondays, and doctors triaging then can be overloaded. In addition, there are different doctor styles, preferences and levels of experience of triage and telephone consultations. GPT, as with any activity involving human behaviour, is not a consistent phenomenon; as this practice manager puts it, all doctors are different:

All doctors are different, aren't they, and it's trying to get them all regimented doing the same thing. So some will do it one way and then some don't really want to do it that way, and it's just knowing their little ways really.

Practice manager, practice 8 (GPT)

Some doctors are perfectly comfortable working on the phone, but for those who are not, perhaps through a lack of confidence and experience of triaging and of telephone consultations in general, triaging can be experienced as laborious, stressful and slow. This could result in subversion of the triage process.

We've got one GP close to retirement and he just didn't really triage at all. He was just a glorified receptionist which was very frustrating. He'd just speak to them and book them in anyway, you know, so [sigh] . . .

Practice manager, practice 7 (GPT)

Some doctors just could not keep up the requisite pace. One practice manager talked despairingly of such a doctor, whose attempts at triaging were as long as full telephone consultations, leaving the whole practice behind with work.

In practices in which there is not the full quota of doctors – such as practice 8, which had a duty doctor system – doctors had a particular burden, exacerbated by a rota system that had the same doctor coping with busy Mondays each week. The duty triage system had the additional challenge that the period when the duty doctor was triaging meant time lost to the surgery, thus reducing the number of appointments available. This placed an additional burden on the doctors not on triage duty, who then had to work extra sessions or take paperwork home.

There was also the problem that demand is uneven and unpredictable with attendant workload effects.

Because some days there's about two pages, on a Monday or whatever, some days. You never know, do you, what days are going to be busy in general practice. And some days you think, 'Oh look at that list, it's just pages!'

Practice manager, practice 8 (GPT)

The integrated triage system, in which triage is combined with surgery appointments, could also cause excessive workload.

*So I think the work load is really quite high with this system and I think the workload has got to the extent where actually it feels, for some GPs, it felt impossible to get the work done in a working day. There are so many calls coming in, **and** seeing patients, **and** doing visits, none of the paperwork was getting done. So the working day became longer, more stressful . . .*

GP 05, practice 5 (GPT)

NT For triaging nurses there was also a reported problem of work overload. Those in the two NT practices that did not appoint extra nurses for the trial were particularly overworked, somehow having to graft hours of triage a day onto their already busy schedule. The nurse in one of these practices, practice 1, described the pressure and stress of having a high workload to be undertaken in conditions of extreme time urgency, in which she could have a list of 26 patients to ring back before her clinic started immediately afterwards. There was concern that the pressure would lead to mistakes.

Even in practices where additional nurses had been appointed for the trial, nurses described being daunted by the columns of triage calls to be made. Getting hold of people was a frustration for nurses as it was for doctors. Another frustration was not being able to hear what patients were saying when there was a lot of background noise, either from the patient's side or in the practice.

Nurses faced the additional challenge of having to perform triage using CDSS. This was found to cause a number of problems, and some nurses, especially those experienced in triage, chose not to use it, or to ignore what it said.

There's some of it that I just think is, the thing with [the CDSS] is that sometimes you can ask something, for instance someone's got a headache and next minute you know it's practically going into them having a brain tumour and that was the part of it that used to annoy all of us. Because it would bring up boxes that were completely unnecessary. And quite frankly I just flip out of them.

Nurse, practice 2 (NT)

Others used it reluctantly under the impression that doing so was mandatory for the trial, rather than just as a support tool. One of the issues with the CDSS was that it appeared not to be sensitive to all of the problems with which patients present when requesting a same-day consultation. It had been designed for acute primary care/urgent needs (such as in out-of-hours primary care) and was not viewed by nurses as wholly comprehensive for in-hours telephone triage in a number of ways. Some nurses felt that it

lacked information on commonly presented problems in primary care such as contraception, medication interactions, withdrawal of medications, medication reviews, moles, for queries from patients with diabetes about whether they should increase their insulin, or on medication for anxiety.

I think the software isn't necessarily the best for primary care. It's very good for out-of-hours but primary care it's not. There's gaps in it. For instance if someone rings up to have a mole looked at, as one of the examples, there's nothing on the [CDSS] system that deals with moles. At all. Not at all. Because in out-of-hours it doesn't occur, and [the software] I understand was built up for out-of-hours . . .

Nurse, practice 1 (NT)

Some of the nurse users felt that triaging with CDSS was laborious and 'long-winded', sometimes taking 10–15 minutes for each patient, in comparison with GPT for which, at least in one practice, only four minutes per patient was allotted. The CDSS sometimes directed nurses to ask what were described by some as 'ridiculously inappropriate' questions that could not be bypassed, for example about sexually transmitted diseases and pregnancy for a patient in her eighties. Some nurses felt the CDSS forced them to give appointments against their own judgement, resulting in them having to field complaints about inappropriate appointments from doctors. Some doctors reported finding the CDSS summaries to be unhelpful, even misleading, and did not use them, preferring instead to start the consultation from scratch. Consequently, no time seemed to be saved in consultations.

When somebody else has taken the history, I think I would feel very uncomfortable about not going over the questions myself.

GP 02, practice 1 (NT)

Nurses had a number of suggestions for improving the conversion rate under NT. One suggested that if they could prescribe certain things under the Patient Group Directive, such as antibiotics for urinary tract infections, a great many GP appointments could be saved.

Challenges of general practitioner-led telephone triage, workload issues: key themes

- GPT, whichever way it is organised, can increase doctors' workload.
- Not all doctors are comfortable or confident with telephone work.
- It can be stressful for doctors trying to keep up with the volume of work.
- There are workload disparities between doctors.
- Demand is unpredictable and therefore hard to manage, which can lead to wasted appointments or, conversely, the need to work late.
- Ring-back can be a source of frustration and wasted time when people cannot be reached.

Challenges of nurse-led telephone triage, workload issues: key themes

- There is work overload, especially when triage is added to usual workload with no extra nursing capacity.
- Ring-back is a source of frustration and wasted time when people cannot be reached.
- Doctors prefer to take the patient's history again rather than using the nurse's summaries; essentially, the patient has his/her history taken twice.
- There are issues associated with using CDSS.
- The CDSS was perceived as going into too much irrelevant detail and generating inappropriate questions.
- It is laborious and long-winded and takes too much time.
- It is not sensitive because it was not developed for primary care.
- Its advice can conflict with nurses' judgement.
- It advises appointments against nurses' judgement, which can cause problems with doctors.

- Its summaries are unhelpful to doctors, who tend to ignore them.
- Not being able to prescribe limits the number of GP appointments nurses can prevent.
- Doctors have different regime preferences for certain conditions, making the judgement about whether an appointment is necessary difficult.

Challenges for members of staff

Triage can represent a particular challenge in practices with a personal list system. In some practices with a duty doctor system, the triaging doctor would triage patients who needed an appointment into the next available slot, rather than into the patient's own or usual doctor's surgery. In practices with a personal doctor system, the close doctor–patient relationship was considered to represent good care and was highly prized. Restricted access to the doctor by his/her own patients was one of the consequences of triage that doctors felt could cause good doctor–patient relationships to disintegrate over time; they found it personally distressing to observe their patients going in to see another doctor while they were on the premises. The loss of routine and advance appointments that resulted from triage also impeded patients' access to their own doctor, and was also seen as detrimental to the personal list system and continuity of care and dysfunctional for accessibility in general.

In the GPT practice with an integrated triage system there were accounts of how 'incredibly stressful' doctors found the combination of triaging and seeing patients in routine appointments. In the duty system, too, the intensity of the workload could have an adverse effect on the quality of consultations.

But also the volume of calls, if there was one triager and there were huge numbers of calls, you know, the quality went down of the triage and people were just trying to book, book, book, book and we're not going into the depth of the consultation because they just couldn't.

GP 01, practice 7 (GPT)

Safety and risk were salient for both doctors and nurses. One doctor who had experienced a serious adverse event in a previous clinical setting had an ever-present fear of a catastrophic mistake, and another felt strongly that 'if you're practising safe medicine people should be seen'. Nurses would err on the side of caution to avoid risk, one arguing that she was risking her registration if she did otherwise. This caution reduced the number of appointments that could be prevented. In practices in which appointments had always been scarce, patients had acquired the habit of exaggerating their symptoms in order to get an appointment, and this made it difficult to make a judgement about whether an appointment was appropriate.

In one practice, for which anxiety about doctors getting 'cross' was a recurrent theme, a nurse informant cited this as an underlying anxiety when deciding whether or not an appointment was necessary, in case she got it wrong.

I suppose you feel the pressure a little bit because you worry about those people who you didn't bring in or you worry that the GP might be annoyed with you about people that you did bring in.

Nurse, practice 4 (NT)

Both doctors and nurses found triaging tiring. Some nurses found triage boring and not a good use of their skills, making them feel like telephone operators, and others found spending hours at a computer screen hard going:

For me personally I suppose I don't want to be sitting on the phone all morning. I'm a nurse practitioner and actually I'd rather be seeing patients.

Nurse, practice 2 (NT)

Receptionists still had to conduct what could be described as informal pretriage triaging, by asking patients what the problem was so that the triaging clinician could prioritise the ring-back list. This was an

optional part of the ESTEEM receptionists' protocol, although receptionists were advised that if patients were happy to provide information, it would assist with prioritisation. Some hated doing this, feeling that it was not their place because they had no clinical expertise, and that it was intrusive and inappropriate. Only one of the practices decided that the receptionists should not ask patients why they wanted an appointment, and they were glad not to have to. Patients could become abusive when asked, which was unpleasant for staff.

It's funny I think [laughter] it's really funny what some people will share without batting an eyelid. I've had two where I've just sort of recently picked up the phone and someone's sort of shared with me, oh I've got this awful pain in my backside, right up my bottom. And then other people who will just not and are very angry that you're asking for information, so again the girls have had to explain that it helps the doctors to prioritise the calls if they have an idea of what the problem is.

Practice manager, practice 5 (GPT)

In practices in which there was not the clinical capacity to provide enough appointments to meet demand, such as in practice 1, or where they were attempting to embrace the Advanced Access model of triage advocated by the Productive Primary Care presentation, no appointments or fewer appointments were available outside the triage system. Prebookable appointments were reduced in most of the GPT practices, because doctors were tied up doing triage duty and this caused difficulties for receptionists, who had no appointments to offer, as some patients could get angry and abusive.

The fact that in some practices patients did not like the triage system caused difficulties for staff, especially the reception team and sometimes for triaging nurses. Patients could see triage as a barrier to access, an unnecessary extra step to getting an appointment, and they made their opinions known. Reception represented the front line in this struggle and staff found it hurtful.

Patients threatened to take their custom to A&E departments and were sometimes abusive to staff. This was a particular issue in practice 1, where staff said they would leave if triage continued after the trial, but other practices also experienced verbal abuse or at least bad temper from patients. Patients made formal and informal complaints.

Challenges for members of staff: key themes

- Not seeing their own patients was distressing for doctors who valued having a personal list.
- The high volume of calls caused distress to triaging clinicians.
- There were concerns about risk and safety because patients were not seen face to face.
- Nurses were anxious about doctors' reactions if they booked an appointment inappropriately. Doing so was often a result of CDSS-generated instructions.
- Receptionists disliked having to ask people why they wanted an appointment. Doing so could result in abuse from patients.
- The loss of or reduction in prebookable appointments was a reduction in both access and continuity, and caused problems for receptionists when patients were annoyed at not being able to book in advance.
- Patients did not like the system and could be abusive.

Challenges for the practice

Despite the fact that practices were remunerated for participation in the trial, there was concern expressed that performing triage was costing them money. Apart from the increase in telephone bills, one doctor argued that in NT patients appeared to be effectively consulting twice for the same condition and had their history taken by two different people, and that this had cost implications for the practice. (It is worth noting that this doctor was from a NT practice whose motivation to participate in the trial was described by staff as financial, and which chose not to use the trial payment to provide any additional nursing capacity.)

Um, well, my inclination is to feel that we have frequently ended up with a patient in touch twice for the same problem, once by the nurse and once by me. And there isn't value added to my consultation by the nurse having spoken to the patient already beforehand and therefore we pay for two consultations rather than one.

GP 03, practice 1 (NT)

Another doctor, from a GPT practice, cited the 51 appointment slots lost per week to duty triage as a cost to the practice, as locums had to be employed and extra sessions worked to fill the gap. The increase in the telephone bill was another cost cited.

Staff in five of the eight intervention practices claimed that there had been no reduction in appointments (practices 1, 2, 5, 6 and 7). Triage was felt to be a lot of effort for very little gain, causing organisational problems and, in one practice in particular, dissent and dissatisfaction among staff and patients alike. Routine advance appointments disappeared as appointments were blocked out for triaged patients and this was a recurring dissatisfaction, reported to be a major cause of patients' complaints. In the two NT practices that had not provided additional nurse time (practices 1 and 3) the knock-on effects of lost nurse appointments were causing difficulties for both practice and patients. Chronic disease management in particular had been neglected, which would impact on practices achieving their Quality and Outcomes Framework targets (NHS Employers 2014⁶⁵), with potential loss of income and reputation. Practice 3 had to have a gap from ESTEEM while nurses caught up.

Another difficulty was the issue of negotiating the timing of surgeries in relation to the timing of triage. Often these took place at the same time of day, and triage appointments were wasted when there was no time for patients to get into the practice in time, and then surgeries ran over time in an attempt to fit people in.

In two of the NT practices some of the doctors were not happy with NT as a model. The CDSS was felt by some to be too 'tick boxy', and 'algorithmic medicine' was in any case considered an affront to good patient care. It was felt that nurses did not have diagnostic or triaging skills and that taking a patient's history was a highly skilled proficiency acquired over years of training and practice by doctors but not nurses.

Finally, somewhat perversely there was a concern that if triage was successful the improved access it provided would not so much manage demand as increase it.

And there is a concern in people that firstly, by triage, you're going to increase demand because the more you answer demand the more people will demand. So that's one concern.

GP 06, practice 7 (GPT)

Challenges for the practice: key themes

- There was concern about the cost of triage.
- In NT practices patients were assessed twice.
- In some GPT practices locums had to be employed to cover surgery time lost to duty doctor triage.
- The telephone bill increased.
- There was no perceived reduction in appointments in some practices.
- The triage system caused dissent and dissatisfaction among staff.
- Patients perceived triage as a barrier to access.
- Some doctors in NT practices were not happy with NT as a model.
- CDSS was seen by some as inappropriate care.
- There was concern that demand would increase as access improved.

Staff perceptions of challenges for patients

Some staff members reported that triage was making patients anxious. The fact that patients who went on to have appointments with someone other than the person who had triaged them and had to have their history taken twice was felt to be a disadvantage. Patients were sometimes uncomfortable about discussing medical problems with nurses, which was an unfamiliar thing to do. If they had ongoing problems with which their own doctor was familiar, triage meant having to tell their story all over again from the beginning to someone new. Waiting for ring-back could be inconvenient, and even stressful, for patients, especially if they were in a state of pain or anxiety or if they had things to do – get to work, get children to school and so on. Patients were often not in a situation where they could be called back; they were driving, or at work, somewhere noisy or not in a situation of privacy. Mobile phone signals were often poor.

Some patients were frustrated by triaging with CDSS.

They just said this is absolutely a waste of time, why are you asking me all these ridiculous questions. Yes and a few of them got a bit abusive.

Nurse, practice 2 (NT)

Some patients hated being asked to tell receptionists what was wrong. Staff also identified patients for whom triage was not and never would be appropriate. These included people with a previously unreported anxiety or depression who may have spent days plucking up the courage to ring, and where the patient has multiple complex comorbidities, as well as patients with hearing disabilities or learning disabilities or dementia or who were very old or confused, who were excluded as per protocol. Some patients just did not like the triage system in general.

Staff perceptions of challenges for patients: key themes

- Some patients were made anxious by triage.
- Patients had their history taken twice.
- Some patients were uncomfortable talking about personal medical matters to a nurse.
- Waiting for ring-back could be inconvenient and stressful for patients.
- Patients were not always in a situation where they could receive a phone call.
- Patients did not always like to tell receptionists why they wanted an appointment.
- Some patients found the CDSS questions to be inappropriate.
- Triage may never be appropriate for some patient groups.

Patients' accounts of their experience of triage

There was evidence from patients' accounts of a certain problem in relation to trial recruitment in one of the practices where triage meant that there were no advance appointments available. Patients phoning for an appointment and being told they could not book ahead were asked if they wanted an appointment that day. If they said they did, they were included in ESTEEM, even although their reason for wanting a same-day appointment was that no other sort of appointment was available. One patient to whom this had happened reported being confused about how the appointment system worked, because she had been permitted to book an advance appointment in the past. She had searched for, and failed to find, an explanation of the appointment system in her practice. She found the triage system a very long-winded way of getting an appointment, particularly as she knew her condition – a changed mole – would need to be looked at. Another patient from this practice – practice 1 – found the triage system unnecessary, inconvenient and time-wasting, and reported that staff had told her they did not like it either.

Yeah, well it was silly. It was unnecessary. It was just unnecessary. You know, it was a waste of everyone's time, a waste of my time, a waste of her time. You know, all it did was push the appointment back you know, and, and, it was very inconvenient.

Patient 06, practice 1 (NT)

Another patient who also found triage using CDSS long-winded expressed concern about whether her elderly mother could have coped with answering so many questions.

There were doubts whether GPT adequately discharged a doctor's duty of care, and the concern that the quality of diagnosis must be adversely affected because triage relies on the patient being able to hear and understand what the doctor says on the phone.

But I think perhaps, it depends whether the fraternity, and by that I mean the doctors' fraternity, feels that they've discharged the duty of care, I think. Without getting legal, I mean duty of care in so far as he's happy that what he's said is understood and is OK, 'cos it seems to me that there is some kind of differentiation in the level of, not only the level of service but in the level of the quality of the diagnosis if you can't see what it is you are looking at.

Patient 03, practice 6, GPT

Patients also reported the difficulties of waiting for the ring-back when they were unlikely to be in a situation where they could receive a call. They described having to sit around waiting, not knowing if they had an appointment or not, not being able to plan their day, not knowing whether they should go to work and, if so, whether they needed to request time off for a doctor's appointment.

Telling the receptionist what was wrong was a common reason for disliking the system, with the view that such things are too personal to discuss with strangers over the phone with fears about confidentiality and who else might hear. There were concerns about how triage would work for patients with poor communication.

Patients' accounts of triage: key themes

- Patients had sometimes not initially requested a same-day appointment.
- Some patients were confused about how the new appointment system worked.
- Some patients considered triage to be unnecessary and inconvenient.
- There were concerns that certain people would not cope with the number of questions emerging from the NT.
- Some patients found it difficult and inconvenient waiting to be rung back.
- Patients disliked telling the receptionist the reason for their appointment request.
- Some patients were concerned about the confidentiality of the telephone conversation.

Acceptability of triage

There was considerable variation in the degree to which triage was acceptable, both between practices and within practices. Accounts of positive experiences of both NT and GPT suggest that triage can, under certain circumstances, be an acceptable model of service delivery. However, the extent of accounts of negative experiences and challenging aspects of triage strongly suggest that triage was not acceptable for a number of staff and patients across the practices. When the criteria of good communication within the practice, supportive staff relations and full consultation with staff regarding trial participation were not present, and especially if there was a culture of resistance to change, triage was less likely to be found acceptable by staff. For some staff and patients, telephone triage challenged beliefs about what constitutes good patient care, in particular the axiom that seeing patients is an essential component of good, safe care. There was anxiety among staff about adverse events and a fear of litigation. In addition, some staff felt performing triage was not an appropriate use of their skills, and it was also experienced as both boring and stressful. Both nursing and reception staff found aspects of telephone triage forced them into a role with which they were uncomfortable. Some doctors and nurses were not comfortable with telephone work. Staff believed there were patient groups for which telephone triage was inappropriate. Triage was felt to compromise continuity of care, more so where advance appointments were no longer available, and GPT could be irreconcilable with a personal list system. All of these factors impacted on the

acceptability of telephone triage. Many patients found triage unacceptable; in particular, it could be experienced as an unnecessary barrier to access to the doctor.

What influences the extent to which the telephone triage interventions are seen to work or not work?

In this section we summarise material from the previous sections to consider the issue of acceptability and perceived effectiveness. There is no doubt whatsoever that staff perceived that a telephone triage system can work to manage demand, increase patient access and reduce clinical workloads. It is also clear that this is not inevitably the case. Although the data are complex, and do not reveal any consistent and predictable patterns of what might predispose triage to success or failure, it is possible to identify factors that are likely to have an impact given certain conditions. This may be at least helpfully suggestive, if not definitively predictive, of what can make the introduction of telephone triage successful or not.

What is seen to work in telephone triage?

Telephone triage delivered by GPs or nurses was seen to be effective for reducing both appointments and consultation times, and therefore for rationalising resources. Telephone triage delivered by nurses can work well. It can be perfectly effective and acceptable to both practice staff and patients, and can be felt to reduce doctors' appointments. In this process evaluation of the eight intervention practices studied, the practice whose telephone triage was perceived to significantly reduce doctors' appointments more than any of the others was a NT practice. NT was felt to potentially create a sustainable long-term reduction in doctors' appointments by teaching self-management skills to patients and the conditions in which a doctor's appointment is and is not necessary. NT is likely to be more effective when performed by a nurse practitioner or a primary care nurse, i.e. a nurse with experience of history-taking and diagnosis, and who can prescribe, but these are not necessary preconditions for success. It may be more likely to effectively reduce doctors' appointments when nurses can prescribe certain medications for commonly presented conditions. Using CDSS can be a useful training tool for nurses learning triage, although some nurses are already experienced at taking histories, diagnosis and assessing appropriate treatment options, and some nurses feel it increases the amount of time taken to triage.

Triage training and experience can encourage more effective triage by both nurses and doctors. Nurses are less likely to have experience in history-taking, diagnosis and treatment, and are likely to require a longer induction period.

If there is already a culture of triage and telephone consultation in the practice the effectiveness of the introduction of a more widely applied triage system will be increased and expedited, as both clinicians and patients are accustomed to a remotely delivered service. Initial resources need to be devoted to setting up an effective triage system that will produce resource savings when established. Staffing needs to be appropriate. Good effective communication within the practice, with formal and informal opportunities for airing and sharing problems and successes, predisposes the practice to more successful introduction of triage.

A preparatory assessment of needs and capacity predisposes practices to more successful introduction of triage, and a run-in period to practise the intervention is likely to be helpful, as it was during the ESTEEM trial. Commitment to the change by all stakeholders is essential, and should be established before triage is set up, and a champion to drive it forward and address problems is helpful. Triage can improve access, and may be more acceptable at times when access has been poor in the past, although previous good access may create easier and more trusting relations between staff and patients. The opportunity to hear the experience of others who have successfully introduced the innovation can inspire and motivate staff to drive the change, but it is probably important that this is delivered by people without an overt commitment to a particular model that may not be the most appropriate for the practice.

What is seen not to work in telephone triage?

Many of the aspects of triage which were perceived not to work were a function of the trial rather than triage per se. An example is randomisation to an arm that is perceived as less effective than the system in place before the trial, for example when effective face-to-face NT was replaced by less effective GPT. Randomisation is not likely to be an effective way of introducing a major organisational change. Another example is when practices were already effectively managing their demand for same-day appointments so that there was no perceived need for change. It is axiomatic that models of triage that do not evolve from the organisational culture and framework of a particular practice, and particularly from a perceived need for change, are less likely to work, at least in the short term. In some practices the timing demands of the trial enforced the introduction of triage at times that were felt to be inappropriate, such as during a 'flu' epidemic. Implementing triage is also more problematic in practices with authoritarian hierarchies and poor staff communication, in which there are no conditions for formally or informally discussing concerns and experiences.

In relation to whether or not NT is seen to work, it is worth noting that some nurses are frequently nervous about triage, which may lie outside their comfort zone of experience and training, especially if they have no diagnostic experience and rely on CDSS. Nurses embark on telephone triage from a very different starting point from doctors, and embedding NT into practice is likely to take longer than perhaps a GPT system might, and require additional preparation, support and resourcing. Launching straight into a NT system in the absence of this preparation is unlikely to be as successful as a carefully managed change. NT is less likely to be successful if practices do not support nurses by ensuring that there is sufficient nurse capacity to undertake this new role, as well as performing their usual roles (such as chronic disease management, vaccination programmes and management of dressings). Nurses need to be proficient in IT, which was not always the case in this trial. Although it was regarded as a useful training tool by some nurses who had no triage experience, the CDSS that was used (which was designed for the different demands of out-of-hours primary care services) was judged by some nurses to be unsuitable for use in daytime primary care triage in a number of ways. Using CDSS may lead some nurses to feel that their judgement is undermined with regard to when an appointment is necessary. This may increase the numbers of patients advised to attend for a face-to-face appointment, as CDSS 'errs on the side of caution' and may also cause a strain on practice relationships. It was seen by some to be 'long-winded' and seen as increasing the time taken to triage.

Doctors without experience and training in triage can lack confidence and take too much time to triage, and doctors with no experience in or desire to practice telephone consultation are less likely to be efficient at triage or to have a commitment to its success. For a few doctors, telephone triage is an insult to their model of good and safe care, and as such can represent an existential challenge to their identity as good doctors. In addition, in the ESTEEM trial, the triage model initially promoted by Productive Primary Care (and sometimes confused with the ESTEEM protocol) was not wholly compatible with triage within ESTEEM, and this may have confounded and confused the delivery of triage in the trial. Superimposing a predesigned (and quite extreme) model of triage – such as the Advanced Access model – onto an organisation without careful consideration of specific aspects of that organisation is unlikely to work, and practices that tried to do this were forced to abandon the attempt. In the Productive Primary Care model receptionists do not have appointments to offer, and advance appointments cannot be made. Both staff and patients at practices who tried to implement this, as well as those in which loss of advance appointments was an adverse effect of the need to prioritise appointments for the triage system, found this uncomfortable. It can spoil good practice or patient relationships and can be embarrassing and disempowering for receptionists. It excludes patients who, for whatever reason, find telephones inaccessible. It is a barrier to continuity, as patients can no longer book ahead with their own doctor who knows their history, thus undermining the relationship of trust between doctor and patient. The Advanced Access system was advocated as a response to the Department of Health target that patients should be able to see a doctor within 48 hours⁶⁶ but was modified, as recommended by Department of Health policy in 2004, following a public confrontation with the Prime Minister by a disgruntled patient in 2006.⁶⁷ Pope *et al.*⁶⁷ highlight a number of publications which identify problems with the acceptability of

Advanced Access to both patients and clinicians, related to its adverse impact on doctors' workload, continuity of care and patients.

There is no one approach to performing GPT which is appropriate for all or most practices. Effectiveness depends on a number of contingencies, for example whether there is a usual doctor system, individual doctors' confidence and competence, whether doctors have prior triage experience and/or experience of telephone consultations, and whether their model of good patient care and safe practice encompasses a commitment to seeing patients face to face. Telephone triage is more acceptable to some types of patients than others, so practice catchment and population will influence effectiveness. Success will also relate to how effective a system for providing same-day appointments exists before a triage system is introduced. GPT can increase workload and exacerbate workload disparities between doctors in a practice.

Ring-back issues are a constant challenge in telephone triage. Delay in ring-back, patients not being able to receive calls, patients who cannot be reached and poor mobile phone signals can make telephone triage less effective, frustrate both staff and patients, and waste time. Practices in which triage was organised in a way that meant patients were waiting for long periods, sometimes hours, for ring-back were doing their patients a disservice and undermining the success of telephone triage as an efficient system. Strategies for timely ring-back should be prioritised. Similarly, noise, either in the practice or in the patient's home, can compromise the safety and effectiveness of telephone triage. Practices need to provide a quiet space for triaging, and receptionists could advise patients waiting for ring-back that they will need to be in a place of privacy and quiet. There are safety and confidentiality issues in ringing people when these conditions cannot be fulfilled. Patients should be assured of privacy and confidentiality when they are rung back; there was concern among patients about who might be 'listening in'.

There are certain patient groups for whom triage is inappropriate, for example some people with mental illness, substance abuse issues, learning disabilities, dementia, hearing disabilities or who are very old, who were excluded from ESTEEM. Policies need to be in place to exempt such patients from a triage system.

Chapter 4 Discussion and conclusions

Summary of findings

Our findings show that introducing either GPT or NT resulted in an increase in the number of primary care contacts in the 28 days following a patient's request for a same-day consultation with a doctor when compared with practices' usual processes for handling such requests. Triage was associated with an overall reduction in GP face-to-face contacts. However, GPT resulted in an increase in GP telephone and face-to-face contacts when combined together, whereas NT resulted in a reduction in this measure. These changes reflect redistribution of GP workload from face-to-face to telephone consultations following the introduction of GPT, and redistribution of workload from GPs to nurses following the introduction of NT.

However, when considering differing patterns and duration of patient care across all three arms, there was no difference in 28-day health-care costs of patients. Triage appeared safe, and no differences in patient health status were observed across the arms. NT was associated with a small reduction in patient satisfaction compared with GPT or UC.

Primary care contacts in the 28 days following a same-day consultation request

Our primary outcome was the number of contacts undertaken in primary care settings (GP practices, out-of-hours primary care services, walk-in centres and A&E departments) in the 28 days following a same-day face-to-face GP consultation request. Our findings in relation to the primary outcome are clear. Implementing triage, whether GP- or nurse-led, resulted in additional, not substituted, workload.

In UC, 51% of patients received only one contact across the 28-day follow-up period following a same-day consultation request. In contrast, this proportion was reduced to 23% in GPT and 12% in NT.

In this study, 16,211 patients generated 39,736 consultations with a clinician in the 28 days following a same-day consultation request. Of these consultations, 22,471 (57%) took place on the index day. Across the 28-day period, only very small numbers of patients (550 in total) were seen in A&E and there was no evidence that either form of triage increased or reduced attendances at A&E.

Our main ITT analysis demonstrated that GPT resulted in a relative increase in the rate of primary care contacts across the 28-day period of 33% when compared with UC; the equivalent increase following NT was 48%. A small increase (4%) in the rate of contacts was observed when NT was compared with GPT. As to be expected, per-protocol analysis of the primary outcome demonstrated an intensification of the treatment effects of both GPT and NT. The findings of the main ITT analysis were robust to the use of imputed data for cases where primary outcome data were not available.

In recognition of the inevitable need for a proportion of patients to be seen face to face following triage, we undertook a sensitivity analysis in which we combined all within-practice contacts on the index day as just one contact. This analysis demonstrated an increased rate of contacts in GPT of 10%, and in NT of 12%, when compared with UC.

Interpreting the ITT analysis of the primary outcome in conjunction with the above sensitivity analysis, it is evident that there was an increased rate of contacts over the 28-day period in both triage arms compared with UC.

Additional sensitivity analyses were undertaken, based on dividing the primary outcome chronologically into contacts taking place on the index day only ('day 1') and those taking place during the remaining 27 days of the 28-day follow-up period (days 2–28). Considering only contacts taking place on the index day, the rate of contacts in GPT compared with UC was increased by 51% and in NT by 72%, largely attributable to the telephone call undertaken in triage. There was an increased rate of contacts in NT compared with GPT of 14% – again, largely attributable to the telephone call undertaken in triage. The increased rate of contacts in the triage arms was greater on the index day than on the subsequent follow-up days.

Care on the index day

Numbers of contacts and disposition of patients

In UC, a substantial majority of patients (87%) received a GP face-to-face contact, with no further primary outcome contacts on the index day. Other patient pathways were represented by substantially smaller proportions of patients.

In contrast, patients in the triage arms had more diverse patterns of patient management when compared with UC, possibly reflecting a more flexible approach to patient assessment and management in the triage arms.

Where GPT was implemented, 46% of patients received only the GP triage contact on the index day, and 36% received a GP face-to-face consultation on the index day following the initial GP triage contact; a smaller proportion (9%) of patients had a nurse face-to-face consultation on the index day following their initial GP triage contact. A small proportion (6%) of patients received a GP face-to-face consultation on the index day instead of being managed under the triage system.

Where NT was implemented, 22% of patients received only the nurse triage contact on the index day, and 56% of patients received a GP face-to-face consultation following the initial nurse triage contact; 9% of patients received a nurse face-to-face contact following nurse triage. For a small proportion of patients (9%), their first contact on the index day was a GP face-to-face consultation instead of being managed under the triage system.

Distribution of clinician time in practice

Considering workload from the perspective of the GP, the introduction of GPT was associated with a 55% reduction in the rate of GP face-to-face contacts on the index day (compared with UC), although there was a 49% increase in the rate of GP overall telephone and face-to-face contacts combined. Across the whole 28-day follow-up period, the reduction in the rate of GP face-to-face consultations was smaller (39%), reflecting the deferral of some workload from the index day to the follow-up period, whereas there was a 38% increase in the rate of GP telephone and face-to-face contacts combined.

Introduction of NT was associated with a 31% reduction in the rate of GP face-to-face consultations on the index day (compared with UC), and a reduction of 28% in the rate of GP telephone and face-to-face contacts combined. Across the whole 28-day follow-up period, there was a 20% reduction in the rate of GP face-to-face consultations, with a 16% reduction in the rate of GP telephone and face-to-face contacts combined.

In considering these observations regarding GP contacts, it is worth noting that the introduction of GPT involved GPs undertaking the work of triage in addition to the (reduced) number of face-to-face contacts; in contrast, where NT was introduced, nurses, not GPs, delivered the triage element – delivering the resulting 'gain' to GPs in reduced numbers of GP face-to-face consultations.

Taking account of both the number and type of contacts in the practice (whether with a GP or nurse, face to face or telephone) and their associated durations provided additional insight into clinician workload, which were somewhat different from those based solely on the number of contacts. In UC, the estimated duration of patient–clinician contact time within the practice on the index day (based on the first two contacts within the practice) was 9.6 minutes, compared with 10.3 minutes when GPT was implemented, and 14.8 minutes where NT was implemented. The distribution of the time by clinician was, however, markedly different across the three arms.

In UC, the substantial majority (9.1 minutes) of the 9.6 minutes estimated contact time on the index day comprised contact with the GP. Where GPT had been implemented, the estimated GP contribution to the overall workload was 9.0 out of the 10.3 minutes, with nurse contact accounting for 1.3 minutes, the latter usually resulting from nurse face-to-face contact. Where NT had been implemented, of the total 14.8 minutes of estimated contact time provided on the index day, overall, around half of this was with GPs (7.7 minutes), with the remainder (7.1 minutes) with nurses.

Resource use

Overall, the substantial majority of service use in all three arms took place in GP practice settings. Only very small numbers of contacts occurred in A&E, walk-in centres or out-of-hours primary care services over 28 days. The similarity of the rates of contact with non-practice-based services across arms provides no evidence to suggest that triage encouraged patients to seek care outside the practice.

Overall, we observed low rates of patient non-attendance compared with that reported in other studies,² thus there was no evidence that triage may advantage practices by reduced non-attendance rates.

There was also no evidence of differences in patient self-reported use of NHS Direct between trial arms – around 4–6% of participants reported using NHS Direct in the 28-day follow-up period.

Economic analysis

Overall, costs incurred were very similar across all three arms across the 28-day follow-up period. Costs incurred on the index day were observed to be lower in both triage arms; the projected cost saving relating to care on the index day (compared with UC) was approximately twice as much in GPT compared with NT, although overall the absolute differences were modest (£5.75 vs. £2.58 per patient).

Although there is a difference in the number of contacts by triage arm over the 28-day follow-up period, mean costs for primary care contacts (primary outcome) are similar for UC, GPT and NT. This indicates that the costs associated with triage are offset over the 28-day follow-up period, with the added cost of triage resulting in fewer GP face-to-face consultations in practice when patients initially request a same-day GP appointment.

The main area of contact, and cost, is associated with GP consultations in the practice, representing almost 90% of the costs in the UC arm, and data indicate few contacts with out-of-hours primary care, walk-in centre or A&E compared with the level of contact with the GP in a primary care setting.

Estimates of intervention cost for the triage interventions are an important factor in the cost analyses. Estimates of the intervention cost for triage contacts are presented in a transparent manner, and follow the methodology for unit costs in health and social care used by the PSSRU,³⁸ which are widely accepted and used within health service research. The main cost component for triage contacts is the GP or nurse time associated with the triage contact, and this has been collected at a participant level across the ESTEEM trial, providing good quality data on which to estimate unit costs.

Base-case cost estimates, over the 28-day follow-up period, used published unit costs for primary care contacts and A&E contacts. These costs reflect the opportunity cost associated with health-care resources; however, these cost estimates may be open to some limitations. Therefore, sensitivity analyses have been presented, in estimates for the triage contact unit costs and for the estimates of 28-day follow-up cost, to address uncertainty in the use of published unit costs. Where different assumptions have been used for unit costs for health-care contacts, we specifically used lower unit cost estimates that did not include components of cost associated with the qualification costs for GP or nurse contacts, and did not include costs associated with the direct support staff for GPs. Although we saw much lower overall estimates of 28-day costs, we did not observe differences between intervention arms.

In base-case analyses, as per the prespecified analysis plan, we used published unit costs for GP contacts. In so doing, we used an assumption that within-practice GP consultations have a mean duration of 11.7 minutes,³⁸ this being based on an allocation of GP consultation time over an estimated number of GP in-surgery consultations. However, the ESTEEM trial has considered use of triage in the subpopulation of primary care patients for whom a same-day request is made to see a GP. Trial findings indicate that the mean time for a within-practice GP consultation is shorter (estimated 9.5 minutes) than that derived for cost estimates across the broader GP patient population. The difference in the ESTEEM trial population and the data reported for the broader primary care patient group (a difference of 2.2 minutes, a difference of £8.14 at base-case cost assumptions), if supported by future research, may influence the interpretation of findings in the cost analyses presented, showing UC to be less costly than the GPT and NT intervention arms. This would be the case even where the difference in time/cost is smaller, given the estimate for the cost per minute of GP time and the relatively small total costs being compared. It is difficult to know from the trial if this estimated mean GP contact duration (UC) of 9.5 minutes will hold in a larger sample or if GP consultations in the two triage arms would have been equally different (from expected mean of 11.7 minutes). New data are required to inform this issue, as these research questions were not specified in the ESTEEM trial.

One hypothesis may be that a proportion of contacts among the requests for same-day GP consultations are indeed relatively short and could be managed via a triage telephone contact, and it is these contacts (when presenting for a face-to-face GP consultation) that drive a mean duration of 9.5 minutes that is less than expected. What we may be seeing in the triage interventions is that less complex presentations are being managed via telephone triage, and represent those contacts that are resolved at triage level and do not need a further primary care contact. However, an alternative explanation may be that the data from ESTEEM on mean contact time are not comparable to those used to estimate the unit costs for GP consultations reported by PSSRU,³⁸ and the timings recorded in the ESTEEM trials may need to be adjusted to include an additional allowance for non-contact activities during the consultation time allocated for GPs.

Safety

Triage appeared safe. There was no evidence of excess hospital admissions within 7 days or A&E attendance within 28-day follow-up period in either triage arm when compared with UC. There was also no evidence of excess deaths within 7 days of the index consultation request.

Patient health status

At the point of completing a questionnaire (around 28 days following the index consultation request), patients reported similar levels of resolution of the original problem in all three arms. Interestingly, around 53–59% of patients across the three arms reported being 'much better' by the time they returned a questionnaire. Thus, around 45% of participants reported significant residual issues related to their original consultation request by the time they returned a questionnaire. Perhaps not surprisingly given these observations on problem resolution, there were no differences in self-reported health status by trial arm, as measured by the EQ-5D. Triage, whether GPT or NT, appeared to achieve similar health outcomes compared with UC.

Patient experience of care

Although, overall, patients' experience of care was met with high levels of satisfaction (in the order of 90% being 'very' or 'fairly' satisfied), patients in the NT arm were somewhat less satisfied than those in the GPT or UC arm. Patients reported that it was easier to get through to the practice on the phone in practices implementing GPT in comparison with UC, and that it was harder to get prompt care in NT by comparison with both GPT and UC. Patients also reported that it was more difficult to see a doctor or nurse if the patient wanted to do so in NT, but not GPT, compared with UC, and that it was harder to get medical advice in NT compared with both GPT and UC. NT was reported as being less convenient than either UC or GPT.

In summary, NT appeared to be somewhat less acceptable to patients than either GPT or UC. There was no significant difference in the acceptability to participants of GPT when compared with UC. GPT overall appeared as acceptable to patients as UC.

Research findings in context

The ESTEEM trial was developed in response to the lack of substantial research evidence concerning the clinical effectiveness and cost-effectiveness of telephone triage in general practice. No substantive new research examining the effects of telephone triage for acute care on in-hours primary care workload, resource use, safety or patient experience has been reported during the course of this trial (although new evidence on the safety of telephone triage in out-of-hours care may be pertinent⁶⁸). Much of this literature is summarised in a systematic review by Bunn *et al.*²⁷ undertaken in 2005. Although there was heterogeneity in reported outcomes, key findings of that review indicated that telephone triage or consultation with either a GP or a nurse reduced the number of face-to-face contacts with GPs. Telephone triage was safe, and patients found triage to be acceptable overall. However, methodological flaws were noted in many of the included primary studies, hence caution was urged when interpreting findings. Our findings contribute important data that extend the existing evidence base around patient experience and the safety of telephone triage and, we believe, provides definitive answers to questions around primary care contacts, workload and NHS resource use.

Workload

General practitioner-led telephone triage has been promoted by the NHS Institute for Innovation and Improvement as a mechanism for improving patient access and reducing GP workload. There has been an increase in the uptake of telephone triage systems despite the lack of a robust evidence base – especially evidence on the impact on general practice and wider primary care workload. Our findings are novel and suggest that telephone triage of patients seeking a same-day appointment increases primary care workload (in terms of number of contacts) within the 28 days following the patient's consultation request, but with no observed changes in overall cost when compared with UC.

A key finding in this study, however, is that any change in workload is restricted to general practice: GPT or NT did not appear to result in overflow into other primary or secondary care services. We found no difference between arms in the number of out-of-hours primary care contacts or A&E attendances. Inconsistent messages have emerged from past research. For example, McKinsty *et al.*¹³ found no difference in out-of-hours primary care contacts following GP telephone and GP face-to-face consultations. However, other research suggested that NT might lead to significantly more out-of-hours primary care and A&E contacts than UC.³

However, because the majority of contacts occurred in general practice, our data provide an important proxy for workload within this context. Past research has shown that telephone triage or consultation by a GP or a nurse was associated with a decrease in same-day GP workload: the number of same-day GP appointments being reduced to between 39% and 44%.^{3,11} However, reconsultation rates within the few weeks after telephone consultation have been shown to increase by a similar magnitude, around 30–50%.^{3,13}

Our findings confirm that a triage system for same-day requests reduces GP face-to-face appointments. Across the 28 days following the initial same-day request there were, on average, 1.46 GP face-to-face contacts per patient in UC. This reduced by 37% to less than one GP appointment (0.92) under a GPT system and by 18% (1.19 average appointments) under a NT system. However, any time saving for GPs due to reduced face-to-face contacts over the 28-day follow-up period would have been offset, to some extent, by the increased time spent performing triage in GPT.

The majority of primary care contacts following a same-day consultation request occurred in the practice on the day of the consultation request. GPs undertaking phone care have been reported to definitively manage around 29% of same-day requests¹¹ and nurses around 26% in this 'first pass'.^{3,69} In our study, in the GPT arm, 21% of patients managed per protocol had no further contacts following triage – the equivalent figure for NT was 8%.

A key issue to consider, however, is the patient's overall clinician contact time – taking into account the number, type (defined by clinician and mode) and duration of contacts. We found that both triage systems increased clinician contact time relative to UC. Similar findings have been reported with respect to NT;³ even though there was an increase in overall clinician contact time, there was a reduction in GP contact time, relative to UC, a reduction which was offset by increased nurse contact time. Our findings mirrored this for NT but do not show any substantial time-saving for GPs following GPT. Rather, GP contact time following GPT was only slightly less than that estimated for UC, although there was an important redistribution of patient contact time with increased use of nurse input.

It might have been assumed that face-to-face consultations would be shorter following triage, on the basis of the history having already been documented during the triage call. However, we observed that in both triage systems, the duration of subsequent GP face-to-face consultations were, in fact, longer than index consultations in UC. A triage system is designed such that the sickest patients are assessed and may be seen face to face in the practice subsequent to triage. This particular subgroup of patients, i.e. those with data on a subsequent GP face-to-face consultation on the day of or day after the index triage consultation, are likely to have more presenting problems and greater health complexities.¹² A triage system might offer flexibility to ensure that patients with complex health needs are provided with the longer consultations necessary.

Safety

Bunn *et al.*²⁷ did not examine patient safety in detail – there was limited evidence in that review to provide a comprehensive assessment on the safety of triage. To date, the message is mixed. For example, RCTs comparing telephone consultations with face-to-face consultations found no difference in the number of A&E attendances^{13,70,71} or hospital admissions.⁷⁰ Two RCTs comparing triage by GPs and nurses in out-of-hours primary care settings found no difference in hospital admissions.^{5,15} Richards *et al.*,³ however, found an increase in A&E admissions following nurse telephone triage in general practice.³ McKinstry *et al.*⁷² suggested that telephone consultations may be less likely to include sufficient information to exclude important serious illnesses and therefore were potentially more likely to compromise patient safety.

Existing evidence specifically relating to NT suggests that it is safe. Richards *et al.*²⁰ conducted research to assess the quality of information-gathering and accuracy of decision-making by nurses undertaking telephone triage.²⁰ Independent GP and nurse assessors rated audio-recordings of nurse telephone triage consultations. Out of 218 consultations only 7 (around 3%) were rated as potentially unsafe, suggesting that, overall, nurse telephone triage is safe. Research into nurse telephone triage in out-of-hours primary care settings also suggests that triage is safe. A recent systematic review by Huibers *et al.*⁶⁸ noted that for observational studies (where real patient outcomes were assessed) triage was safe in 97% of all patients contacting out-of-hours primary care and in 89% of patients with high urgency.

Like most of the studies noted above:

... our trial was not powered to inferentially test safety outcomes and we cannot conclusively rule out differences between study groups. In view of this factor and of other evidence [20,72] caution is needed before firm conclusions can be drawn from these results, and further studies, possibly including different study methods such as significant event audit, might provide useful additional evidence about the safety of triage.

*Campbell et al.*⁶⁴

Patient experience

There is limited evidence on the 'acceptability' of telephone triage in primary care.²⁷ Generally, patients have found telephone consultations with GPs and nurses to be acceptable,^{11,13,70,73–75} although one systematic review included primary studies that, overall, showed patient dissatisfaction with telephone consultations.⁷⁶ Our findings are consistent with past research with respect to GPT.^{11,13} NT, on the other hand, was somewhat less acceptable than both GPT and UC. Perhaps this should not be surprising given that patients originally called the practice wanting to see a GP, but were in fact offered a telephone consultation with a nurse – considered in that context, the small reduction in satisfaction we observed seems predictable.

Process evaluation

A great deal of information that is interesting and useful has emerged from the process evaluation of the ESTEEM trial. However, as the previous sections have demonstrated, triage has not emerged as a phenomenon with predictable patterns from which rules and guidelines can be produced which would be applicable regardless of context. How triage was operationalised, how it was experienced, its acceptability, and whether it 'worked' for practices, staff and patients emerged as highly contingent, idiosyncratic and variable. The findings in this process evaluation echo those of a qualitative study of the Advanced Access model of telephone triage,⁶⁷ which also found wide variance in interpretation and implementation of the model, and that informal organisational behaviour resulted in its adaptation to practice contexts, norms and values.

Our process evaluation was undertaken within the context of delivering a RCT; it was not triage alone that was being evaluated, but rather triage introduced and delivered under trial conditions. This adds another complication to interpreting the findings. The demands of the trial introduced confounding factors, additional pressures and strictures to be negotiated at the same time as introducing a major organisational change. A further difficulty is that our quantitative data were collected at a particular point in time. In almost all cases this was fairly soon after the trial began, when telephone triage was new and, for most practices, challenging, depending on people's individual, collegial and institutional situations. Yet innovations take time to settle down. How a phenomenon is experienced is mediated by the dynamics of time and experience and by its gradual normalisation. Findings may have been different if data had been collected at a different stage in the process, perhaps when staff had had time to become more accustomed to telephone triage and felt more competent and confident. Indeed, by the time process evaluation fieldwork was undertaken, participants were already contrasting what they or others felt 'at first' with how they felt at the time of the interview.

Perhaps the most striking finding from this extensive qualitative study is that there is no strong and compelling narrative about what works and what does not work in relation to telephone triage in primary care. There are no predictable patterns of causality from which messages about effectiveness can be abstracted in relation to this particular organisational change. Although GPT 'works' to manage demand for some doctors in some practices, other doctors, sometimes in the same practices, see themselves transformed by triage into 'glorified booking clerks' with an even higher workload. Similarly, while NT 'works' in some practices it is not well regarded elsewhere. One can interrogate the data in vain in search of causal explanations.

Where NT was judged not to have been successful, we witnessed evidence of poor communication practices. In one particular setting, there was anxiety and uncertainty on account of recent substantial medical staff changes. In addition, difficulty with implementation was sometimes observed to be associated with poor perceptions of pre-existing appointment systems, problems in internal staffing and relationships (including poor relationships with patients), lack of consensus about trial participation among the practice team, and questions about the motivation to participate in the study. On occasions, the process evaluation team observed that relationships between staff and patients appeared strained and that some patients had to pay phone charges to make appointments. In one setting where NT was judged not to be successful, no extra resources had been provided to the nurses performing the triage, and nursing staff felt that the doctors had not discussed the trial with them during the course of implementation. However, many of the same factors – a chaotic appointment system before the trial, poor communication within the practice, continuity issues and no additional nursing resources – prevailed in other NT practices, at which NT was judged to have been successful, to have reduced doctors' appointments by half and to have been experienced positively by most staff and patients. How can this be explained?

One approach would be from the theoretical framework of complexity theory. Here innovation within organisations is seen as an organic and adaptive process in a continuous state of flux and adaptation as the organisations adapt and innovations interact with each other in perpetual iteration.⁷⁷ Such an approach rejects the notion that relationships of cause and effect can be identified by breaking the organisation and the innovation down into component parts.⁷⁸ Instead, health-care organisations are understood as complex adaptive systems, which can be made sense of only by understanding the interplay of relationships within microcontexts that are not controllable by an external agent. Methodologies that enable the capture of microcontextual factors in time and place, such as ethnographic, ethnomethodological and linguistic approaches, are recommended to complement the more traditional approach to gathering information of the meaning and experience of this intervention for participants. Process evaluations and other research into organisations may be enriched by a more diverse methodology.

The findings from the process evaluation demonstrate the complexity of primary care organisations, and the significance of individual practice cultures. No widely applicable causal conclusions about what works and what is acceptable have emerged. The process evaluation has provided rich data on the experience of different models of triage, and how these models were adapted to local circumstances. Our findings demonstrated that both models of triage were sometimes experienced positively by staff and patients. It has also raised issues that any practice considering implementing triage might want to reflect on before proceeding.

Strengths and limitations

The ESTEEM trial has a number of strengths. The trial was a rigorously conducted multicentre cluster RCT, using both quantitative and qualitative methods and in full conformity with the CONSORT guidelines.⁵⁷ It was a definitive study, including a large number of both practices and patients with few exclusion criteria and a spread of geographical locations, increasing the external validity of the trial and the generalisability of the findings. Although we succeeded in recruiting large numbers of practices and patients from diverse geographical locations, enhancing the generalisability of our results, our findings may be less applicable to practices serving populations with greater ethnic diversity, or those located in inner-city areas with very high levels of deprivation. Additionally, the organisation of the practices to run the interventions in three 'waves' allowed us to undertake the trial over a spread of seasons over 2 years. The study used a remote web-based randomisation system for the allocation of practices and we conducted the analyses according to a predefined analysis plan agreed with the Trial Steering Committee. The trial was fully powered on a prespecified primary outcome and we achieved our recruitment target. The cluster randomised design prevented contamination.

Our assessment of feasibility was based on a conservative estimate of just 20% of patients requesting same-day consultations. In the event, and although based on rather crude data (see *Table 8*), around 40% of patients probably presented same-day consultation requests, more in line with relevant published literature⁴ (of which we were aware but had considered as an estimate only).

The design of the study to include a pilot and feasibility phase with a built-in observational study prior to the main trial phase proved invaluable for developing the trial materials and procedures to be rolled out across the four centres. PPI contributed to both the development of patient recruitment materials and membership of the Trial Steering Committee.

Our original patient recruitment target in gaining access to the primary outcome data was exceeded owing to the introduction of a process, borne out of the pilot study, involving an initial verbal consent followed by a written communication giving patients the opportunity to opt out. We were grateful for the understanding and support of the regional REC in this regard, and see this understanding as the demonstration of critically important insight for future similar research initiatives. With this amendment, we were able to document our primary outcome in 77% of trial participants, rather than the 53% we originally hoped to seek.

Despite the challenge of the collection of our primary outcome data being undertaken by different study personnel across the four centres, we repeatedly demonstrated a high level of agreement between the four trial researchers with respect to the collection of primary outcome-related material extracted when undertaking the case note reviews at different study time points.

A further strength of the study is the parallel process evaluation undertaken to provide context to the main outcomes for the study and to assess the reasons why telephone triage is seen to work or not work.

Practical challenges were encountered in conducting ESTEEM. Recruitment of general practices to this complex trial was a particular challenge. Even when recruited, practice retention – notably in NT – proved problematic with 12 out of 54 practices withdrawing from the study, 10 of these having been randomised to deliver NT. Time and practical constraints of operating the trial limited our ability to exactly match replacement practices on the stratification variables of practices that chose to withdraw. We did not have a large pool of substitutes from which to select replacements (sometimes only one replacement was available). It is important to note that no practices withdrew once the intervention had begun and ‘live’ data collection established; our sense was that practices fully appreciated the size of the challenge of delivering NT only when forced to consider the practical realities of staffing, nurse availability and system redesign following randomisation. However, there were considerable time and cost implications of practices deciding to withdraw after having undertaken training and preparation work. Considerable care was expended to ensure that replacement practices remained blind to their allocation right up to and including the point at which they agreed to participate, thus ensuring allocation concealment. Given all of these considerations, we thus feel that our methods of replacing practices, although pragmatic, did not introduce bias.

Our estimates of patient eligibility at entry to the trial were based on data from Receptionist’s Log Sheets. This proved a challenging part of our data collection; reception staff members were observed to be extremely busy, and we have no reliable objective measure of total receptionist workload (and thus can only provisionally estimate the proportion of eligible patients). The proportion of patients excluded from the trial was observed to be highest in NT, raising the possibility that reception staff in some way provided an informal ‘filter’ with respect to recruitment to the trial. Being alert to this, we invested substantially in this area, conducting a series of ‘integrity checks’ prior to ‘live’ data collection to ensure that at least 75% of eligible patients were receiving the relevant intervention.

The trial placed a considerable burden on practice staff, including relatively short but intense periods of completing data collection forms and liaising with the research team. The NT practices experienced the burden of incorporating CDSS into their daily work. In considering the fidelity of the trial a key issue to consider is what 'dose' of triage was actually delivered. We did attempt to clarify the extent to which the CDSS had been used within the nurse telephone triage process (see *Appendix 14*). For example, we explored the number of questions that had been used and the duration for which the CDSS was 'open' during the course of a consultation, and derived a measure of 'intensity' of CDSS use combining the number of questions asked and the duration of the consultation. These are areas that would benefit from further investigation – an important consideration would be the possibility that 'NT' was seen to work or not to work without having a full or clear understanding of the extent to which CDSS was used or not used.

In addition, the deliberate recruitment of practices with little or no triage experience to undertake the trial in order to assess the impact of orchestrating a triage system may have meant that lack of confidence in triaging or organising the system, combined with the relatively short-term commitment to complete the trial, may have influenced the way the interventions were implemented within practices.

We identified a potential issue regarding practices running triage for a relatively short data collection period (approximately 5 weeks, or until the patient recruitment target was reached) and our proposed 28-day follow-up period for these patients. Patients entering the trial in the initial weeks of the data collection period may have had subsequent care within 28 days that was managed under the triage system (i.e. subsequent care dealt with by triage if further same-day consultations were requested, possibly for the same problem), although patients entering towards the end of the data collection period would do so only if the practice opted to continue the triage arrangement beyond the trial. If a difference was observed between arms in the tendency to revert to the previous 'UC' arrangement after the end of the data collection period, there would be concern over whether this would introduce bias to the primary outcome. It would not have been possible to fund the resultant NHS excess treatment costs of all triage practices to continue triaging for the whole of the follow-up period. As such, we requested that all triage practices revert to their UC arrangements after the end of the data collection period, at least until the final outcome data were measured. All but two of the triage practices agreed to do this; however, we have taken this into account within our sensitivity analyses and found it to have no effect on our results.

When complex interventions, such as we have outlined, might be introduced across the wider health-care system, it is of importance to recognise the need for a reasonable length of follow-up before providing a determination of 'success' or 'failure' with respect to the intervention itself. Taking a view of resource use just with respect to the index day would have led to very different conclusions from the broader '28-day view' that we took with respect to determining trial outcomes. On a related matter, it would have been of interest and potential importance to undertake a longer period of follow-up within practices but we felt that this was not practical on account of the uncertain nature of any benefit for patients or practices, and of the very substantial organisation implications incurred by practices across all their systems when introducing triage or agreeing to participate in the research. Ideally, however, a follow-up period of 6 months or 1 year would have been useful and it is possible that this may have led to different conclusions, either supporting or opposing the use of triage systems.

In collection of our primary outcome data, it was not possible for the trial researchers to be blinded, as the trial researchers provided training and support to practice staff during the trial. However, all trial analyses were undertaken by an independent statistician who was blind to practice allocation and followed a statistical analysis plan that was elaborated a priori.

Implications for future research

The core design involved a pilot and feasibility phase, with clear stopping rules elaborated a priori, before moving to the full trial phase. The pilot phase provided important learning, but a longer pilot phase, and, in particular, a longer gap between the pilot phase and the main trial, would have been helpful. No substantial gap had been scheduled between these two phases of the research. The two-phase approach taken represents, in our opinion, an excellent opportunity and model for public support of trials but does create problems in relation to staffing and resourcing models for the trial.

We have previously referred to the flexible, pragmatic and helpful support we obtained from the REC with whom we had regular exchanges, especially in the early days of the study. Their agreement to a revised consenting process underpinned the success of this study. Without the approach they adopted, it is likely that the proportion of patients for whom we were able to perform case note review would have been much smaller (and thus we would not have secured our POM). The revised consenting process to which the REC agreed was possibly somewhat unusual in the UK, involving, as it did, verbal consent to case note review being recognised as 'consent'. We supplemented this process with an opportunity to opt out using follow-up written consent, but the use of verbal consent is important and underpinned the success and generalisability of the study. Ethics committees face challenges when receiving requests such as this from researchers. But when research is carefully designed and has had relevant PPI in the design and implementation of the research, and where the work is undertaken in close collaboration with the REC, we believe that this pragmatic approach supports publicly funded research in a way that is commendable. This is an area where guidance might usefully be given to ethics committees. An alternative decision not to grant the amendment supporting the revised consent process would have been potentially very damaging to this study and incurred major staffing and academic disruption as well as significant loss of public finances.

Our trial involved a cluster randomisation design, with clusters being designated at the level of general practices. Although the RCT design is seen as providing potentially high-quality evidence, when undertaking evaluations of complex organisational interventions in primary care settings (as accomplished in this trial), alternative approaches are possible – possibly based on ethnographic or realist approaches. Our parallel process evaluation incorporated important elements of ethnographic investigation in both the pilot and main trial phases. Were we to undertake a similar evaluation in the future, we would propose adopting a similar research design to that which we adopted here, using a more diverse methodology.

Our estimates of sample size were based on a number of assumptions including the proportion of patients that might be anticipated to present same-day consultation requests. Our actual estimate (around 40% of patients presenting consultation requests) was based on reception staff completing a log sheet. In general, we found that the ability of receptionists to undertake such tasks, while continuing to deliver an effective service in what we acknowledge is a pressured environment, was limited, as reflected in the number of missing data encountered in such log sheets. In retrospect, however, a similar future study might reasonably estimate this proportion as around one-third (33%). In addition, although we had reimbursed practices for receptionist time in contributing to the study, we might have more actively ensured that dedicated receptionist time was indeed made available to support the trial processes.

The outcomes selected were based on previous research and in the event proved generally acceptable, workable, and, we believe, appropriate to the investigation. Our primary outcome – the number and rate of contacts incurred in primary care settings in the 28 days following an index consultation request for a same-day, face-to-face appointment with a GP – was based on previous research undertaken by one of the co-applicants.³ Extracting data informing the POM involved undertaking a large number of reviews of patient case notes. This involved prior consent from the patient with agreement to review their case notes – provided in around 77% of cases. A range of secondary outcomes were also investigated including outcomes relating to patient safety (deaths, A&E attendances, hospital admissions) and patient experience of care. These secondary measures also proved generally acceptable and workable. We identified important

inconsistencies with respect to the documentation of walk-in centre attendance – this documentation varying between regions. Considering that there was no reason why this should be systematically different across the three trial arms, we elected to use walk-in centre attendance based on case note review data. Walk-in centre attendance overall, however, was low among trial participants. We were also unable to effectively capture any contacts with the national nurse-led telephone advice service NHS Direct, these contacts not being routinely recorded in the patient record. Once again, we did capture information on patient self-reports of NHS Direct usage; further work is required to investigate the validity and reliability of these self-reported measures of resource use.

With regard to the inferential analyses, the appropriate ICC or cluster-level SD (approximating the coefficient of variation) is reported, which may prove useful for other researchers undertaking similar work.

Our process evaluation identified that practices participated in this trial for a range of reasons. One important reason that may be of relevance to future research was that the joint engagement in a research project provided a degree of cohesion and 'common purpose' with respect to the practice team. For participating practices at least, the research question was seen as relevant and of importance – the research design appeared of interest, and the research process itself seemed to offer specific advantages to potential participant practices.

We used an external consultant to support practices in the early stages of reviewing workload, appointment systems and internal practice processes with respect to delivering care to patients making same-day appointment requests. The external consultant who supported this study visited the practices, reviewed practice workload in detail and provided advice on reconfiguring appointment systems to allow for the implementation of triage (where this was appropriate). In the early stages of the trial, we did recognise that the external consultant may have been at risk of promoting a slightly different model of care from that we intended. The consultant's communication style was persuasive and attractive to a number of the practices, and, for this reason, to ensure fidelity with respect to the intervention, we felt it appropriate to discuss this specific issue with the consultant. That discussion was well received and we are confident that the changes introduced by practices were both relevant and appropriate to the design and delivery of the ESTEEM intervention.

We encountered problems for receptionists in relation to the routine recording of information regarding patients who were attending. Asking extremely busy receptionists to undertake routine documentation of workload is extremely challenging, and possibly simply non-viable on anything but a relatively small scale. It would have been desirable to define the study population in detail throughout the duration of data collection, but this would involve routine documentation of all patients requesting appointments with a doctor, and then defining those who were requesting same-day appointments in a consistent way. Although this may appear a straightforward task, given the busyness of the 'front end' of primary care and the considerable responsibilities that receptionists bear, sometimes under extremely stressful and challenging conditions, routine data capture involving receptionists may be extremely difficult. To address this, our intention was to explore the possibility of using electronic computerised means of defining the sample population using a single key stroke entry attributed to a patient requesting a same-day appointment. Primary care computer systems are both diverse and complex and, despite our best efforts, we were unable to achieve this within the time frame of the trial.

We did explore electronic capture of the above information and related timings information with respect to individual consultations but found this approach not sufficiently reliable across various IT platforms to result in effective data capture. We were grateful for the financial support of the Comprehensive Local Research Network in delivering and exploring this. Following the investment of considerable effort with a commercial academic partner, we believe we almost achieved this but could not deliver a sufficiently robust system within the time frame of the trial to support the trial processes. This would be an important area for future development, and one that could be achieved, potentially, at only modest cost. The routine capture and identification of samples of relevant patients by receptionists, and the capture and extraction

of timings data from routine electronic patient records, are both important and relevant issues for future research.

The ESTEEM trial has specifically addressed triage-based approaches to managing the consultation requests of individuals seeking same-day appointments. Two further related areas are worthy of further research:

1. Extending the reach of the research to incorporate all individuals presenting consultation requests, such as is proposed in 'Doctor First' models of care (patient access: www.patient-access.org.uk/2013; productive primary care: www.productiveprimarycare.co.uk/2013).
2. Exploring the use of current alternatives to telephone-based consultation for such individuals, possibly exploiting the potential of video (Skype™; Skype Ltd, Rives de Clausen, Luxembourg), text (e-mail) or social media-based approaches [such as Facebook (Cambridge, MA, USA), or Twitter (San Francisco, CA, USA)].

Although ESTEEM has highlighted the overall acceptability to patients and practitioners of triage-based approaches, further research might explore the advantages/disadvantages to specific population groups [e.g. on the basis of age, ethnicity, language, educational attainment or setting (urban/rural)] of remote or IT-based approaches to managing health-care requests.

Implications for health care

The ESTEEM trial has, we believe, provided important research evidence regarding the clinical effectiveness and cost-effectiveness of telephone triage for managing same-day consultation requests in general practice. The specific context is important – our target population was individuals, or their representatives, making requests for same-day face-to-face appointments with their GP. This particular population represents around 30% of general practice-based primary care workload,⁴ and, for the reasons we have outlined, we believe the findings are generalisable to a large section of the UK primary care population. However, they do not focus on other patients requesting appointments with a GP, such as those wishing to book appointments in advance, or those attending GP-led chronic disease management clinics, etc. There may be important potential to explore the use of telephone triage in consultation for this larger group of patients, although the design and implementation of any such study would potentially be usefully informed by the research findings we present.

Two major models of 'GP First' care are currently promoted in the UK by quasi-commercial organisations.^{79,80} Interest in these models has also been expressed by the 'Productive Primary Care' workstream of the NHS Institute for Innovation, and further investigation of the potential for such models has recently been proposed by the NIHR Health Services Delivery Research Board.⁸¹

The summary of our research findings suggests that, overall, triage, whether performed by a GP or by a nurse using CDSS, should be introduced with full awareness of the whole-system implications arising from the decision to implement such a process. Although GP face-to-face contacts on the index day were reduced, taking a broader, 28-day perspective, overall primary care contacts increased, even after accounting for the inevitable additional consultation incurred by the introduction of triage on the index day. Despite these changes in workload and in work distribution overall, NHS costs per patient across the three arms of the trial were almost identical. Triage of both types appears safe and no differences in patient health status were observed. Furthermore, NT appeared somewhat less acceptable to patients than either GPT or UC. We feel that the marginal reduction in satisfaction in NT is understandable given that patients requesting a face-to-face consultation with a GP actually received a telephone consultation with a nurse. Having said that, it is important to recognise the potential benefits that some nurses reported following their engagement in the nurse-triage arm of the trial. Some nurses reported positive professional benefits, describing the use of skills that they felt were otherwise underexploited. We did not incorporate a measure of job satisfaction in our study but this might be an important consideration for any similar work. In some

settings, the adoption of GPT or NT processes may be seen to offer a solution to managing workload pressures, although careful consideration needs to be given to address patients' experience of care, and to address potential knock-on effects that might arise elsewhere in primary care following introduction of a triage system.

When whole-system redesign is anticipated, careful consideration needs to be given to both the human resource management and the electronic systems support required to effectively deliver the intervention. Computer systems – as they relate to primary care – are complex and diverse; there are at least eight major primary care computer systems commonly in use across the UK. Working with system developers to exploit the potential of both the electronic record and practice computer administrative system represents an important but underexploited opportunity in the UK. The CDSS we adopted has been widely used across the UK in supporting out-of-hours primary care and NHS Direct. It has not previously been widely adopted or researched within the context of in-hours primary care. Overall, the CDSS integrated reasonably well with established practice computer systems but further work may need to be done in this regard and across a wider range of primary care computer systems than we were able to exploit in this study.

The data that we have reported relate to the original research questions posed and the a priori statistical and health-economic analyses defined. Many other areas of investigation might be addressed within a secondary analysis of this data set, for example examining the potential access issues or patient experience outcomes for subgroups of the sample population. Triage systems appeared generally acceptable and, in the broadest terms, similar to UC with respect to patients' reports of their experience of care; however, we have not yet had the opportunity to examine the experience of important subgroups of our sample population – for example younger or older people, parents or individuals who were unable to attend the practice within regular hours (or found it difficult to do so) owing to work commitments. All of these might be expected to have substantially different views from the overall population with respect to their experience of care, and this represents just one of many further analyses that might be undertaken with respect to the current data set.

General practitioners and their teams might wish to ask 'Should we introduce triage?'. The evidence from ESTEEM is summarised in *Table 36* for the main effects that might reasonably be anticipated following the introduction of either GP- or nurse-led triage. The findings from the process evaluation suggest that the data presented in *Table 36* should be carefully considered in the light of practice organisation and the particular local circumstances prevailing. Introducing triage, whether GPT or NT, was associated with an increased primary care contact rate of 33% in GPT, and 48% in NT (10% and 12%, respectively, after accounting for the inevitable additional contacts incurred as a result of triage). However, the nature and professional distribution of the work is likely to change, although overall costs to the NHS (over 28 days) should remain unchanged.

In particular, introducing GPT is likely to reduce the number of GP face-to-face consultations by around 40% over a 28-day period, partially offset by an increase in the number of telephone consultations. Introducing NT is likely to result in reduced numbers of GP face-to-face contacts by around 20% over 28 days but with a substantial increase in nurse contact time on the index day. These findings might have different relevance for different practices. For example, if the priority is to reduce demand on GPs then NT might be appropriate. If the priority is to control demand for GP face-to-face appointments then GPT might be appropriate; although it incurs a similar amount of work in total, the timing is more in the GPs' control. Unanticipated and whole-system effects should be taken into account when considering the introduction of clinician triage. Clinician triage of patients seeking same-day consultations with a GP may offer substantial advantages in supporting the delivery of patient care, and potentially offers a useful approach in the armamentarium of tools facilitating the delivery of effective NHS primary care.

TABLE 36 Summarising evidence from ESTEEM considered in reference to UC

	GPT	NT
Primary care contact rates (%)		
Counting all contacts on index day as separate consultations	↑ 33	↑ 48
Counting all within-practice contacts on index day as a single consultation	↑ 10	↑ 12
GP face-to-face contacts on index day only	↓ 55	↓ 31
GP face-to-face contacts over 28 days following index consultation request	↓ 39	↓ 20
GP telephone and face-to-face contacts on index day only	↑ 49	↓ 28
GP telephone and face-to-face contacts over 28 days following index consultation request	↑ 38	↓ 16
Patient–clinician contact time (minutes) on index day (estimate)		
Overall GP and nurse contact time (for a patient making a same-day consultation request in UC, 9.6 minutes)	↑ 0.7	↑ 5.2
Overall GP contact time (for a patient making a same-day consultation request in UC, 9.1 minutes)	↓ 0.1	↓ 1.4
Overall nurse contact time (for a patient making a same-day consultation request in UC, 0.6 minutes)	↑ 0.7	↑ 6.5
Safety		
A&E attendance in 28 days following index consultation request	→	→
Emergency hospital admission rate in 7 days following index consultation request	→	→
Mortality rate in 7 days following index consultation request	→	→
Patient experience of care		
Overall satisfaction	→	(↓) ^a
Health economics		
Cost per patient	→	→
↓, decreased; ↑, increased; →, unchanged.		
a Marginal reduction only.		

Acknowledgements

Contributions of authors

Professor John L Campbell (Professor of General Practice and Primary Care; chief investigator) took overall responsibility for the study; supervised trial conduct; was involved in all stages of the ESTEEM trial, including conception, design, analysis and interpretation of data; drafted and critically revised the report for important intellectual content and approval of the final version.

Mrs Emily Fletcher (trial manager) was responsible for the day-to-day management of the study and acquisition of data; coordinated research staff across sites; contributed to interpretation of data; drafted and critically revised the report for important intellectual content and approval of the final version.

Professor Nicky Britten (Professor of Applied Healthcare Research; process evaluation lead) took overall responsibility for the process evaluation and was involved in all stages of the ESTEEM trial, including conception, design, interpretation of data and critically revising the report for important intellectual content and approval of the final version.

Professor Colin Green (Professor of Health Economics; economic evaluation lead) took overall responsibility for the economic evaluation; led the economic analysis; and was involved in all stages of the ESTEEM trial, including conception, design, interpretation of data, and drafting and critically revising the report for important intellectual content and approval of the final version.

Dr Tim Holt (NIHR academic clinical lecturer; principal investigator for general practices near Warwick) supervised the trial and research staff in this region and was involved in all stages of the ESTEEM trial, including conception, design, interpretation of data, and critically revising the report for important intellectual content and approval of the final version.

Professor Valerie Lattimer (Head of School of Nursing Sciences, Professor of Health Services Research; principal investigator for general practices near Norwich) supervised the trial and research staff in this region and was involved in all stages of the ESTEEM trial, including conception, design, interpretation of data, and critically revising the report for important intellectual content and approval of the final version.

Professor David A Richards (Professor of Mental Health Services Research; co-applicant) was involved in all of the stages of the ESTEEM trial, including conception, design, interpretation of data and critically revising the report for important intellectual content and approval of the final version.

Dr Suzanne H Richards (senior lecturer; co-applicant) was involved in all stages of the ESTEEM trial, including conception, design, interpretation of data, and drafting and critically revising the report for important intellectual content and approval of the final version.

Professor Chris Salisbury (Professor in Primary Health Care; principal investigator for general practices near Bristol) supervised the trial and research staff in this region and was involved in all stages of the ESTEEM trial, including conception, design, interpretation of data and critically revising the report for important intellectual content and approval of the final version.

Professor Rod S Taylor (Professor of Health Services Research; statistical lead) took overall responsibility for the main trial analysis and was involved in all stages of the ESTEEM trial, including conception, design, interpretation of data, and drafting and critically revising the report for important intellectual content and approval of the final version.

Dr Raff Calitri (research fellow; trial researcher in Exeter) operationalised the protocol, was involved in acquisition of data, contributed to trial management responsibilities and contributed to additional analyses, interpretation of data, and drafting and critically revising the report for important intellectual content and approval of the final version.

Ms Vicky Bowyer (research fellow; trial researcher in Warwick) operationalised the protocol, was involved in acquisition of data and contributed to critically revising the report for important intellectual content and approval of the final version.

Dr Katherine Chaplin (research fellow; trial researcher in Bristol) operationalised the protocol, was involved in acquisition of data and contributed to additional analyses and critically revising the report for important intellectual content and approval of the final version.

Miss Rebecca Kandiyali (research fellow; health economics researcher) was involved in acquisition of data and contributed to the economic analysis and the drafting/approval of the final version of the report.

Dr Jamie Murdoch (senior research fellow; trial researcher in Norwich) operationalised the protocol, was involved in acquisition of data and contributed to critically revising the report for important intellectual content and approval of the final version.

Ms Linnie Price (associate research fellow; process evaluation researcher) operationalised the protocol, was involved in acquisition of data, undertook the qualitative analysis and interpretation of data and contributed to the drafting and approval of the final version of the report.

Ms Julia Roscoe (associate research fellow; trial researcher in Warwick) operationalised the protocol, was involved in acquisition of data and contributed to critically revising the report for important intellectual content and approval of the final version.

Mrs Anna Varley (research fellow; trial researcher in Norwich) operationalised the protocol, was involved in acquisition of data and contributed to critically revising the report for important intellectual content and approval of the final version.

Dr Fiona C Warren (lecturer; statistician) contributed to the development of the statistical analysis plan, undertook the main trial analysis and contributed to drafting and critically revising the report for important intellectual content and approval of the final version.

We would like to thank the external members of the Trial Steering Committee for their advice and support for the project: Professor Brian McKinstry (University of Edinburgh), Professor Sandra Eldridge (Queen Mary University of London), Professor Brendan Delaney (King's College London), Dr Louise Locock (University of Oxford), Mrs Barbara Tilbury (Patient Representative) and Mr Geoff Barr (Patient Representative). Our thanks also go to the Data Monitoring Committee comprising Professor Bruce Guthrie (Dundee University), Dr Ben Carter (Cardiff University) and Professor Paul Kinnersley (Cardiff University). We are also grateful to our Patient and Public Involvement group for their advice and input throughout the trial.

We would like to thank both the staff and the patients of our participant general practices who gave generously of their time to support the study and complete data collection materials. We also thank staff within South West 2 Research Ethics Committee, primary care trusts, Comprehensive Local Research Networks and PCRN across the four regions, who provided valuable assistance to us throughout the study.

For their administrative support during the trial, we would like to thank Donna Poade, Amy Gratton, Julia Carver, Gillian Blunt, Vivienne Owen, Lynn Green and Lorna Burchell. Likewise, we thank Claire Telford, Frances Carpenter and Sue Rugg for their contribution as researchers during the course of the study.

We extend our thanks also to Plain Healthcare for providing the Odyssey PatientAssess CDSS to support the nurses' triaging decisions, and for delivering the training in using this software and technical support to practices. Likewise, we thank Dr Steve Clay and Mr Dillon Sykes from Productive Primary Care for providing training and advice to intervention practices in organising the triage systems and in GP triage skills. Thanks also to Jane Vickery, Laura Cocking and Elliot Carter at the Peninsula Clinical Trials Unit for their guidance and support in building and maintaining our study database and to CFEP-UK for their support with double data entry.

Publication

Campbell JL, Fletcher E, Britten N, Green C, Holt TA, Lattimer V, *et al.* Telephone triage for management of same-day consultation requests in general practice (the ESTEEM trial): a cluster-randomised controlled trial and cost-consequence analysis. *Lancet* 2014;**384**:1859–68.

References

1. Department of Health. *Delivering Investment in General Practice: Implementing the New GMS Contract*. London: Department of Health; 2004.
2. Salisbury C, Montgomery AA, Simons L, Sampson F, Edwards S, Baxter H, *et al*. Impact of Advanced Access on access, workload, and continuity: controlled before-and-after and simulated-patient study. *Br J Gen Pract* 2007;**57**:608–14.
3. Richards DA, Meakins J, Tawfik J, Godfrey L, Dutton E, Richardson G, *et al*. Nurse telephone triage for same day appointments in general practice: multiple interrupted time series trial of effect on workload and costs. *BMJ* 2002;**325**:1214–17. <http://dx.doi.org/10.1136/bmj.325.7374.1214>
4. Stoddart H, Evans M, Peters TJ, Salisbury C. The provision of 'same-day' care in general practice: an observational study. *Fam Pract* 2003;**20**:41–7. <http://dx.doi.org/10.1093/fampra/20.1.41>
5. Lattimer V, George S, Thompson F, Thomas E, Mullee M, Turnbull J, *et al*. Safety and effectiveness of nurse telephone consultation in out of hours primary care: randomised controlled trial. The South Wiltshire Out of Hours Project (SWOOP) Group. *BMJ* 1998;**317**:1054–9. <http://dx.doi.org/10.1136/bmj.317.7165.1054>
6. Marsh GN, Horne RA, Channing DM. A study of telephone advice in managing out-of-hours calls. *J R Coll Gen Pract* 1987;**37**:301–4.
7. Marklund B, Bengtsson C. Medical advice by telephone at Swedish health centres: who calls and what are the problems? *Fam Pract* 1989;**6**:42–6. <http://dx.doi.org/10.1093/fampra/6.1.42>
8. Dale J, Crouch R, Lloyd D. Primary care: nurse-led telephone triage and advice out-of-hours. *Nurs Stand* 1998;**12**:41–5. <http://dx.doi.org/10.7748/ns1998.08.12.47.39.c2520>
9. Office for National Statistics. *Living in Britain*. London: Office for National Statistics; 2008.
10. NHS Information Centre. *GP Workload Survey 2007*. 2007. URL: www.ic.nhs.uk/pubs/gpworkload (accessed 12 November 2014).
11. Jiwa M, Mathers N, Campbell M. The effect of GP telephone triage on numbers seeking same-day appointments. *Br J Gen Pract* 2002;**52**:390–1.
12. Richards DA, Meakins J, Godfrey L, Tawfik J, Dutton E. Survey of the impact of nurse telephone triage on general practitioner activity. *Br J Gen Pract* 2004;**54**:207–10.
13. McKinstry B, Walker J, Campbell C, Heaney D, Wyke S. Telephone consultations to manage requests for same-day appointments: a randomised controlled trial in two practices. *Br J Gen Pract* 2002;**52**:306–10.
14. Richards DA, Godfrey L, Tawfik J, Ryan M, Meakins J, Dutton E, *et al*. NHS Direct versus general practice based triage for same day appointments in primary care: cluster randomised controlled trial. *BMJ* 2004;**329**:774. <http://dx.doi.org/10.1136/bmj.38226.605995.55>
15. Thompson F, George S, Lattimer V, Smith H, Moore M, Turnbull J, *et al*. Overnight calls in primary care: randomised controlled trial of management using nurse telephone consultation. *BMJ* 1999;**319**:1408. <http://dx.doi.org/10.1136/bmj.319.7222.1408>
16. Lattimer V, Sassi F, George S, Moore M, Turnbull J, Mullee M, *et al*. Cost analysis of nurse telephone consultation in out of hours primary care: evidence from a randomised controlled trial. *BMJ* 2000;**320**:1053–7. <http://dx.doi.org/10.1136/bmj.320.7241.1053>
17. Brown A, Armstrong D. Telephone consultations in general-practice: an additional or alternative service. *Br J Gen Pract* 1995;**45**:673–5.

18. Poole SR, Schmitt BD, Carruth T, Peterson-Smith A, Slusarski M. After-hours telephone coverage: the application of an area-wide telephone triage and advice system for pediatric practices. *Pediatrics* 1993;**92**:670–9.
19. Holmstrom I. Decision aid software programs in telenursing: not used as intended? Experiences of Swedish telenurses. *Nurs Health Sci* 2007;**9**:23–8. <http://dx.doi.org/10.1111/j.1442-2018.2007.00299.x>
20. Richards DA, Meakins J, Tawfik J, Godfrey L, Dutton E, Heywood P. Quality monitoring of nurse telephone triage: pilot study. *J Adv Nurs* 2004;**47**:551–60. <http://dx.doi.org/10.1111/j.1365-2648.2004.03132.x>
21. Crouch R, Woodfield H, Dale J, Patel A. Telephone assessment and advice: a training programme. *Nurs Stand* 1997;**11**:41–4.
22. Derkx HP, Rethans JJ, Muijtjens AM, Maiburg BH, Winkens R, van Rooij HG, et al. Quality of clinical aspects of call handling at Dutch out of hours centres: cross-sectional national study. *BMJ* 2008;**337**:a1264. <http://dx.doi.org/10.1136/bmj.a1264>
23. Car J, Freeman GK, Partridge MR, Sheikh A. Improving quality and safety of telephone based delivery of care: teaching telephone consultation skills. *Qual Saf Health Care* 2004;**13**:2–3. <http://dx.doi.org/10.1136/qshc.2003.009241>
24. Currell R, Urquhart C, Wainwright P, Lewis R. Telemedicine versus face to face patient care: effects on professional practice and health care outcomes. *Nurs Times* 2001;**97**:35.
25. Giesen P, Ferwerda R, Tijssen R, Mokkink H, Drijver R, van den Bosch W, et al. Safety of telephone triage in general practitioner cooperatives: do triage nurses correctly estimate urgency? *Qual Saf Health Care* 2007;**16**:181–4. <http://dx.doi.org/10.1136/qshc.2006.018846>
26. Plain Healthcare. Odyssey Clinical Decision Support Software. 2008. URL: www.advancedcomputersoftware.com/ahc/ (accessed 12 November 2014).
27. Bunn F, Byrne G, Kendall S. The effects of telephone consultation and triage on healthcare use and patient satisfaction: a systematic review. *Br J Gen Pract* 2005;**55**:956–61.
28. Venning P, Durie A, Roland M, Roberts C, Leese B. Randomised controlled trial comparing cost effectiveness of general practitioners and nurse practitioners in primary care. *BMJ* 2000;**320**:1048–53. <http://dx.doi.org/10.1136/bmj.320.7241.1048>
29. Kinnersley P, Anderson E, Parry K, Clement J, Archard L, Turton P, et al. Randomised controlled trial of nurse practitioner versus general practitioner care for patients requesting 'same day' consultations in primary care. *BMJ* 2000;**320**:1043–8. <http://dx.doi.org/10.1136/bmj.320.7241.1043>
30. Shum C, Humphreys A, Wheeler D, Cochrane MA, Skoda S, Clement S. Nurse management of patients with minor illnesses in general practice: multicentre, randomised controlled trial. *BMJ* 2000;**320**:1038–43. <http://dx.doi.org/10.1136/bmj.320.7241.1038>
31. NHS Institute for Innovation and Improvement. *Stour Access System: A New Way to Manage GP Appointments. Better for GPs, Better for Patients, Better All-round*. 2008. URL: www.stoursurgery.co.uk/website/J00700/files/StourAccessSystemBrochure.pdf (accessed 12 November 2014).
32. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008;**337**:a1655.
33. Campbell JL, Britten N, Green C, Holt TA, Lattimer V, Richards SH, et al. The effectiveness and cost-effectiveness of telephone triage of patients requesting same day consultations in general practice: study protocol for a cluster randomised controlled trial comparing nurse-led and GP-led management systems (ESTEEM). *Trials* 2013;**14**:4. <http://dx.doi.org/10.1186/1745-6215-14-4>

34. Snooks HA, Carter B, Dale J, Foster T, Humphreys I, Logan PA, *et al.* Support and Assessment for Fall Emergency Referrals (SAFER 1): Cluster Randomised Trial of Computerised Clinical Decision Support for Paramedics. *PLOS ONE* 2014;**9**:e106436. <http://dx.doi.org/10.1371/journal.pone.0106436>
35. Campbell MK, Snowdon C, Francis D, Elbourne D, McDonald AM, Knight R, *et al.* Recruitment to randomised trials: strategies for trial enrolment and participation study. The STEPS study. *Health Technol Assess* 2007;**11**(48). <http://dx.doi.org/10.3310/hta11480>
36. Ukoumunne OC, Gulliford MC, Chinn S, Sterne JA, Burney PG. Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. *Health Technol Assess* 1999;**3**(5).
37. Association of Public Health Observatories (APHO). 2013. URL: <http://fingertips.phe.org.uk/profile/general-practice> (accessed 12 November 2014).
38. Curtis L. *Unit Costs of Health and Social Care 2012*. Canterbury: PSSRU; 2012.
39. EuroQol group. EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy* 1990;**16**:199–208. [http://dx.doi.org/10.1016/0168-8510\(90\)90421-9](http://dx.doi.org/10.1016/0168-8510(90)90421-9)
40. McHorney CAJ, Ware JE Jr, Lu JFR, Sherbourne CD. The MOS 36 item Short Form Health Survey (SF-36):III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 1994;**32**:40–66. <http://dx.doi.org/10.1097/00005650-199401000-00004>
41. Baker R. Development of a questionnaire to assess patients' satisfaction with consultations in general practice. *Br J Gen Pract* 1990;**40**:487–90.
42. Mercer SW, Howie JGR. CQI-2: a new measure of holistic interpersonal care in primary care consultations. *Br J Gen Pract* 2006;**56**:262–8.
43. NHS England/Ipsos MORI. *GP Patient Survey*. 2013. URL: www.gp-patient.co.uk/ (accessed 12 November 2014).
44. Bower P, Roland M, Campbell JL, Mead N. Setting standards based on patients' views on access and continuity: secondary analysis of data from the general practice assessment survey. *BMJ* 2003;**326**:258. <http://dx.doi.org/10.1136/bmj.326.7383.258>
45. Campbell JL, Roland MO, Richards S, Dickens A, Greco M, Bower P. Users' reports and evaluations of out-of-hours healthcare and the UK national quality requirements. A cross sectional study. *Br J Gen Pract* 2009;**59**:e8–15. <http://dx.doi.org/10.3399/bjgp09X394815>
46. Campbell JL. Patients' perceptions of medical urgency: does deprivation matter? *Fam Pract* 1999;**16**:28–32. <http://dx.doi.org/10.1093/fampra/16.1.28>
47. Howie JGR, Heaney DJ, Maxwell M. *Measuring Quality in General Practice. Occasional Paper No. 75*. London: Royal College of General Practitioners; 1997.
48. Adams G, Gulliford MC, Ukoumunne OC, Eldridge S, Chinn S, Campbell MJ. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *J Clin Epidemiol* 2004;**57**:785–94. <http://dx.doi.org/10.1016/j.jclinepi.2003.12.013>
49. Hippisley-Cox J, Fenty J, Heaps M. *Trends in Consultation Rates in General Practice 1995 to 2007: Analysis of the QRESEARCH database*. Nottingham: QRESEARCH; 2008.
50. Jacoby A. Possible factors affecting response to postal questionnaires: findings from a study of general practitioner services. *J Public Health Med* 1990;**12**:131–5.
51. Smith WC, Crombie IK, Campion PD, Knox JD. Comparison of response rates to a postal questionnaire from a general practice and a research unit. *Br Med J (Clin Res Ed)* 1985;**291**:1483–5. <http://dx.doi.org/10.1136/bmj.291.6507.1483>

52. Salisbury C. Postal survey of patients' satisfaction with a general practice out of hours cooperative. *BMJ* 1997;**314**:1594–8. <http://dx.doi.org/10.1136/bmj.314.7094.1594>
53. McColl E, Jacoby A, Thomas L, Soutter J, Bamford C, Steen N, *et al.* Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients. *Health Technol Assess* 2001;**5**(31). <http://dx.doi.org/10.3310/hta5310>
54. UK Government. *Indices of Multiple Deprivation 2010*. URL: www.gov.uk/government/publications/english-indices-of-deprivation-2010 (accessed 12 November 2014).
55. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012;**345**:e5661. <http://dx.doi.org/10.1136/bmj.e5661>
56. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, *et al.* Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;**337**:a2390. <http://dx.doi.org/10.1136/bmj.a2390>
57. Bennett S, Parpia T, Hayes R, Cousens S. Methods for the analysis of incidence rates in cluster randomized trials. *Int J Epidemiol* 2002;**31**:839–46. <http://dx.doi.org/10.1093/ije/31.4.839>
58. Pacheo GD, Hattendorf J, Colford JM, Mäusezahl D, Smith T. Performance of analytical methods for overdispersed counts in cluster randomized trials: Sample size, degree of clustering and imbalance. *Stat Med* 2009;**28**:2989–3011. <http://dx.doi.org/10.1002/sim.3681>
59. White IM, Royston P, Wood AM. Multiple imputations using chained equations: issues and guidance for practice. *Stat Med* 2011;**30**:377–99. <http://dx.doi.org/10.1002/sim.4067>
60. Lyratzopoulos G, Elliott M, Barbieri JM, Henderson A, Staetsky L, Paddison C, *et al.* Understanding ethnic and other socio-demographic differences in patient experience of primary care: evidence from the English General Practice Patient Survey. *BMJ Qual Saf* 2012;**21**:21–9. <http://dx.doi.org/10.1136/bmjqs-2011-000088>
61. Craig D, Rice S. *CRD Report 6: NHS Economic Evaluation Database Handbook*. York: Centre for Reviews and Dissemination, University of York; 2007.
62. Green J, Thorogood N. *Qualitative Methods for Health Research*. London: Sage; 2004.
63. Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. *BMJ* 2000;**320**:114–16. <http://dx.doi.org/10.1136/bmj.320.7227.114>
64. Campbell JL, Fletcher E, Britten N, Green C, Holt TA, Lattimer V, *et al.* Telephone triage for management of same-day consultation requests in general practice (the ESTEEM trial): a cluster-randomised controlled trial and cost-consequence analysis. *Lancet* 2014;**384**:1859–68. [http://dx.doi.org/10.1016/S0140-6736\(14\)61058-8](http://dx.doi.org/10.1016/S0140-6736(14)61058-8)
65. NHS Employers. 2014/15 General medical Services (GMS) Contract Quality and Outcomes Framework (QOF). *Guidance from GMS contract 2014/15*. 2014. www.nhsemployers.org/PAYANDCONTRACTS/GENERALMEDICALSERVICESCONTRACT/QOF/Pages/ChangestoQOF201415.aspx (accessed 1 November 2014).
66. Department of Health. *The NHS Plan*. London: HMSO; 2000.
67. Pope C, Banks J, Salisbury C, Lattimer V. Improving access to primary care: eight case studies of introducing Advanced Access in England. *J Health Serv Res Policy* 2008;**13**:33–9. <http://dx.doi.org/10.1258/jhsrp.2007.007039>
68. Huibers L, Smits M, Renaud V, Giesen P, Wensing M. Safety of telephone triage in out-of-hours care: a systematic review. *Scand J Prim Health Care* 2011;**29**:198–209. <http://dx.doi.org/10.3109/02813432.2011.629150>

69. Gallagher M, Huddart T, Henderson B. Telephone triage of acute illness by a practice nurse in general practice: outcomes of care. *Br J Gen Pract* 1998;**48**:1141–5.
70. Darnell JC, Hiner SL, Neill PJ, Mamlin JJ, McDonald CJ, Hui SL, *et al*. After-hours telephone access to physicians with access to computerized medical records. Experience in an inner-city general medicine clinic. *Med Care* 1985;**23**:20–6. <http://dx.doi.org/10.1097/00005650-198501000-00003>
71. Vedsted P, Christensen MB. The effect of an out-of-hours reform on attendance at casualty wards. The Danish example. *Scand J Prim Health Care* 2001;**19**:95–8. <http://dx.doi.org/10.1080/028134301750235303>
72. McKinstry B, Hammersley V, Burton C, Pinnock H, Elton R, Dowell J, *et al*. The quality, safety and content of telephone and face-to-face consultations: a comparative study. *Qual Saf Health Care* 2010;**19**:298–303. <http://dx.doi.org/10.1136/qshc.2008.027763>
73. Stirewalt CF, Linn MW, Godoy G, Knopka F, Linn BS. Effectiveness of an ambulatory care telephone service in reducing drop-in visits and improving satisfaction with care. *Med Care* 1982;**20**:739–48. <http://dx.doi.org/10.1097/00005650-198207000-00009>
74. Kelly M, Egbunike JN, Kinnersley P, Hood K, Owen-Jones E, Button LA, *et al*. Delays in response and triage times reduce patient satisfaction and enablement after using out-of-hours services. *Fam Pract* 2010;**27**:652–63. <http://dx.doi.org/10.1093/fampra/cmq057>
75. Smits M, Huibers L, Oude BA, Giesen P. Patient satisfaction with out-of-hours GP cooperatives: a longitudinal study. *Scand J Prim Health Care* 2012;**30**:206–13. <http://dx.doi.org/10.3109/02813432.2012.735553>
76. Leibowitz R, Day S, Dunt D. A systematic review of the effect of different models of after-hours primary medical care services on clinical outcome, medical workload, and patient and GP satisfaction. *Fam Pract* 2003;**20**:311–17. <http://dx.doi.org/10.1093/fampra/cm313>
77. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q* 2004;**82**:581–629. <http://dx.doi.org/10.1111/j.0887-378X.2004.00325.x>
78. Kernick D. Wanted: new methodologies for health service research. Is complexity theory the answer? *Fam Pract* 2006;**23**:385–90. <http://dx.doi.org/10.1093/fampra/cml011>
79. GP Access. *PatientAccess*. 2013. URL: www.patient-access.org.uk/ (accessed 12 November 2014).
80. Productive Primary Care. *Doctor First*. 2013. URL: www.productiveprimarycare.co.uk/doctor-first.aspx (accessed 12 November 2014).
81. National Institute for Health Research (NIHR). *Health Services and Delivery Research Programme*. 2013. URL: www.nets.nihr.ac.uk/programmes/hsdr (accessed 12 November 2014).
82. Chalder M, Wiles NJ, Campbell J, *et al*. A pragmatic randomised controlled trial to evaluate the cost-effectiveness of a physical activity intervention as a treatment for depression: the treating depression with physical activity (TREAD) trial. *Health Technol Assess* 2012;**16**(10). <http://dx.doi.org/10.3310/hta16100>

Appendix 1 ESTEEM trial: pilot study report

A 12-month, preliminary pilot cluster RCT and parallel process evaluation was undertaken in six practices, and aspects of this work were informative to the main trial protocol. The key learning outcomes included confirmation of the feasibility of recruiting practices, the feasibility and acceptability of the triage interventions, alterations required for the proposed data collection methods and the level of clustering of outcomes.

Methods

Six pilot practices were recruited across Devon and Bristol, and were randomly allocated to each of GP-led triage (GPT), nurse-led triage (NT) and usual care (UC). Practices received training on the triage interventions, as described for the main trial. Practices collected outcome data for a 2-week period of patient recruitment. During this period, practices were each asked to recruit 200 patients.

Data collection

Capturing the patient sample

Receptionists completed a handwritten log sheet, recording all telephone calls received on each day of the data collection period. In order to capture patient information, the receptionist completed a row of data for each patient who called to request a same-day appointment. For each call, the receptionist recorded their own initials, and the date and time of the call, and then completed four response options to determine eligibility for ESTEEM ('appointment request?' yes/no; 'same-day request?' yes/no; 'patient well enough for trial?' yes/no; 'patient communicates without difficulty?' yes/no). If all four response options were 'yes', the receptionist went on to complete the patient's unique practice identification number, date of birth and gender. A further column was completed to indicate if the eligible patient was added to the triage list (yes/no) and, if not, a reason was recorded for no triage.

The function of this daily log sheet was twofold:

1. *To provide a sample frame – an indication of the proportion of patients who were eligible for triage as well as those who were eligible and actually entered into the trial* Key demographic variables, including date of birth and gender, were captured so that we could assess qualitative differences between patients eligible and entered, and those eligible and not entered (as well as comparing against ineligible patients).
2. *To serve as a database of patients from which to manage research activities (e.g. questionnaire mail-outs)* A member of practice staff collected all log sheets together at the end of each day and assigned a study-specific Read Code to the electronic record of all patients who were marked on the log sheets as eligible for ESTEEM. This facilitated electronic searches of the practice systems to generate patient lists for the purpose of mailing follow-up study questionnaires and conducting case note reviews.

Clinician data collection forms

During both the pilot study and main trial, clinicians completed a short data collection form ('Clinician Form') at the time of the initial consultation following a patient's same-day request (triage or usual care contact). This form captured details of the consultation, including which health professional undertook the consultation, whether the patient 'did not attend', a measure of case mix, treatment and management options chosen (e.g. ordering of tests, recommending subsequent appointment or referral), and the start and end time of the consultation.

Questionnaires

Patients were sent a questionnaire 4 weeks after their index consultation. The questionnaire included questions on satisfaction and health status (described in *Secondary outcomes*). The covering letter contained a section for patients who did not wish to participate or receive any further contact in relation to the study. These patients were asked to send back a blank questionnaire in a prepaid envelope as an indication that they wished to have no further contact. These patients did not receive any further communication in relation to the study. The questionnaire contained a section at the end, which offered the opportunity for patients who were willing to complete a questionnaire to opt out of having their case note reviewed. Those who did not respond were sent a reminder letter with another copy of the questionnaire 2 weeks after the previous mail-out.

Transition from pilot to main trial

Certain important aspects of the methodology were altered as a result of observations during the pilot study. The key learning outcomes related to 'stopping rules', which included confirmation ahead of the main trial of the feasibility of recruiting practices, the feasibility and acceptability of the triage interventions, alterations required for the proposed data collection methods and the level of clustering of outcomes.

Stopping rules

Progression from the pilot study to the main trial was dependent on the achievement of predefined stopping rules. These were as follows:

1. *Recruitment of general practices* $\geq 70\%$ of target number (i.e. 29/42) of main trial practices signed up to participate in the main trial by the end of the pilot.
2. *Collection of outcomes* $\geq 55\%$ of patient-reported outcomes (questionnaire response rate) and $> 70\%$ of patient questionnaire respondents providing written consent to a case note review.
3. *Clustering* Design effect attributable to clustering should not increase the proposed sample size by $> 20\%$. The latter may have been satisfied through modification of the estimated ICC based on pilot data, altering the number of clusters (practices) recruited into the study, or a combination of both measures.

Table 37 illustrates the performance of the pilot study against these stopping rules and action taken to ensure progression to the main trial phase when any stopping rule was not met.

Main trial practice recruitment

A total of 29/42 (70%) practices were to be recruited by the end of the feasibility phase.

Figure 10 outlines progress with main trial practice recruitment at the time of completing the pilot study. We had recruited 38 participant practices, most of which are prepared to begin the trial in either spring or autumn of 2011. Recruitment to this complex trial was challenging, as participant practices were required to reorganise delivery of services at the reception desk, which was reliant on the commitment and understanding of all members of practice staff.

TABLE 37 Summary table of pilot study outcomes against stopping rules

Pilot study outcome	Stopping rule	Pilot study performance	Action taken
Main trial practice recruitment	70% (29/42 practices)	90% (38/42 practices)	None required
Questionnaire response	55%	38%	Shortened questionnaire (use of EQ-5D rather than SF-36 measure of health status) Included an incentive for respondents Included a second reminder mailing
Consent to case note review	70%	86%	Approval obtained from REC for clinicians to record patients' initial verbal consent to notes review at the end of the same-day contact, to be followed up by postal questionnaire at 4 weeks allowing opt-out
ICC	Pilot ICC requiring an increase in sample size for main trial by < 20%	Pilot ICC = 0.03 (vs. original estimate of 0.05)	No increase in sample size required

Common reasons for practices declining participation included:

- not being interested in the research question or running a triage system ($n = 13$)
- limited staffing capacity to take on trial or to operate the trial if randomised to a particular arm ($n = 11$)
- recent or imminent change to IT system ($n = 1$)
- partner surgeries getting involved in the trial ($n = 2$)
- declining without specifying a reason ($n = 16$).

Questionnaire response

The lower-than-anticipated questionnaire response rate observed during the pilot study (38% against the stopping rule of 55%) prompted a revision of the patient questionnaire, including incorporation of the EQ-5D measure of health status rather than the SF-36, to shorten the overall length and to incorporate a second reminder mailing. After the initial mailing of the questionnaire 4 weeks after the same-day consultation request, two reminders were sent to non-respondents at 6 and 8 weeks. In addition, the ethical approval granted at the beginning of the pilot included an option to provide an incentive for questionnaire respondents. As such, main trial patients received an additional leaflet within the questionnaire pack offering the chance of being entered into a prize draw of 20 prizes of £25 worth of shopping vouchers.

During the pilot study, obtaining patient consent to the case note review was dependent on the return of the questionnaire, as we had not yet incorporated a stage for clinicians to request initial verbal consent at the end of the index consultation.

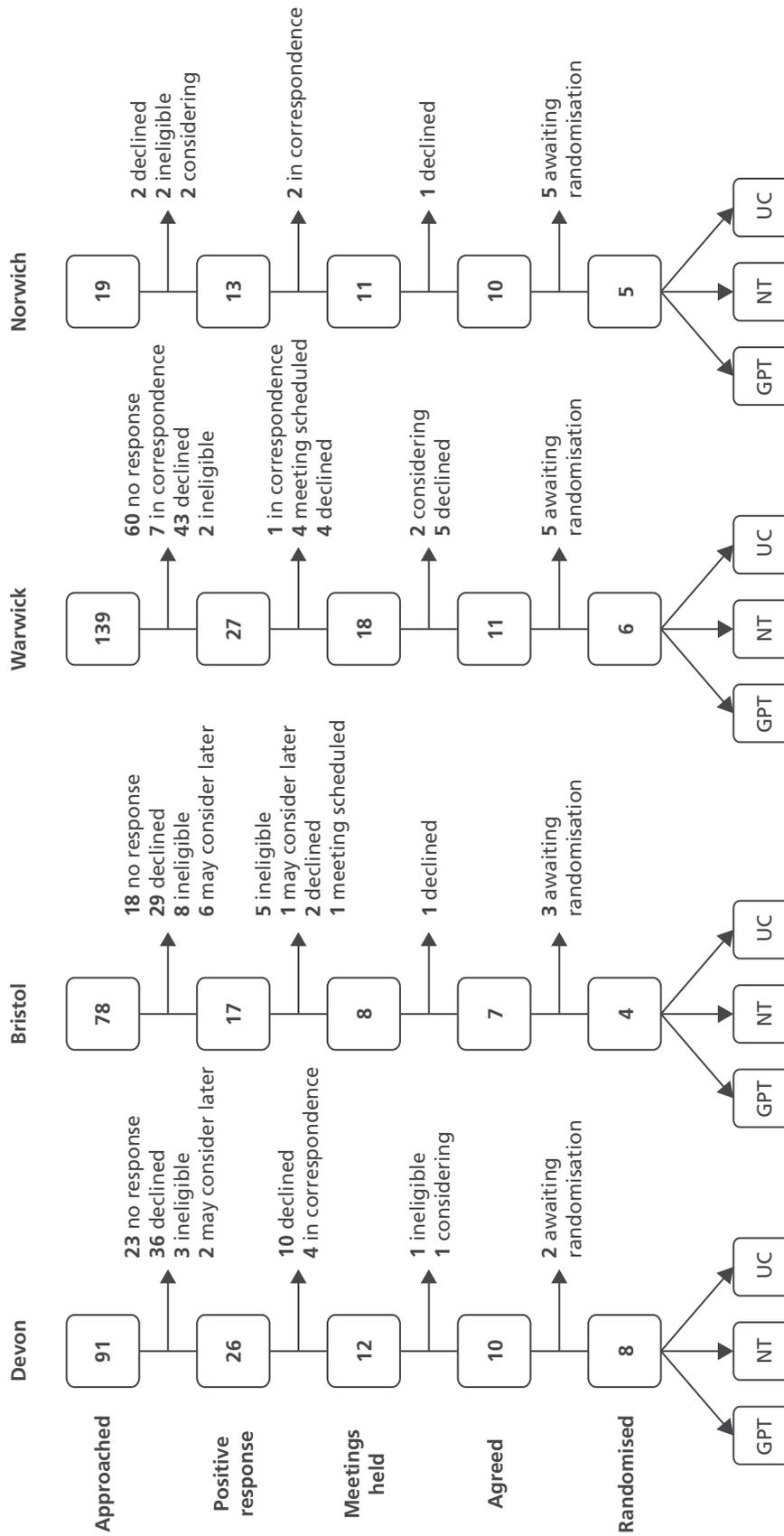


FIGURE 10 Main trial practice recruitment at end of pilot (July 2010 to February 2011).

Consent to case note review

The acquisition of initial verbal consent to the case note review by the clinician (with subsequent opt-out option within the postal questionnaire) is a departure from the original pilot methodology. The lower-than-anticipated pilot questionnaire response rate indicated that we may not be able to achieve sufficient numbers of main trial patients providing consent to the case note review if we were to be dependent on patients returning a questionnaire and providing written consent. The introduction of the verbal consent stage at the end of the index consultation ensured that we would secure sufficient numbers of patients for whom we could collect our primary outcome data.

Intracluster correlation coefficient

Against our original estimate of ICC = 0.05, proposed in our pilot study outline, the ICC based on the primary outcome is calculated as 0.03 (95% CI 0.00 to 0.08, see Stata output at end of this short report); our statisticians (Taylor, Warren) note: 'The original stopping rules were based on 55% questionnaire response, followed by 70% consent to notes review, i.e. approx. 38.5% of the original sample consenting to notes review. Even if we required a sample size of 5819 per arm (according to the [worst-case scenario] ICC of 0.08), this would have produced 2240 respondents consenting to notes review. With the current arm size of 3751, if we assume 85% consent to notes review based on verbal consent and questionnaire consent, this gives a total of 3188 consenting to notes review per arm, which is well above the worst-case ICC sample size.' We were therefore reassured by these calculations and saw this as attaining the fourth of the original stopping rules.

Method of capturing the patient sample

The pilot study allowed the development of a method to ensure correct patient identification. The pilot study initially involved receptionists completing a handwritten log sheet, recording all telephone calls received each day of the data collection period.

It became evident during the pilot study that despite standardised training in completing the log sheets, they were not completed consistently by receptionists within each practice and also across practices. Constraints on receptionist time, especially during periods of heavy demand, made completing the log sheet a real challenge. Accordingly, in the main trial the study database itself was subsequently populated directly from practice appointment records, which directly captured the patients being identified for the study, rather than being reliant on the daily handwritten receptionists' log sheets. We also introduced monitoring procedures to ensure high levels of patient identification (see *Chapter 2, Integrity checks*).

Issues reported to Health Technology Assessment at end of pilot

A number of factors experienced during the feasibility phase and some subsequent proposals are important to mention here, as follows:

Delay to start in Warwick and Norwich

The establishment of research staff in the recruitment sites of Warwick and Norwich has been affected by the extension to the feasibility phase, to allow for the development work for the pilot study in Exeter and Bristol to proceed. All appointments were delayed by 1 month, such that a researcher and administrator at each site would have been in post during June and December 2010, respectively. Difficulty in appointing to these posts has meant that the Warwick researcher (Vicky Bowyer) and the Norwich administrator (Anna Varley) started in October 2010. The post of administrator in Warwick is yet to be filled, as a result of organisational changes within the University of Warwick, and the post of researcher in Norwich is being covered in the interim by Jamie Murdoch, a Research Fellow in Urgent Care at University of East Anglia (UEA). In addition to this, Professor Lattimer's move to UEA from the University of Southampton resulted in the transferral of all trial arrangements from Southampton to UEA and the reapplication for research and development (R&D) and funding approvals in PCTs around Norwich. Recruitment of general practices in

both Warwick and Norwich has presented a challenge as a result of these factors. However, staff at these sites have worked very hard and with enthusiasm to produce encouraging progress.

Excess treatment costs from Primary Care Trusts

Significant difficulty has been experienced in agreeing Excess Treatment Costs from a number of PCTs from whom R&D approval has been obtained. Without secure agreement of this funding, recruitment of practices and operation of the trial will be at risk. Where funding has been agreed, an approach has been made to each PCT for reassurance that the agreed funds will be made available to the research team, and responses are awaited. A significant shortfall remains (£88,000) and a subvention application is being prepared for submission to DH, following advice from Trudi Simmons (DH R&D).

Election and White Paper

During the feasibility phase, the election of the coalition government, particularly with the release of the White Paper in July 2010, is likely to affect ongoing recruitment and may influence operation of the trial during 2011 and 2012.

Ethics

We undertook substantial pretrial work with respect to ethics (which was funded internally at cost to the Department but at no cost to the HTA). This meant that we secured ethical agreement for the trial at a very early stage in the process. Following full ethical approval, we have submitted and had approved 4 amendments to our protocol – we have a minor amendment currently under consideration. Each of these submissions does of course require significant attention to detail and relevant documentation using standard Integrated Research Application System (IRAS) procedures.

Capacity: possible need for extension in time and resource

This study started on 1 November 2009, and was funded for a period of 45 months. The original design involved a 12-month feasibility phase, followed by a 33-month 'main trial' phase. Clear stopping rules were set to demarcate progress to the main trial phase. It was perhaps a mistake that there was no flexible time to allow for careful review of data accumulated during the feasibility phase prior to commencement of the main trial phase.

Following personnel problems encountered within the first few weeks of the trial, and the subsequent departure of the trial manager, we were granted a 3-month extension to the feasibility phase. At the time this was discussed, we did not request an extension to the overall length of the trial itself, although we did note that the extension in feasibility may have to be matched by an extension to the overall trial.

Our experience of conducting the feasibility work has identified the very substantial complexity of this trial. Although we have completed all of the recruitment we set out to achieve, we have in addition collected, collated, and are in the process of analysing the feasibility data. This report represents an important communication with HTA, but ideally, we would have built in some time at this stage to ensure that all the lessons from the feasibility phase have been carefully learnt, incorporated within the main trial documentation, and disseminated out to the researchers and principal investigator across the four sites. Recognising this, we have built in an additional 1-month period to allow the lessons from the feasibility to be learnt and disseminated. However, some additional time would have been extremely helpful, allowing us to further refine our intervention manuals, standard operating procedures, and supporting data collation and analytical materials.

In summary therefore, although we may be able to achieve the main trial within the original resource, there is also the serious possibility that, simply on account of the complexity of the work being undertaken, an extension of 4–6 months may be necessary. This is an area I would wish to discuss with HTA.

Personnel

In general, the trial is reasonably well resourced. Processes around supporting practices, undertaking the (highly complex) research training procedures, for example targeting reception and clinical staff, is, in its own right, an important and complex process requiring careful support from research staff. At the same time, those research staff are undertaking mailing of questionnaires, and the review of clinical notes. These processes are proving challenging to manage within the resource available.

Given the difficulties encountered, we would like to propose a solution (which requires an additional staff member to be costed into our work). We believe that, with the allocation of this modest additional resource, we may be able to complete the trial within the original time frame. The data collection is scheduled to take place in three separate 'waves' within each of the four recruitment centres. Each wave of the intervention is associated with a period in which a review of clinical notes takes place. It would be of great help to allow us to employ somebody – a supplementary research administrator post – on an intermittent basis. That individual would be specifically trained in clinical note review and undertake the bulk of that work, but supported by local research staff. Essentially they would be employed for two months, half time, for each of the three waves in each of the four centres. This amounts to the equivalent of a 1.0 full-time research administrator over 24 months (although this resource will be distributed between the four centres on an intermittent basis).

Travel

We have been outstandingly successful in recruiting practices associated with each of the four research centres. Recruitment, however, has been somewhat sluggish in some centres and this has necessitated 'casting the net wider' in terms of the geography. We do therefore anticipate the requirement for an increase in the travel budget to allow research staff to visit the practices on a regular basis, supporting practice staff personally and face to face, ensuring continuing compliance with the intervention itself, undertaking data collection, and generally maintaining the profile of the research and research processes within the practice. Across the four centres we believe that an additional £15,000 would allow for that resource to be met.

ESTEEM pilot study ICC analysis (17 February 2011)

All NHS primary health care contacts (including 'did not attends')
 . loneway sum4weekcontactsincludingdnas practiceid.

One-way Analysis of Variance for s~includin~s: Sum 4-Week contacts
 (INCLUDING D

Number of obs = 346
 R-squared = 0.0311

Source	SS	df	MS	F	Prob > F
Between practiceid	29.40243	5	5.880486	2.18	0.0560
Within practiceid	917.21318	340	2.6976858		
Total	946.61561	345	2.7438134		

Intraclass Asy.
 correlation S.E. [95% Conf. Interval]
 0.02082 0.02457 0.00000 0.06898

Estimated SD of practiceid effect .2395275
 Estimated SD within practiceid 1.642463
 Est. reliability of a practiceid mean 0.54125
 (evaluated at $n = 55.48$)

All NHS primary health care contacts (excluding 'did not attends')
 . loneway sum4weekcontactsexcludingdnas practiceid

One-way Analysis of Variance for s~excludin~s: Sum 4-week Contacts (Excluding D

Number of obs = 346
 R-squared = 0.0364

Source	SS	df	MS	F	Prob > F
Between practiceid	32.832854	5	6.5665708	2.57	0.0269
Within practiceid	870.10356	340	2.5591281		
Total	902.93642	345	2.617207		

Intraclass Asy.
 correlation S.E. [95% Conf. Interval]
 0.02745 0.02871 0.00000 0.08372

Estimated SD of practiceid effect.2687723
 Estimated SD within practiceid 1.599728
 Est. reliability of a practiceid mean 0.61028
 (evaluated at $n = 55.48$)

EQ-5D index score

. loneway eq51_index practiceid

One-way Analysis of Variance for eq51_index:

Number of obs = 266
R-squared = 0.0950

Source	SS	df	MS	F	Prob > F
Between practiceid	2.0487013	5	5.40974027	5.46	0.0001
Within practiceid	19.520951	260	0.07508058		
Total	21.569652	265	0.08139491		

Intraclass Asy.
correlation S.E. [95% Conf. Interval]

0.09593	0.07234	0.00000	0.23772
---------	---------	---------	---------

Estimated SD of practiceid effect .0892562
Estimated SD within practiceid .2740084
Est. reliability of a practiceid mean 0.81676
(evaluated at $n = 42.01$)

Sensitivity analyses – assuming data is dichotomous

. loneway sum4wk_dna_cat practiceid

One-way Analysis of Variance for sum4wk_dna~t:

Number of obs = 346
R-squared = 0.1069

Source	SS	df	MS	F	Prob > F
Between practiceid	6.0930857	5	1.2186171	8.14	0.0000
Within practiceid	50.924255	340	0.14977722		
Total	57.017341	345	0.16526766		

Intraclass Asy.
 correlation S.E. [95% Conf. Interval]

 0.11398 0.07784 0.00000 0.26653

Estimated SD of practiceid effect.1388056
 Estimated SD within practiceid.3870106
 Est. reliability of a practiceid mean 0.87709
 (evaluated at $n=55.48$)

. loneway eq51_index_cat practiceid

One-way Analysis of Variance for eq51_index~t:

Number of obs = 266
 R-squared = 0.0482

Source	SS	df	MS	F	Prob > F
Between practiceid	3.0432896	5.60865793	2.63	0.0241	
Within practiceid	60.073252	260.23105097			
Total	63.116541	265.23817563			

Intraclass Asy.
 correlation S.E. [95% Conf. Interval]

 0.03745 0.03855 0.00000 0.11301

Estimated SD of practiceid effect .0948105

Estimated SD within practiceid .4806776
 Est. reliability of a practiceid mean 0.62039
 (evaluated at $n = 42.01$)

The process evaluation pilot study

Aims and objectives

The process evaluation consisted of two discrete phases; an initial 1-year pilot study conducted alongside, and feeding into the pilot of the main trial, followed by the main process evaluation study. This is the report from the process evaluation component of the pilot.

The specific aims of the process evaluation pilot were to (1) ensure that the triage training and interventions could be feasibly delivered by practices and (2) test and fine-tune data collection methods.

Methods

The pilot study was conducted in six general practices in Bristol and Devon between 2010 and 2011. It was designed to inform both the subsequent process evaluation and the main ESTEEM trial. The practices had been randomly allocated to two each of GPT, NT and UC. They received training on the triage interventions and collected outcome data as they would in the main trial, but using only a 2-week period of patient recruitment.

Data collection for the pilot consisted of systematic observation and in-depth interviews. A total of 49 hours was spent observing reception and triage staff activities in the six practices. In addition, 47 semistructured interviews were held with a purposively selected sample of patients and staff. The week-long observations of practice staff, patient actions and interactions, and other naturally occurring events were undertaken during the implementation period of the study interventions. Data were recorded in the form of brief paper-based field notes. Observation took place mainly in communal areas of the practices, such as patient waiting rooms, but also in more private spaces, such as receptionists' areas. Locations were chosen in order to gain a good appreciation of the variety of actions, interactions and events occurring in each practice, while remaining sensitive to issues of participant trust.

Observation focused on factors known to influence the implementation and incorporation of complex health interventions, aiming to illustrate the context and mechanisms operating and interacting in the management of patients' same-day consultation requests. Written or audio-taped field notes recorded activities relating to the running of the study and the delivery of the interventions rather than the usual business of the practices. Interview data were transcribed verbatim and analysed thematically across trial arms. Observations and findings were summarised systematically and reported at regular intervals to the ESTEEM Trial Management Group. Successful aspects of trial protocol and procedure were noted and retained and the team developed solutions to any identified problems, which were trialled and re-reported as an iterative process of refining procedures.

Semistructured interviews were conducted with six patients and 37 practice staff members, using an interview schedule developed by the Research Team and the Patient Group (*Table 38*). The patients were from three practices, two providing NT and one providing GPT. Participant patients were of both sexes across a range of ages. Patients were asked to reflect on their experiences of accessing and receiving care. Staff interview participants comprised practice managers, receptionists, and nurses or doctors delivering the interventions in six practices: two NT practices, two GPT practices and two UC practices.

TABLE 38 Staff interview participants in the pilot study

Practices	GPs	Practice nurses/nurse practitioners	Receptionists	Practice/reception managers	Totals
GPT	5	1	5	3	14
NT	1	4	5	3	13
UC	2	2	3	3	10
Totals	8	7	13	9	37

Fifteen clinicians and 22 administrative staff from across six trial practices participated. Staff were asked about the feasibility and practicality of the data collection methods and the implementation of the intervention.

Field notes and interview notes were analysed systematically to identify successful and less successful aspects of the trial protocol. Notes were made of what did and did not work well: problems noted by each group of practice staff were recorded, with reasons for any problems, if discernible, and things that went well were also noted. These notes and examples were discussed by the research team, and potential problems, and solutions to problems, were identified. Where appropriate, such solutions were fed back to participating practice teams for consideration, comment and ratification before proceeding to the full trial.

Results

The report on the substantive findings from the pilot research interview data on how triage and the trial were experienced can be seen at the beginning of this report. This section reports on how findings from the pilot informed the development of the trial protocol, procedures and paperwork in preparation for the main trial.

Engagement in the trial was a problem for some practice staff. Some receptionists in particular had anxieties about their role in telephone triage, perceiving it as forcing them into the role of gate-keepers, representing erecting a barrier to access to doctors' appointments. This could lead to them inadvertently subverting the protocol, for instance by putting patients down for triage as well as booking them an appointment; reverting to usual care; omitting to inform patients they were part of a trial and that they would receive questionnaires; and not explaining triage in a clear and effective way. It was also observed that log sheets were not being completed with accuracy. These issues were addressed by encouraging managers to involve and communicate effectively with receptionists, and in particular to stress the positive aspects of the trial, and by developing, with the receptionists' collaboration, a 'script' for explaining the trial to patients. The trial flow chart was improved to make it easier to understand, and these measures were integrated into the training strategy. Training was also amended to address motivational problems, particularly for receptionists, whose role is crucial because they facilitate the point of patient entry into the trial. Emphasis was placed on such positive benefits of triage as patient equity and the reduction of pressure on receptionists themselves in not having to find appointments. As completion of the logs was reviewed, the need was identified for them to be amended to differentiate between patients who are given same-day appointments, patients who request same-day appointments but do not get one, and patients requesting a telephone consultation but not a face-to-face appointment. The relevant changes were made.

It became clear that not all practice staff members had received adequate information about the trial, and consequently felt alienated or uninvolved. Procedures were revised to ensure that staff members were each sent a personally addressed, role-specific information pack. It was observed that to increase the fluency and ownership of the trial with receptionists and other staff then practices need to have a key senior person, a 'champion', to direct and drive it and address staff queries and concerns, and it was agreed that this would be encouraged by the researchers when establishing the trial in the practices. A number of other training issues were identified. GPs needed to be confident with the system for booking appointments. Triage nurses should be computer literate and confident in using the decision support software, which itself should be ensured as working well before commencement of triaging.

The pilot revealed ambiguities for which guidelines needed to be developed. One such was the reluctance of some practices to include infants and young children in the triage protocol, and it was decided that practices could exclude them if they wished. Policies were also developed for the procedure if patients refused to be triaged. Another ambiguity was defining what exactly constitutes a same-day appointment request, for example when the patient does not specify. Policy on whether use of the decision support software was mandatory for triaging nurses also had to be defined, and it was decided that nurses would be encouraged, but not compelled to use the software.

Processes for checking compliance with trial procedures were also developed. Effective communication channels between the practices and trial researchers were put in place, in particular a named key person to contact in each organisation, following suggestions from practice managers. Guidance and clear definitions, such as on 'what is the end time' were put in place to ensure harmonisation of the completion of Clinician Forms, which had been shown to be done somewhat haphazardly. Clinicians' concerns about the eliciting of verbal but not written consent for the team to access patient notes were addressed by reassurance and evidence of ethical approval for this, and, in one case, a personal visit from a senior member of the study team. Combining triage with the duty doctor role was observed to potentially compromise safety, and practices were advised in training not to do this, although some continued to do so. Practices were advised to triage consistently during set periods, which had been observed not to be happening.

Conclusion

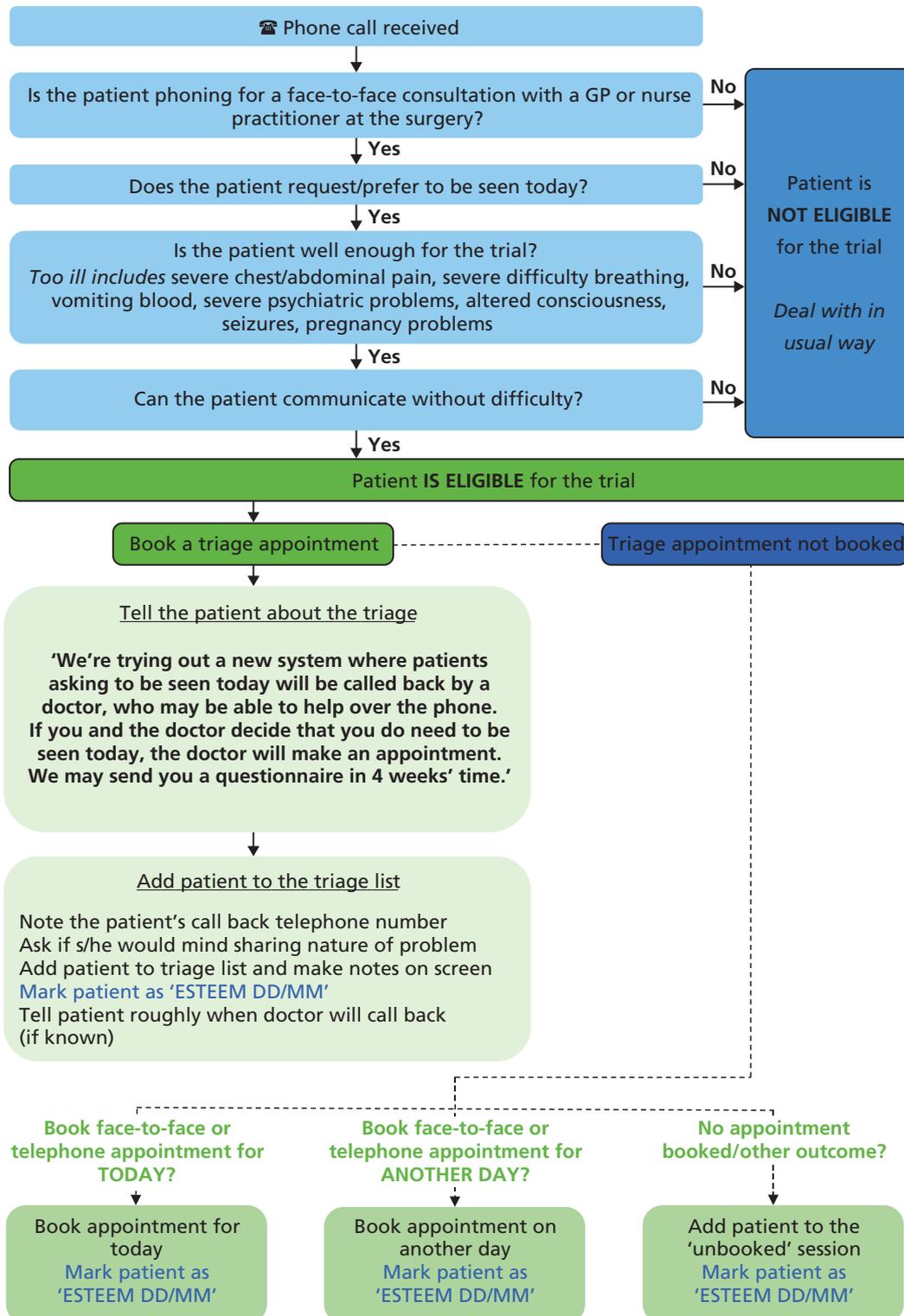
The qualitative pilot process evaluation provided insights into the processes and experiences of establishing a trial in practice microcontexts, which enabled amendments and adjustments to be made to the trial protocol, procedures and materials to facilitate the smooth introduction and conduct of the main RCT. It enabled protocol amendments to be submitted for ethical approval in a timely manner, which did not delay or compromise the smooth running of the trial once it was under way.

Appendix 2 Summary of changes to original ESTEEM protocol

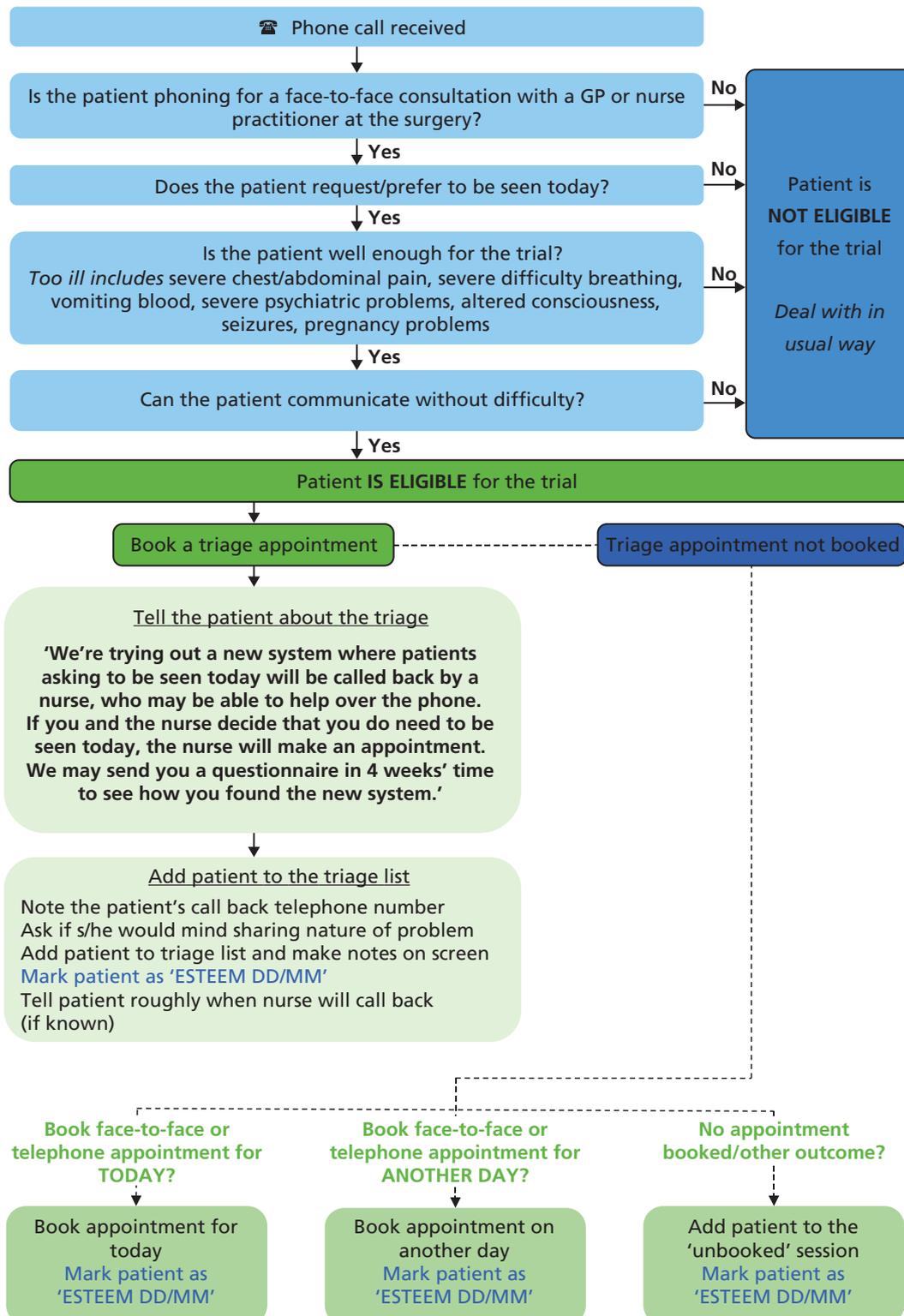
Change to protocol	Date
Before pilot study	
Amendments to the original patient information leaflet [Trial], patient invitation letter [process evaluation] and the 'thanks but not needed letter' to patients [process evaluation] and inclusion of an incentive for questionnaire respondents	16 October 2009
Between pilot study and main trial	
Amendments to process evaluation recruitment materials advised by PPI group	2 August 2010
Amendments to Main Trial recruitment materials advised by PPI group	26 August 2010
Use of verbal consent to case note review	29 December 2010
Transfer of patient names and addresses	18 March 2011
Amendments to patient documentation following end of pilot and use of Waiting Room Posters	1 June 2011
During main trial	
Use of amended mail-out materials for 'missed' patients at one practice	30 July 2012
Use of postcode data to calculate a deprivation score	3 September 2012
Change of study Sponsor from Devon PCT to Royal Devon & Exeter NHS Foundation Trust	30 October 2013

Appendix 4 Receptionist flow charts

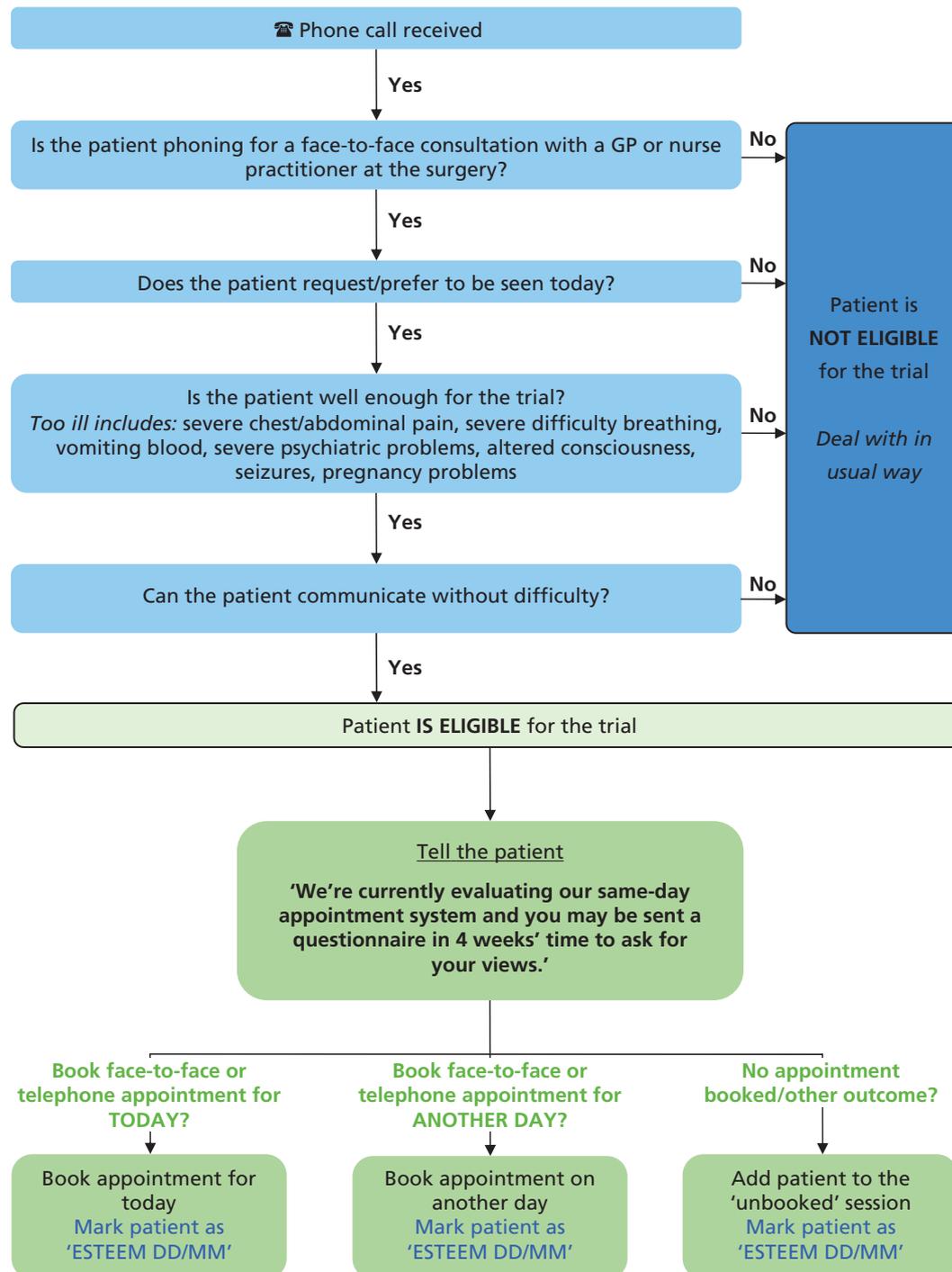
Receptionist flow chart (general practitioner triage)



Receptionist flow chart (nurse triage)



Receptionist flow chart (usual care)



Appendix 5 Questionnaire for withdrawn practices

QUESTIONNAIRE FOR WITHDRAWN PRACTICES

To help us understand the reasons why practices were unable to continue participation in ESTEEM please complete the form below:

Practice Name: _____ Randomised to: _____

Completed by: _____ Date: _____

What were the main reasons for withdrawing from ESTEEM? (tick all that apply):

1. Dissatisfaction with the Trial arm randomised to.

2. The practice was under-resourced to cope with the demands of ESTEEM (tick all that apply):
 - Difficulty recruiting additional staff
 - Difficulty extending staff hours as needed
 - Other, **please provide details**

3. Your practice experienced unexpected extenuating circumstances (tick all that apply):
 - Long-term sickness of key staff member(s)
 - Retirement/resignation of key staff member(s)
 - Other, **please provide details**

4. Insufficient 'buy-in' from all members of practice staff

5. Lack of support from the research team

6. Change of practice IT system

7. Change of practice premises

8. Triage system implemented between being recruited and starting the trial

9. Other, please state:

Please add any further comments overleaf

Appendix 6 Covering invitation letters

Covering letter to patients, on practice-headed note paper

<Surgery address>

<Patient address>

<Date>

<patient computer ID>

Dear <Name of patient>,

About four weeks ago you telephoned our surgery requesting a same-day consultation. The doctor or nurse you saw or spoke to may have briefly mentioned that our surgery is involved in a research study. The study, led by the Peninsula Medical School in Exeter, looks at how we provide care for patients wanting a same-day consultation. In particular it looks at how we might continue to provide high-quality care when there are not enough same-day consultations to meet demand.

Our surgery is fully in support of this important piece of research. The results could help to improve the way that health services are delivered by the NHS, locally and nationally.

I have enclosed a brief questionnaire about the above arrangements. Please help us by completing it and returning it in the pre-paid envelope supplied.

If you saw or spoke to one of our doctors or nurses, at the end of your consultation they may have asked for your permission to allow one of the researchers to look at your medical records. At the end of the questionnaire you will find a consent form. If you have now changed your mind, and would or would not like a researcher to look at your medical records, you can let us know by completing this form. If you did not have the opportunity to agree or disagree to your records being reviewed, please complete the consent form now.

The enclosed information leaflet contains further information about the study. If you have any queries, please do not hesitate to contact the researchers directly, using the details given at the end.

If you do not wish to take part in the study, I would be grateful if you would still return the blank questionnaire in the envelope provided. We will then know that you do not wish to receive any further contact from us about this study.

Your assistance would be of great value and very much appreciated.

Yours sincerely,

<Name of GP or Practice Manager>

Covering letter to parents, on practice headed note paper

<Surgery address>

<Patient address>

<Date>

<patient computer ID>

Dear parent/guardian of <Name of patient>,

About four weeks ago you telephoned our surgery requesting a same-day consultation for your child. The doctor or nurse you saw or spoke to may have briefly mentioned that our surgery is involved in a research study. The study, led by the Peninsula Medical School in Exeter, looks at how we provide care for patients wanting a same-day consultation. In particular it looks at how we might continue to provide high quality care when there are not enough same-day consultations to meet demand.

Our surgery is fully in support of this important piece of research. The results could help to improve the way that health services are delivered by the NHS, locally and nationally.

I have enclosed a brief questionnaire about the above arrangements. Please help us by completing it on your child's behalf and returning it in the pre-paid envelope supplied.

If you saw or spoke to one of our doctors or nurses, at the end of your consultation they may have asked for your permission to allow one of the researchers to look at your child's medical records. At the end of the questionnaire you will find a consent form. If you have now changed your mind, and would or would not like a researcher to look at your child's medical records, you can let us know by completing this form. If you did not have the opportunity to agree or disagree to your child's records being reviewed, please complete the consent form now.

The enclosed information leaflet contains further information about the study. If you have any queries, please do not hesitate to contact the researchers directly, using the details given at the end.

If you do not wish to take part in the study on your child's behalf, I would be grateful if you would still return the blank questionnaire in the envelope provided. We will then know that you do not wish to receive any further contact from us about this study.

Your assistance would be of great value and very much appreciated.

Yours sincerely,

<Name of GP or Practice Manager>

Appendix 7 Patient information leaflet



HTA Trial No. 08/53/15



**A study about your experience of
getting care from a GP or Nurse**

PATIENT INFORMATION

LEAFLET

This leaflet contains an invitation to take part in the above study. It tells you why the study is being done and what it would involve for you. Please take time to read it through carefully, so that you can decide whether or not you would like to take part.

THE RESEARCH STUDY

Patients often ring their surgeries asking for same-day consultations. Not all such patients want or need to see a doctor or nurse face-to-face on each occasion. Some people, such as those who work during the day or have caring responsibilities, may find it easier to get prompt advice over the phone, rather than having to visit the practice in person. This system of assessing how best to meet patients' needs is called '**telephone triage**'.

Our study looks at how telephone triage systems led by doctors and nurses work. We are particularly interested in how these systems affect patients' experiences, safety and health, as well as practices' workloads and costs. Your surgery is one of 42 across the country that has kindly offered to help us with this study. Each of the 21,000 or so patients from these surgeries who are invited to participate will have asked for a same-day consultation.

This leaflet invites you to join the study. If you decide to take part, we would like you to fill in the short questionnaire enclosed with this leaflet. This will tell us how you got on with the system used by your surgery to manage same-day consultations.

This study is being carried out by a team of experienced researchers based in universities in Bristol, Exeter, Norwich and Warwick. We are available to answer your questions. You will find your local research teams' contact details at the end of this leaflet.

Thank you for taking the time to read this leaflet

What is the purpose of this study?

This study looks at new ways to improve patients' access to services when they ask for same-day consultations. In particular, it aims to find out how well telephone triage systems led by doctors and nurses work. We are especially interested in any differences that these systems make to patients' experiences of care, safety and health, as well as to surgeries' workloads and costs.

Your surgery is taking part in this study. While the study is running, when you phone to ask for a same-day consultation, a doctor will phone you back. S/he will talk to you about your needs, to see if s/he can help you over the phone. If you and the doctor decide that you still need to come into the surgery, then s/he will make an appointment for you.

Who is taking part in this study?

All adult patients or parents of children who have contacted their doctors' surgeries asking for same-day consultations with a doctor or nurse during the period of the study may be asked to take part.

What will happen to me if I decide to take part in this study?

If you decide to take part in this study, we would like you to **complete and return the brief questionnaire enclosed**. This shouldn't take more than about 15 minutes of your time. The questionnaire asks for your views on your surgery's same-day consultation service.

On the day of your consultation, a healthcare worker may have asked if you would give permission for one of our researchers to look at your medical records. The enclosed questionnaire asks the same question. If you originally said that that it would be OK for us to look at your records, but have now changed your mind, you can 'opt out'. Similarly, if you originally said 'no' to us seeing your records, and have now changed your mind, you can 'opt in'. If you let us look at your records, we will only collect information on the number and length of any visits that you made to primary care services (e.g. your surgery, A&E, a walk-in centre) within a month of making your same-day consultation request. Doing this will let us compare the effects of the different systems that we are studying.

The questionnaire allows you to say 'yes' or 'no' to having your records reviewed. **No sensitive information about your detailed medical history will be recorded, either way.**

Do I have to take part? Will my decision affect the services that I receive?

We do hope that you will be willing to complete the enclosed questionnaire, but your decision about whether to take part in the study will in no way affect the services that you receive. It is entirely up to you whether or not you decide to take part in the study. You may also withdraw from the study at any time, without giving any reason.

What are the possible risks or disadvantages in taking part?

As far as we know, there are no particular risks or disadvantages involved in taking part in this study. The surgeries taking part in the study have processes in place to ensure that

any problems that might arise from the triage systems being used are spotted early on and resolved.

What are the possible benefits of taking part?

Taking part in this study would allow you to have your say about your healthcare services, helping us to ensure that such services are provided in the best possible way. Your contribution could therefore help to improve things for other people at your surgery and/or using general practice services nationally.

Will the information I provide be kept confidential?

Yes. All information about you will be handled in strict confidence in accordance with ethical and legal practice. Your decision to take part in the study will also be kept confidential. The only time that we would pass on information, for example to your doctor, would be if something in your questionnaire made us think that you, or someone else, was at risk of serious harm. Normally, we would only pass on information in this way after discussing the situation with you. Your medical records will only be reviewed by a study researcher if you give us permission to do so.

All personal information will be stored securely in line with the Data Protection Act (1998) for up to a year and then destroyed. Completed questionnaires will also be destroyed after 5 years.

What will happen if I don't want to carry on with the study?

If you decide to take part in the study and then change your mind, you may stop taking part at any time without having to give a reason. If this should happen, general information collected up to that time, which does not identify you (such as your gender and age) might still be kept. This would let us see if patients who left the study differed in some way from those who stayed. It would be important to know this because, if the people who left and stayed were different, it might bias the study results. If you do decide to withdraw from the study then please call a member of the research team – our contact details are at toward the end of this booklet.

What if there's a problem?

Any complaint about the way that you have been dealt with during this study will be promptly and thoroughly addressed. If you have any complaints or concerns about any aspect of the study itself or the research team's work, please feel free to discuss them with us. You will find our contact details at the end of this leaflet. If you have more general concerns about the way that you have been approached or treated during the study, you can access the NHS complaints mechanism via the Patient Advice and Liaison Service (PALS) on 0800 073 0741. Calls made to this number are free. Alternatively, you can visit the PALS website at: <http://www.pals.nhs.uk/>

How will the results of the research study be used?

The results of this study will interest many different people and organisations. We will share them with the surgeries involved and with other NHS bodies, to help them to

improve their services. We will also write them up for the study's funding body, for academic publications and conferences as well as for the general public. The results will also be available on the following website: <http://wdbdev.pcmd.ac.uk/esteem/>

No individual patient will be identifiable in anything we write.

Who is organising/funding the study? Who has reviewed it?

Our study is funded by the National Institute for Health Research Health Technology Assessment Programme <http://www.hta.ac.uk/>. The study has been reviewed, and its quality assured, by the Institute's team of scientific experts and the South West 2 NHS Research Ethics Committee (Ref. 09/H0202/53). The research is sponsored NHS Devon Primary Care Trust.

Which organisations are carrying out the study?

The study is being led by the Peninsula Medical School, based in Exeter, and carried out by research teams based in universities in Bristol, Exeter, Norwich and Warwick.

Want to know more about the study or questionnaire?

If you have any questions about this study, please do not hesitate to contact us.

Dr. Raff Calitri
Researcher (ESTEEM)
Peninsula Medical School
Smeall Building
St. Luke's Campus
University of Exeter
EX1 2LU

Mrs Emily Fletcher
Trial Manager (ESTEEM)
Peninsula Medical School
Smeall Building
St. Luke's Campus
University of Exeter
EX1 2LU

Email: raff.calitri@pcmd.ac.uk
Tel: 01392 722762

Email: emily.fletcher@pcmd.ac.uk
Tel: 01392 722826

Thank you for taking the time to read this information leaflet



Still not sure whether to take part?

For independent advice about taking part in research studies in general you can contact 'INVOLVE', an organisation that offers advice and information on patient/public research in the NHS, at:

Wessex House,
Upper Market Street,
Eastleigh,
Hampshire,
SO50 9FD

Telephone: 02380 651088
Textphone: 02380 626239
Fax: 02380 652 885

Email: admin@invo.org.uk
Website: <http://www.invo.org.uk/index.asp>

If you live in Devon, 'Folk-Us' offers similar advice & information at:

Room 407,
Noy Scott House,
Royal Devon & Exeter Hospital,
Barrack Road,
Exeter,
Devon,
EX2 5DW.

Telephone: 01392 403049
Fax: 01392 406936

Email: folk.us@exeter.ac.uk
Website: <http://www.folkus.org.uk/>

Appendix 8 Patient questionnaire



A study about your experience of getting care from a GP or nurse



To help us improve the care you receive please answer the questions below by putting an x in ONE BOX for each question unless more than one answer is allowed (these questions are clearly marked). Leave any questions you can't answer blank.

We will keep your answers completely confidential.

Q1 Are you completing this questionnaire for...?

- Yourself
 Someone else → If you are completing the questionnaire on behalf of someone else please answer all the questions based on the care they received.

CARE RECEIVED ON THE DAY YOU CALLED

You contacted your GP surgery around 4-6 weeks ago asking to see a GP that same day.

Q2 What happened when you initially contacted the GP surgery?

Please **X** all the boxes that apply to you

I spoke to a receptionist and ...

- Had a call back from a GP later that day
 Had a call back from a nurse later that day
 Made an appointment with a GP or nurse
 Can't remember

Q3 How easy or difficult was it to....?

Please put an **X** in one box for each row

	Very easy	Fairly easy	Neither easy nor difficult	Fairly difficult	Very difficult
Get through to the practice on the phone	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Receive prompt care	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
See a GP or nurse if you wanted to	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Get medical help or advice for the problem you presented.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Q4 How convenient was the care provided by your GP surgery on that day?

- Very convenient
 Fairly convenient
 Not very convenient
 Not at all convenient

Q5 Thinking about the reason why you contacted the GP surgery or health centre that day, is the problem now...?

- Much better
 A bit better
 The same
 A bit worse
 Much worse
 Don't know

Q6 Overall, how satisfied or dissatisfied were you with the care received on that day?

- Very satisfied
 Fairly satisfied
 Neither satisfied nor dissatisfied
 Fairly dissatisfied
 Very dissatisfied

CARE RECEIVED SINCE THAT DAY

The next questions are about what has happened since the day you called.

Q7 In the 4 weeks since you called to see a GP the same day, how many times, if any, have you...?

Please do not include any care you received on the day you first called.

	None	1	2	3	4	5	6 or more
Visited an NHS walk-in centre	<input type="checkbox"/>						
Phoned NHS Direct for health advice	<input type="checkbox"/>						

YOUR HEALTH TODAY

If you are completing this questionnaire for a child aged 11 or below please go to Q9.

If you are completing this questionnaire for yourself or another adult, please answer the following question.

Q8 By placing an **X** in one box in each group below, please indicate which statements best describe your health state today.

Mobility

- I have no problems in walking about
 I have some problems in walking about
 I am confined to bed

Self-Care

- I have no problems with self-care
- I have some problems washing or dressing myself
- I am unable to wash or dress myself

Usual Activities (e.g. work, study, housework, family or leisure activities)

- I have no problems with performing my usual activities
- I have some problems with performing my usual activities
- I am unable to perform my usual activities

Pain/Discomfort

- I have no pain or discomfort
- I have moderate pain or discomfort
- I have extreme pain or discomfort

Anxiety/Depression

- I am not anxious or depressed
- I am moderately anxious or depressed
- I am extremely anxious or depressed

SOME QUESTIONS ABOUT YOU

We would now like to ask some questions to help us see how experiences of GP consultations vary between different groups of people.

Remember that if you are completing the questionnaire for someone else please answer the questions about the person who was the patient.

Q9**Are you male or female?**

- Male
- Female

Q10**How old are you?**

- | | |
|-----------------------------------|-------------------------------------|
| <input type="checkbox"/> Under 18 | <input type="checkbox"/> 55 to 64 |
| <input type="checkbox"/> 18 to 24 | <input type="checkbox"/> 65 to 74 |
| <input type="checkbox"/> 25 to 34 | <input type="checkbox"/> 75 to 84 |
| <input type="checkbox"/> 35 to 44 | <input type="checkbox"/> 85 or over |
| <input type="checkbox"/> 45 to 54 | |

Please turn over 

Q11 What is your ethnic group?

- White
 Mixed / multiple ethnic groups
 Asian / Asian British
 Black / African / Caribbean / Black British
 Other ethnic group

Q12 If you need to see a GP at the GP surgery during typical working hours, can you take time away from your work to do this?

- Yes, I can easily take time away from work
 Yes, I can take time away from work but with some difficulty
 No, I can't take time away from work
 Not relevant as I do not work (unemployed, retired)
 Not relevant as this questionnaire is being completed on behalf of a child or an unemployed or retired adult

Q13 Do you have any long-standing health problems?

- Yes
 No

THE PATIENT'S CONSENT

On the day you requested a same-day consultation, a clinician may have asked if you would give permission for a researcher to look at your medical records to see whether you've needed any further help or advice about the original problem. To check whether our appointment system is working, we need to know about any consultations you have had since you first called to see a GP on the same day.

Your name will not be associated with the information we collect and all information will be treated with the strictest confidence and will not be passed on to any third party. We hope you will feel able to allow us to look at your notes to extract this information. However, if you do not now wish to have your notes examined, please indicate this below.

Q14 Please place an x in only one box, then sign and date below. Please do not cross either box on anyone else's behalf, unless you are a parent or legal guardian providing consent on behalf of your child.

- I AGREE to the review of my medical notes
 I DO NOT AGREE to the review of my medical notes

Signature _____

Date _____

Thank you for your time. Please return this questionnaire in the reply paid envelope provided.

Any and all copyrights for Q8 (including layout) vest in the EuroQol Group. The EuroQol Group reserves all rights.
 ©1992 EuroQol Group. EQ-5D™ is a trademark of the EuroQol Group.

Appendix 10 Adverse events reporting procedure

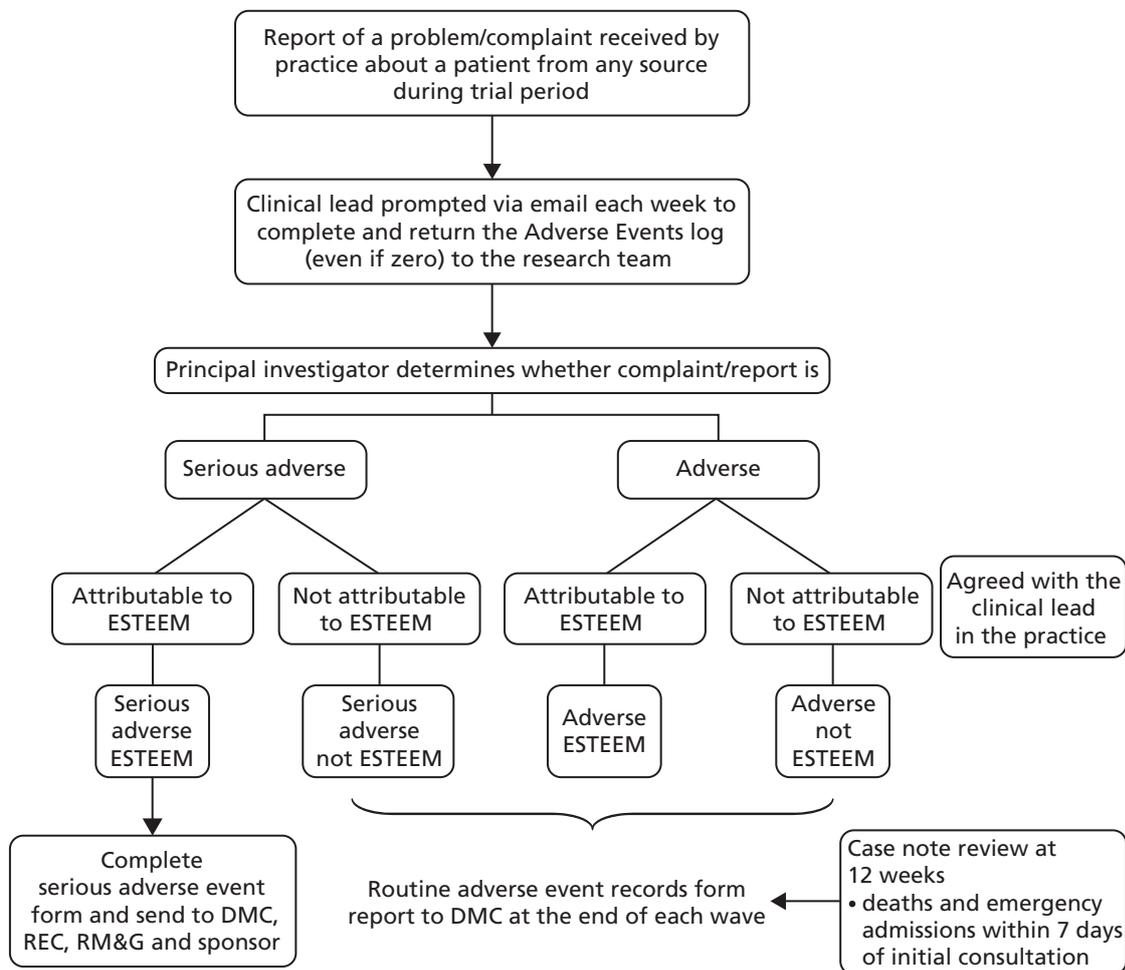


FIGURE 11 Flow diagram representing the adverse events reporting procedure for the ESTEEM trial. DMC, Data Monitoring Committee; REC, Research Ethics Committee; RM&G, Research Management & Governance.

The purpose of this process is to detect any problems during the period that the intervention is running, through reports of problems/complaints received by the practice. All reports of events will be collected by the research team, prompting the Clinical Lead to review problems/complaints recorded by the practice and to complete (even if to state 'no events') and return the Adverse Events log. Practices will be asked to report anything that is of major concern to the research team in real time. The assessment of seriousness and attribution of problems will be undertaken by the principal investigator using the following rules:

Seriousness

A serious adverse event includes:

- i. death
- ii. serious illness or substantial deterioration in health leading to A&E attendance and/or emergency hospital admission.

An adverse event includes:

- i. delay in treatment with no direct resultant harm
- ii. a patient or health professional making a complaint to the practice regarding process of care.

Attribution

Events attributable to ESTEEM

Events that occur as a suspected result of the mode or timing of the initial consultation could include:

- misdiagnosis/assessment of the severity of a condition or delay in treatment due to not seeing a patient face to face (e.g. skin malignancy, giving advice/prescribing antibiotics over the telephone for a chest infection which is actually more severe, meningitis labelled as viral illness, dyspepsia that turns out to be cardiac ischaemia, abdominal pain that becomes a ruptured appendix, delay in treatment for a stroke/meningitis/myocardial infarction)
- failure to manage infections adequately over the telephone, for example not recognising the need for antibiotics or prescribing antibiotics unnecessarily
- misunderstanding of medication instructions over the telephone, leading to side effects
- patient confusion over changes to the appointment system, for example patients not turning up for appointments
- delay in treatment of patients at the practice caused by safeguarding the time of a clinician, for example a nurse who is responsible for cervical smears may not have as much time to do them if time is taken up triaging for ESTEEM
- increase in clinician stress, deterioration in health and/or absence from work owing to initial experience with an intense workload of telephone calls
- complaint reported by a patient or health professional regarding the initial consultation, for example timing or content.

Events not attributable to ESTEEM

Events that are unrelated could include:

- death, A&E attendance or hospital admission due to a cause unrelated to that discussed during telephone triage
- an event that would have happened regardless of the mode of consultation, for example drug-related reaction.

If a serious adverse event is deemed to be attributable to ESTEEM, the Trial Manager will report this within 24 hours, via fax, using the serious adverse event form, to the Research Ethics Committee (REC), Research Management & Governance (RM&G) units, the Data Monitoring Committee (DMC), and sponsor.

A retrospective search of practice records will collect data on patient deaths during the trial, to undergo the same process of determining attribution.

A separate process of case note review will be undertaken by the research team 12 weeks after patients are entered into the trial, which will assess A&E attendances and emergency hospital admissions.

Case note review at 12 weeks

The study protocol states 'deaths within 7 days of same day consultation request (from practice records), and attendance at A&E within 4 weeks and number and length of stay of emergency hospital admissions within 7 days of index consultation (from primary care records examined 12 weeks after same day consultation request)'.

The events listed above that are found at case note review will not be subject to the same assessment of seriousness and attribution as detailed for the events reported weekly by practices to the research team. Count data of events recorded at case note review will be presented to the DMC as the trial progresses and reported in full at the end of the trial.

Recording emergency hospital admissions

The researcher conducting the case note review will record all hospital admissions within 7 days of the index consultation. An admission will be detected by the presence of a discharge summary in the patient's record.

Notes:

Patient admitted to a hospital ward via A&E

If there is only one discharge summary = one contact (admission)

If a separate discharge summary relating to the A&E contact has not been generated in addition to the ward admission, this may indicate that the admission via A&E was arranged by a GP or admitting team.

If there are two discharge summaries = two contacts (one A&E plus one admission)

A separate discharge summary generated for the A&E contact and for the ward admission may indicate that the patient presented themselves to A&E and was subsequently admitted to a ward.

Patient admitted directly to a hospital ward, Medical Assessment Unit or equivalent (not via A&E)

This counts as one contact (admission), even if the patient was admitted and discharged on the same day.

Patient admitted via A&E, but did not progress to a hospital ward before discharge

Even if a discharge summary has been generated to suggest that a hospital admission occurred, this counts as one contact (A&E), even if the patient was admitted and discharged on the same day or on different dates.

Defining emergency hospital admissions

The researcher will present all admissions found in the notes to the lead GP in the practice, for advice on whether an admission was an 'emergency', using the following definition:

An '*emergency*' hospital admission for the trial is an UNPLANNED admission, irrespective of clinical need.

An '*unplanned*' admission is any admission (through any route, e.g. A&E, Bed Manager, GP contacting the hospital team), irrespective of clinical need, which was arranged on the same day as the admission occurred and not before that date.

An admission on the same date following an outpatient's appointment would count as 'unplanned'.

In cases in which a patient is seen (e.g. by out-of-hours service) and admitted the next day, but within 24 hours, these scenarios would count as 'unplanned'.

A '*planned*' admission is any admission arranged for further into the future, even if the person is acutely unwell.

Appendix 11 Clinician Form

		CLINICIAN FORM		Practice ID: (Pre-printed)	
Patient computer ID: _____		Date: ___/___/___			
Patient DOB: ___/___/___		Gender: Female <input type="checkbox"/> Male <input type="checkbox"/>			
1. Is this consultation: <i>(tick one)</i> Telephone <input type="checkbox"/> Face-to-face <input type="checkbox"/>					
<i>Please tick box if patient did NOT attend/ Could NOT be contacted by phone</i>					
2. Health Professional: Patient's GP <input type="checkbox"/>					
<i>(tick one)</i> Duty GP <input type="checkbox"/>					
Other GP <input type="checkbox"/>					
Nurse <input type="checkbox"/>					
Nurse Practitioner <input type="checkbox"/>					
Other _____					
Health Professional Initials: _____					
3. START TIME ___ __ : ___ __ (from computer screen)					
4. Content <i>(please tick one box for each line)</i> 0= no content; 1= some; 2= substantial			5. Further information <i>(tick as many as apply)</i>		
Physical	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	Prescription given/arranged?	yes <input type="checkbox"/> no <input type="checkbox"/>
Social	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	Investigation ordered?	yes <input type="checkbox"/> no <input type="checkbox"/>
Psychological	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	If yes: <input type="checkbox"/>	Bloods <input type="checkbox"/>
Admin	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	(tick as many as apply)	Urine <input type="checkbox"/>
					Swabs <input type="checkbox"/>
					XR <input type="checkbox"/>
					Other <input type="checkbox"/>
6. Follow-up at practice advised?			7. Referred to hospital?		
yes <input type="checkbox"/> no <input type="checkbox"/>			yes <input type="checkbox"/> no <input type="checkbox"/>		
If yes: <input type="checkbox"/>			If yes: <input type="checkbox"/>		
Same day <input type="checkbox"/>			Same day <input type="checkbox"/>		
Next day <input type="checkbox"/>			2 week wait <input type="checkbox"/>		
Within 3-7 days <input type="checkbox"/>			Routine <input type="checkbox"/>		
Later than 7 days <input type="checkbox"/>					
With whom:			8. Other referral?		
GP <input type="checkbox"/>			yes <input type="checkbox"/> no <input type="checkbox"/>		
Nurse <input type="checkbox"/>					
Other <input type="checkbox"/>					
What type:					
Face-to-face <input type="checkbox"/>					
Telephone <input type="checkbox"/>					
Home visit <input type="checkbox"/>					
9. LAST QUESTION!					
"Thank you. As part of our research, we will send you a short questionnaire about your experience of care today. We would be grateful if you would fill it in and return it to the research team we're working with. We would also like a researcher to look at your notes in about 12 weeks' time to examine whether you've needed any further help or advice. Would that be OK?"					
Yes <input type="checkbox"/>					
No <input type="checkbox"/>					
10. END TIME: ___ __ : ___ __ (from computer screen)					

Appendix 12 Case note review

Case Notes Review Form

Patient ID Number Practice ID

DOB Gender Review Date

Is the patient still registered with the surgery? Yes No If no, has the patient:
 Moved away? (date) / /
 Died? (date) / /

Date of same day appointment request / /

Contacts within 4 weeks up to and including: / /

Emergency Hospital Admissions within 7 days up to and including: / /

No contacts after index consultation

	Date	Contact 1 - Index	Contact 2	Contact 3	Contact 4	Contact 5	Contact 6	Contact 7	Contact 8	Contact 9	Contact 10	Contact 11
GP Surgery Contacts												
GP: In Surgery												
GP: Telephone												
GP: Home Visit												
GP: Unspecified												
Nurse: In Surgery												
Nurse: Telephone												
Nurse: Home Visit												
Nurse: Unspecified												
General Unspecified												
DNA												
Doctor												
Nurse												
Unspecified												
Doctor: In Surgery												
Doctor: Telephone												
Doctor: Home Visits												
Nurse: In Surgery												
Nurse: Telephone												
Nurse: Home Visit												
Unspecified												
A&E Contacts												
A&E contact												
Emergency Admission												
Hospital Admission**												
Bed days												

Researcher Comments

** Hospital Admission Codes: Unrel = Unrelated admission; R-P = Related, planned admission; R-U = Related, unplanned admission

Appendix 13 Overview of practice integrity checks

Practice	Run-in period		Data collection period			
	Check 1	Check 2	Check 3	Check 4 ^a	Check 5	Check 6
0101	x	✓	✓			
0104	x	✓	✓			
0107	✓	✓	✓			
0203	x	✓	✓			
0204	x	x	✓	✓		
0205	x	✓	✓			
0206	✓	✓	✓			
0302	✓	✓	✓			
0307	✓	✓	✓			
0308	✓	✓	✓			
0403 ^b	✓	✓	x			
0406	x	x	✓	✓	✓	
0410	x	✓	✓			
0103	✓	✓	✓			
0105	✓	✓	✓			
0110	✓	✓	✓			
0111	x	✓	✓			
0202	x	x	x	✓	✓	✓
0207	✓	✓	✓			
0210	✓	✓	✓	✓		
0301	x	✓	✓			
0304	x	✓	✓			
0305	✓	✓	✓			
0309 ^b	x	✓	x			
0310	x	✓	✓			
0401 ^b	✓	✓	x			
0407	✓	x	✓	✓		
0408	✓	✓	✓			
0102	✓	✓	x			
0106	✓	✓	✓			
0108	x	✓	✓			
0109 ^b	x	✓	x			
0201	✓	✓	✓			
0208	✓	✓	✓			

Practice	Run-in period		Data collection period			
	Check 1	Check 2	Check 3	Check 4 ^a	Check 5	Check 6
0209	✓	✓	✓			
0303	x	✓	x			
0306	x	x	✓	x	✓	
0310 ^b	x	✓	x			
0402	✓	✓	✓			
0404	✓	✓	✓			
0405	✓	✓	✓			
0409	x	✓	✓			

✓, the integrity check was passed; x, the integrity check was failed on one or more aspects.

a When Check 2 was failed, practices remained in run-in and further checks were arranged.

b When a practice proceeded to data collection and failed its final integrity check during the first week of data collection, it was permitted to continue and the trial researcher had further communication with the practice regarding research procedures.

Appendix 14 Assessment of nurses' use of computerised decision support software during ESTEEM

The aim of ESTEEM was to compare GPT with NT and both of these with UC. As part of the NT arm, a CDSS package was installed in practices. All nurses involved in the triage were provided with full training on the use of the software and telephone triage consultation skills. They were told that the software was there as a support tool for them to use and, although the aim of the trial was not to evaluate the software, they should at least open the software for all ESTEEM patients. The aim of this paper is to review the use of this software by the nurses involved in the trial.

We have CDSS data for 14 of the 15 NT practices; at one practice the software was removed before it was possible to run the reports needed (*Table 39*).

The software was installed in all of the NT practices. Where possible, the CDSS software was embedded into the practices' current system. If this was possible the nurse was able to open the patient's record and then open the CDSS, meaning that the patient's demographic information was automatically entered into CDSS from his/her medical records. For these patients we are able to match them to the ESTEEM database using their unique computer ID number.

Owing to the different systems used by practices it was not always possible to embed the software for all of them. For these patients we therefore do not have their computer ID number. When opening the software the nurse was required to manually enter the patient's demographic information. This included either his/her age or date of birth, along with his/her gender. These patients have been matched to the ESTEEM database using date of birth where possible. In addition, the date of call was compared to ensure that the correct patient was identified. When only age was entered this was used alongside patient's

TABLE 39 Summary of practices and patients for whom CDSS data were available

ID	Location	List size	ESTEEM size group	Patients recruited	Complete data	Incomplete data	Incorrect date	No data
103	Urban	7981	Large	524	418	3	0	103
305	Urban	5974	Medium	517	132	4	10	371
407	Rural	6037	Medium	498	401	29	5	63
401	Urban	10,500	Large	570	415	13	13	129
111	Rural	11,267	Large	586	524	7	7	48
110	Urban	15,800	Small	372	2	351	0	19
207	Urban	9233	Large	577	539	17	0	21
301	Rural	5949	Medium	492	359	6	8	119
304	Rural	3227	Small	350	113	10	8	219
309	Inner city	5300	Medium	411	195	0	0	216
105	Urban	12,000	Medium	526	469	21	4	32
311	–	12,500	Large	569	472	34	8	55
210	Urban	9400	Large	555	78	10	3	464
408	Rural	3100	Small	147	135	5	0	7

gender and date of call. Occasionally, it was the case that, for example, two 50-year-old females would ring on the same day; in this case their Clinician Forms were used to identify the start time of the consultation to differentiate between them. In a small number of cases the nurses' initials were also used if more than one nurse was triaging on any particular day.

Any duplicate contacts (calls after date of the index consultation) were removed from the database along with any non-ESTEEM patients. Therefore, each patient had a single row of data in the spreadsheet.

The reports pulled from the software included a wide range of variables, which were used to calculate the level of urgency of the call and therefore indicate to the nurses how quickly the patient needed to be seen. The majority of these variables were purely computational and have little or no meaning by themselves. In addition, the outcome of the call was based on whether the nurses used the software and, if so, to what extent. However, these reports did include the length of the consultation and the number of questions asked by the nurse. The aim is to use these variables to devise a measure of intensity and a proxy variable for use of the software. The length of assessment/number of questions asked is used as a measure of intensity. In addition, we could group consultations into one of three groups: software opened and used, software opened but not used, and software not opened or used.

'Complete data' means that the reports contained both length of assessment (in seconds) and number of questions asked (this implies that the CDSS was opened and used). 'Incomplete data' means the report contained the length of the assessment but as no presenting complaint was entered no information is available regarding the number of questions asked (this implies the software was opened but not used). 'Incorrect date' means that although there are complete data available for this patient the data do not relate to the index consultation. Finally, 'No data' relates to patients for whom the CDSS was not opened or used.

Table 40 shows the average length of call, number of questions asked and intensity score for each practice, along with the percentage of patients within each practice that the software was used for.

The average length of calls was 251.77 seconds or just over 4 minutes. During this time an average of 14 questions was asked, meaning that the nurses spent approximately 23 seconds on each question. However, there does not appear to be any correlation between usage of the software and the length of call or number of questions asked.

This shows that overall the CDSS was at least opened for 71% of patients, although these values varied from 95% down to just 16%. Although the nurses were asked to open the software for all patients, the software was provided as a decision support tool and therefore it was at each nurse's discretion how much they used the software.

Feedback from the nurses who participated in the trial indicated that sometimes the software was not opened as it could not be used for a particular patient. The software was initially developed for out-of-hours primary care services as opposed to a primary care setting, and, therefore, there was no option for medication queries and potential interactions. One example of this was contraception, the only question set within the software related to emergency contraception as opposed to problems with current medication, i.e. forgetting to take a tablet.

The CDSS provided considerable support to the nurses, especially at the beginning of the trial; however, if it was to be used as part of a wider implementation of telephone triage it would be worth investing in software that is specifically designed for a primary care setting.

TABLE 40 Average length of call, number of questions, intensity score and percentage of patients for whom CDSS was used by practice

Practice	Average length of call	Average number of questions asked	Intensity	% used	% opened, not used	% not used
207	283.26 (n = 556)	17.25 (n = 539)	17.73 (n = 536)	93.4	2.9	3.6
408	248.24 (n = 140)	17.24 (n = 135)	14.79 (n = 134)	91.8	3.4	4.8
111	351.75 (n = 538)	15.17 (n = 531)	27.2 (n = 516)	89.4	1.2	9.4
105	179.19 (n = 494)	14.52 (n = 473)	16.62 (n = 467)	89.2	4.0	6.8
311	223.60 (n = 514)	12.00 (n = 480)	24.38 (n = 468)	83.0	6.0	11.1
407	159.73 (n = 435)	7.63 (n = 406)	27.54 (n = 398)	80.5	5.8	13.7
103	416.62 (n = 421)	12.84 (n = 418)	34.92 (n = 400)	79.8	0.6	19.7
301	257.99 (n = 373)	13.10 (n = 367)	24.01 (n = 365)	73.0	1.2	25.8
401	191.74 (n = 441)	13.72 (n = 428)	16.39 (n = 421)	72.8	2.3	24.9
309	388.69 (n = 195)	9.96 (n = 195)	67.61 (n = 190)	47.4	0.0	52.6
304	187.24 (n = 131)	13.60 (n = 121)	16.05 (n = 121)	32.3	2.9	64.9
305	114.66 (n = 146)	12.83 (n = 140)	9.66 (n = 140)	25.5	0.8	73.7
210	143.03 (n = 91)	13.83 (n = 81)	13.81 (n = 77)	14.1	1.8	84.1
110	79.76 (n = 353)	9.5 (n = 2)	15.05 (n = 1)	0.5	94.4	5.1
Mean	251.77 (n = 4828)	14.27 (n = 4316)	23.28 (n = 4234)	62.3	9.1	28.6

Appendix 15 Practice Profile Questionnaire

ESTEEM TRIAL PRACTICE PROFILE QUESTIONNAIRE

The following information will be used to help the research team check whether you are eligible to participate in the ESTEEM trial. The information will only be used by the research team for the purpose of the trial. Please complete all the questions as accurately as possible

1. PRACTICE CONTACT DETAILS

Practice name:

Practice Manager's name:

Practice ID:

Address:

Telephone:

Post code:

Email:

PCT :

Alternative contact name:

Fax:

Telephone:

Email:

2. LIST SIZE AND STAFFING

Q1. What is your current practice list size?

Q2. Please fill in the numbers of staff currently employed within the practice:

(include all health professionals working mainly in the practice, whether employed by the practice or PCT)

	Full time	Part time		Full time	Part time
GPs (principals or salaried)	<input type="text"/>	<input type="text"/>	District nurses	<input type="text"/>	<input type="text"/>
Practice nurses	<input type="text"/>	<input type="text"/>	Midwives	<input type="text"/>	<input type="text"/>
Nurse practitioners	<input type="text"/>	<input type="text"/>	Other	<input type="text"/>	<input type="text"/>
Healthcare assistants	<input type="text"/>	<input type="text"/>	Other	<input type="text"/>	<input type="text"/>
Physiotherapists	<input type="text"/>	<input type="text"/>			

3. PRACTICE CHARACTERISTICS

Please put a cross to indicate your answers ☒

Q3. Would you classify your practice as:

- Rural
- Urban
- (Inner) city

Q4. What computer system is used for patient records?

- EMIS LV
- EMIS PCS
- Microtest
- Vision
- Synergy
- SystemOne
- Other

Q5. Please indicate whether your practice offers any of the following services aimed at improving access to primary care:

- Nurse-led walk-in clinic
- Minor injuries clinics
- Other service (please specify)
- Other service (please specify)

Q6. Please describe the number of sessions that the practice regards as a 'Whole Time Equivalent (WTE)' doctor:

1 session = 4 hours of clinical activity

- Less than 7
- Seven
- Eight
- Nine
- Ten
- More than ten

Please put a cross to indicate your answers ☒

Q8. Does the practice allocate time in the diary for doctors or nurses to carry out telephone triage?

- Yes
 No (if no, please go to Q12)

Q9. How many doctors or nurses currently carry out telephone triage at the practice?

- Doctors
 Nurses

Q10. Which one of the following best describes how telephone triage is used by doctors/nurses at your practice:

- All patients seeking a consultation are triaged on the telephone
 All patients seeking a same-day consultation are triaged on the telephone
 Patients seeking a same-day consultation are triaged on the telephone only if all the appointment slots are taken
 Telephone triage is not used across the whole practice but at least one doctor triages his/her own patients

Q11. What proportion of patients seeking a same-day consultation are triaged on the telephone (please estimate)?

- < 25%
 26-50%
 51-75%
 76-100%
 Cannot estimate

Q12. Please describe below the prescribing responsibilities and other activities that nurses at the practice are currently involved in:

5. CONSULTATIONS/ APPOINTMENTS

Please put a cross to indicate your answers

Q13. How would you describe your consultation management / appointment system?

- Strictly personal list (patients may only see their named doctor)
- Patients are encouraged to see their named doctor but may see other doctors if they wish
- Patients may see any available doctor
- Other arrangement – please describe below

6. TYPE OF APPOINTMENT SYSTEM

Q14. Please estimate what proportion of patients is seen with the following arrangements:

	%
Turn up and wait to be seen	
Pre-booked appointment (booked at least one day in advance)	
Appointments booked on the day	
Other (specify)	
Other (specify)	
TOTAL	100 %

Thank you for your time

**Please return this questionnaire to the Peninsula Medical School in the
reply paid envelope provided.**

OR

Fax to: 01392 432223

THE UNIVERSITY OF
WARWICK

UNIVERSITY OF
EXETER



 University of
BRISTOL

UEA
University of East Anglia

Appendix 16 Availability of patient information throughout the trial, consent status and initial patient management

Owing to the different sources of data, and the fact that not all patients had each source of data available, there were varying combinations and numbers of data available for patients (*Table 41*).

TABLE 41 Patient data availability, intervention status and consent status through the trial

	UC (N = 7283; 34.7%)	GPT (N = 6695; 31.9%)	NT (N = 7012; 33.4%)	Total N = 20,990
Demographic data only (no case note review, Clinician Form or non-blank questionnaire^a): n (%)				
Yes	608 (8.4)	325 (4.9)	209 (3.0)	1142 (5.4)
No	6675 (91.7)	6370 (95.2)	6803 (97.0)	19,848 (94.6)
Clinician Form present: n (%)				
Yes	5984 (82.2)	6032 (90.1)	6598 (94.1)	18,614 (88.7)
No	1299 (17.8)	663 (9.9)	414 (5.9)	2376 (11.3)
Case note review present: n (%)				
Yes	5574 (76.5)	5177 (77.3)	5468 (78.0)	16,219 (77.3)
No	1709 (23.5)	1518 (22.7)	1544 (22.0)	4771 (22.7)
Case note review and Clinician Form present: n (%)				
Yes	5052 (69.4)	4924 (73.6)	5317 (75.8)	15,293 (72.9)
No	2231 (30.6)	1771 (26.5)	1695 (24.2)	5697 (27.1)
Patients returning non-blank^a questionnaire: n (%)				
Yes	4182 (57.4)	4113 (61.4)	3837 (54.7)	12,132 (57.8)
No	3101 (42.6)	2582 (38.6)	3175 (45.3)	8858 (42.2)
Patients responding to at least 50% of 24 questionnaire items^{b,c} (excluding consent): n (%)				
Yes	4124 (98.6)	4060 (98.7)	3745 (98.6)	11,929 (98.7)
Patients responding to at least 75% of 24 questionnaire items^c (excluding consent): n (%)				
Yes	4067 (97.3)	4007 (97.4)	3695 (97.3)	11,769 (97.3)
Case note review and non-blank^a questionnaire present: n (%)				
Yes	3361 (46.2)	3340 (49.9)	3022 (43.1)	9723 (46.3)
No	3922 (53.9)	3355 (50.1)	3990 (56.9)	11,267 (53.7)
Case note review, Clinician Form and non-blank^a questionnaire present: n (%)				
Yes	2844 (39.1)	3092 (46.2)	2877 (41.0)	8813 (42.0)
No	4439 (61.0)	3603 (53.8)	4135 (59.0)	12,177 (58.0)
Patients giving verbal consent to case note review: n (% total N; % with Clinician Form ^d)	5512 (75.7; 91.7)	5204 (77.7; 86.0)	5653 (80.6; 85.3)	16,369 (78.0; 87.6)

continued

TABLE 41 Patient data availability, intervention status and consent status through the trial (*continued*)

	UC (N = 7283; 34.7%)	GPT (N = 6695; 31.9%)	NT (N = 7012; 33.4%)	Total N = 20,990
Patients providing written consent to case note review: n (% total N; % with questionnaire returned ^e)	3306 (45.4; 78.7)	3267 (48.8; 79.3)	2970 (42.4; 77.0)	9543 (45.5; 78.4)
Total patients providing final consent (written or verbal) ^f to case note review: n (%)	5589 (76.7)	5202 (77.7)	5488 (78.3)	16,279 (77.6)
Patients withdrawing consent to case note review (verbal consent provided, written consent refused): n (% total N; % patients with both Clinician Form ^d and questionnaire returned ^e)	608 (8.4; 17.3)	573 (8.6; 15.1)	639 (9.1; 17.4)	1820 (8.7; 16.6)
Patients opting in to consent to case note review on questionnaire (verbal consent refused/missing, written consent provided): n (% total N)	685 (9.4)	571 (8.5)	474 (6.8)	1730 (8.2)
n (including patients with both Clinician Form ^d and questionnaire returned ^e , % patients with both Clinician Form ^d and questionnaire returned ^e)	175 (5.0)	332 (8.7)	337 (9.2)	844 (7.7)
Patients treated per protocol^g according to case note review: n (%)				
Yes	5574 (100)	4802 (92.8)	4860 (88.9)	15,236 (93.9)
No	0 (0)	375 (7.2)	608 (11.1)	983 (6.1)
Patients treated per protocol^g according to Clinician Form: n (%)				
Yes	5984 (100)	5665 (93.9)	5661 (85.8)	17,310 (93.0)
No	0 (0)	367 (6.1)	937 (14.2)	1304 (7.0)
<p>a At least one question (excluding consent) completed.</p> <p>b Minimum number of questions answered was 4/24 (132 patients ≈ 1.1%).</p> <p>c Excludes 108 patients who received short version of questionnaire, 40 of whom returned a non-blank questionnaire.</p> <p>d Includes 71 patients with a Clinician Form dated prior to index day, not otherwise included.</p> <p>e Includes 32 patients with consent status positively recorded but return of questionnaire not otherwise recorded.</p> <p>f Written consent or verbal consent if written consent missing; 60 patients who consented to case note review did not have case note review performed.</p> <p>g By case note review: patients in GPT/NT treated per protocol if the first contact after the index consultation request was a telephone triage according to intervention on day of index consultation request. By Clinician Form: patients in GPT/NT treated per protocol if the contact type was a telephone triage according to intervention on day of index consultation request. All patients in UC were per protocol by both definitions.</p>				

Case note review data were available for 77% (16,219/20,990) of patients (similar across the three arms). There was some variation across trial arms with regard to the rates of completion of Clinician Forms and return of a completed questionnaire; these variations inevitably led to imbalances in the numbers of patients with a full set of trial data (case note review, Clinician Form and questionnaire) available.

The complexity of the consent procedure was compounded by the differing availability of information at different stages of the trial. Only patients with a Clinician Form available were able to provide verbal consent, and although the UC arm had the lowest rate of availability of Clinician Forms, the UC arm had the highest proportion of patients with a Clinician Form providing verbal consent (92%, 5512/5984). Across the three arms, there was little difference in the proportions of patients who provided written consent to case note review having returned a questionnaire; also, the proportions of patients by arm providing final consent to case note review were similar.

Patients were categorised in the triage arms by whether or not they were treated per protocol according to both case note review data and Clinician Form data (although it should be remembered that these data sources could conflict if both were available for an individual patient).

Appendix 17 Practice Profile Questionnaire: additional data

Practice profile item	UC (N = 14)	GPT (N = 13)	NT (N = 15)
Number of nurse practitioners at practice (including full- and part-time): ^a mean (SD), n	0.6 (0.8), 14	0.9 (1.6), 13	0.8 (0.7), 15
Does the practice have a nurse-led walk-in clinic? n (%)			
Yes	1 (7.1)	0 (0)	2 (13.3)
No	13 (92.9)	13 (100)	13 (86.7)
Total N	14	13	15
Does the practice have minor injuries clinics? n (%)			
Yes	4 (28.6)	4 (30.8)	0 (0)
No	10 (71.4)	9 (69.2)	15 (100)
Total N	14	13	15
Number of sessions regarded as WTE: n (%)			
< 7	0 (0)	0 (0)	1 (7.7)
7	3 (23.1)	0 (0)	0 (0)
8	5 (38.5)	6 (46.2)	8 (61.5)
9	5 (38.5)	6 (46.2)	2 (15.4)
10	0 (0)	1 (7.7)	1 (7.7)
> 10	0 (0)	0 (0)	1 (7.7)
Total N	13	13	13
How many doctors currently carry out telephone triage at the practice? n (%)			
0	1 (25.0)	4 (50.0)	3 (37.5)
1–3	1 (25.0)	1 (12.5)	0 (0)
≥ 4	2 (50.0)	3 (37.5)	5 (62.5)
Total N	4	8	8
How many nurses currently carry out triage at the practice? n (%)			
0	1 (33.3)	5 (71.4)	4 (66.7)
1	2 (66.7)	0 (0)	0 (0)
2	0 (0)	2 (28.6)	1 (16.7)
≥ 3	0 (0)	0 (0)	1 (16.7)
Total N	3	7	6

Practice profile item	UC (N = 14)	GPT (N = 13)	NT (N = 15)
How would you describe your consultation management/appointment system? n (%)			
Patients are encouraged to see their named doctor but may see other doctors if they wish	7 (50.0)	5 (38.5)	5 (33.3)
Patients may see any available doctor	6 (42.9)	8 (61.5)	10 (66.7)
Other arrangement	1 (7.1)	0 (0)	0 (0)
Total N	14	13	15
Estimated proportion (%) of patients seen with the following arrangements: n (%)			
<i>Turn up and wait to be seen</i>			
0–24	13 (100)	12 (100)	15 (100)
25–49	0 (0)	0 (0)	0 (0)
50–74	0 (0)	0 (0)	0 (0)
≥ 75	0 (0)	0 (0)	0 (0)
Total N	13	12	15
<i>Prebooked appointment (at least 1 day in advance)</i>			
0–24	1 (7.7)	0 (0)	1 (6.7)
25–49	1 (7.7)	2 (16.7)	5 (33.3)
50–74	7 (53.9)	7 (58.3)	6 (40.0)
≥ 75	4 (30.8)	3 (25.0)	3 (20.0)
Total N	13	12	15
<i>Appointment booked on the day</i>			
0–24	3 (23.1)	4 (33.3)	5 (33.3)
25–49	5 (38.5)	6 (50.0)	4 (26.7)
50–74	3 (23.1)	2 (16.7)	5 (33.3)
≥ 75	2 (15.4)	0 (0)	1 (6.7)
Total N	13	12	15
<i>Other</i>			
0–24	13 (100)	11 (91.7)	15 (100)
25–49	0 (0)	1 (8.3)	0 (0)
50–74	0 (0)	0 (0)	0 (0)
≥ 75	0 (0)	0 (0)	0 (0)
Total N	13	12	15

WTE, whole-time equivalent.

a If data missing for either/or full- and part-time nurse practitioners, total assumed to be 0.

Appendix 18 Predictors of case note review and questionnaire return

Patient-level demographic characteristics for patients who had a case note review performed, and patients who returned a non-blank questionnaire, were not substantively different from those for the full cohort (Table 42). Among patients with case note review, there were higher proportions in the two oldest age groups in GPT; also, GPT had fewer patients in deprivation quintiles 1 and 2. There was a slight tendency for patients in the NT arm to be more likely to have case note reviews [odds ratio (OR) 1.08, 95% CI 1.00 to 1.17] than UC patients (Table 43), although NT patients had lower odds of returning a completed questionnaire than UC patients (OR 0.88, 95% CI 0.82 to 0.94). Patients in GPT, however, were more likely to return a questionnaire than those in UC (OR 1.10, 95% CI 1.03 to 1.19).

TABLE 42 Baseline patient demographics by availability of case note review data and return of completed questionnaire

Patient demographic item	UC (N = 7283)	GPT (N = 6695)	NT (N = 7012)
Individual patient characteristics for patients with case note review data available			
	UC (N = 5574)	GPT (N = 5177)	NT (N = 5468)
Gender: ^a n (%)			
Male	2286 (41.0)	2142 (41.4)	2179 (39.9)
Female	3288 (59.0)	3035 (58.6)	3289 (60.2)
Age ^a (years): mean (SD)	42.5 (23.9)	45.8 (25.2)	41.9 (25.4)
Age by category, years; n (%)			
< 5	513 (9.2)	460 (8.9)	644 (11.8)
5–11	359 (6.4)	291 (5.6)	360 (6.6)
16–24	597 (10.7)	467 (9.0)	562 (10.3)
25–59	2537 (45.5)	2161 (41.7)	2334 (42.7)
60–74	1107 (19.9)	1093 (21.1)	970 (17.7)
≥ 75	461 (8.3)	705 (13.6)	598 (10.9)
Deprivation (^b IMD 2010 ⁵⁴ score): mean (SD), n	17.4 (10.3), 5536	16.9 (11.7), 5162	17.8 (11.6), 5405
Deprivation (IMD 2010 ⁵⁴ quintile based on rank): n (%)			
Quintile 1 (most deprived)	352 (6.4)	383 (7.4)	492 (9.1)
Quintile 2	1225 (22.1)	735 (14.2)	1085 (20.1)
Quintile 3	1417 (25.6)	1577 (30.6)	1320 (24.4)
Quintile 4	1502 (27.1)	1486 (28.8)	1402 (25.9)
Quintile 5 (least deprived)	1040 (18.8)	981 (19.0)	1106 (20.5)
Individual patient characteristics for patients who returned a non-blank^c questionnaire			
	UC (N = 4182)	GPT (N = 4113)	NT (N = 3837)
Gender: ^a n (%)			
Male	1610 (38.5)	1633 (39.7)	1465 (38.2)
Female	2572 (61.5)	2480 (60.3)	2372 (61.8)
Age ^a (years): mean (SD)	47.0 (24.0)	49.9 (24.8)	47.8 (25.4)

continued

TABLE 42 Baseline patient demographics by availability of case note review data and return of completed questionnaire (*continued*)

Patient demographic item	UC (N = 7283)	GPT (N = 6695)	NT (N = 7012)
Individual patient characteristics for patients who returned a non-blank^c questionnaire			
	UC (N = 4182)	GPT (N = 4113)	NT (N = 3837)
Age by category, years: n (%)			
< 5	349 (8.4)	328 (8.0)	387 (10.1)
5–11	225 (5.4)	212 (5.2)	216 (5.6)
16–24	303 (7.3)	230 (5.6)	225 (5.9)
25–59	1787 (42.7)	1610 (39.1)	1523 (39.7)
60–74	1077 (25.8)	1077 (26.2)	942 (24.6)
≥ 75	441 (10.6)	656 (16.0)	544 (14.2)
Deprivation (^bIMD 2010⁵⁴ score): mean (SD), n	16.6 (9.6), 4157	16.2 (10.6), 4104	16.2 (10.5), 3801
Deprivation (IMD 2010⁵⁴ quintile based on rank): n (%)			
Quintile 1 (most deprived)	210 (5.1)	244 (6.0)	240 (6.3)
Quintile 2	844 (20.3)	603 (14.7)	698 (18.4)
Quintile 3	1107 (26.6)	1227 (29.9)	897 (23.6)
Quintile 4	1161 (27.9)	1238 (30.2)	1067 (28.1)
Quintile 5 (least deprived)	835 (20.1)	792 (19.3)	899 (23.7)

a From baseline demographic data.

b IMD 2010⁵⁴ score and rank derived from residential postcode data mapped to LSOA.

c At least one question complete (excluding consent).

TABLE 43 Demographic associations with consent to case note review and questionnaire response

Demographic variable	Case note review available: OR (95% CI)	Questionnaire response (non-blank ^a questionnaire returned): OR (95% CI)
Arm (reference: UC)		
GPT	1.01 (0.94 to 1.10)	1.10 (1.03 to 1.19)
NT	1.08 (1.00 to 1.17)	0.88 (0.82 to 0.94)
Age (years; reference: 25–59 years)		
0–4	1.02 (0.91 to 1.14)	0.93 (0.85 to 1.03)
5–11	1.08 (0.94 to 1.23)	0.90 (0.80 to 1.01)
16–24	0.87 (0.78 to 0.96)	0.45 (0.41 to 0.49)
60–74	1.43 (1.31 to 1.58)	3.54 (3.24 to 3.87)
≥ 75	1.47 (1.30 to 1.66)	2.82 (2.53 to 3.14)
Gender (reference: female)		
Male	1.10 (1.03 to 1.17)	0.82 (0.77 to 0.87)
Deprivation (^bIMD 2010⁵⁴ quintile based on rank; reference: quintile 5, least deprived)		
Quintile 1 (most deprived)	0.93 (0.81 to 1.06)	0.49 (0.43 to 0.55)
Quintile 2	0.93 (0.84 to 1.03)	0.69 (0.63 to 0.76)
Quintile 3	1.07 (0.97 to 1.18)	0.83 (0.76 to 0.90)
Quintile 4	1.10 (1.00 to 1.22)	0.97 (0.89 to 1.05)

a At least one question complete (excluding consent).

b IMD 2010⁵⁴ score and rank derived from residential postcode data mapped to LSOA.

Young adults aged 16–24 years had lower odds of having a case note review than the reference age group (OR 0.87, 95% CI 0.78 to 0.96). Young adults also had lower odds of returning a completed questionnaire (OR 0.45, 95% CI 0.41 to 0.49). Older patients in the age groups 60–74 years and > 75 years were more likely to have a case note review than adults in the reference category; older patients also had considerably higher odds of returning a questionnaire (see *Table 43*). Male patients had higher odds of having a case note review performed than female patients (OR 1.10, 95% CI 1.03 to 1.17), but had lower odds of returning a questionnaire (OR 0.82, 95% CI 0.77 to 0.87).

Patients in quintiles 1 and 2 had lower odds of having a case note review performed than patients in the least deprived quintile (quintile 5), whereas patients in quintiles 3 and 4 had higher odds of having a case note review performed; however, for all four comparator quintiles, the CIs were wide and included '1' (see *Table 42*). With regard to return of a completed questionnaire, however, there was a clear trend evident, with odds of returning a completed questionnaire being lower in the three most deprived quintiles; quintiles 1, 2 and 3 all had lower odds (vs. quintile 5) of returning a completed questionnaire and a 95% CI with an upper bound of < 1.

These analyses indicate clear associations with demographic factors (and treatment arm) and the availability of outcome data, most importantly with the availability of primary outcome data and safety outcome data (A&E attendances and emergency hospital admissions), both derived from case note review and patient satisfaction data derived from the questionnaire.

Appendix 19 Total primary care and accident and emergency contacts within 28 days of index day (primary outcome) across trial arm: per-protocol analysis

	UC (N = 5572)	GPT (N = 4796)	NT (N = 4860)
Total primary care contacts: ^{a,b} total contacts, mean (SD)	10,616, 1.91 (1.43)	12,963, 2.70 (1.75)	14,116, 2.90 (1.66)
RR (95% CI)			
	GPT vs. UC	NT vs. UC	NT vs. GPT
Poisson GLLAMM ^c with robust residuals; adjusted ^d (n = 15,134)	1.39 (1.34 to 1.45)	1.51 (1.46 to 1.57)	1.11 (0.99 to 1.23)
	Cluster-level SD: ^e 0.07		

GLLAMM, Generalized Linear and Latent Mixed Models.

a Patients who died within 28 days of index day excluded.

b Includes all 20 contact types constituting the primary outcome.

c All models include random effect on practice.

d Adjusted models adjust for practice variables: location (Devon, reference; Bristol; Warwick; Norwich), practice list size [large (> 8000 patients), reference; medium (3500–8000 patients); small (< 3500 patients)], practice deprivation [non-deprived (at or below average deprivation for England by APHO ratings), reference; deprived (above average for England by APHO ratings)] and patient variables: age (categorised, 0–4; 5–11; 16–24; 25–59, reference; 60–74; ≥ 75 years), gender (reference: female), IMD 2010⁵⁴ deprivation based on residential postcode categorised into quintiles by rank (reference: least deprived, quintile 5).

e Approximate coefficient of variation for Poisson model with log link.

Appendix 20 Interactions between intervention and practice-level variables, and between intervention and age category/gender/ethnic group, for primary outcome

Interaction term ^{a-d}	RR (95% CI)	p-value
Arm × location		
GPT vs. UC × Bristol	1.01 (0.92 to 1.11)	0.08
GPT vs. UC × Warwick	1.01 (0.81 to 1.26)	
GPT vs. UC × Norwich	0.92 (0.81 to 1.04)	
NT vs. UC × Bristol	0.86 (0.81 to 0.91)	< 0.001
NT vs. UC × Warwick	0.96 (0.91 to 1.02)	
NT vs. UC × Norwich	0.91 (0.77 to 1.07)	
NT vs. GPT × Bristol	1.05 (0.97 to 1.14)	< 0.001
NT vs. GPT × Warwick	1.57 (1.40 to 1.75)	
NT vs. GPT × Norwich	1.48 (1.36 to 1.62)	
Arm × practice size		
GPT vs. UC × medium	1.10 (0.74 to 1.63)	0.56
GPT vs. UC × small	0.93 (0.81 to 1.07)	
NT vs. UC × medium	1.05 (0.40 to 2.76)	< 0.001
NT vs. UC × small	1.39 (0.32 to 5.93)	
NT vs. GPT × medium	0.96 (0.25 to 3.62)	< 0.001
NT vs. GPT × small	1.49 (0.39 to 5.62)	
Arm × practice deprivation		
GPT vs. UC × practice deprived	0.99 (0.93 to 1.06)	0.75
NT vs. UC × practice deprived	0.92 (0.86 to 0.99)	0.02
NT vs. GPT × practice deprived	0.93 (0.86 to 1.02)	0.12
Arm × patient age (years)		
GPT vs. UC × 0–4	1.09 (0.98 to 1.22)	< 0.001
GPT vs. UC × 5–11	1.12 (1.03 to 1.23)	
GPT vs. UC × 16–24	1.01 (0.93 to 1.09)	
GPT vs. UC × 60–74	0.95 (0.88 to 1.03)	
GPT vs. UC × ≥ 75	0.92 (0.83 to 1.03)	

Interaction term ^{a-d}	RR (95% CI)	p-value
NT vs. UC × 0–4	1.06 (0.95 to 1.17)	0.01
NT vs. UC × 5–11	1.11 (1.02 to 1.19)	
NT vs. UC × 16–24	1.01 (0.94 to 1.09)	
NT vs. UC × 60–74	0.97 (0.89 to 1.06)	
NT vs. UC × ≥ 75	0.92 (0.85 to 0.99)	
Arm × patient gender		
NT vs. GPT × 0–4	0.97 (0.89 to 1.06)	0.86
NT vs. GPT × 5–11	0.99 (0.90 to 1.07)	
NT vs. GPT × 16–24	1.00 (0.93 to 1.08)	
NT vs. GPT × 60–74	1.02 (0.95 to 1.10)	
NT vs. GPT × ≥ 75	0.99 (0.91 to 1.08)	
Arm × patient ethnicity^e		
GPT vs. UC × 'Other ethnic group'	0.95 (0.91 to 1.00)	0.04
NT vs. UC × 'Other ethnic group'	1.00 (0.96 to 1.05)	0.87
NT vs. GPT × 'Other ethnic group'	1.05 (1.01 to 1.10)	0.01
Arm by patient ethnicity^e		
GPT vs. UC × 'Other ethnic group'	0.90 (0.75 to 1.08)	0.25
NT vs. UC × 'Other ethnic group'	1.05 (0.85 to 1.28)	0.66
NT vs. GPT × 'Other ethnic group'	1.16 (1.00 to 1.36)	0.06

a Patients who died within 28 days of index day excluded.

b All models include random effect on practice.

c Model adjusts for practice variables: location (Devon, reference; Bristol; Warwick; Norwich), practice list size [large (> 8000 patients), reference; medium (3500–8000 patients); small (< 3500 patients)], practice deprivation [non-deprived (at or below average deprivation for England by APHO ratings), reference; deprived (above average for England by APHO ratings)] and patient variables: age (categorised, 0–4; 5–11; 16–24; 25–59, reference; 60–74; ≥ 75 years), gender (reference: female), IMD 2010⁵⁴ deprivation based on residential postcode categorised into quintiles by rank (reference: least deprived, quintile 5).

d Interaction terms derived by fitting a series of models with an interaction term for arm with covariate of interest (practice, deprivation, location, list size and patient age): interactions between arm and only one covariate were tested in each model; models were fitted twice, once with UC as reference for treatment arm and once with GPT as reference.

e Model fitted with standard adjustment variables and ethnic group dichotomised as White or Other ethnic group (comprising ethnic groups: Mixed/multiple ethnic groups; Asian/Asian British; Black/African/Caribbean/Black British; and Other ethnic group).

Appendix 21 Patient-reported health status (European Quality of Life-5 Dimensions): individual question frequencies

EQ-5D item	UC (N = 6123) ^a	GPT (N = 5711) ^a	NT (N = 5719) ^a
Q1: mobility: n (%)			
1	2531 (72.3)	2413 (69.8)	2223 (71.5)
2	965 (27.6)	1031 (29.8)	883 (28.4)
3	6 (0.2)	11 (0.3)	5 (0.2)
Total N	3502 (100.0)	3455 (100.0)	3111 (100.0)
Q2: self-care: n (%)			
1	3149 (90.5)	3052 (88.6)	2774 (89.7)
2	312 (9.0)	360 (10.5)	307 (9.9)
3	20 (0.6)	34 (1.0)	13 (0.4)
Total N	3481 (100.0)	3446 (100.0)	3094 (100.0)
Q3: usual activities: n (%)			
1	2317 (66.1)	2201 (63.5)	2056 (66.2)
2	1059 (30.2)	1116 (32.2)	929 (29.9)
3	131 (3.7)	148 (4.3)	123 (4.0)
Total N	3507 (100.0)	3465 (100.0)	3108 (100.0)
Q4: pain/discomfort: n (%)			
1	1632 (46.6)	1631 (47.3)	1501 (48.5)
2	1629 (46.5)	1589 (46.1)	1391 (45.0)
3	239 (6.8)	225 (6.5)	200 (6.5)
Total N	3500 (100.0)	3445 (100.0)	3092 (100.0)
Q5: anxiety/depression: n (%)			
1	2431 (70.4)	2421 (70.9)	2253 (73.4)
2	906 (26.2)	880 (25.8)	722 (23.5)
3	117 (3.4)	113 (3.3)	93 (3.0)
Total N	3454 (100.0)	3414 (100.0)	3068 (100.0)

^a Includes only patients aged ≥ 16 years.

Appendix 22 Patient experience of care: individual question frequencies

Patient experience item	UC (N = 7283)	GPT (N = 6695)	NT (N = 7012)
How easy or difficult was it to get through to the practice on the phone? n (%)			
Very easy	1379 (33.9)	1850 (46.1)	1266 (34.2)
Fairly easy	1497 (36.8)	1459 (36.3)	1416 (38.3)
Neither easy nor difficult	372 (9.2)	290 (7.2)	340 (9.2)
Fairly difficult	580 (14.3)	318 (7.9)	468 (12.7)
Very difficult	235 (5.8)	100 (2.5)	211 (5.7)
Total N	4063 (100.0)	4017 (100.0)	3701 (100.0)
How easy or difficult was it to receive prompt care? n (%)			
Very easy	1885 (49.4)	1888 (50.7)	1282 (37.0)
Fairly easy	1363 (35.7)	1273 (34.2)	1344 (38.8)
Neither easy nor difficult	373 (9.8)	352 (9.4)	484 (14.0)
Fairly difficult	134 (3.5)	151 (4.1)	259 (7.5)
Very difficult	61 (1.6)	63 (1.7)	92 (2.7)
Total N	3816 (100.0)	3727 (100.0)	3461 (100.0)
How easy or difficult was it to see a GP or nurse if you wanted to? n (%)			
Very easy	1836 (47.3)	1607 (44.4)	1273 (36.4)
Fairly easy	1341 (34.5)	1166 (32.2)	1253 (35.8)
Neither easy nor difficult	339 (8.7)	421 (11.6)	449 (12.8)
Fairly difficult	272 (7.0)	276 (7.6)	369 (10.5)
Very difficult	95 (2.5)	150 (4.1)	157 (4.5)
Total N	3883 (100.0)	3620 (100.0)	3501 (100.0)
How easy or difficult was it to get medical help or advice for the problem you presented? n (%)			
Very easy	2158 (56.5)	2208 (58.1)	1634 (46.7)
Fairly easy	1168 (30.6)	1137 (29.9)	1233 (35.2)
Neither easy nor difficult	318 (8.3)	274 (7.2)	376 (10.7)
Fairly difficult	123 (3.2)	120 (3.2)	178 (5.1)
Very difficult	52 (1.4)	59 (1.6)	81 (2.3)
Total N	3819 (100.0)	3798 (100.0)	3502 (100.0)
How convenient was the care provided by your GP surgery on that day? n (%)			
Very convenient	2699 (66.3)	2602 (64.7)	2106 (57.1)
Fairly convenient	1181 (29.0)	1139 (28.3)	1231 (33.4)
Not very convenient	148 (3.6)	212 (5.3)	251 (6.8)
Not at all convenient	44 (1.1)	69 (1.7)	101 (2.7)
Total N	4072 (100.0)	4022 (100.0)	3689 (100.0)

Patient experience item	UC (N = 7283)	GPT (N = 6695)	NT (N = 7012)
Thinking about the reason why you contacted the GP surgery or health centre that day – is the problem now ... ? n (%)			
Much better	2178 (53.4)	2352 (58.7)	2186 (58.7)
A bit better	952 (23.3)	808 (20.2)	740 (19.9)
The same	790 (19.4)	664 (16.6)	616 (16.6)
A bit worse	79 (1.9)	91 (2.3)	82 (2.2)
Much worse	55 (1.4)	50 (1.3)	51 (1.4)
Don't know	25 (0.6)	45 (1.1)	48 (1.3)
Total N	4079 (100.0)	4010 (100.0)	3723 (100.0)
Overall, how satisfied or dissatisfied were you with the care received on that day? n (%)			
Very satisfied	2710 (66.2)	2630 (65.2)	2199 (59.4)
Fairly satisfied	1044 (25.5)	1008 (25.0)	1063 (28.7)
Neither satisfied nor dissatisfied	208 (5.1)	217 (5.4)	237 (6.4)
Fairly dissatisfied	95 (2.3)	129 (3.2)	134 (3.6)
Very dissatisfied	36 (0.9)	50 (1.2)	71 (1.9)
Total N	4093 (100.0)	4034 (100.0)	3704 (100.0)

Appendix 23 Case complexity

Clinician Form available ^a	UC (N = 5984)	GPT (N = 6032)	NT (N = 6598)
% missing (excluding Clinician Forms indicated as DNA): n (%)			
Physical component	130 (2.2)	113 (1.9)	152 (2.4)
Social component	1003 (16.9)	1320 (22.5)	982 (15.2)
Psychological component	938 (15.8)	1265 (21.6)	953 (14.8)
Administrative component	958 (16.2)	1322 (22.5)	984 (15.3)
Overall complexity ^b	1082 (18.2)	1444 (24.6)	1047 (16.2)
Total N	5931	5868	6450
Overall case complexity ^{a,b} (scale of 0–8): mean (SD), n	2.32 (0.88), 4854	1.92 (0.73), 4431	2.20 (1.04), 5409
Physical component:^a n (%)			
0	262 (4.5)	257 (4.5)	204 (3.2)
1	946 (16.3)	1580 (27.4)	1571 (24.9)
2	4607 (79.2)	3937 (68.2)	4538 (71.9)
Total N	5815 (100.0)	5774 (100.0)	6313 (100.0)
Social component:^a n (%)			
0	4204 (85.2)	4264 (93.5)	4682 (85.4)
1	647 (13.1)	248 (5.4)	728 (13.3)
2	85 (1.7)	49 (1.1)	71 (1.3)
Total N	4936 (100.0)	4561 (100.0)	5481 (100.0)
Psychological component:^a n (%)			
0	4055 (81.1)	3946 (85.5)	4511 (81.9)
1	605 (12.1)	469 (10.2)	724 (13.1)
2	340 (6.8)	201 (4.4)	275 (5.0)
Total N	5000 (100.0)	4616 (100.0)	5510 (100.0)
Administrative component:^a n (%)			
0	4044 (81.2)	4096 (89.9)	4799 (87.7)
1	745 (15.0)	337 (7.4)	525 (9.6)
2	191 (3.8)	121 (2.7)	148 (2.7)
Total N	4980 (100.0)	4554 (100.0)	5472 (100.0)

a Includes 365 Clinician Forms indicated as 'DNA', 18 of which had total case complexity data available.

b Sum of the four individual complexity components.

Appendix 24 ESTEEM intervention resource use and cost estimates

Number of attendees by practice

Organisational training: descriptive detail of training courses attended

General practitioner practices

Practice ID ^a	GP	Nurse	Nurse practitioner	Practice manager	Reception	Duration (minutes)
Organisational training: GP practices						
1	2	0	1	2	2	60
2	2	0	0	1	1	60
3	4	0	0	1	7	60
4	2	0	0	1	2	90
5	2	0	0	1	2	60
6	2	0	0	1	2	60
7	2	0	0	1	2	75
8	2	0	0	1	3	120
9	2	0	0	1	2	120
10	4	0	0	1	5	120
11	4	0	0	2	2	90
12	1	0	0	1	3	60
13	4	1	0	2	2	120
Skills training: GP practices						
1	2	0	1	2	2	75
2	2	0	0	1	1	60
3	4	0	0	1	7	60
4	2	0	0	0	0	60
5	2	0	0	1	2	60
6	2	0	0	1	1	60
7	2	0	0	0	0	60
8	2	0	0	1	3	45
9	2	0	0	1	2	120
10	3	0	0	1	2	120
11	4	0	0	2	2	90
12	1	0	0	1	3	60
13	4	2	0	2	2	120

a Study practice ID removed for presentation of data.

Nurse triage

Practice ID ^a	GP	Nurse	Nurse practitioner	Practice manager	Reception	Duration (minutes)	
Organisational training: NT practices							
14	0	1	0	2	7	60	
15	1	2	2	2	0	60	
16	2	3	0	1	0	120	
17	0	0	0	2	1	120	
18	0	2	0	1	2	90	
19	0	2	0	1	1	90	
20	0	2	0	1	1	75	
21	1	0	0	1	2	90	
22	1	2	0	1	2	120	
23	1	0	0	1	3	120	
24	0	1	0	0	2	120	
25	1	1	0	1	2	120	
26	1	3	0	1	3	180	
27	1	1	0	0	2	120	
28	1	1	0	1	3	120	
Skills training: NT practices							
						Session 1	Session 2
14	0	4	0	1	0	90	360
15	0	2	2	0	0	195	195
16	0	3	0	0	0	210	210
17	1	6	1	2	1	180	195
18	0	2	0	0	0	140	120
19	0	2	0	0	0	180	120
20	0	2	0	0	0	180	120
21	0	3	0	0	0	420 (combined)	
22	0	2	0	0	0	165	120
23	0	2	0	0	0	420 (combined)	
24	0	1	0	0	0	420 (combined)	0
25	0	3	0	0	0	420 (combined)	
26	0	4	0	1	0	60	30
27	0	4	0	0	0	60	60
28	0	2	0	0	0	180	180
a Study practice ID removed for presentation of data.							

Remote one-to-one training, nurse triage

Practice ID ^a	Remote training		Duration (minutes)
	Nurse	Nurse practitioner	
14	4	0	180
15	2	2	180
16	3	0	180
17	6	1	180
18	2	0	180
19	2	0	180
20	2	0	180
21	3	0	180
22	2	0	180
23	2	0	180
24	1	0	180
25	3	0	180
26	4	0	180
27	4	0	180
28	2	0	180

^a Trial practice ID removed for publication.

Estimation of unit cost for non-medical general practice staff

Component	Practice manager/other band 6 non-clinical staff	Admin/other band 4 non-medical staff
Salary (£)	29,464 ^a	20,433 ^b
Salary oncosts (£) ^c	7071	4933
Overheads: non-staff (£) ^d	15,334	10,646
Overheads: capital (£) ^e	3654	2536
Estimated total cost/annum (£)	55,523	38,548
Working hours/weeks ^a	37.5/42	37.5/42
Estimated unit cost/hour (£)	35	24
Estimated unit cost/minute (£)	0.59	0.41

a Based on a worker at the mid-point (increment) of Agenda for Change band 6 [www.nhs Careers.nhs.uk/working-in-the-nhs/pay-and-benefits/agenda-for-change-pay-rates/ (accessed 1 November 2014)].

b Based on a worker at the mid-point (increment) of Agenda for Change band 4 [www.nhs Careers.nhs.uk/working-in-the-nhs/pay-and-benefits/agenda-for-change-pay-rates/ (accessed 1 November 2014)].

c Based on 24% of salary costs, for employment oncosts (pension, National Insurance contribution), as seen in estimates for staff grades/similar salary structures reported in PSSRU report (Curtis³⁸).

d Based on allowance of 41.97% of salary costs (including oncosts), as the framework for costing applied by Curtis,³⁸ in PSSRU-estimated staff costs for similar salary structures.

e Based on an assumption of 10% of salary costs, applying similar framework to that reported by Curtis,³⁸ in PSSRU-estimated staff costs for similar salary structures.

Note

Published unit costs were/are not reported for the non-medical GP staff above, therefore cost estimates have been set out using reported salary levels by staff grade and assumptions to estimate salary oncosts, overheads and subsequent cost per hour/minute. Methods are based on the framework set out by Curtis³⁸ in PSSRU cost estimates for staff with similar salary structures/levels.

Appendix 25 Data used to estimate staff contact time for triage contact (source: Clinician Form data)

Descriptive flow of Clinician Form timings data for health-economic analysis

General practitioner triage arm

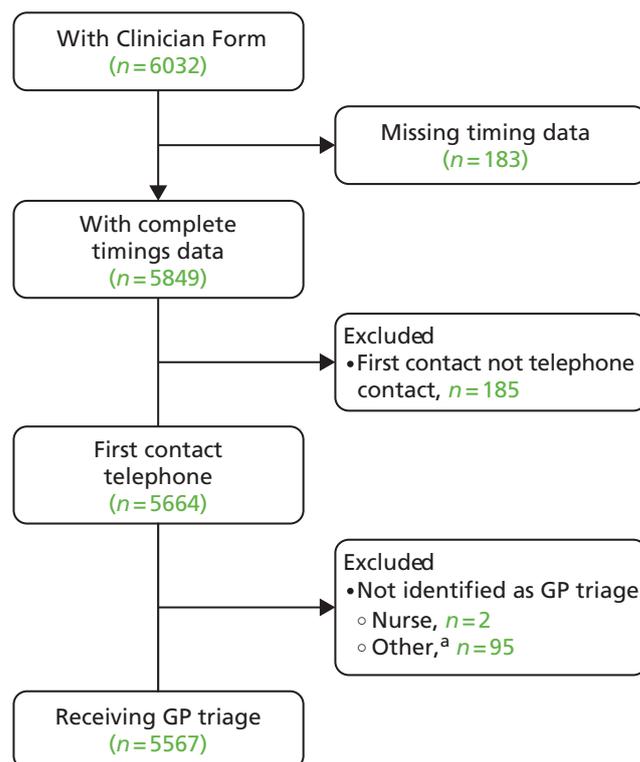


FIGURE 12a Descriptive flow of Clinician Form timings data for health economics analysis for GP triage arm. a, Coded as 'other'. Includes data entries that are 'missing/spoiled/other'. Note that 'GP' includes any type of GP (e.g. locum, GP registrar), foundation doctor (Year 2) and physician assistant. 'Nurse' includes nurse practitioner.

Nurse triage arm

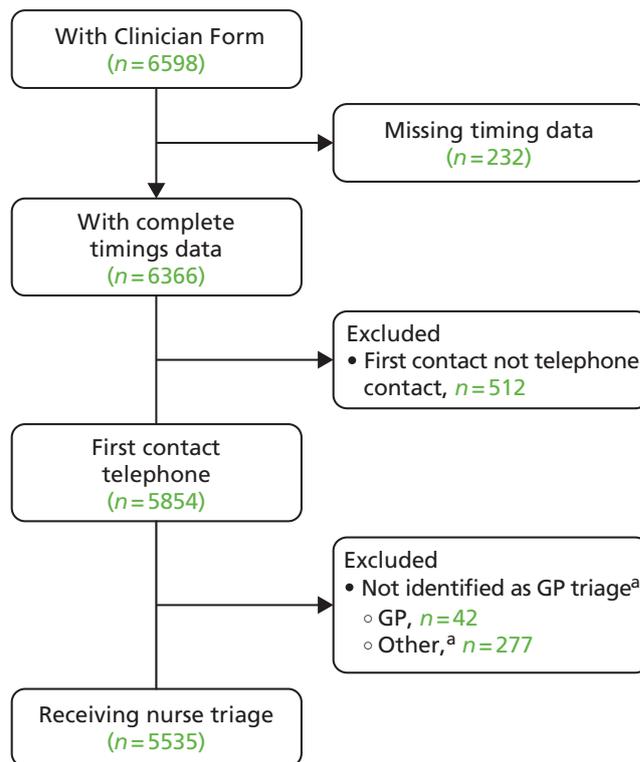


FIGURE 12b Descriptive flow of Clinician Form timings data for health economics analysis for nurse triage arm. ^a, Coded/cleaned as 'other'. Includes data entries that are 'missing/spoiled/other'. Note that 'GP' includes any type of GP (e.g. locum, GP registrar), foundation doctor (Year 2) and physician assistant. 'Nurse' includes nurse practitioner.

Appendix 26 Mean triage contact time by practice

General practitioner-led telephone triage: mean time (minutes) taken by general practitioners in delivering triage

Practice ID ^a	Mean	SD	Min.	Max.	<i>n</i>
1	4.00	2.36	1.00	15.00	490
2	3.02	3.03	1.00	36.00	290
3	4.73	2.86	1.00	22.00	404
4	3.88	2.76	1.00	21.00	487
5	4.36	2.60	1.00	22.00	280
6	3.50	3.50	1.00	58.00	460
7	3.22	2.10	1.00	15.00	534
8	4.24	2.83	1.00	22.00	449
9	3.16	1.68	1.00	11.00	342
10	3.90	2.81	1.00	23.00	492
11	4.54	2.93	1.00	29.00	489
12	4.54	2.67	1.00	18.00	355
13	4.74	3.29	1.00	47.00	495
Total	4.00	2.83	1.00	58.00	5567

Max., maximum; min., minimum.

a Trial practice ID removed for publication.

Nurse-led (computer-supported) telephone triage: mean time (minutes) taken by nurses in delivering triage

Practice ID ^a	mean	SD	Min.	Max.	<i>n</i>
14	9.86	4.36	1	32	454
15	7.09	3.66	1	42	472
16	6.22	5.65	1	57	333
17	8.83	3.17	1	21	393
18	5.91	3.58	1	47	352
19	6.12	2.08	1	16	501
20	6.60	4.22	2	46	337
21	6.01	2.72	1	17	410
22	5.00	2.70	1	28	296
23	3.62	1.87	1	13	300
24	4.15	1.72	1	11	279
25	6.39	3.21	1	37	476
26	5.82	4.65	1	60	381
27	7.32	3.32	1	35	414
28	8.20	2.82	2	17	137
Total	6.56	3.83	1	60	5535

Max., maximum; min., minimum.

^a Trial practice ID removed for publication.

Appendix 27 Alternative unit costs for general practitioner and nurse consultations

Alternative per-minute unit cost rates (with and without qualifications) for GPs and nurses were explored in the sensitivity analysis. In addition, we considered the GP cost of triage inclusive of direct care staff costs. The base-case unit costs (£3.40 GPs, £0.88 nurses) have previously been presented but are illustrated here for comparative purposes.

Resource-use unit	Unit cost (£)	Source
GP time (minutes) [base case]	3.40	Per-minute unit cost from Curtis <i>et al.</i> , ³⁸ table 10.8b, p. 183 (includes qualifications, excludes direct care staff costs)
GP time (minutes)	3.70	Per-minute unit cost from Curtis <i>et al.</i> , ³⁸ table 10.8b, p. 183 (includes qualifications and direct care staff costs)
GP time (minutes)	2.80	Per-minute unit cost from Curtis <i>et al.</i> , ³⁸ table 10.8b, p. 183 (excludes qualifications and direct care staff costs)
GP time (minutes)	3.10	Per-minute unit cost from Curtis <i>et al.</i> , ³⁸ table 10.8b, p. 183 (excludes qualifications and includes direct care staff costs)
Nurse time (minutes) [base case]	0.88	Calculated from the cost per hour of face-to-face contact time reported in Curtis <i>et al.</i> , ³⁸ table 10.6, p. 180 (includes qualifications)
Nurse time (minutes)	0.75	Calculated from the cost per hour of face-to-face contact time reported in Curtis <i>et al.</i> , ³⁸ table 10.6, p. 180 (excludes qualifications)
Nurse practitioner time (minutes) [base case]	1.52	Calculated from the cost per hour of client contact reported in Curtis <i>et al.</i> , ³⁸ table 10.7, p. 181 (includes qualifications)
Nurse practitioner time (minutes)	1.35	Calculated from the cost per hour of client contact reported in Curtis <i>et al.</i> , ³⁸ table 10.7, p. 181 (excludes qualifications)

Note

Where reference is made to direct care costs associated with GPs, these relate to costs for 0.43 full-time equivalent practice nurse per GP.

Appendix 28 Sensitivity analyses using alternative unit costs for out-of-hours contacts

Resource use unit	Unit cost	Source
Out-of-hours in-surgery consultation (doctor/nurse) ^a	£25.15	Chalder <i>et al.</i> ⁸²
Out-of-hours telephone (doctor/nurse) ^a	£16.39	
Out-of-hours home visit (doctor/nurse) ^a	£83.88	

^a Units costs updated to 2012 costs using index from Curtis *et al.*³⁸

Using this level of unit cost demands an assumption about how to distribute the out-of-hours contacts where the contact setting was not identified. These will be distributed as per frequency density of specified contacts.

Appendix 29 Total number of contacts by consultation type (from case note review)

Number of contacts over 28 days by consultation type

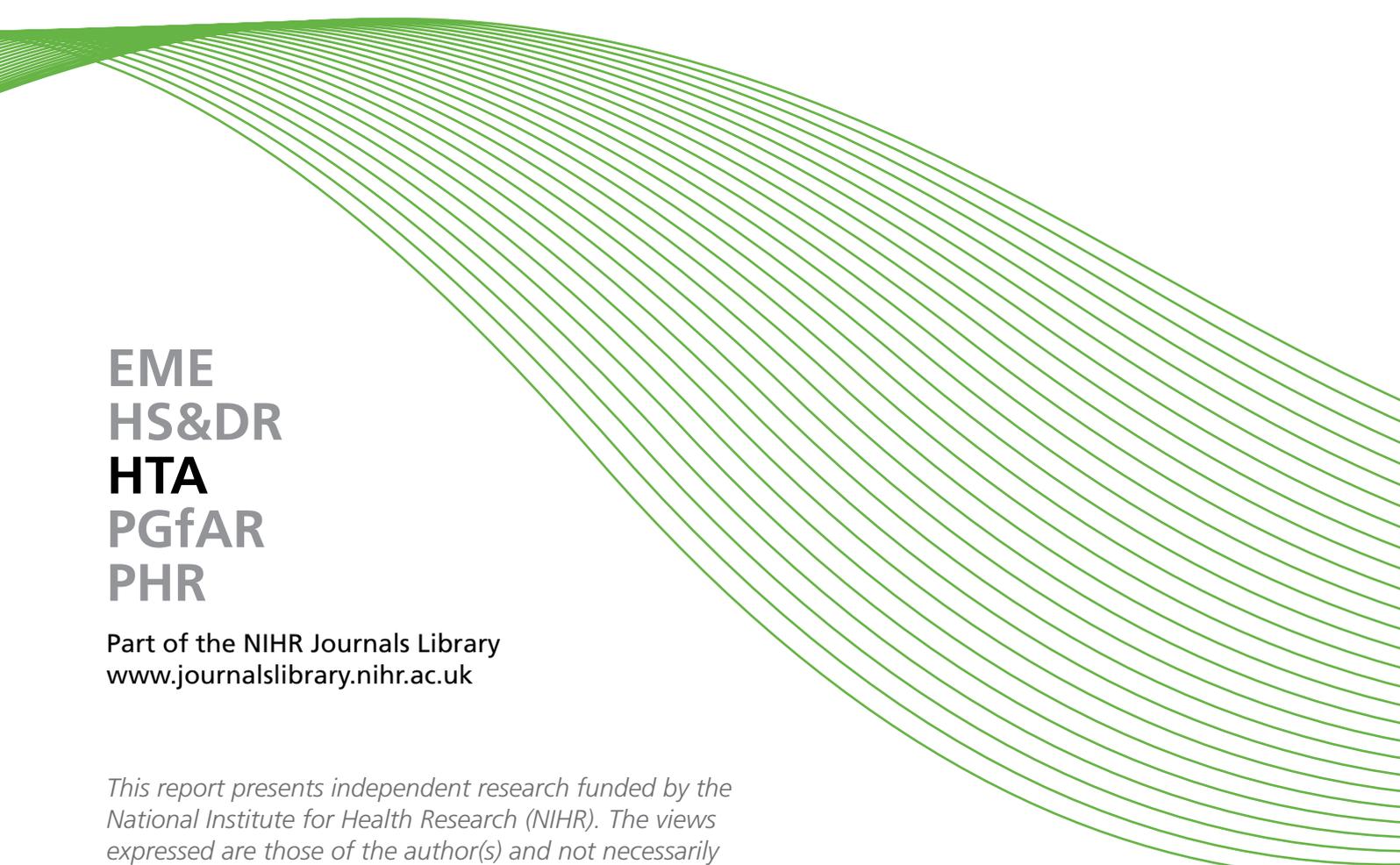
Contact type	<i>n</i>	% of total
Triage		
GP	4818	12.13
Nurse	4860	12.23
Total	9678	24.36
GP		
In surgery	19,375	48.76
Telephone	3190	8.03
Home	298	0.75
Unspecified	24	0.06
General unspecified	6	0.02
Nurse		
In surgery	4962	12.49
Telephone	766	1.93
Home	5	0.01
Unspecified	15	0.04
Total out of hours	772	1.94
Walk-in centre		
Doctor	18	0.05
Nurse	58	0.15
Unspecified	19	0.05
A&E	550	1.38
Total	39,736	100.00

Data from complete case analyses, *n* = 16,211 (from case note review).

Same-day contacts by contact category

Contact type	<i>n</i>	%
Triage (GP/nurse)	9678	43.07
GP all	11,202	49.85
Nurse all	1452	6.46
Out of hours (total)	37	0.16
Walk-in centre	5	0.02
A&E	97	0.43
Total	22,471	100

From case note review, complete case analysis (*n* = 16,211).

A decorative graphic consisting of numerous thin, parallel green lines that curve from the left side of the page towards the right, creating a sense of movement and depth.

**EME
HS&DR
HTA
PGfAR
PHR**

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

This report presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health

Published by the NIHR Journals Library