Ambulance Response Programme

Evaluation of Phase 1 and Phase 2

Final Report

Janette Turner

Richard Jacques

Annabel Crum

Joanne Coster

Tony Stone

Jon Nicholl

School of Health and Related Research (ScHARR)

University of Sheffield

j.turner@sheffield.ac.uk

July 2017

Contents

Executive Summary

Background

Time based standards have been used as a key performance measure for ambulance services both nationally and internationally, despite a lack of evidence that they actually lead to good clinical care. They have been effective in driving improvements and maintaining response times to the most critically ill and injured patients. However, efforts to meet these standards in the face of rising demand have led to a range of operational behaviours that may be inefficient, with the potential to distort the system away from a central focus on patient care and outcome.

The aim of the NHS England Ambulance Response programme (ARP) is to review ambulance response performance standards and explore strategies that can reduce operational inefficiencies and improve the quality of care for patients, their relatives, and carers. To support these objectives three initiatives have been developed to try and improve the clinical response:

1. Phase 1 - Exploration of changes to the triage of calls. The 999 call handling process has been modified to incorporate questions that immediately identify the most urgent calls. For all other calls additional time is allowed to support targeting the right resource to the right patient and reducing allocation of multiple resources in order to "stop the clock". This is Dispatch on Disposition (DoD).
2. Phase 2 - A review of call categories and development of a new set of categories that align clinical and resource allocation requirements and response options for 999 dispatch codes.
3. Phase 3 – A review of the current Ambulance Quality Indicators (AQI) and development of a revised set of indicators linked to the revised call categories, and with a focus on patient focussed reporting of performance indicators.

A formal trial of Phase 1 began in October 2015 and a trial of the Phase 2 revised call categories began in 3 services in April 2016. NHS England commissioned the School of Health and Related Research to conduct an independent evaluation of ARP and this report describes the findings of this evaluation.

Evaluation approach

We have used a mixed methods design comprising:

- Quantitative analysis of trends for a broad range of operational performance and resource utilisation indicators. For Phase 1 we conducted a controlled before and after time series study comparing changes in pilot sites to control sites over the time period October 2014 – March 2016. For Phase 2 we have conducted a descriptive analysis of trends for the same indicators in the 3 trial services.
- A description of weekly reporting of any adverse incidents or patient safety issues.
- A descriptive analysis of Nature of Call (NoC) capture for the most urgent and cardiac arrest calls.

- A descriptive commentary on the distribution of time to arrive at a final triage disposition and the effects of extended call triage times.
- A survey of staff to assess the impact of DoD on their work and elicit their views on the effectiveness of the operational changes.
- A consensus study to review ambulance quality indicators.
- Qualitative feedback from services on their experiences of implementing ARP.

Findings from Phase 1

- Of the 30 indicators measured, and excluding effects that were an inevitable consequence of the additional call assessment time, 23 showed an improvement, 5 no significant change and 2 a deterioration for green calls. All three areas of deterioration were associated with Green 2 (lower priority) incidents. There has been a modest effect on key performance measures with a 6.6% increase in 8 minute response time performance for Red 2 and a reduction in the time to first resource on scene for red 1 and Red 2.
- The most marked effect is on efficiency with a clear and measurable benefit in resource use across all 12 relevant indicators. We have estimated that nationally there is potential to gain 10243 whole resources per week which would then be available to respond to other incidents. Differences between resource allocation and arrival on scene indicate fewer resources are being assigned to calls and then cancelled before arrival.
- There was no change in the proportion of calls managed by hear and treat in the pilot sites. This may be due to case-mix and more serious calls during the winter months, or it may mean that other measures (for example the deployment of experienced clinical staff to manage these calls) are also needed before the full benefits can be achieved.
- Throughout the DoD pilot periods there have been no identified serious or adverse patient incidents associated with the implementation of DoD. Around 75% of the most urgent (Red 1) and cardiac arrest calls are identified and a resource sent before triage commences in most services although there are some between service differences. This may be a feature of the two different call assessment systems in use.
- Analysis of call completion times identified up to 240 seconds as the optimal triage time before a resource is dispatched to support intelligent dispatching but minimise clinical risk.
- ARP is generally viewed by Emergency Operations Centre (EOC) staff, clinical hub staff and frontline operational staff as a positive development that improves the ability to dispatch the right resource to the right patient and reduces stand downs for operational staff. .

Findings from Phase 2

The first iteration of the call category review (Phase 2.1) revealed one category was too large and not sufficiently discriminatory to support prioritisation of resource allocation. The categories were further reviewed and a second iteration (Phase 2.2) implemented. For Phase 2.2 the main findings are;

- The proportion of calls requiring an 8 minute response (Category 1) has, as intended, substantially reduced compared to the current model although there were differences between sites in the proportion (6-11%). The highest proportion of calls are assigned to category 2 (42 -56%). Variation in the proportion of calls assigned to each category is most likely a feature of the two call assessment systems and there is scope to further refine allocation. Overall, there is not a very clear and obvious distinction in conveyance rates between the Response and Transport categories with substantial conveyance rates in each. It may be that the current call assessment systems are not sufficiently discriminatory enough to allow decisions about which incidents may or may not need transport at the time of the call. Consideration should be given to whether this distinction is worth retaining.
- The percentile response times show that, for all 3 services, 90% of Category 1 calls received a response within 13 minutes and 95% within 16 minutes. For other categories the figures are more varied and were noticeably shorter across all categories in one service. The percentile response times for each category in each service provide a picture of what can be achieved operationally.
- Adjusted trends for a range of response intervals and resource utilisation indicators for phase 2.2. show a trend towards improving response performance in some services and no obvious degradation in performance in any service. Small but significant additional efficiency gains were also made in one service. Analysis of changes in key performance measures for the 999 population across the 3 ARP phases found that although, apart from Category 1, there is more flexibility in response timeframes, the service delivered to the whole 999 population only changes by a small amount. Further efficiency gains were made during phase 2.1, equivalent to making an additional 4500 resources per week available for response across England. Examination of response performance between urban and rural areas found a complex picture. There was evidence of a reduction in 95th response time and arrival at hospital times in rural areas in two services but also clear pressure in urban areas. More detailed investigation is needed to better understand the factors that influence performance in different geographical areas.
- There was no indication of reported patient safety issues during Phase 2 and a high proportion of the most urgent and cardiac arrest calls were identified by NoC in the two services using NHS Pathways. In the AMPDS service, on average, calls with cardiac arrest codes are identified in around 60 seconds.
- The staff surveys showed that the changes were broadly welcomed by EOC staff who considered further improvements had been made in allowing them to manage resources effectively. For operational staff there were mixed views so although the reduction in the proportion of calls requiring an 8 minute response was welcomed there were some concerns around whether the right types of calls have been allocated to the right categories.
- Overall, phase 2 appears to have brought a degree of stability to operational performance during a period of high demand and significant operational challenges. Evidence from the services also highlights that implementation of the revised call categories requires substantial operational changes such as fleet configuration. Potential benefits may not be immediately apparent as there will be a process of continual review and readjustment of a whole range of functions that underpin operational delivery before the right balance is found to maximise effectiveness. Clearer changes may become more apparent when this new model of service has been operating for a longer period.

- Alongside Phase 2.2 the Ambulance Quality Indicators have been reviewed and a revised set constructed aligned to the new call categories. These indicators, although still processes, provide a more focussed set of measures aligned to the intentions of ARP. They reflect more meaningful and transparent reporting of response time performance for all 999 calls, not just the most urgent. There remains a significant need to replace at least some operational process indicators with better and more patient focussed clinical outcome measures. There is real enthusiasm and innovative ideas on how this can be developed from both the ambulance sector and wider NHS but to be successful it needs to be supported by a co-ordinated, national, long term work programme that is properly resourced.

Summary

The evaluation of Phase 1 has provided strong evidence that the introduction of longer call assessment times produces clear benefits for operational efficiency and this is translated in to better response time performance for the most seriously ill patients. There is no evidence to suggest that patients, particularly those with time critical conditions, are disadvantaged by this initiative and the processes put in place to ensure early detection and dispatch to patients with cardiac arrest or potential cardiac arrest are working well. Overall, the evidence supports the Phase 1 changes as providing benefits to services that allows them to better manage the resources they have which becomes increasingly important as demand continues to rise.

Phase 2.2 is a much more substantial change and has only been in operational practice for a relatively short period of time. It has also not been possible to conduct a comparative analysis so we cannot say if the new model is “better” or “worse” than the existing model, only that it is different. The trends in phase 2.2 indicate a period of operational stability in the 3 services trial services during a period of high demand and when response time performance continued to deteriorate in services operating the current national model. The relative whole service performance stability emerging from the combined DoD and revised call category initiatives in the Phase 2 trial sites suggests that the more flexible approach to call assessment, resource dispatch and response intervals may be helping to reduce further deterioration in performance and maintain a consistent service.

Modern day ambulance services in England provide a response to a large and heterogeneous population of patients with health problems of variable clinical acuity ranging from life-threatening emergencies to the complex health problems of an ageing population and minor problems. The right solution is expected from a single service that is providing both emergency and urgent care. These problems are not unique to England and alternative delivery models are being explored and implemented in other countries. The call category trial has attempted to address this issue and potentially provides a realistic but clinically appropriate operational model that can be put in place to help deliver a service in an environment where demand will continue to increase but financial resources will be constrained. The alternative is to retain the current model but the evidence over the last few years has shown that provision of an 8 minute response to at least half of 999 callers is simply not achievable and creates inefficiencies that are not compatible with providing an equitable or clinically appropriate service. It is too early to say if the phase 2.2 model is the “right model”. However, the current evidence shows that it has produced benefits without compromising patient safety and supports provision of a consistent response despite substantial fluctuations in demand and so is a realistic alternative.

1. Introduction

1.1 Background

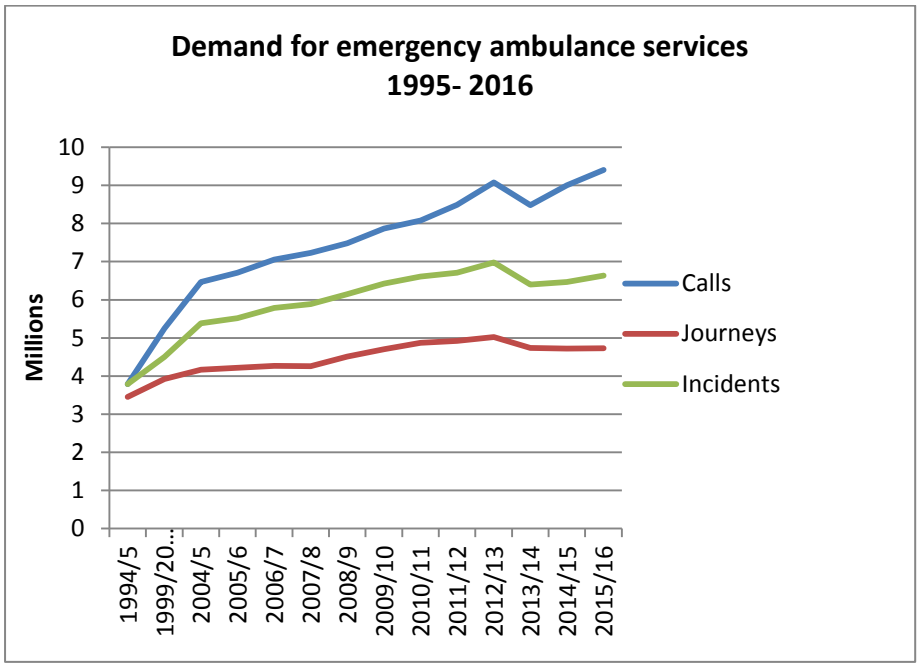
Demand for Emergency and Urgent care

Demand for emergency and urgent care increases year on year creating a source of considerable pressure within the NHS. In particular, problems in achieving expected time based performance targets for emergency departments and ambulance services attract a great deal of scrutiny and media attention. The reasons for increases in demand are a complex mix of changing demographic, health and social factors but historically the way urgent and emergency care is delivered has remained broadly the same.

Recognition of these issues and a need to re-think how services are delivered prompted the NHS England review of urgent and emergency care in 2013. This review set out five key elements considered to be core to the development of urgent and emergency care[1] - Help people to stay well and self care; right advice or treatment first time; urgent care closer to home; hospital care in centres with best expertise and development of networks to connect services. Following publication of the review a number of work programmes have been developed to support implementation of these key principles and the Ambulance Response Programme (ARP) is one of these initiatives.

Ambulance services in England have seen a remarkable increase in demand for their services (Figure 1).

Figure 1: Demand for emergency ambulance services



In the 20 years from 1994/5 to 2014/15, 999 calls have more than doubled from 4million to 9million per year. During the same period ambulance incidents (calls where the ambulance service responds either by dispatching a vehicle or by providing telephone advice and referral) have increased by 65% and ambulance journeys (where a patient is taken to hospital) by 40%. The more modest increase in journeys reflects the substantial changes that have taken place during this period and which have seen the ambulance service develop from a transport service to a significant provider of emergency and urgent care in its own right. As a consquence, a far greater proportion of 999 calls are now managed either at the time of the call (hear and treat) or by skilled ambulance clinicians at an incident scene or in patients homes (see and treat). There is a growing body of evidence supporting the development of the paramedic workforce to provide care closer to home[2] and this expansion of the ambulance service role was identified as a key contributor to achievement of the NHS England review principle of providing care outside acute hospitals.

Ambulance Performance Standards

Response time performance for emergency calls has been used as an indicator of ambulance service quality in England since 1974. These were revised in 1996 with variable standards set for different categories of calls to reflect differences in urgency and clinical need. The use of response time standards as a benchmark for performance is not unique to the UK. Internationally it has been the predominant measure for Emergency Medical Services (EMS) performance.

The rationale for using response time performance as a quality measure is based in research evidence on the relationship between time and patient outcome for very specific clinical conditions, predominately out of hospital cardiac arrest. For this group the relationship between time and outcome is well documented with an inverse relationship between delay in resuscitation, in particular defibrillation in ventricular fibrillation arrests, and survival[3]. Ambulance response time is a related factor and shorter response time is significantly associated with increased probability of survival[4]. The current "Red" calls requiring an 8 minute response do not only include cardiac arrest but also other conditions where timely intervention may affect outcome including severe chest pain, breathing difficulties, bleeding and unconsciousness. However the evidence on the relationship between ambulance service response time and outcome for this broader group of patients is limited. Several studies from the USA have explored the relationship between response time and survival and have found that there is no relationship between adjusted survival for patients with traumatic injury and response time,[5] or for the broader EMS population, with benefit only found for response times less than 4 minutes[6]. Similarly, a UK study of response time and outcomes in a cohort of 12,319 patients considered to have life-threatening emergencies found no difference in mortality rates with response time longer than 8 minutes after adjusting for a range of patient and service characteristics[7]. Another US study found no significant difference in survival for patients with life threatening conditions who received a response time that exceeded the 10.59 minute standard for that service when compared to patients who were responded to within standard[8]. None of the available evidence has demonstrated any positive relationship between shorter response times and a decrease in mortality across all emergency patients or those with life threatening conditions other than out of hospital cardiac arrest. The value of a response time standard as a measure of the impact and quality of ambulance service care is therefore questionable but, not just in the UK but internationally, the organisation and operational design of ambulance services have been dominated by the need to meet these standards. However, the of population patients who call emergency

ambulance services need help for increasingly complex and diverse health problems. At the same time the range of responses an ambulance service can make – for example by using enhanced telephone assessment and advanced paramedic practitioners as well as providing emergency transport - has also grown. This means that there is a pressing need to improve the process of matching the right response to clinical need at the time someone calls for help but this process is hampered by the need to also meet response time targets.

1.2 Ambulance Response Programme Initiatives

The current time-based ambulance response standards have been effective in driving improvements and maintaining response times to the most critically ill and injured patients. However, efforts to comply with these standards in the face of steadily rising demand have led to a range of operational behaviours that appear increasingly inefficient, and which have the potential to create a system which is unduly influenced by response-time targets. This means that some 999 calls receive multiple vehicle responses whilst others experience long delays and calls that could be managed through alternative pathways (e.g. "hear and treat") are not always recognised. As a consequence there is potential for unevenly distributed clinical risk across the 999 population.
The NHS England Ambulance Response programme (ARP) has been established to review ambulance response performance standards and explore strategies that can reduce operational inefficiencies whilst focussing on the clinical need to maintain a very rapid response to the most seriously ill patients, reduce overall clinical risk in the ambulance system and improve the quality of care (effectiveness, safety, experience) for patients, their relatives and carers.

The ARP comprises three broad initiatives

1. Exploration of the potential of allowing additional time for triage of calls other than those that are most urgent – Dispatch on Disposition (DoD). Currently up to 60 seconds is allowed to assess calls and establish the urgency of the problem and the type of response required at which point the "clock" starts for performance measurement regardless of whether triage is complete and triggering the allocation of a resource to maximise the chances of arriving within target time. It has been argued that a number of improvements could be obtained if there was additional time to triage calls including;
   - providing a more clinically appropriate response by targeting the right resource (clinician skills and vehicle type) to the right patient
   - reducing allocation of multiple resources whether suitable or not in order to "stop the clock" even though urgency has not been established (thus improving efficiency and keeping more resources available for response)
   - increasing "Hear and Treat" rates as the additional time would allow these calls to be identified and managed appropriately whilst at the same time reducing the number of potential hear and treat calls where a vehicle is allocated and sometimes arrives on scene before telephone assessment is complete.

2. To further support these initiatives a review and trial of the categorisation of calls and the associated response time standards. Currently, a relatively small proportion (~3%) of 999

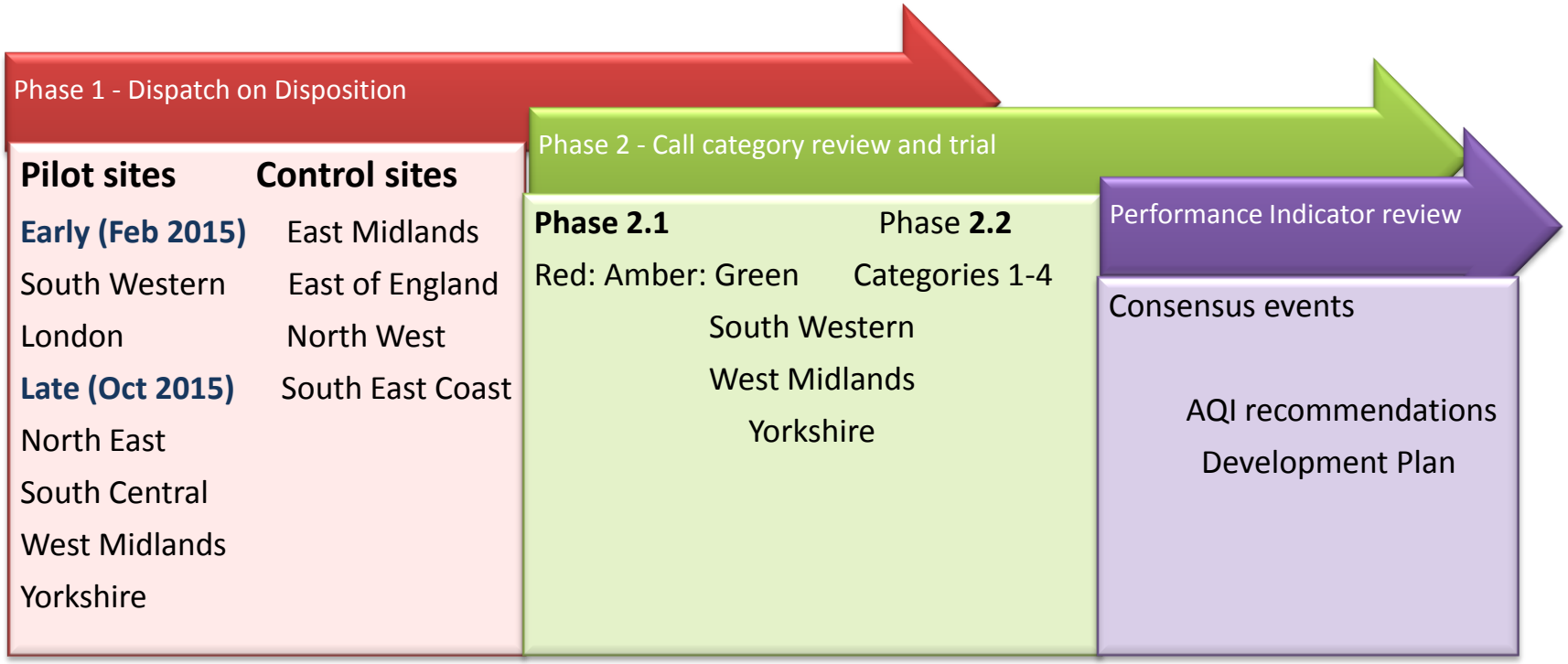
calls are classified as requiring the most urgent response (Red 1). An additional 40- 45% are classified as requiring a Red 2, 8 minute response, although many of these calls are subsequently considered to not need this level of emergency response. A requirement to respond to almost half of the 999 call volume within 8 minutes reinforces some of the operational practices driven by targets rather than clinical need. The current call category system fails to take account of both the increasingly diverse range of health problems, both in terms of type and urgency, that people request ambulance service help for. A revision of the call categories can start to address provision of responses that are a better fit between urgency, clinical need and the most appropriate response to those needs.

3. Much of ambulance service operations are driven by the current Ambulance Service Quality Indicators (AQI's) and, although these comprise a range of process and clinical indicators, there is an emphasis on time based measures that do not reflect clinical care delivered. Alongside the initiatives to improve both call assessment and dispatch and call categorisation so they are better aligned to the clinical needs of patients, the AQI's also require review so that they better represent any changes made as a consequence of the other ARP initiatives.

In January 2015 approval was received from the Secretary of State to pilot "dispatch on disposition" (DoD). The existing requirements for measuring ambulance service response times was that the "clock start" for the most life-threatening calls (Red 1) begins when a 999 call is connected to the ambulance service switchboard (call connect) and for all other calls the earliest of when a disposition is reached (that is, assessment is complete); a vehicle is assigned to respond to the call or 60 seconds after the call is answered. From February 2015 two services: London Ambulance Service (LAS) and South Western Ambulance Service NHS Foundation Trust (SWAST) were allowed up to an additional 120 seconds of triage time for all but the most urgent of 999 calls before clock start and resource allocation if a disposition had not been reached (180 seconds in total). An early evaluation of the effects of this additional time showed that there were trends towards achieving some of these objectives, with numbers of resources allocated decreasing and hear and treat rates increasing. Importantly no patient safety issues were reported during the initial pilot phase. Following this early pilot work a decision was made to continue DoD and expand the number of pilot areas as part of the broader Ambulance Response Programme. This would allow a fuller assessment of the effects of DoD in services using both NHS Pathways and AMPDS triage systems. This extended pilot began in October 2015 and is Phase 1 of ARP.

During Phase 1 of ARP the work to review the current call categories was also carried out and approval was agreed to conduct a trial of these revised call categories, initially in two ambulance services and then a third. This is Phase 2 of the programme which began in April 2016. Once these trials were under way the third element of the programme, the review of Ambulance Quality Indicators was initiated in October 2016. Figure 2 provides an overview of the ARP phases.
At the time the decision was made to expand the ARP to further test DoD and develop the revision of call categories and AQI's, NHS England commissioned the School of Health and Related Research at the University of Sheffield to conduct an independent evaluation of the programme. This report describes the methods and findings of this evaluation.

Figure 2: Ambulance Response Programme work programme



1.3 Overview of the evaluation design

The implementation of ARP has potential for impact on a range of operational and patient related factors. To address this we have assessed effects on relevant activities using a mixed methods design incorporating quantitative, qualitative and survey approaches. Phase 1 and Phase 2 have been examined separately with differences in evaluation design although there are some common features to both.

Phase 1 – Dispatch on Disposition

We have undertaken five separate analyses to assess the effects of DoD:

i) A controlled before and after time series study. This design allows trends and changes in operational processes to be measured over a suitable period before the introduction of DoD and during the pilot phase, and compare changes in pilot sites to control sites taking in to account other factors such as call volume fluctuations and seasonal effects.
ii) Weekly reporting of any adverse incidents or patient safety issues.
iii) A descriptive commentary on the distribution of time to arrive at a final disposition (NHS Pathways DX code or AMPDS determinant) and the effects of extended call triage times.
iv) A descriptive analysis of data reported on Nature of Call (NoC) capture for Red 1 and cardiac arrest calls.
v) A survey of staff to assess the impact of DoD on their work and elicit their views on the effectiveness of the operational change.

The detailed description of the implementation of DOD, methods and the results of these analyses are presented in section 2.

Phase 2 – Call category review and trial

For phase 2 a controlled comparison with other services was not possible as the revised call categories are substantially different from the existing categories and therefore no equivalent comparisons can be made. We have conducted a descriptive quantitative analysis of operational performance measures for all 3 services conducting the Phase 2 trial supplemented by additional information provided by individual trial services. The analyses undertaken were:

i) Trends in key operational performance indicators on response timings and resource use
ii) Trends in whole service measures of call cycle times and resource use before and after the introduction of phase 2
iii) Weekly reporting of any adverse incidents or patient safety issues
iv) An updated descriptive analysis of data reported on Nature of Call (NoC) capture for the most urgent category calls and cardiac arrest calls
v) Qualitative accounts of internal operational measures conducted by the 3 trial services
vi) An additional staff survey to assess their views on the call category changes and their effects on service operation.

The implementation of Phase 2 and results of these are analyses are described in section 3.

2. Evaluation of Phase 1 - Dispatch on Disposition

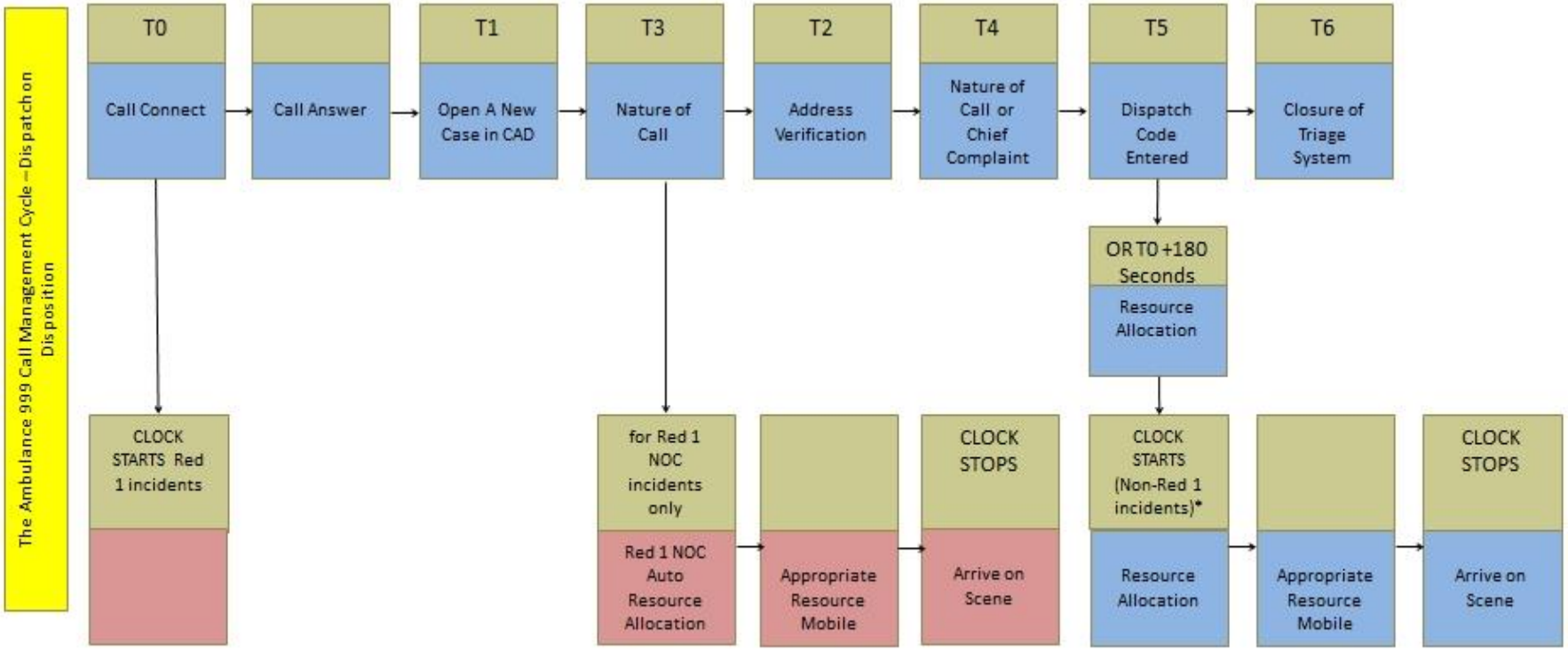
2.1 Implementation of the extended pilot phase

Following the positive indications from the initial trial of DoD in SWAST and LAS In October 2015 four additional services - North East Ambulance Service (NEAS); South Central Ambulance Service (SCAS); West Midlands Ambulance Service (WMAS) and Yorkshire Ambulance Service (YAS) joined SWAST as DoD pilot sites. LAS continued to use the additional time for call triage but, as a number of significant other changes occurred in this service during 2015, LAS has to be considered separately to the other pilot sites. Four other services in England – East of England Ambulance Service (EEAS); East Midlands Ambulance Service (EMAS); North West Ambulance Service (NWAS) and South East Coast Ambulance Service (SECAMB) – agreed to participate as control sites which allowed a comparative analysis of DoD effects.

A key principle of this operational change was that those patients with a potentially life-threatening time dependent health problem (predominantly out of hospital cardiac arrest) should, wherever possible, still get an immediate resource dispatch. To mitigate the risk of delaying help arriving in these cases 3 pre-triage filter questions on consciousness and breathing status were added to the call handling process. An immediate (Red 1) response was allocated to any call where patients were report as not breathing or unconscious with noisy breathing so there is no delay in sending help. These calls are unaffected by the additional triage time. In addition a list of broad conditions or events was created and the response to the first triage question on the main problem recorded if it related to one of these conditions. This process is referred to as Nature of Call (NoC). Figure 3 provides an overview of the different time components during the call triage and resource allocation process in one participating ambulance service (South Western): other services will differ slightly but the underlying principles are consistent.

The four new pilot sites were all allowed an additional 120 seconds call triage time (180 seconds in total) before clock start and resource allocation if a disposition had not been reached. As some calls take longer than 180 seconds to assess there remained a question about the optimum time that could be allowed for call assessment and allocation of a response balanced against the risk of delaying response for calls that may be potentially, but not immediately obviously, life-threatening. To address this, the ARP programme agreed that SWAST also test total call triage times of 240 seconds and subsequently 300 seconds before clock start and allocation of resource if no disposition had been reached.

Figure 3: Call triage and resource allocation process



• DoD Trial clock starts for Non-Red 1 incidents at the earliest of: Dispatch Code/Final Disposition (T5) OR 180 seconds after call connect OR an allocation. At 165 seconds after call connect the incident will be transferred to Dispatch for allocation for Non-Red 1NOC where there has been no T5.

2.2 Time series analysis

2.2.1 Analysis methods

From the inception of the extended pilot site study in October 2015 all 10 regional ambulance services in England returned weekly data for a comprehensive dataset that included:

- Activity – total numbers of calls, incidents and volumes within each response category (Red 1, Red 2 and Green 1-4)
- A range of mean and percentile (50, 90, 95%) time intervals by category – call to allocation, arrival on scene, clinician on scene
- Mean number of resources allocated by the ambulance service for each call category for both all resources (this includes external resources such as community first responders) and core resources, that is ambulance service only
- Proportion of calls managed by hear and treat
- Proportions of calls (excluding Red 1) with clock start times within 3 categories (Chief Complaint/DX code reached; 1st assign of a response; Triage time (60/180 seconds)
- Proportion of calls in NHS pathways sites with a code of DX 014 – that is, calls where a resource arrives on scene before triage is complete
- Additional information for factors that may have an impact on operational performance including calls passed from NHS 111, hours lost at hospital
- Services also supplied the same weekly data for the historical period October 2014 to pilot start providing a full year of operational data before DoD was implemented.

An interim analysis of the first 6 months of the extended DoD was reported to NHS England in May 2016. This showed that overall there were positive effects with a clear gain in efficiency through a reduction in resources per incident, improved response times to Red 2 calls, no detrimental effect on the response to Red 1 calls and that the change was well supported by staff. As a result DoD was extended to the remaining services in England in October 2016. All of these analyses are described in this report. The time series analysis has been updated and now presents the results of operational trends for all services including those who implemented DoD later in 2016 although these only reflect a relatively short period of operation.

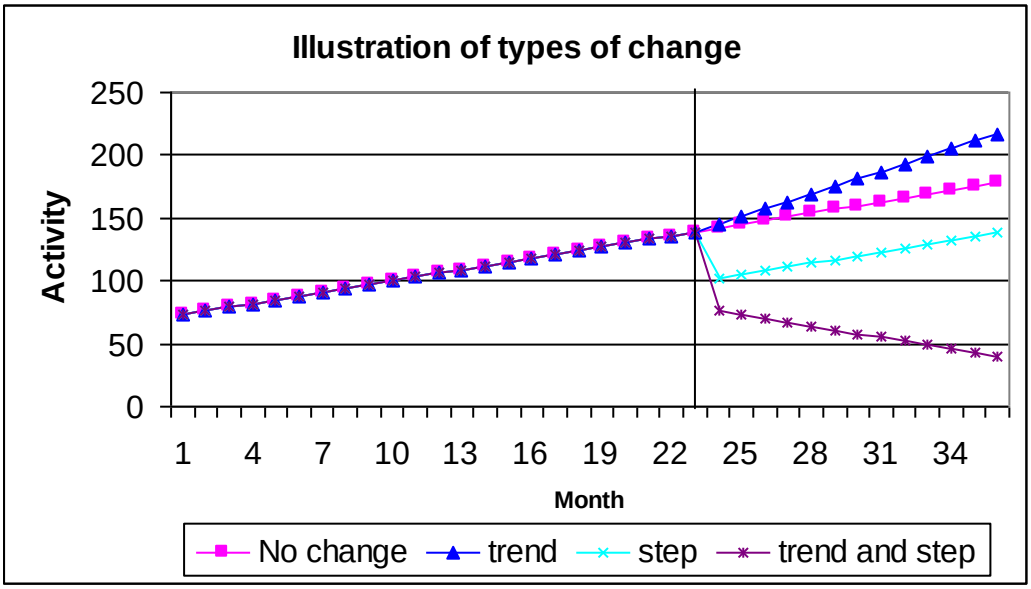
The times series analysis has been conducted in three steps;

1. We plotted weekly activity using the raw data for each service from 6thOctober 2014 to the start of Phase 2 in the three services conducting the call category trial (April and June 2016) and for the other 7 services up to the week beginning 2nd January 2017 so that change over time could be seen graphically. Four services were control sites and did not introduce DoD until October 2016.

2. We then fitted a number of time series regression models to each site to test for evidence of an immediate step change, a change in the overall trend, or both, following the introduction of DoD. A

step change is an abrupt rather than linear change in the mean level of the time series following the introduction of DoD. Figure 4 illustrates the three different change models relative to the no change model.

Figure 4: Illustration of different types of change in a time series



This allowed us to identify if DoD produces an immediate change in a variable being tested (such as Red 1 8 minute performance), a change in trend (indicated by the slope of the graph) or a combination of the two.

In order to account for other factors which may have an impact on the variables being measured the models consisted of terms to adjust for seasonality, an overall trend, the total number of emergency incidents, the total number of calls answered, and the number of hours lost at hospital. A forward step-wise procedure was used to identify if variables for the step change or change in trend should be included in the final model. Models were compared using likelihood ratio tests, Akaike information criterion (AIC) and the p-value of individual variables. Final models were compared with the plotted weekly activity to see if they made sense visually. When analysing the data from SWAST, we tested for a step change or change in trend at the four dates where changes to the time were made (week beginning 02/02/15 180 seconds; 05/10/15 240 seconds; 07/12/15 300 seconds; 29/02/16 240 seconds). The same forward step-wise procedure was used but with variables entered in chronological order.

3. We then tested for changes in the pilot sites compared to the control sites using time series regression to test for the impact of the changes. This comparative analysis was conducted for the period October 2014 to April 2016 when all 10 services could be included (that is before 3 services moved to Phase 2). These models also consisted of terms to adjust for seasonality, an overall trend, the difference between pilot and control sites, the total number of emergency incidents, the total number of calls answered, and the number of hours lost at hospital but also included a random intercept to allow for differences between ambulance trusts. We tested for differences in the pilot and control sites after the introduction of extra time by adding interaction terms between the pilot/control variable and variables for a step change and change in trend. A similar procedure to (2)

was used to select the final model. This approach allowed us to test if there were changes in pilot sites (either immediate as a step change, or over time as a change in trend) that were significantly different from the underlying trends identified in the control sites.

The pilot versus control site analyses were conducted twice, once using a model comprising all 4 control sites and the four pilot sites who began DoD in October 2015 (NEAS, SCAS, WMAS and YAS). The analyses were then repeated including SWAST and LAS in the model. Both of these services add complexity which may affect results as a) they have both been operating DoD for a longer time b) LAS faces some unique challenges and have implemented a range of other initiatives during the pilot phase that may impact on the items measured and c) the additional time allowed for call assessment has changed 4 times during the study period in SWAST. As both services began using the additional call assessment time in February 2015 there was also a shorter "before" period of historical data from which to estimate underlying trends in these 2 services before DoD.

There are potentially four categories of Green calls but as these are locally determined there are differences between services in terms of the number of categories used (some services only use 2 of the 4), types of calls included in each category and any performance standard associated with it. We have used the Green 2 category in this analysis as this category is used by all 10 services and comprises face to face responses requiring an ambulance resource to be dispatched.

The detailed results of the time series statistical analysis of DoD effects for a range of measures for both individual pilot site models and the pilot versus control sites models are provided in the supplementary file. Here we present a summary of the main findings of the comparative analysis of DoD pilot sites versus control sites.

2.2.2 Time series analysis results

National trend in 999 calls and incidents

For context on the operating environment during the trial period we have summed weekly totals of 999 calls and incidents for all 10 services for the whole study period (October 2014 – December 2016; Figure 5). Pressure on the wider emergency and urgent care system also has an impact on ambulance service operations, particularly delayed handovers and queuing at hospitals as this makes resources unavailable for emergency responses. Figure 6 shows the combined weekly hours lost at hospital for the 10 English regional ambulance services during the same period.

Figure 5: Weekly 999 call demand (England) October 2014-December 2016

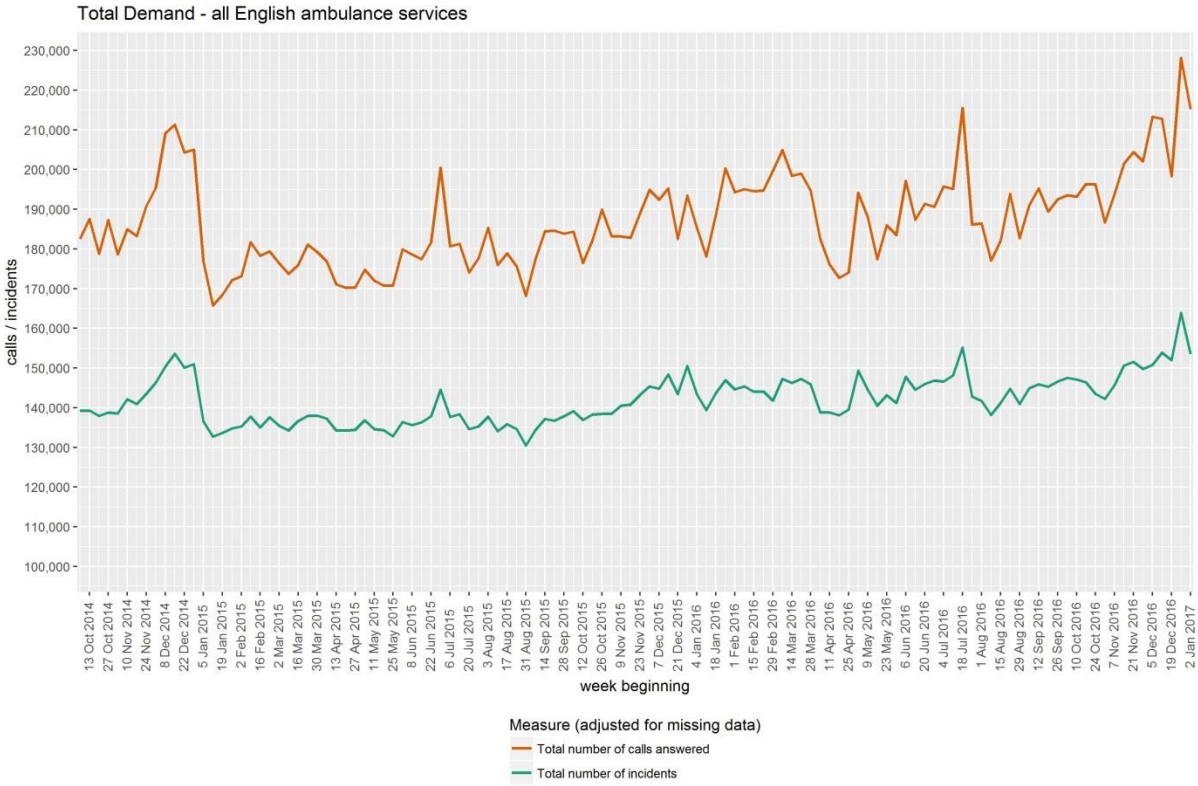
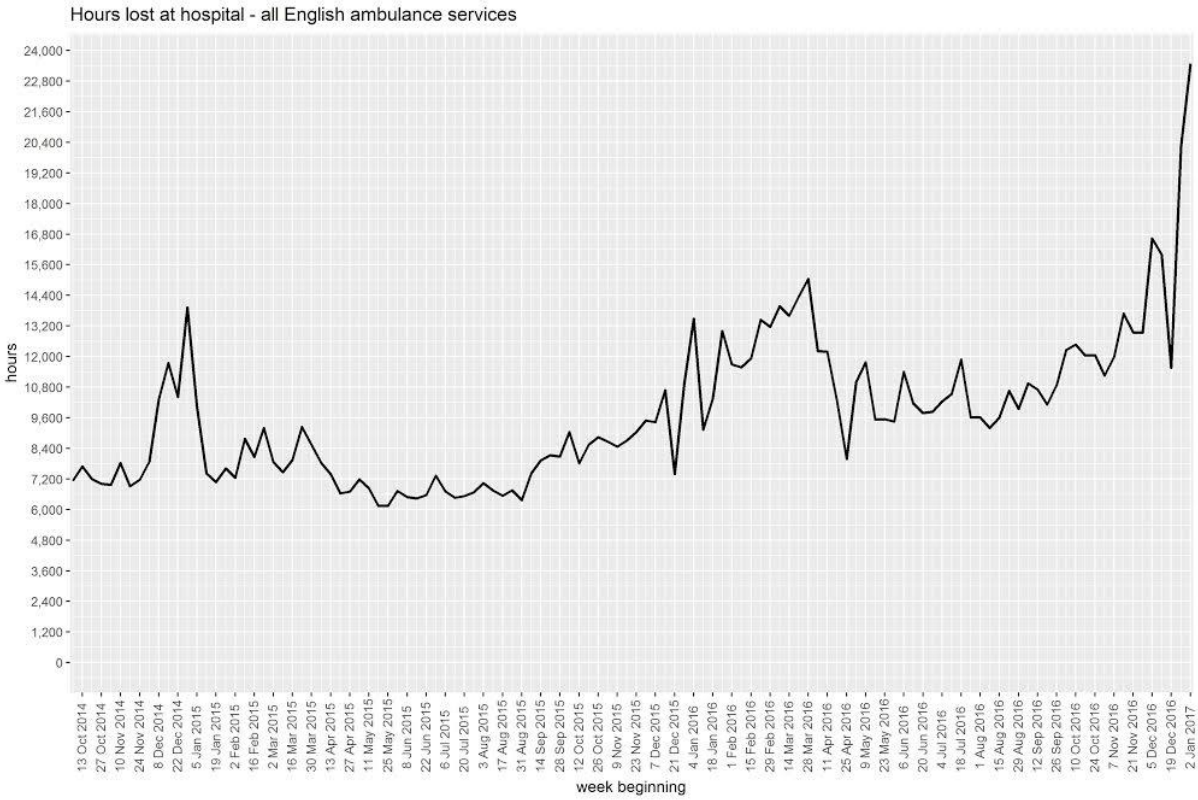


Figure 5 clearly shows the particularly difficult winter period in November and December 2014 followed by a relatively stable period with lower demand from January 2015 through to August 2015 (with a one week spike in July). However, from September 2015 there was a steady upward trend in demand and, with the exception of a few weeks during the summer of 2016, no downward seasonal trend towards levels seen for much of the earlier 2015 period. This upward trend has continued to a peak in demand during December 2016. Alongside this there is the same pattern for hours lost at hospital with an upward trend from August 2015 which has continued (Figure 6). This sustained high level and continually increasing demand creates significant pressure on services and potentially limits the impact that service changes such as DoD can deliver. Over a 12 month period January to December 2016 the 10 English ambulance services lost at total of 606,000 ambulance service hours waiting to hand over patients at hospitals.

Figure 6: Weekly lost hours at hospital (England) October 2014 – December 2016



Summary results of pilot versus control site analysis

Pilot site models

The results of the individual pilot site analyses are provided in detail in the supplementary file. Overall these illustrate individual variation between services in terms of the effects of implementing Dispatch on Disposition. For each variable we have examined, with the exception of the two clearly related measures of DX014 calls in NHS Pathways sites and clock start triggers, there has been no clear pattern of change in a single direction across all the pilot sites and instead a mixed picture of effect between services. The graphs for each measure show clearly the weekly volatility in operations and performance that all ambulance services face and are required to manage. However the individual site analyses provide a useful review for each service.

Pilot and control site comparisons

The pilot site versus control site comparisons provide a more robust measure of the effects of DoD as they take account of concurrent changes that may be occurring across all services. The main findings are summarised in Table 1. Results from both models (Model one excluding SWAST and LAS and Model 2 including SWAST and LAS) are presented. All results where values are given are statistically significant. Where no significant difference has been identified this is recorded as "no change". For step changes just the value is recorded. For slope changes the value is recorded as per week indicating that there was an upward or downward trend. A simpler summary of the main direction of change for key indicators is provided in Figure 7.

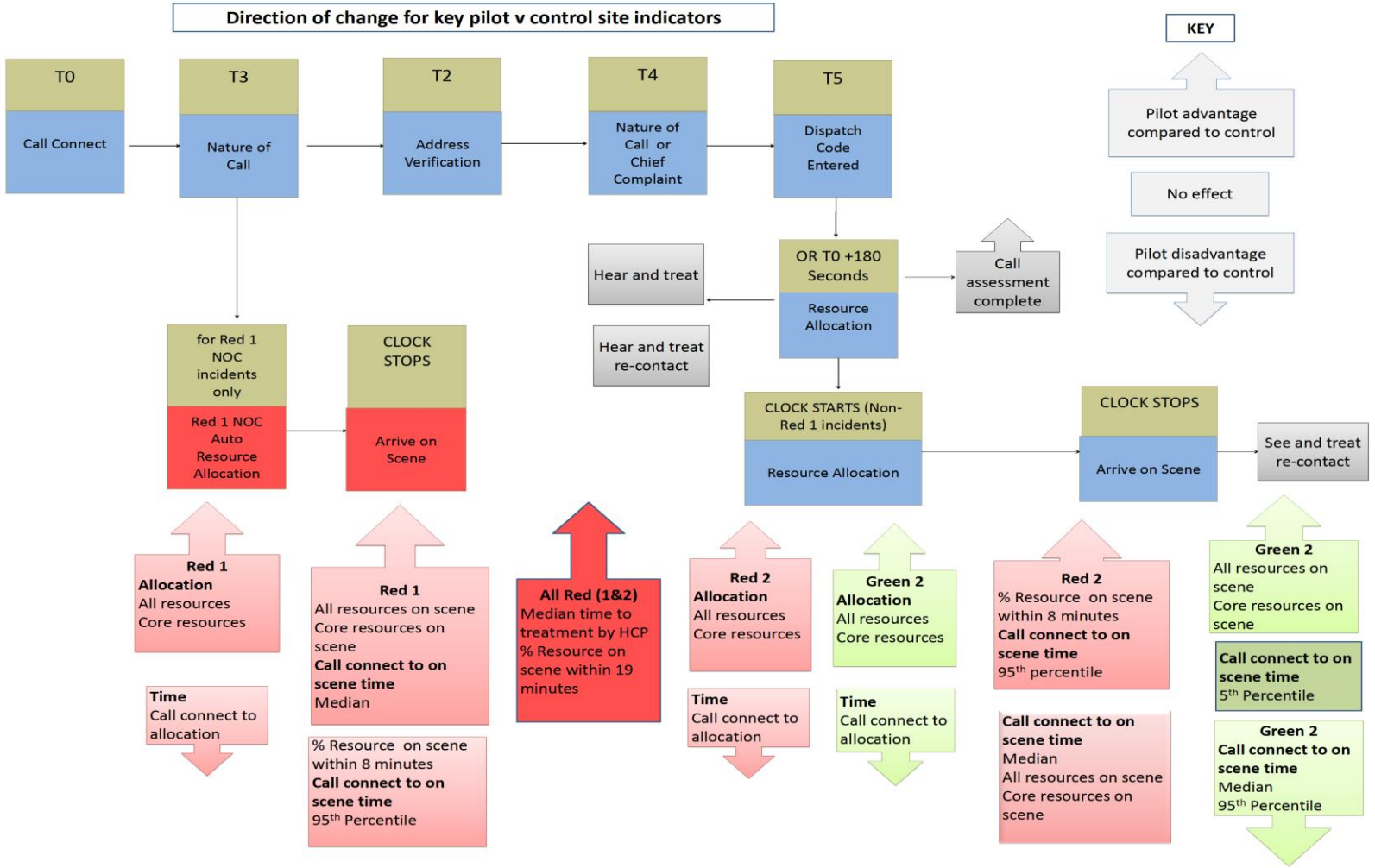
Table 1 – Changes in key process measures in pilot sites compared to control sites

Measure	Model 1 - Pilot vs Control site model (excluding LAS & SWAST) Change % (95% CI, P)	Model 2 - Pilot vs Control site model (including LAS & SWAST) Change % (95% CI, P)	Effect of DOD
Percent of incidents by category			No effect model 1. Model 2 proportion of Red calls decreased in pilot sites.
Red Incidents vs Green incidents	No change	↓-1.66 (95% CI: -2.95 to -0.36, P=0.013)	
Red 1 Incidents v Red 2 incidents	No change	No change	No change in the proportions of Red incidents
Percent of incidents with a resource on scene within 8 minutes			No change in Red 1 response time performance
Red 1	No change	No change	Both models identified an improvement of 6.6 - 5.8 percentage points in R2 response time performance
Red 2	↑6.6% (95% CI: 3.4 to 9.8, P<0.001)	↑5.8% (95% CI: 2.9 to 8.7, <0.001)	
Percent Red incidents where a conveying resource arrives within 19 minutes	↑2.2% (95% CI: 0.8 to 3.6, P=0.003)	No change	Model 1 A19 response time performance improvement of 2.2%
Median time to treatment for red incidents (seconds)	↓ -1.7 seconds per incident per week, (95% CI: -2.74 to -0.79, P<0.001)	↓-1.32 Seconds per incident per week, (95% CI: -2.35 to -0.30, P=0.012)	In both models median time was increasing each week in pilot and control sites but the increase was smaller in pilot sites
Percent of incidents that were resolved by Hear and Treat	No change	No change	
Average allocations - all resources			Model 1 & 2 indicate a weekly reduction in average resource allocations per Red 1 incident per week in the pilot sites.
Red 1 incidents	↓ -0.011 allocations per incident per week, (95% CI: -0.015 to -0.007, P<0.001)	↓-0.010 allocations per incident per week, (95% CI: -0.013 to -0.007, P<0.001)	
Red 2 incidents	Step ↓ -0.06 allocations per incident (95% CI: -0.08 to -0.03, P<0.001)	Step ↓-0.04 allocations per incident (95% CI: -0.07 to -0.01, P<0.001)	Model 1 & 2 indicate a weekly reduction in average resource allocations per Red 2 incident per week in the pilot sites.
Green 2 incidents	Step ↓-0.10 allocations per incident (95% CI: -0.14 to -0.07, P<0.001)	Step ↓-0.07 allocations per incident (95% CI: -0.11 to -0.03, P<0.001)	The same pattern was found for Red 1 & 2 calls also applied to Green 2 calls
Average allocations - core resources			Both models identified a step decrease in average allocations per incident for Red 1, Red 2 and Green 2 incidents
Red 1 incidents	↓-0.1 allocations per incident (95% CI: -0.15 to -0.05, P<0.001)	↓-0.11 allocations per incident (95% CI: -0.61 to -0.71, P<0.001)	

Measure	Model 1 - Pilot vs Control site model (excluding LAS & SWAST) Change % (95% CI, P)	Model 2 - Pilot vs Control site model (including LAS & SWAST) Change % (95% CI, P)	Effect of DOD
Red 2 incidents	↓-0.06 allocations per incident (95% CI: -0.09 to -0.03, P<0.001)	↓-0.05 allocations per incident (95% CI: -0.08 to -0.02, P<0.001)	
Green 2 incidents	↓-0.12 allocations per incident (95% CI: -0.17 to -0.0, P<0.001)	Step ↓--0.09 (95% CI: -0.14 to -0.03, P=0.002)	
Average responses on scene - all resources			For Red 1 incidents both models identified decreasing responses on scene per incident per week. For Red 2 incidents there was no change in average responses on scene. Both models identified a decrease per incident for Green 2 incidents
Red 1 incidents	↓ (-0.004 per week, 95% CI: -0.007 to -0.0015, P=0.006)	↓-0.004 per week, (95% CI: -0.006 to -0.0001, P=0.004)	
Red 2 incidents	No Change	No change	
Green 2 incidents	↓ -0.03 responses on scene per incident (95% CI: -0.05 to -0.02, P<0.001)	↓-0.02 responses on scene per incident (95% CI: -0.04 to -0.01, P=0.013)	
Average responses on scene - core resources			Both models identified a reduction in average responses on scene for Red 1. There was no difference for Red 2. Both models identified a reduction for Green 2 incidents
Red 1 incidents	↓-0.06 responses on scene per incident (95% CI:-0.01 to -0.02, P=0.003)	↓-0.004, responses on scene per incident per week (95% CI: -0.006 to -0.0002, P<0.001)	
Red 2 incidents	No change	No change	
Green 2 incidents	↓-0.02 responses on scene per incident (95% CI: -0.04 to -0.01, P=0.004)	↓-0.02 responses on scene per incident (95% CI: -0.03 to -0.002, P=0.024)	
Median time from call connect to resource allocation (seconds)			Overall there was an increase in median time to resource allocation across all categories with the biggest change for Green 2 incidents. For Red 2 model 1 showed an initial increase followed by a weekly decrease compared to a weekly increase in control sites.
Red 1 incidents	↑ 11.7 (95% CI: 6.0 to 17.4, P=0.001)	↑11 (95% CI: 4.9 to 17.1, P<0.001)	
Red 2 incidents	Step:↑ 59.6 (95% CI: 48.2 to 71.1, P<0.001) Slope: ↓-1.2 seconds per week (95% CI: -2.4 to -0.10, P=0.031)	↑53.7 (95% CI: 43.2 to 64.3, P<0.001)	
Green 2 incidents	↑ 287.7 (95% CI: 176.0 to 399.3, P<0.001)	↑233 (95% CI: 128.5to 337.5, P<0.001)	
Median time from call connect to resource on scene (seconds)			Both models no reduction in median time to resource on scene for Red 1 or Red 2. Both models identified an increase in time to resource on scene for Green 2 incidents
Red 1 incidents	No Change	No Change	
Red 2 incidents	No Change	No Change	

Measure	Model 1 - Pilot vs Control site model (excluding LAS & SWAST) Change % (95% CI, P)	Model 2 - Pilot vs Control site model (including LAS & SWAST) Change % (95% CI, P)	Effect of DOD
Green 2 incidents	↑ 149.4 seconds (95% CI: 7.5 to 291.3, P=0.040)	↑139.1 seconds (95% CI: 9.9 to 268.2, P=0.035)	
95th percentile time from call connect to resource on scene (seconds)			Both models indicated a reduction in the 95th percentile time from call connect to resource on scene for Red 1 and Red 2 incidents There was no change in 95th percentile times for Green 2 incidents
Red 1 incidents	↓-9.45 seconds per week, (95% CI: -13.0 to -5.9, P<0.001)	↓-9.30 seconds per week, (95% CI: -12.29 to -6.33, P<0.001)	
Red 2 incidents	Step↓-166.6 seconds, (95% CI: -273.2 to -60, P=0.002) Slope↓-8.3 seconds per week (95% CI: -14.9 to -1.6, P=0.015)	↓-12.7 seconds per week, (95% CI: -17.5 to -8.0, P<0.001)	
Green 2 incidents	No change	No change	
Percentage of clock start trigger for Red 2 calls (CC/Initial DX)	↑ 18.1% (95% CI: 13.7 to 22.4, P<0.001)	Not available	There was an increase in the proportion of calls with a clock start at completion of assessment
Re-contact rates			There was no change detected in rcontact rates
Hear and treat	No change	No change	
See and treat	No change	No change	

Figure 7: Direction of change for key indicators



2.2.3. Summary of time series analysis findings

The main changes in the pilot sites when compared to the control sites during the evaluation period were;

- There was no change in the proportion of 999 calls allocated to Red and Green categories or between Red 1 and Red 2 categories in the basic model (model 1). Model 2 did show a change in the proportion of Red calls increasing each week in both pilot and control sites but the rate of change was smaller in the pilot sites. There was no change detected in the proportions of Red 1 and Red 2 calls. It can be argued that DoD could produce a shift from Red to Green as the additional time for call assessment may allow more detailed questioning to identify lower acuity calls. However, there may be case-mix differences with more high acuity illness occurring. This would also be expected in the control sites and the trend of increasing Red 2 calls and decreasing Green calls was evident in the time series graphs for both pilot and control sites. No or modest shifts in the proportions of calls assigned to different call categories may be as much a feature of case mix differences as any effects of DoD. If acuity is higher DoD is unlikely to generate more Green calls.
- There was a statistically significant increase in the proportion of calls with a clock start at chief complaint or initial DX code decision indicating that more calls are receiving a completed assessment within the allowed triage time.
- DoD has had no significant impact on the key response time targets for Red 1 calls, but there is a significant increase in the proportion of Red 2 calls responded to within 8 minutes in the pilot sites compared to the control sites in both models (6.6% and 5.8% respectively). DoD sites have an additional 120 seconds in which to “start the clock”, but this does not in itself translate to a response time performance gain because pilot sites are waiting until assessment is complete (or 180 seconds has lapsed) to allocate resources, rather than dispatching on obtaining an address or pre-alerting vehicles to start moving before clock start as can currently happen in the control sites. This is therefore a real gain in performance in terms of the measured standard of “clock start to response on scene” and the additional triage time affords no advantage as a component of this measure. Model 1 showed an increase of 2.2% in Red 19 minute performance but no effect was seen in model 2. There was a reduction in median time to treatment (that is by a healthcare professional) for Red calls in both models although the difference was only just under 2 seconds per incident per week. Nevertheless this demonstrates an advantage in the pilot sites and in an environment where this measure was deteriorating at a greater rate in the control sites.
- There was no change in the proportion of calls resolved by hear and treat in the pilot sites compared to the control sites or any consistent pattern of change in the individual pilot sites, although the graphs in the supplementary file show an increase in 2 sites. An increase in hear and treat rate is seen as a potential benefit of DoD as the additional call assessment time should allow better identification of suitable calls. The lack of impact may be simply that DoD alone cannot produce this effect and that other strategies, for example increasing the number and availability of clinical staff to provide enhanced clinical assessment, is needed. It may also be the case that a benefit has not been realised because of changes in case-mix. A safe service will only increase the hear and treat rate when the call case-mix

allows this and calls are received where this type of management is appropriate. If the case-mix is of higher acuity then the opportunity to provide hear and treat for more cases is diminished. Given the indications within the data of a shift in call proportions towards higher acuity red calls that will not be suitable for hear and treat, it is possible that this is one explanation for the lack of effect.

- There was a clear and consistent pattern of a reduction in average allocations of all resources and core resources per incident, and for calls with a response arriving on scene, across all call categories in the pilot sites compared to the control sites. The gains appear small, for example, using the model 1 results, a reduction of 0.1 in allocation of core resources per incident for Red 1 calls. However, when considered in the context of call volumes the benefit becomes more obvious. For allocation of core resources this equates to a gain of 100 resources (a vehicle available for response) per 1000 incidents for Red 1, 60 per 1000 incidents for Red 2 and 120 per 1000 incidents for Green 2 incidents. To estimate the potential impact nationally we have used the weekly data returns to calculate the average number of incidents for each call category per week and multiplied this by the relevant reduction in allocation of core resources (that is, resources owned and financed by the ambulance service so excluding, for example, volunteer first responders). The results are given in Table 2 and show that implementation of DoD nationally could potentially produce an additional 10243 resources which would be available at the time of a 999 call each week. This does not mean that this number of actual responses will be made as this will depend on how each one of those available resources is used and for how long – one whole resource available is not equivalent to one unit hour – but this measure does provide a strong indication that DoD has produced some of the intended efficiencies as the cumulative effect of reductions in the average allocations per incident releases resources that are then available for dispatch. This in turn creates potential to reduce waiting times for patients, particularly when calls waiting ("stacking") are greater than resources available to respond. It also increases the likelihood that the right resource is available. This has been achieved over a period when demand and lost hours at hospital have been rising so resources are more intensively utilised.

Table 2: Estimated weekly gains from reduced core resource allocation in England

Call category	Average weekly incidents	Change in pilot site allocation of core resources	Estimated weekly whole resource available for response
Red 1	3570	-0.1 per incident	357
Red 2	64995	-0.06 per incident	3900
Green 2	49883	-0.12 per incident	5986
Total	118448		10243

- There was also a reduction in the number of responses arriving on scene in Red 1 and Green 2 calls but not Red 2 and the reduction was substantially smaller than that found for resource allocation. This indicates that, as intended, there has been a reduction in the allocation of multiple resources to incidents which are subsequently stopped before arriving on scene.

- There was an increase in median time from call connect to resource allocation across all call categories. This would be less expected for Red 1 calls as these still generate an immediate dispatch, however the additional time was small (11 seconds) and may reflect the use of the pre-triage nature of call questioning. For Red 2 calls there was an immediate increase of 59.6 seconds and for Green 2 an increase of 288 seconds. This would be expected as a feature of the additional call triage time in that resources are not allocated until triage is complete or the maximum time allowed is reached. For Red 2 calls the increase indicates that having an additional 120 seconds triage time does not lead to an equivalent delay in resource allocation as these calls are more likely to be identified early in the triage process. The model did show that after the initial step increase of 59.6 seconds median allocation time decreased by 1.12 seconds per week. This may be the result of changes back to original operational practice or could indicate that as DoD becomes embedded the process becomes increasingly efficient. The increase in median allocation time for Green 2 calls was more substantial (4.8 minutes). This may in part reflect the additional questioning time needed to reach urgent rather than emergency dispositions. It may also be the case that as dispatchers wait for call assessment to be completed before allocating a resource priority will be given to Red 2 calls, resulting in a wait for allocation to Green 2 calls.
- Median time from call connect to resource on scene was unchanged for Red 1 and Red 2 incidents. For both Red 1 and 2 incidents the 95th percentile time from call connect to resource on scene showed a trend of weekly reduction compared to the control sites and this was most marked for Red 2 incidents in model 1 where the reduction is almost 3 minutes. Overall this indicates that, for the population of Red 999 calls, DoD and the additional time allowed for better call assessment has not had a detrimental effect on the response interval and hence the timeliness of patients receiving help, and is beginning to produce some small gains. Median time from call connect to resource on scene was increased more substantially for Green 2 incidents (149 seconds) although this is still within the additional time allowed for call triage and there was no difference between pilot and control sites in the 95th percentile time for Green 2 incidents.
- There was no difference between the pilot and control sites in re-contact rates for hear and treat or see and treat calls.

2.3. Time to complete call assessment

The DoD pilot has explored a number of options around how much additional time is needed to complete call triage before clock start and allocation of a resource. The starting point for the early pilot sites (SWAST & LAS) was 120 seconds in addition to the 60 seconds already allowed before clock start (180 seconds total). For the extended pilots, the additional 4 services also utilised up to 180 seconds for call triage. There remained a question about the optimum time allowed for call triage that balances the need to have sufficient time to establish clinical need and appropriate response, and at the same time limits the clinical risk associated with delaying response to calls that may be ultimately be life-threatening even though this is not immediately apparent. To address this, SWAST trialled two additional extensions of triage time (240 seconds and 300 seconds) during the extended trial period (October – December 2015).

Three services (SWAST – NHS Pathways; LAS and YAS – AMPDS) have provided data on the time taken from receiving calls (T0) to completion of call and arrival at a final disposition (NHS Pathways DX code or AMPDS determinant). Table 3 provides a comparison of the cumulative distribution of proportions of calls at T5 in one minute intervals up to 5 minutes (300seconds) for these 3 services.

Table 3 – Cumulative % distribution of proportions of calls T0- T5

Time period (seconds	SWAST R1 (%)	SWAST R2 (%)	YAS Red (%)	LAS Red (%)	YAS Green (%)
0-60	19.9	5.2	9.7	2.1	8.4
61-120	65.8	31.9	57.2	42.2	48.7
121-180	86.3	55.8	83.6	76.9	80.5
181-240	93.8	71.6	93	90.8	91.9
241-300	96.9	81.2	96.3	95.8	95.7
>300 (% calls in this category)	3.1	18.8	3.7	4.2	4.3

There are some differences between services which most likely reflect the differences in the two call triage systems. AMPDS in general is a more linear system and tends towards a faster process in arriving at a final determinant. Although there are differences between the two AMPDS sites both have reached a final determinant in 90% or more of Red calls by 240 seconds. Between 180 seconds and 240 seconds the marginal gain is 9.4 – 13.9% and from 240 seconds to 300 seconds there is a much smaller gain of 3.3 – 5%. There is little difference for Green calls with over 90% completing triage by 240 seconds. In both AMPDS services the proportion of calls taking more than 300 seconds to reach a final determinant was similar for Red and Green calls at around 4%.

In contrast, the pattern of distribution was different in SWAST (using NHS Pathways) for some call categories. A higher proportion of calls were completed in 0-120 seconds although we cannot distinguish Red 1 & 2 differences. Proportions of calls with completed triage were similar to the AMPDS sites for Red 1 calls at 180, 240 and 300 seconds with a marginal gain of 7.5% between 180 seconds and 240 seconds and 3.1% between 240 seconds and 300 seconds. This is not surprising as Red 1 calls are those where a rapid resource allocation is most likely to occur. For Red 2 calls a smaller proportion of calls had reached a final DX code at 240 seconds but it is difficult to compare with the AMPDS site data as these were for Red 1 and Red 2 combined. The marginal gains at 240 seconds and 300 seconds were greater (15.8% and 9.6% respectively) and there is a substantially greater proportion of calls in Red 2 (18.8%) where a final DX code had not been reached by 300 seconds (Figure 8). NHS Pathways has a more complex architecture with bigger scope to question a caller through linked decision trees and this may be one reason why it potentially takes longer to reach a DX code for Red 2 calls. Complete data for combined green calls was not available but Figure 9 shows that most gains for Green calls have been made by 240 seconds – 300 seconds. The longer tail is most likely indicative of the more complex questioning of urgent rather than emergency calls within NHS Pathways in order to establish whether alternative dispositions to ambulance dispatch can be reached.

Figure 8: SWAST - Time to establish T5 (final DX code) from T1/T0.

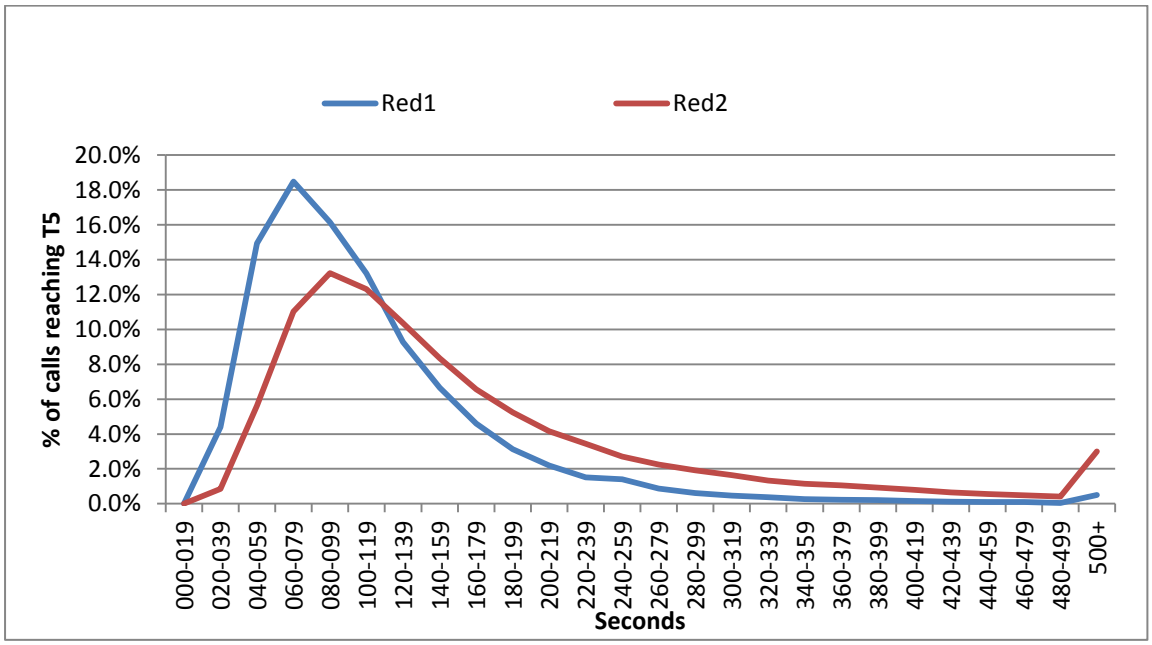
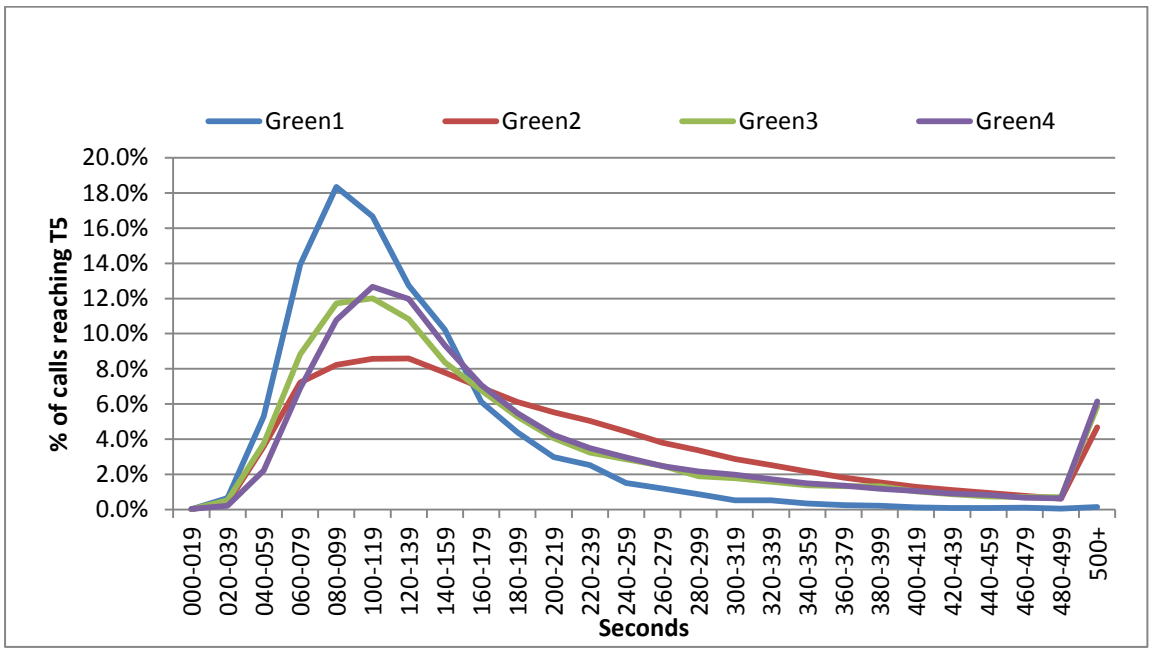


Figure 9: SWAST - Time to establish T5 (final DX code) from T1/T0



In summary, for Red calls there seem to be only small marginal gains from extending call triage time beyond 240 seconds regardless of the call triage system used. For green calls there are also only small marginal gains for triage times over 240 seconds where AMPDS is the triage system but there may be a larger gain for NHS Pathways sites. However, there does not appear to be a substantial step between 240 seconds and 300 seconds in the NHS Pathways managed distribution curve and there would remain a flat tail past this point (most likely reflecting very complex calls). Assuming call

volumes are low in this extended triage time tail call triage times over 240 seconds may have minimal operational impact if resources are sent but manages the potential clinical risk that within these complex calls there may be a small number of patients who may be disadvantaged by a delayed response.

We have also examined the results of the individual SWAST model statistical analysis described in section 2 to see if there were any statistically significant changes in key measures associated with the increases of call assessment time to 240 and 300 seconds. The majority of changes identified in SWAST occurred after 10th February when DoD was introduced and for most measures effects were unchanged after increasing call assessment times.

Changes that were detected after the introduction of 240 seconds on 5th October were 1) small weekly increases in the proportion of red calls; 2) small weekly reductions in core resources arriving on scene for Red 1 calls; 3) weekly increases in the proportion of clock start trigger CC/Initial DXcode.

Changes that were detected after the introduction of 300 seconds on 14th December were 1) a decrease in the percentage of Red incidents responded to within 19 minutes; 2)an increase (5 seconds per week) in median time to treatment for Red calls; 3) a decrease in time to resource allocation for Red 1 incidents; 4) a weekly increase in the 95thpercentile time to resource on scene for Red 2 incidents; 5) a weekly increase in median and 95th percentile time to resource on scene for Green 2 calls.

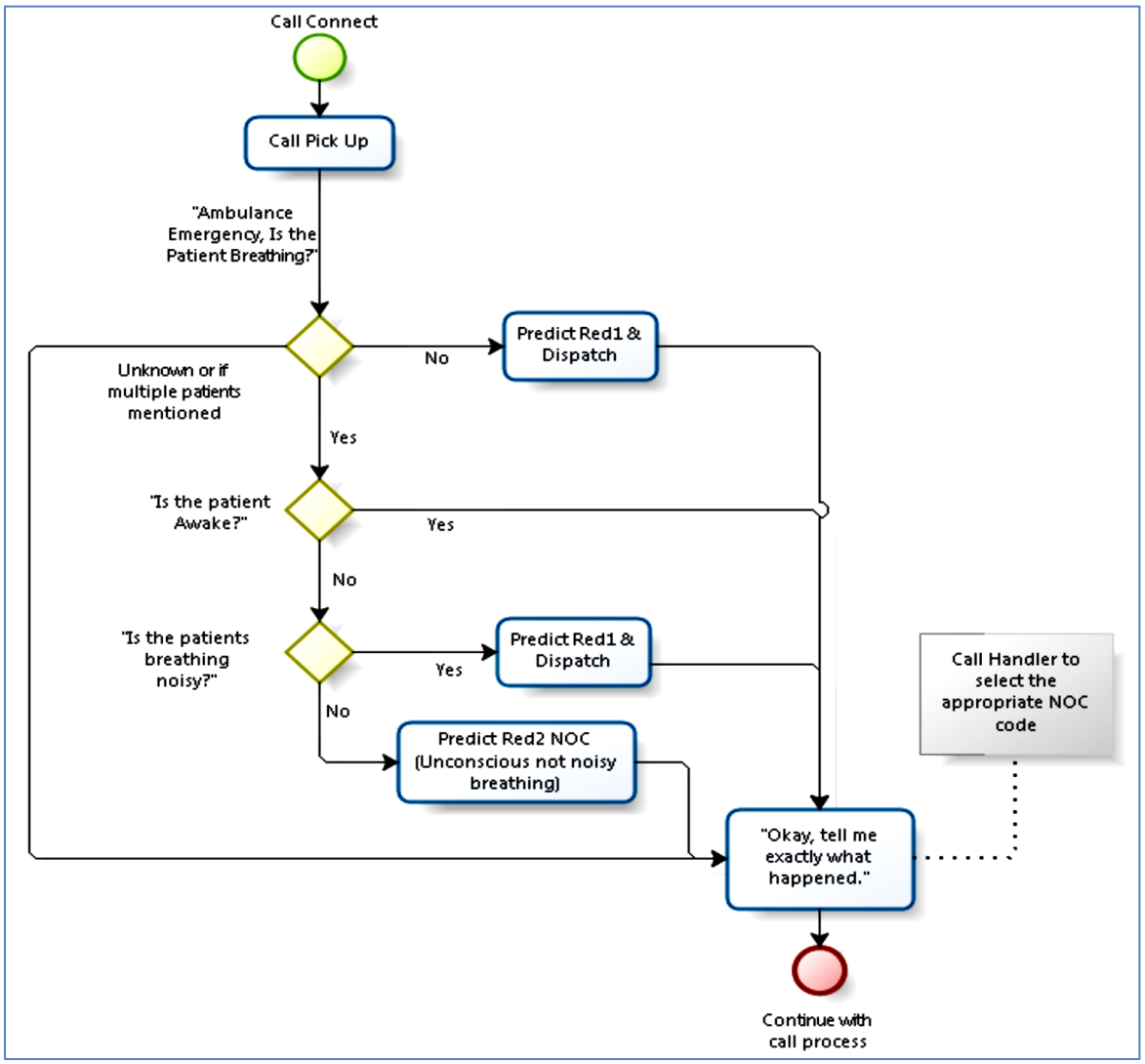
The change to 240 seconds potentially provided small benefits in additional reduction in resource use and an increase in calls with completed assessment. The change to 300 seconds suggests some of the benefits achieved with 180 and 240 seconds start to be reversed, particularly in relation to response times. Most of the benefits of DoD are realised with 180 seconds for call assessment, however there may be a small additional benefit from increasing this to 240 seconds with no loss of effect in the key measures tested. There appears to be no advantage to a call assessment time of 300 seconds and benefits may be reduced. The SWAST model supports the earlier conclusion that optimum call assessment time is up to 240 seconds, but not beyond this.

2.4. Pre-triage questions and Nature of Call identification of emergency calls and cardiac arrest

Identification of calls for potentially life-threatening conditions

A component of the Dispatch on Disposition initiative and subsequent call category trial has been the introduction of 3 pre-triage sieve (PTS) questions and nature of call identification using a pre-defined list of problems (collectively NoC) at the beginning of the call management process. The purpose of the NoC is to facilitate the early identification of patients with a potentially life threatening emergency in order that immediate dispatch of an appropriate resource can take place at the earliest possible point in the call cycle. These immediately life threatening calls are a subset of 999 emergencies that may benefit from early dispatch despite full details of the emergency not being available at that point. It is a necessary “safety net” to minimise the risk of delaying sending help with the introduction of additional call assessment time . Figure 10 shows a simple overview of the revised call process.

Figure 10 Revised call process incorporating pre-triage questions and Nature of Call



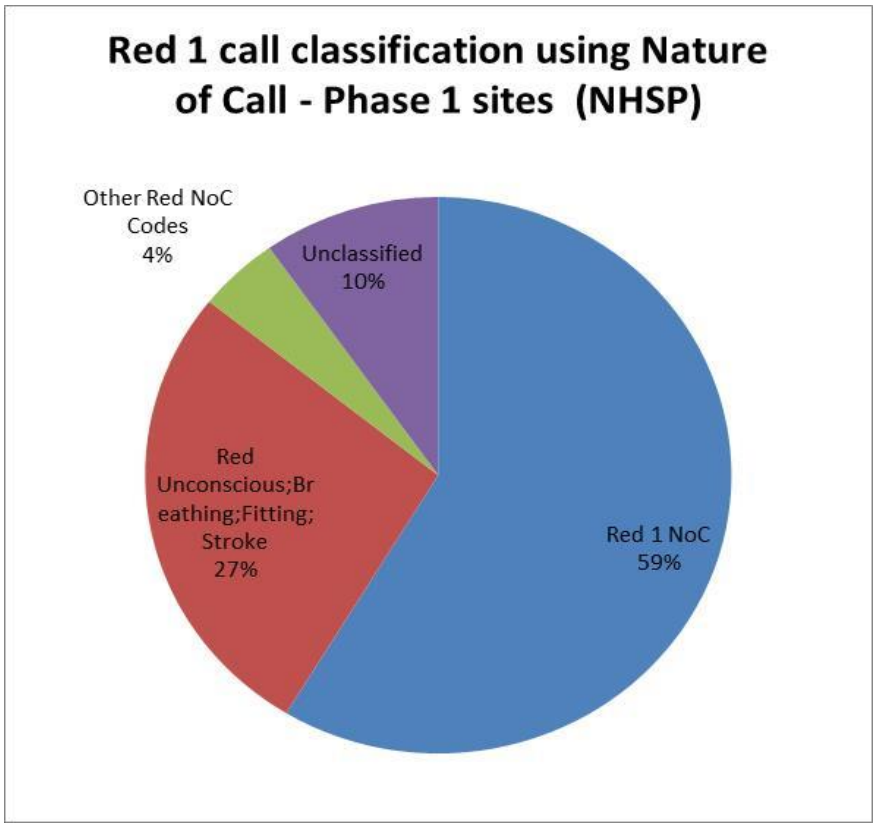
In England two call assessment systems are used – NHS Pathways and the Advanced Medical Priority Dispatch System (AMPDS). The two systems broadly assess calls in the same way but there are some differences in architecture. In particular, the pre-triage questions and NoC are very similar to the key questions and chief complaint identification process at the beginning of the AMPDS call assessment process which creates potential duplication of questioning. The majority of DoD pilot services use NHS Pathways for call triage and so the NoC process has been developed and implemented as a core process for all NHS Pathways users.

Within NHS Pathways, as seen in Figure 10, if the patient is reported as not breathing or unconscious with noisy breathing a Red 1 alert is generated and the call sent for immediate dispatch. At the next step "tell me exactly what happened" a Nature of Call descriptor is chosen from a pre-defined list (see appendix 1 for the current list used in Phase 1 services). Choking and drowning/water incident also generate a Red 1 dispatch. All other descriptors are assigned as either likely Red 2 or Green calls. The call then proceeds using the normal NHSP assessment process. We have assessed the likely rate with which red 1 calls are identified early in the call cycle by using the pre-triage questions and NoC descriptors – the Red 1 capture rate. Table 4 and Figure 11 show the Red 1 capture rates for four Phase 1 services using NHS Pathways.

Table 4 : Red 1 call capture rates using NoC call processes – phase 1

	SCAS	**SECAMB**	**NEAS**	**WMAS**	**combined**
	Nov15-Dec16	Nov16-Jan17	Jan-17	Oct15-June16	
TOTAL Red 1 Incidents	14083	4493	937	15841	35354
Red 1 NoC (including pre-triage questions)					
Choking	556	135	26	714	1431
Drowning/Water incident	18	12	1	11	42
Red 1	5857	1989	396	11071	19313
Total	6431	2136	423	11796	20786
% Red 1 Identified	45.7%	47.5%	45.1%	74.5%	58.8%
Red NoC codes autodispatch					
Unconscious (Not Red1 scenarios)	2136	895	198	2316	5545
Breathing	811	190	36	173	1210
Fitting	1242	304	61	797	2404
Stroke/Neurological	292	60	14	62	428
Total	4481	1449	309	3348	9587
% additional calls with autodispatch	31.8%	32.3%	33.0%	33.0%	27.1%
Total Red 1 and autodispatch NoC	10912	3585	732	15144	30373
% Red 1 and autodispatch NoC	77.5%	79.8%	78.1%	95.6%	85.9%
Red 1 allocated to other Red NoC codes	935	258	93	304	1590
% identified by other Red NoC	6.6%	5.7%	9.9%	1.9%	4.5%
Red 1 allocated to other NoC or unclassified	2236	650	112	393	3391
% identified by other NoC	15.9%	14.5%	12%	2.5%	9.6%

Figure 11 : Red 1 call classification using nature of call



It can be seen that WMAS have consistently recorded a high red 1 capture rate just using the pre-triage questions and the choking and drowning descriptors. The other three services have a lower capture rate of around 45%. During the course of the DoD pilot South Central ambulance service monitored which calls within the remaining NoC descriptors most frequently reached a Red 1 disposition as triage progressed. They found four additional descriptors which were highly likely to generate a Red 1 disposition (unconscious not Red 1; breathing problems; fitting and stroke/neurological). In this service these 4 NoC descriptors now also generate an autodispatch on selection rather than waiting for call assessment to complete to minimise delays (see Box 1). This improves the Red 1 capture rate to 77% in this service and would generate the same rate in the other two similar services if calls assigned to this descriptor generated early dispatch as shown in Table 4. There is a substantially higher number of calls not assigned a NoC descriptor in three services when compared to WMAS. The unassigned calls are predominately calls sent from NHS111 which already have a disposition code assigned to them. WMAS allocate a NoC to these calls whereas the other services do not. If these calls were excluded the capture rate would increase to 86% in these services.

It is unclear why WMAS generate a much higher Red 1 capture rate at the initial stage (an earlier analysis showed a similar rate in SWAST) compared to the other services which show a smaller but similar rate and this is worth further exploration. The addition of additional NoC descriptors to generate an immediate dispatch does also increase the likelihood of calls being over-triaged. Nevertheless other services have addressed capture rate, made modifications to improve it and have shared practice with other services. There is a recognition that a collaborative effort in auditing and

refining the NoC process so that the "right" calls are captured early without introducing substantial amounts of over triage (assigning early as potentially Red 1 and allocating a resource to calls that after full assessment are not Red 1 calls) will be an ongoing process to continue improvements.

Box 1 – South Central Ambulance Service

Review of Nature of Call Effectiveness - Supporting Appropriate Clinical Response

South Central Ambulance Service (SCAS) initiated the pilot in October 2015 using the agreed ARP set on Nature of Call (NoC) codes. Initial analysis showed less than 60% of Red 1 incidents were receiving a Red1 NoC, down from the levels seen using a similar approach by SCAS (key words) before the start of the trial. The analysis confirmed that there were a significant proportion of Red1 incidents recording against specific Red 2 NoC codes. SCAS approach of using the 3 minute threshold to complete the Pathways assessment meant that for these cases, there was a risk of a delayed despatch. Analysis showed that 75% of the remaining Red1 incidents were recorded against only four of the Red2 NoCs. Whilst it was recognised that a number of Red2 patients would receive an earlier despatch than intended through the trial, it was agreed to treat these four codes (Breathing Problems, Fitting/Seizure, Stroke/Neurological and Unconscious (NOT noisy breathing) as a Red1 NoC, leading to an immediate despatch. As a result of this change, 85% of Red1 incidents are treated as a Red following the entry of the NoC, with a 15% improvement in the time taken to despatch to a Red1 incident. Of the remaining Red2 incidents, 68% receive a Red2 NoC, suggesting a high level of accuracy between the selected NoC and assessment outcome.

Implementation of the pre-triage questions and NoC is more difficult in AMPDS as there is potential for duplication of questioning which may be counterproductive, and complex technical changes to CAD systems are needed to incorporate the pre-triage questions and NoC descriptors outside the AMPDS system. In addition the majority of services using AMPDS have only begun using DoD since October 2016 when Phase 1 was extended to all services in England. Nevertheless AMPDS sites have begun exploring the use of pre-triage questions and a modified version of NoC to see if there is potential to enhance the early identification of the most urgent calls. East of England ambulance service has also introduced pre-triage questions and a modified NoC and their early experience so is described in box 2.

Box 2 – East of England Ambulance Service

"With the introduction of ARP in EEAST we can now predict 75% of Red 1 cardiac arrests. In EEAST we have been able to improve our ability to dispatch the right resource first time. Introduction of the new initial questions (pre-triage) has enabled us to identify patients in cardiac arrest at a very early stage in the call.

Early identification of cardiac arrest and other life threatening calls allows dispatch of closest resource and provision of pre-arrival advice to patients such as Cardio Pulmonary Resuscitation (CPR)"

London Ambulance Service (LAS) introduced the pre-triage sieve questions in January 2017 and have conducted an audit of the effect on the time taken to recognise Red 1 and potential cardiac arrest calls and allocation of a resource. They compared key process measures for 6 months before and 4 months after PTS implementation and found that early identification of Red 1 calls has increased from 17% to 45% and potential cardiac arrest cases (AMPDS card 9) from 45% to 75% (Figure 12). The time taken from call connect (T0) to identification of potential cardiac arrest has reduced on average by 30 seconds from 65 seconds to 35 seconds (Figure 13) with a similar reduction in time from T0 to allocation of a resource.

Figure 12: London Ambulance Service – effect of pre-triage questions on change in rate of early identification of potential cardiac arrest

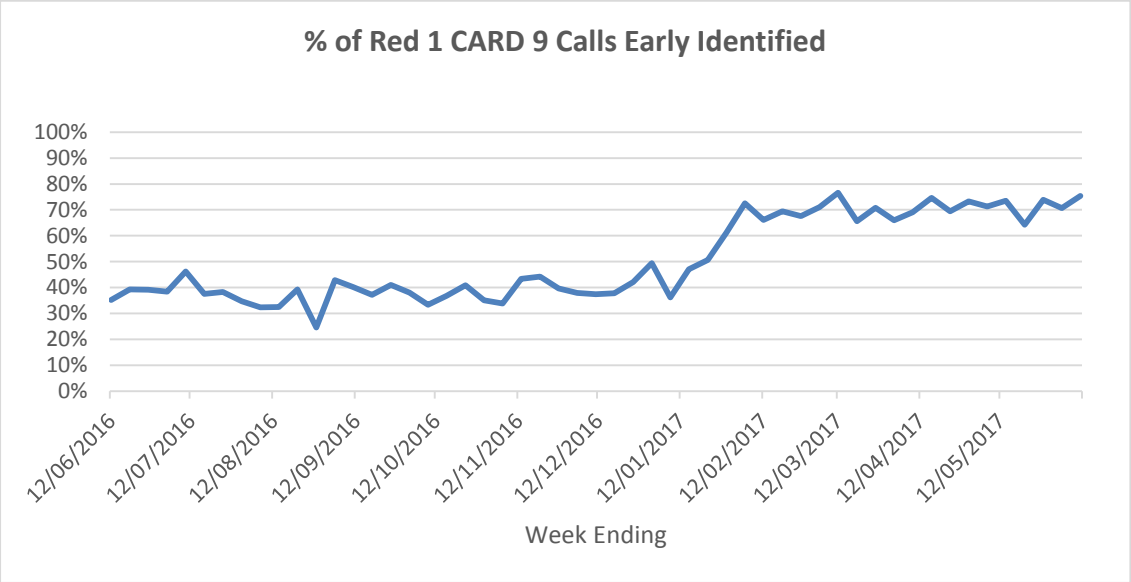
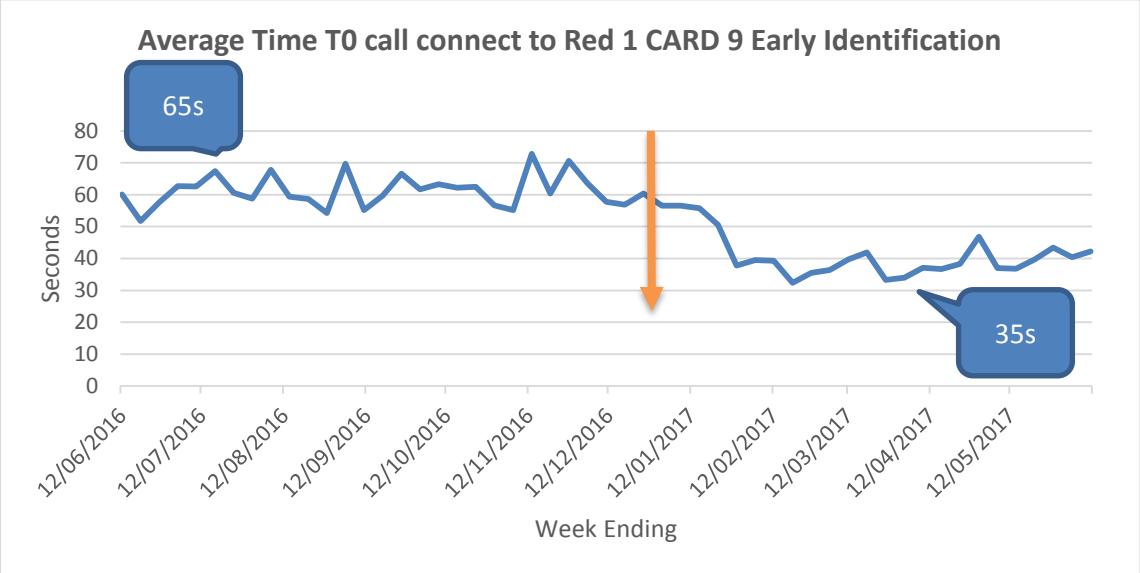


Figure 13: London Ambulance Service – effect of pre-triage questions on change in time to early identification of potential cardiac arrest



This finding is in contrast to the phase 1 statistical analysis where no reduction in allocation time for Red 1 calls was detected although that analysis combined the effects of both PTS questions and the additional triage time. However, a reduction in time to recognise and allocate a resource for potential cardiac arrest has important clinical implications. Research evidence shows that for each 1 minute reduction response time cardiac arrest survival increases by 24%[9] so a 30 second reduction could increase survival by 12%. The most recent figures on cardiac arrest in England report current 28,729 cases per year with a survival rate of 7.9%,[10] that is 2,269 survivors. The LAS data shows an average reduction of 30 seconds to identification and resource allocation but for some cases this may be greater, which will increase the survival advantage, and for others shorter so there is little or no advantage. In the absence of more detailed information on the actual allocation time distribution a cautious estimate of the potential national benefit based on a 12% increase in survival if this magnitude of response time reduction could be replicated across all services would be an additional 272 lives saved each year. Further research will be needed to establish whether this improvement can be replicated, if the reduced allocation time translates in to an equivalent reduction in response times, and the impact on known rather than potential cardiac arrests, to assess whether this potential benefit is realised.

The use of pre-triage questions and nature of call descriptors to identify the most life-threatening calls is more mature in services using NHS Pathways than in services using AMPDS. Within NHSP sites there are differences in the proportion of Red 1 calls identified very early using pre-triage questions and a small number of NoC descriptors but services have looked for solutions to improve the identification of Red1 calls and hence early dispatch of help to this group of patients. Similar work is underway in AMPDS sites and proposals are being developed to explore integration of the pre-triage and NoC components within the AMPDS system itself in order to make the process more efficient. Collectively, across all services, there is recognition that development of NoC is an ongoing process that will need continual refinement and review – particularly if the ARP call category changes are extended to other services since identification of the smaller proportion of calls getting the quickest response needs to be reliable and accurate. A common feature acknowledged for both systems is the importance of call advisor and EMD training to support and reinforce accurate identification of high priority calls, particularly around interrogation and recognition of important cues in answers to the pre-triage questions on consciousness and noisy breathing. There are also problems to be overcome in managing third party calls, for example those from the police, so that prioritisation of these calls can be better aligned to the NoC process without delaying the collection of other critical information such as incident address.

Identification of out of hospital cardiac arrest

During phase 1 of the ARP programme four services using NHS Pathways for 999 call assessment returned weekly data on the number of cardiac arrests in each of the Nature of Call descriptors (3 services for October 2015 – March 2016 and one service October-December 2015). The number and proportion of cardiac arrests identified for each phase 1 NoC descriptor for the pooled data from 4 services is provided in Appendix 1. Table 5 summarises the number and proportion of cardiac arrests identified for the most common descriptors.

Table 5 : Number and proportion of cardiac arrests identified by pre-triage questions and NoC

	Total of NoC	Total arrests	% arrests	% of all arrests
Total	845350	4913	0.6%	100.0%
Red 1	20622	3353	16.3%	68.2%
Unconscious (Not Red1 scenarios)	32425	419	1.3%	8.5%
Breathing	91136	240	0.3%	4.9%
Death unexpected all ages	588	201	34.2%	4.1%
Chest Pain/Cardiac Prob/Back Pain (Upper)	91823	124	0.1%	2.5%
Other	608756	576	0.09%	11.7%
Total	845350	4913	0.6%	100.0%

In all 4 services the majority of cardiac arrest calls are assigned to the descriptor Red 1 although there was some between service variation with the proportion identified as Red 1 reported for the 4 services as 78.6%, 64.3%, 68.4% and 39.9%. As a general principle, the cardiac arrest calls are mainly identified by the three pre-triage sieve questions - that is, patients reported as not breathing or unconscious with noisy breathing, with some additional cases assigned to Red 1 when the initial nature of call descriptor is identified followed by a breathing verification (for example where the caller is not with the patient and is asked to check). Overall 68.2% of cardiac arrest calls were identified as Red 1 and 88.2% were identified within just 5 NoC descriptors. The NoC descriptors other than Red 1 identifying the most cardiac arrests were also consistent across the 4 services with the unconscious (not Red 1) and breathing NoC descriptors being the next two highest in all 4 services. The other two descriptors were chest pain and death unexpected (all ages). The latter is assigned as “amber” in the NoC list and although there were a high number of cardiac arrests within this descriptor (34%) the “amber” designation for this group indicates that additional information collected at the time of the call would suggest that this is predominantly a group of unsalvageable patients where a rapid response would be unlikely to change outcome, although this needs to be verified with outcome data.

Outcomes from cardiac arrest

Overall, the majority of cardiac arrests are detected from a small number of NoC descriptors. However it would be useful to determine if there are other descriptors which include cases of cardiac arrest with a likelihood of successful resuscitation as not all calls reported as, or suggestive of, cardiac arrest are for patients who have a potentially survivable condition. It is also important to consider whether there are call descriptors where there may be low numbers of cardiac arrest but with a high likelihood of successful resuscitation. The Nature of Call categorisation early in the call process provides valuable information on the types of calls which may identify potentially survivable cardiac arrest (and also the types of call where the likelihood of successful resuscitation is negligible). During Phase 1 South Western Ambulance Service (SWAST) collected information of the number of calls assigned to each NoC category, the number of cardiac arrests within each category, timing of arrest (pre-or post, ambulance arrival) and an early indication of potential survivability using the return of spontaneous circulation (ROSC) outcome.

Table 6 provides a summary of the aggregated data provided by SWAST for the time period February 2015 (when DoD was first implemented in this service) to March 2016.There were just 8 NoC categories where at least 5 cases of ROSC were reported. Categories with fewer than 5 ROSC cases were excluded as they can produce disproportionately high percentage ROSC rates where the number of cardiac arrests is also small. The table shows the total number of calls, cardiac arrests and ROSC in each of the 8 call categories and the corresponding ROSC rate, proportion of cardiac arrests occurring after arrival of an ambulance response and the rank of each of these 3 variables (1 highest through to 8 lowest). Proportion of calls where cardiac arrest after arrival of an ambulance response is important as this group of EMS witnessed arrests has a much greater likelihood of successful resuscitation. For comparison the rank for each of these 3 variables has been summed to provide an overall picture of the combined contribution of each of these factors to ROSC for each category.

Table 6: Return of spontaneous circulation (ROSC) rates by call category descriptor

Category	Total calls	Total cardiac arrests	Total ROSC	% ROSC of arrests (rank)	% Arrests crew on scene (rank)	% Arrests en route to hospital (rank)	Rank score (final rank)
Fitting	11892	31	6	19.4 (1)	32.3 (5)	3.2 (6)	12 (=3)
Chest/cardiac/upper back pain	44060	118	22	18.6 (2)	48.3 (3)	22.0 (1)	6 (1)
Collapse	11414	56	10	17.7 (3)	25.0 (7)	8.9 (4)	14 (6)
Medical	61169	82	11	13.4 (4)	57.3 (1)	9.8 (2)	7 (2)
Breathing	40700	234	20	8.6 (5)	50.9 (2)	8.5 (5)	12 (=3)
111/Unknown	Missing	73	6	8.2 (6)	32.9 (4)	9.6 (3)	13 (5)
Unconscious	18053	383	27	7.1 (7)	26.6 (6)	3.1 (7)	20 (7)
Red 1	14910	2728	119	4.4 (8)	3.9 (8)	1.0 (8)	24 (8)

Seven of the 8 call descriptors in this analysis were also in the highest ten call descriptors for volumes of cardiac arrest. The exception is the 111/unknown descriptor. Unsurprisingly the lowest ranked descriptor for ROSC is Red 1 as this group is most likely to comprise cases where cardiac arrest has already occurred. The descriptors with the highest ROSC rates are those where a relatively high proportion of cases have their cardiac arrest in the presence of the attending ambulance crew. Of other descriptors in the top 10 for call volumes within the ROSC data there were 4/50 (8%) ROSC for stroke/neuro; 0/18 for choking; 0/452 for death unexpected all ages and 4/71 (5.6%) for concern for welfare. This confirms that the amber “death unexpected all ages” category has cases with a very low likelihood of successful resuscitation however there were a small number of ROSC cases in the other amber category “concern for welfare”.

Summary of evidence for early identification of emergency calls and cardiac arrest

The nature of call data highlights an important fact, which is that cardiac arrest is a rare event even within an emergency population. Only 0.6% of 999 calls are for cardiac arrest and, in the most recent data available, only 8% of calls identified as cardiac arrest or peri-arrest at the time of the call subsequently prove to be patients in cardiac arrest. Given the challenges of both detecting a rare event and on the basis of what limited information can be extracted from a telephone call with no

sight of the patient, it is important to recognise that identification of around 70% of cardiac arrests early enough in the call assessment process to elicit a Red 1 (and 91.6% a Red 1 or Red 2 response) indicates a high rate of capture. Ideally the capture rate of cardiac arrests in Red 1 and 2 categories should be compared to the capture rates before the introduction of NoC but we have no historical data for this from the 4 sites using NHS Pathways and NoC. London Ambulance Service have provided details of the numbers of cardiac arrest cases within each of the broad groups Red 1 Red 2 and C (or Green) for one year before and one year after the introduction of DoD in February 2015. Prior to DoD 63.9% of cardiac arrests were identified as Red 1 and 29.9% as Red 2 (93.8%) and after 65% Red 1 and 29.4% Red 2 (94.4%) so capture rates were similar each year. The rate is slightly higher for all red calls (3%) than that found in the NoC analysis but as LAS uses AMPDS for call assessment rather than NHS Pathways this may partly account for the difference, which is small. We have also examined some historical (2013)linked data from another service (East Midlands Ambulance Service) which also used AMPDS and found similar results with 64% of cardiac arrests identified as Red 1 and an additional 13.5% as Red 2.

The introduction of NoC shows that a high proportion of cardiac arrests are identified early in the call process and the identification of call types that have a higher ROSC rate provides useful information that could be used to further refine early questioning in the call assessment process. It is also the case that attempts to identify cardiac arrest come at an operational cost as, because the event is rare and can be difficult to detect, it is accompanied by a significant amount of over triage. Only 16 % of Red 1 calls in the phase 1 NoC data were for cardiac arrest and although there is potential for further work to reduce this, given the low frequency of the event it is not clear by how much this over-triage can be effectively reduced.

2.5 Patient safety and adverse incidents

Throughout the pilot phase of DoD all services have monitored and reported weekly on any potential patient safety incidents that may be related to DoD. Services have reported using a number of indicators to assess patient safety including;

- Daily review of DATIX (a serious incident software reporting system) reports for identification and investigation of potential DoD related incidents, particularly delays in care with full investigation of potential incidents
- Review of Serious Incident Report Investigations (SIRI) for potential DoD related incidents with full investigation of incidents
- Regular clinical call audits of 999 calls in line with NHS Pathways/AMPDS quality assurance and compliance processes with an emphasis on potential DoD related problems
- Review of cardiac arrest data and NoC selection particularly Green NoCs. Green NoCs that raise concerns are referred for audit.
- Inclusion of NoC in clinical audit process
- Monitoring of time to respond using the ARP weekly submission data.

Throughout the DoD pilot periods (first and extended pilots) there have been no identified serious or adverse incidents associated with the implementation of DoD. One service investigated a delayed response issue but this was deemed to be a consequence of call taker action and not DoD.

2.6 Staff survey

Survey process

The introduction of Dispatch on Disposition involves a change in working practices for frontline staff. The main changes are for staff in Emergency Operations Centres (EOC) and Clinical Hubs as they are the people directly involved in the call triage and dispatch functions DoD has been designed to change. However, changes in EOC/Clinical Hub will also have an impact on operational road staff as call categorisation and dispatch procedures change, for example by affecting the frequency with which crews are alerted to calls and subsequently stood down or diverted. An important component of the DoD evaluation was an assessment of the impact on staff in terms of changes to their working practices and to explore their views on the potential of the ARP to improve service delivery.

The early pilot sites conducted a short survey of EOC/Clinical Hub staff. We adapted this survey and added additional questions relevant to operational staff. The staff survey was available online to all EOC/Clinical Hub and operational staff in the Phase 1 ARP extended pilot site services from 16th December 2015 to 17th January 2016. There were separate questions for EOC/Hub staff and operational staff with two general questions for all staff. Space for free text comments on perceived advantages and disadvantages was also included. Responses to the survey were entered in to the statistical software package SPSS version 19. Free text comments were tabulated and key themes identified from the narrative responses. The staff survey is provided in Appendix 2.

Survey results

Results presented here are aggregated data for all responses. Some questions were not applicable to all staff (e.g. questions about dispatching were not answerable by EMD's/Call advisors) so all results include only those who responded to each question.

584 staff responded to the survey. Of these 29 (5%) were EMD/call advisor; 48 (8.2%) dispatchers; 11 (1.9%) EOC/Hub clinical advisor; 29 (5%) EOC/Hub manager; and 467 (80%) were operational staff. There was a high level of awareness with 507 (86.8%) of respondents reporting they were aware their service was taking part in the ARP.

Nature of call, triage & call type identification

EOC/Hub staff were asked two questions about their views on effectiveness of pre-triage and NoC questions and allocation of response. Operational staff were asked similar questions with an emphasis on whether they thought triage had improved (Table 7).

Table 7: Responses to triage and NoC questions

Question (EOC/Hub)	Responses (n%)	Question (Operational)	Responses (n%)
How often do you think the pre-triage and nature of call questions identify immediately life threatening calls		**How do you think the Ambulance Response Programme has affected triage and accurate identification of Red 1 calls**	
All the time	5 (4.6)	Much Better	47 (10.1)
Most of the time	68 (63)	A little better	130 (27.9)
Some of the time	31 (28.7)	No change	188 (40.3)
Never	2 (1.9)	A little worse	34 (7.3)
Don't know	2 (1.9)	Much worse	22 (4.7)
Total	108	Don't know	45 (9.7)
		Total	466
What effect do you think the additional time for call assessment has had on patient triage and allocation of the right response category		**How do you think the Ambulance Response Programme has affected triage and accurate identification of other calls**	
		Much Better	52 (11.3)
Much more effective	45 (46.9)	A little better	159 (34.6)
A little more effective	30 (31.3)	No change	157 (34.1)
The same as before	15 (15.6)	A little worse	28 (6.1)
A little less effective	5 (5.2)	Much worse	29 (6.3)
Much less effective	1 (1)	Don't know	35 (7.6)
Total	96	Total	460

A greater proportion of EOC/Hub staff than operational staff thought the pre-triage and NoC questions had improved the identification of life-threatening calls, but in both groups responses indicating that identification was worse were low (1.9% and 12%). Similarly EOC/Hub staff were more likely to respond that allocation of the right response category had improved and responses that this was worse were again small. This is borne out by some of the comments recorded in the survey with a number of statements from operational staff that they thought too many Red calls were not emergencies and were over triaged although they did qualify this by expressing a view that they saw this as a shortcoming of the triage systems rather than the ARP or EOC practices.

Working practices

EOC/Hub staff were also asked additional questions on hear and treat and dispatching processes. 41.7% of EOC/Hub staff thought the additional time for triage had increased hear and treat, 27.2% thought there was no change and 2.9% that it had decreased (28% answered don't know to this question as they may not have been able to answer it). 74.7% thought the changed triage process, with more information available, had improved the ability to dispatch the most appropriate resource and 67.9% thought the new triage process had allowed them to reduce the number of multiple resource dispatches.

Operational staff were asked questions about their perceptions of any change in their working practices resulting from the ARP. Staff reported a modest improvement in the amount of information available to them with 33% responding information was much or a little better and 67% that it was the same. 64.5% responded they thought there had been a decrease in the number of

times they had been stood down since the ARP had started, although 35.5% thought they had been stood down more often. ARP seems to have little effect on the timeliness of meal breaks with only 6.7% reporting that this had improved and most (66.5%) reporting there was no change.

Impact of ARP on demand management and effectiveness

All staff were asked two questions relating to an overall view of the impact of the ARP on managing demand and their ability to be able to do their job effectively. The majority (52.4%) responded that they thought the ARP had allowed their service to manage demand much or a little better, 38% thought there was no change and 9.6% that it was a little or much worse. EOC/Hub staff were more likely to report this had got better than operational staff.

30% of staff thought the ARP had allowed them to do their job much more or a little more effectively with 60.4% reporting no change and 9.7% that they worked a little or much less effectively.

Comparison of NHS Pathways and AMPDS sites

In the extended pilot phase there were 4 sites using NHS Pathways and 1 using AMPDS. This means that responses are higher for NHS Pathways sites than AMPDS (65 versus 45 respondents for EOC/Clinical Hub and 351 versus 116 operational staff respectively). We have compared responses for NHS Pathways and the AMPDS site and responses to the questions were, overall, very similar. EOC/Hub staff in the AMPDS site were more likely to respond positively to the questions on improvements on accuracy of triage and allocation of the right resource (83.9% v 75.4% for NHS Pathways; ability to dispatch appropriate resources (88.9% v 65.6%) and reducing multiple dispatches (75.7% v 61.7%) but more respondents in the NHS Pathways sites thought hear and treat rates had increased (44% v 37%).

Operational staff in NHS Pathways sites were more likely to think call triage was more accurate (50.3% v 32.8%) but responses on decreasing stand downs were identical. Similarly responses to the questions for all staff on the effect of ARP on service ability to better manage demand (52% in both types of service) and the ability to do their job more efficiently (28.6% NHS Pathways v 32.9% AMPDS) were very similar.

Key themes from free text comments

- Overall, the comments reflect a view that additional time for call assessment is seen as a positive step forward and anything that can try and improve the accuracy of triage is seen as a good thing.
- There were some comments on triage systems ability to do this effectively and particularly a view from operational staff that many red calls are not red including calls passed from NHS 111 which services cannot re-triage.
- ARP appears to supports better allocation of the right resource. EOC staff were very positive about this aspect of the programme and this was supported by comments from practitioners in cars who stated that they feel they are being targeted to more appropriate calls.
- Some EOC staff value waiting to allocate vehicles but there are also frustrations on holding back allocating when they “know” it’s going to be a red call. There were mixed views on moving towards a call before allocation – some think waiting is right but others see

advantages for practical reasons (e.g. operational staff can see they may get a call but are moving in the opposite direction while waiting which then has an impact on response time as they turn around).

- Issues remain that despite extra time crews still arrive on scene before a call has "turned green".
- Operational staff commented that stand downs have decreased and they see this as a positive step not just from a crew perspective but also from a public safety and public perception perspective.
- There were comments from operational staff that they thought NoC is working, particularly in identifying peri-arrests, but also frustration that many Red 1 calls (reported as unconscious and noisy breathing) are intoxicated people.
- There were a small number of negative comments that this is just another way of manipulating response performance figures.
- Several comments that ARP and extra time will not have much impact until there are more resources to meet demand.
- EOC/Hub staff (and operational staff) identified the need to have sufficient numbers of clinical advisors in EOC if improvements in hear and treat rates are to be maximised – if there is not enough clinical support then suitable calls will still get an ambulance response.

Summary of staff survey findings

Overall, the staff survey has shown that the ARP is generally viewed by both EOC & clinical hub staff and operational staff as a positive development, with responses indicating a substantial proportion of staff thought there had been some improvements for all questions. The most immediate effect on working practices is with EOC/Hub staff and they tended towards more positive responses to questions with responses indicating substantial improvements in hear and treat rates and the ability to dispatch the right resource to the right patient. Operational staff were also positive, and in particular almost two thirds thought there had been a reduction in the number times they were stood down. Both groups of staff reported an improvement in the accuracy of call triage and half of the respondents thought the ARP was allowing their service to better manage demand. Importantly, responses indicating a negative impact were very small with less than 10% responding that ARP had had a negative effect for most questions.

3. Phase 2 - Call Category Review

3.1 Call Category Review Process

Within the current call category operating framework a substantial proportion of 999 calls (40-46%) are categorised as requiring a Red 2. However, the majority of patients currently coded as Red 2 do not derive clinical benefit from the arrival of an ambulance resource within 8 minutes, but the Red 2 target leads to a range of behaviours that undermine the efficiency of the ambulance service with substantial variation in patient care and unevenly distributed clinical risk. As part of the Ambulance Response Programme (ARP) a commitment was made to review the Red 2 code set to determine whether the current system can be improved and trial possible changes during spring 2016 within the current ARP evaluation plan.

Early discussion with the ARP Expert Reference Group (ERG) concluded that in order to effectively review the current Red 2 code this could not be done in isolation and instead all ambulance dispositions within both AMPDS and NHS Pathways should be reviewed. The review therefore started with a "blank sheet of paper" and developed a call review process that, where possible, took an evidence-based approach by utilising existing data, collecting new data and applying and analysing this data in a transparent and consistent way.

A clinical sub-group was formed to conduct the call review on behalf of the ARP with input and oversight from the Programme Expert Reference Group. The call review process was conducted using the following steps:

1. Creation a new set of definitions for emergency calls based on condition/symptom type, clinical response needed (assessment; treatment/management; transport) and operational response required.

2. A review all call descriptors (AMPDS code or final NHS Pathways symptom group and symptom discriminator[SG/SD] combination) and a category from the revised definition framework set out in(1) allocated to each descriptor.

3. A validity check of the assigned categories using available data to establish, as far as possible, the match between clinical acuity and need for the call descriptor and assigned category using available descriptive call data. Where there is insufficient data to confidently assign a category to a call descriptor a higher response category has been allocated until sufficient data becomes available.

4. A check of the proportions of calls within each type of call assessment/prioritisation system assigned to each of the new defined categories to ensure that the same type of patient is assigned to the same category regardless of the system used.

5. Agreement of the final call descriptor & category assignment with the ARP Expert Reference Group.

A detailed description of the call category review process is provided in Appendix 3. Table 8 provides a description of the first set of revised call categories.

Implementation of the Phase 2.1 call category trial

The revised call categories were approved by the national Emergency Call Prioritisation Advisory Group (ECPAG) in February 2016 and approval was given to begin a trial of the new categories in Spring 2016. Three services (SWAST, YAS and WMAS) were selected to conduct this trial and, following a period of technical reconfigurations within Emergency Operations centres and staff training programmes, the trial began in April 2016 in YAS and SWAST and June 2016 in WMAS. The 3 services incorporated use of both call assessment systems (AMPDS in YAS, NHS Pathways in WMAS and both in SWAST) and represented different vehicle fleet configurations which was considered important given the greater emphasis placed on dispatching the right resource rather than any resource.

During phase 2.1 performance and safety were monitored on a weekly basis to ensure clinical safety, assess the operational feasibility and assess call volumes within a real environment so that adjustments could be made as needed.

The early results from the trial sites showed, as expected, a substantial reduction in the proportion of calls requiring an 8 minute (Red) response (6-8%) when compared to the previous Red 1 and Red 2 proportions of 50% or more. However the group of "Amber" calls posed considerable operational challenges as these comprise a large proportion of calls (>70%). Management of a volume of calls of this magnitude, with inevitable variability in clinical acuity and need, is problematic in terms of prioritising dispatch of available resources as there is insufficient clinical discrimination to help dispatchers. Table 9 shows the proportions of calls assigned to each category for the period 18th April/6th June – 3rd October 2016.

Table 9: Proportion of calls assigned to revised call categories – phase 2.1

Category	YAS	SWAST	WMAS
Red	8.4%	5.5%	6.1%
A Response	43.0%	44.0%	45.7%
A Transport	21.7%	12.0%	6.1%
A Face to Face	8.2%	20.1%	34.3%
All Amber	72.9%	76.1%	86.1%
G Face to Face	6.7%	1.9%	0.5%
G Transport	11.5%	7.8%	8.2%
G Hear and Treat	0.4%	8.6%	6.4%
All Green	18.6%	18.3%	15.1%

Table 8: Revised call category definitions Phase 2.1

<table>
<tr><th>Call type definition</th><th>Response</th><th>Resource</th></tr>
<tr><td rowspan="1">Red– Life-threatening
Time critical life-threatening event needing immediate intervention and/or resuscitation e.g. cardiac or respiratory arrest; airway obstruction; ineffective breathing; unconscious with abnormal or noisy breathing; hanging.
Mortality rates high; a difference of one minute in response time is likely to affect outcome and there is evidence to support the fastest response.
Time interval & performance target– 75% within 8 minutes, Ambulance response within 19 minutes</td><td>Defibrillator
Person trained to use defibrillator
Ambulance clinician who can assess and deliver advanced life support</td><td>Operational response plan to deliver fastest suitable resource</td></tr>
<tr><td rowspan="4">Amber– Emergency
Potentially serious conditions (ABCD problem) that may require rapid assessment, urgent on-scene intervention and/or urgent transport.
Mortality rates are lower; a difference of an extra 15 minutes response time is likely to affect outcome and there is evidence to support early dispatch.
(Call that need conveying clock stop is by the vehicle that actually conveys)</td><td colspan="2">All categories need face to face assessment by a suitably qualified clinician plus</td></tr>
<tr><td>AR(Y1)Assess; Treat; Transport
e.g. Probable MI, serious injury</td><td>Suitably qualified clinician who can assess and treat and vehicle that can transport</td></tr>
<tr><td>AT(Y2)Assess; Transport
e.g. Stroke</td><td>Vehicle that can transport</td></tr>
<tr><td>AF(Y3) Assess; Treat
e.g. Fits; diabetic hyper/hypoglycaemia; overdose; unconscious with normal breathing</td><td>Nearest available resource (any type) with suitably qualified clinician who can assess and treat</td></tr>
<tr><td rowspan="3">Green– Urgent
Urgent problem (not immediately life-threatening) that needs transport within a clinically appropriate timeframe or a further face to face or telephone assessment and management.
Mortality rates are very low or zero; a difference of one hour or more might affect outcome and there is evidence to support alternative pathways of care.</td><td>GF(Z1)Face to face assessment and management which may include transport</td><td>a)Suitably qualified clinician who can assess & manage
b)Transporting vehicle where needed</td></tr>
<tr><td>GT(Z2) Transport only required</td><td>Transporting vehicle with suitable HCP (within specified timeframe)</td></tr>
<tr><td>GH(Z3) Calls which do not require an ambulance response but do require onward referral or attendance of non-ambulance provider in line with locally agreed plans or dispositions, or can be closed with advice (Hear & Treat)</td><td>Suitably qualified clinician in EOC who can assess & manage</td></tr>
<tr><td>Type S – Specialist response
Incidents requiring specialist response i.e. hazardous materials; specialist rescue; mass casualty</td><td>Locally agreed plans apply</td><td></td></tr>
</table>

The clinical coding subgroup therefore undertook a further review of the call categories. In particular it was agreed that the assignment of call codes to the Amber categories was too large and did not sufficiently discriminate between calls for emergency and less conditions which require a response within 19 minutes or less and urgent conditions. The separation of Amber calls into 3 categories based on need for treatment and transport, transport or face to face assessment also appeared to be too complex both in terms of the ability to discriminate with a high degree of specificity at the time of the call and in managing call stacks and appropriate allocation of resources. This could be improved by a simpler transport or assessment split.

Taking these factors in to account the subgroup revisited the first iteration of call category definitions and created a set of revised categories that may better reflect the response required for different types of conditions. The revision incorporated a further differentiation in timeframes to enable better discrimination between different types of calls and a reduction in the number of categories to support operational implementation but retaining the principle of allocating the right type of response rather than any response. The revised categories (Phase 2.2) are set out in Table 10. A detailed description of the second call category review processes is provided in Appendix 2.

Following approvals phase 2.2 of the call category trial commenced in October 2016 in YAS and WMAS, and November 2016 in SWAST. It is the evaluation of this current phase that is predominantly described here.

Table 10: Revised call categories – Phase 2.2

Call type definition	Response and Resource
Category 1 -Life-threatening Time critical life-threatening event needing immediate intervention and/or resuscitation e.g. cardiac or respiratory arrest; airway obstruction; ineffective breathing; unconscious with abnormal or noisy breathing; hanging. Mortality rates high; a difference of one minute in response time is likely to affect outcome and there is evidence to support the fastest response. **Time interval & performance target– 75% within 8 minutes, Ambulance response within 19 minutes**	Defibrillator Person trained to use defibrillator Ambulance clinician who can assess and deliver advanced life support Transporting vehicle where transport required Operational response plan to deliver fastest suitable resource
Category 2 - Emergency Potentially serious conditions (ABCD problem) that may require rapid assessment, urgent on-scene intervention and/or urgent transport. Mortality rates are lower; a difference of an extra 15 minutes response time is likely to affect outcome and there is evidence to support early dispatch. **(Calls that need conveying clock stop is by the vehicle that actually conveys)**	**C2T** Assess; Treat; Transport e.g. Probable MI, serious injury, stroke, sepsis, major burns Suitably qualified clinician who can assess and treat **and** vehicle that transports where needed
	C2R Assess; Treat e.g. Fits; unconscious with normal breathing Nearest available resource (any type) with suitably qualified clinician who can assess and treat
Category 3 – Urgent Urgent problem (not immediately life-threatening) that needs treatment to relieve suffering (e.g pain control) and transport or assessment and management at scene with referral where needed within a clinically appropriate timeframe. Mortality rates are very low or zero; a difference of one hour or more might affect outcome and there is evidence to support alternative pathways of care. **(Calls that need conveying clock stop is by the vehicle that actually conveys)**	**C3T** Assess; Treat; Transport e.g. serious injury modalities without systemic compromise; burns (not major); non-emergency late pregnancy/childbirth problems.
	C3R Assess; Treat Calls within scope of advanced clinical practice and suitable for treat and leave. E.g. uncomplicated diabetic hyper/hypoglycaemia; not immediately at risk drug overdoses; non-emergency injuries; abdominal pain.
Category 4 – non-urgent Problems that are not urgent but need assessment (face to face or telephone) and possibly transport within a clinically appropriate timeframe. **Onward management is locally agreed including transport times for HCP calls**	**C4T** Assess; Treat; Transport • 999 or 111 calls that may require a face to face ambulance clinician assessment • Requests for transport by health care professionals
	C4H Non-ambulance response Calls which do not require an ambulance response but do require onward referral or attendance of non-ambulance provider in line with locally agreed plans or dispositions, or can be closed with advice (Hear & Treat)
Type S – Specialist response Specialist response incidents i.e. hazardous materials; specialist rescue; mass casualty	**Locally agreed plans apply**

3.2 Phase 2 evaluation methods

The revised call categories represent a substantive change and wholly different operating model to the current Red 1, Red 2 and Green categories with a broader range of response interval options that reflect more accurately the acuity present in the 999 call population. There have also been changes in the reporting of response times in the trial sites to align with the principle of dispatching the right resource. For patients who are conveyed to hospital the "clock stop" for response time performance measurement is the conveying vehicle. This is in contrast to the current measure of first resource on scene regardless of type. As a result a controlled comparison with other services was not possible as no equivalent comparisons can be made between current and new categories.

As in Phase 1, the 3 trial sites returned weekly data for a comprehensive dataset for phases 2.1 and 2.2 that included:

- Activity – total numbers of calls, incidents and volumes within each response category
- A range of mean and percentile (50, 90, 95%) time intervals by category – call to allocation, arrival on scene, discharged at scene or arrived at hospital
- Mean number of resources allocated for each call category for both all resources (this includes external resources such as community first responders) and core resources, that is ambulance service only
- Proportion of calls managed by hear and treat
- Proportion of calls conveyed to hospital for each category
- Re-contact rates with the ambulance service within 24 hours for hear & treat and see & treat
- Additional information for factors that may have an impact on operational performance including calls passed from NHS 111, hours lost at hospital and available staff hours.

We have used this weekly data to provide a descriptive analysis of operational performance measures such as numbers of calls by category and response times for phase 2.2 only as this is the most relevant iteration of the call category review. We have also conducted a statistical analysis to identify trends during phase 2.2. For this we have plotted the weekly activity for each trial ambulance service so that change over time during Phase 2.2 could be seen graphically. We then fitted a time series regression model to each ambulance service to test for evidence of a trend over time. The models consisted of terms to adjust for the total number of emergency incidents, the total number of calls answered, the number of hours lost at hospital and the planned staff hours.

In addition to the routine weekly data we have conducted two other analyses.

a) Although we were unable to conduct a controlled comparison between the Phase 2 trial services and Phase 1 services we have attempted to explore any broad effects by examining a small number of whole system measures, that is operational indicators applied to all 999 calls rather than individual categories. Each trial service returned weekly data for the period 1st January – 30th April 2017 on 9 operational indicators;

- Median and 95th percentile time from call connect to first resource on scene

- Median and 95th percentile time from call connect to arrival at hospital (conveyed patients)
- Median and 95th percentile time from call connect to left scene (non- conveyed patients)
- Average core resources per incident for all attended incidents
- Average core resources per incident – all conveyed incidents
- Average core resources per incident – all non-conveyed incidents

We have measured differences in trends between Phase 1, 2.1 and 2.2 for these indicators adjusted for total number of incidents, total number of calls, hours lost at hospital, staff hours and seasonality.

b) Each of the 3 trial services have provided weekly data on a small set of performance measures separated for urban, mixed urban and rural and rural areas for the time period January 2016 to April 2017. This allows an assessment of whether the phase 2.1 and 2.2 changes have had any impact on service equity across different geographical areas.

As for phase 1, we also assessed the identification of the most serious calls using pre-triage questions and NoC and conducted two staff surveys.

3.3 Results of the quantitative analysis of operational indicators

Proportions of incidents assigned to each phase 2.2 call category

Table 11 shows the proportions of incidents assigned to each of the revised phase 2.2 call categories. The proportion of calls assigned to category 1 is a third higher in YAS compared to the other 2 services. In this service a higher proportion of incidents are assigned to Category 2 (Transport) than in the other 2 services and in contrast a much greater proportion are assigned to Category 3 in WMAS and SWAST.

Table 11: Proportions of incidents assigned to each phase 2.2 call category

	WMAS[1]	SWAST[2]	YAS[1]
Category 1	6.9%	6.1%	11.2%
Category 2 Response	9.0%	6.3%	3.5%
Category 2 Transport	32.5%	37.7%	50.0%
Total Category 2	**41.5%**	**44.0%**	**53.5%**
Category 3 Response	34.3%	25.0%	8.3%
Category 3 Transport	8.6%	12.4%	14.3%
Total Category 3	**42.9%**	**37.4%**	**22.6%**
Category 4 Transport	3.1%	4.8%	9.5%
Category 4 Hear and Treat	5.6%	4.5%	3.2%
Total Category 4	**8.7%**	**9.3%**	**12.7%**

[1]WMAS and YAS figures are Oct 2016 – May 2017. [2]SWAST figures are Oct 2016 to Jan 2017 as after this date they were unable to isolate hear and treat calls.

Overall the proportions within each category are more closely aligned between WMAS and SWAST and one obvious explanation for the differences is that different call assessment systems are being used within the services. SWAST uses both systems and for total Category 2 and 3 calls does have a value between the other 2 services but is more closely aligned to WMAS. In YAS (using AMPDS) more calls are allocated to Category 1 and 2 but also category 4 (transport). Category 4 may be influenced by how calls from health care professionals are managed within the triage systems. Although differences in the allocation of calls to different categories may be a feature of how each call assessment has aligned call descriptors (AMPDS codes and NHSP SG/SD descriptors) each category this is not clear and warrants further scrutiny.

A feature of the call category review was an attempt to discriminate between calls that were likely to need transport and those that either needed transport and early treatment that could be initiated by the ambulance service, or could be suitable for assessment and treatment at scene but would not need taking to hospital. Table 12 shows the proportions of incidents that were transported to hospital within each of the revised call categories for each of the 3 trial services.

Table 12: Proportions of incidents that were transported to hospital - Phase 2.2

Volume of transported incidents	WMAS	SWAST	YAS
Category 1	64.0%	58.8%	77.0%
Category 2 R	63.8%	58.8%	67.8%
Category 2 T	69.9%	64.4%	79.9%
Category 3 R	53.4%	46.7%	67.0%
Category 3 T	72.2%	64.4%	67.7%
Category 4 T	39.3%	41.2%	74.8%

The results show some variation in the conveyance rate for Category 1 calls with the highest rate in YAS and the lowest in SWAST. For Category 2 the conveyance rate was higher in the transport category than the response category although the differences are small (6-10%). For Category 3 the difference in conveyance rate between the response and transport categories was bigger in 2 services (19% in WMAS and SWAST) but the same in YAS. The conveyance rate for category 4T is noticeably lower in WMAS and SWAST but this may be a consequence of differences between services in the inclusion of calls from Health Care Professionals (HCP) transport requests in this data. A more detailed analysis of the nature of calls within this category (e.g. 999 calls versus HCP calls) may shed more light on this finding. The conveyance rate was lowest across categories 1-3 in SWAST which historically has one of the highest non-conveyance rates in the country so some of the differences may be explained by operational practices that are unconnected to the call categories. There is not a very clear and obvious distinction in conveyance rates between the Response and Transport categories and so it is worth reflecting on whether the current call assessment systems are sufficiently discriminatory to allow decisions about which incidents may or may not need transport at the time of the call and hence if the added complexity of multiple categories is worthwhile.

Trends in response performance and resource use

Table 13 provides a summary of the average response time performance for each service by category during Phase 2.2. WMAS and YAS figures are Oct 2016 – May 2017. SWAST figures are Oct 2016 to Jan 2017 as after this date they were unable to isolate hear and treat calls. Response time performance for Category 1 calls is consistent across all 3 services with half of these calls receiving a response within 6.5 minutes and 90% in less than 13 minutes. For categories 2-4 there is more between service variation. The response times in these categories reflect the revised clock stop of conveying vehicle for patients who are transported to hospital and so include any waiting time for a conveying resource. Services also have more flexibility in response timeframes for categories other than C1. There is a consistent pattern of shorter times across all categories in WMAS. This may be the result of many internal factors around how the response model has been implemented but to some extent is likely to reflect the shift in fleet changes made in this service with a higher emphasis on provision of ambulances that can both respond to and convey patients is needed and a much lower use of single response vehicles. However, it is likely that other operational processes have also played a part and this would be worth more detailed examination to identify strategies that can

support the implementation of any changes to call categories nationally. For category 2 calls YAS has the highest proportion in this category and demonstrates shorter times for the Category 2 T (transport) than R (response) for the higher percentiles (85-99)indicating that the strategy of allocating the right resource to patients highly likely to need transport is having some success in reducing longer waits.

Table 13: Average response time performance for each service Phase 2.2.

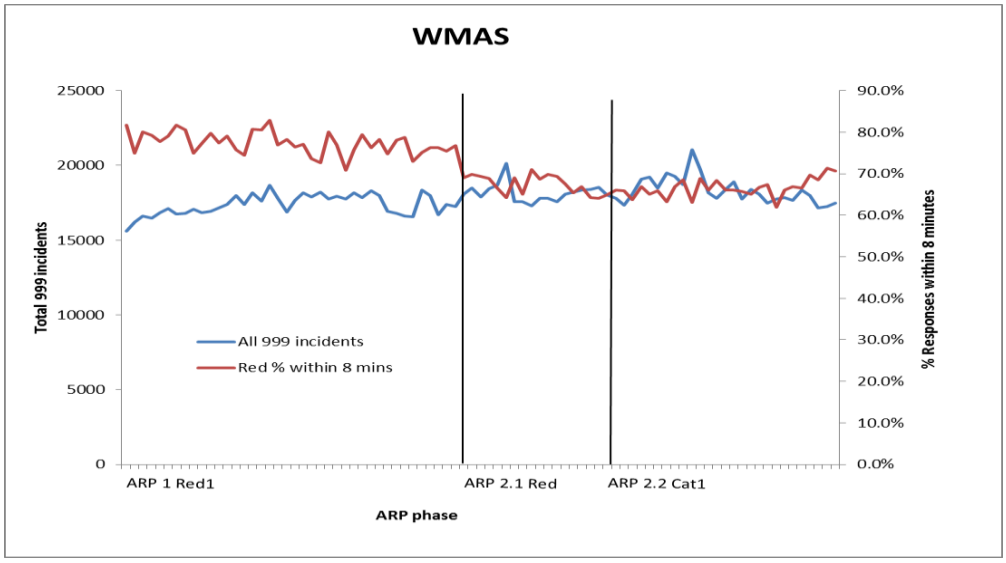
Category	Time from call to resource on scene (hh:mm:ss)	SWAST	YAS	WMAS
Category 1	Mean	00:07:10	00:06:54	00:07:13
	50th centile	00:06:09	00:06:22	00:06:35
	70th centile	00:07:52	00:08:15	00:08:39
	75th centile	00:08:32	00:08:52	00:09:23
	80th centile	00:09:28	00:09:36	00:10:20
	85th centile	00:10:45	00:10:37	00:11:12
	90th centile	00:12:35	00:12:04	00:12:16
	95th centile	00:15:48	00:14:34	00:14:11
	99th centile	00:23:41	00:21:15	00:20:03
Category 2 R	Mean	00:19:04	00:14:42	00:10:42
	50th centile	00:14:03	00:10:38	00:09:33
	70th centile	00:21:39	00:16:35	00:12:52
	75th centile	00:24:23	00:18:25	00:14:01
	80th centile	00:27:51	00:20:47	00:15:14
	85th centile	00:32:34	00:24:01	00:17:07
	90th centile	00:39:59	00:29:26	00:19:04
	95th centile	00:54:39	00:42:01	00:21:00
	99th centile	01:46:53	01:20:29	00:29:26
Category 2 T	Mean	00:21:11	00:15:52	00:11:30
	50th centile	00:18:42	00:12:26	00:10:16
	70th centile	00:28:31	00:16:54	00:14:19
	75th centile	00:32:14	00:18:35	00:15:31
	80th centile	00:37:04	00:20:43	00:16:54
	85th centile	00:43:47	00:23:35	00:18:38
	90th centile	00:54:18	00:27:58	00:21:29
	95th centile	01:14:33	00:38:51	00:23:10
	99th centile	02:13:44	01:05:41	00:32:55
Category 3 R	Mean	00:44:47	00:29:03	00:20:14
	50th centile	00:28:20	00:18:21	00:14:58
	70th centile	00:47:12	00:31:32	00:28:23
	75th centile	00:54:47	00:36:49	00:32:39
	80th centile	01:04:54	00:43:33	00:38:19
	85th centile	01:19:14	00:52:27	00:45:03
	90th centile	01:40:43	01:05:52	00:57:16
	95th centile	01:59:13	01:28:26	00:53:18
	99th centile	03:40:12	02:29:14	01:24:24

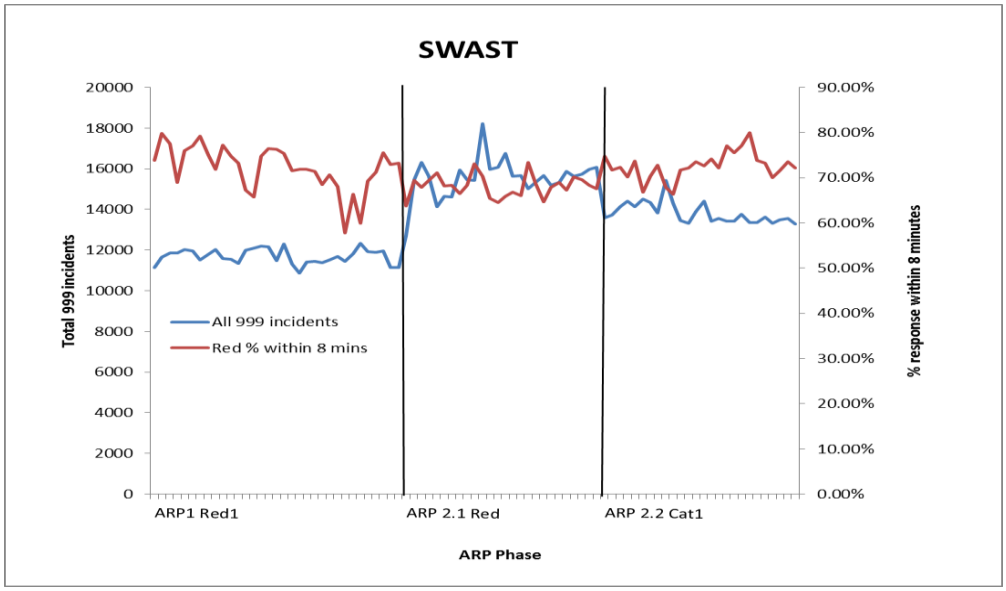
Category	Time from call to resource on scene (hh:mm:ss)	SWAST	YAS	WMAS
Category 3 T	Mean	00:47:16	00:37:15	00:21:19
	50th centile	00:32:18	00:22:24	00:16:24
	70th centile	00:57:59	00:39:43	00:28:08
	75th centile	01:07:22	00:46:42	00:34:41
	80th centile	01:19:14	00:55:39	00:40:16
	85th centile	01:34:45	01:07:29	00:48:20
	90th centile	01:58:01	01:25:36	01:01:14
	95th centile	02:36:48	01:58:43	00:53:36
	99th centile	04:42:14	03:32:22	01:23:49
Category 4 T	Mean	01:20:15	01:20:18	00:34:16
	50th centile	00:58:38	00:51:35	00:21:43
	70th centile	01:38:45	01:31:43	00:13:06
	75th centile	01:53:37	01:47:11	00:14:07
	80th centile	02:12:17	02:06:27	00:15:11
	85th centile	02:36:14	02:31:33	00:16:52
	90th centile	03:12:30	03:08:27	00:19:17
	95th centile	04:16:15	04:08:55	01:46:50
	99th centile	07:00:23	06:38:30	02:53:07
Category 4 H	Mean	00:21:13	00:20:05	00:10:58
	50th centile	00:10:25	00:11:54	00:09:47
	70th centile	00:20:35	00:22:45	00:48:02
	75th centile	00:25:22	00:27:04	00:57:51
	80th centile	00:31:48	00:33:32	01:11:31
	85th centile	00:39:49	00:39:59	01:33:02
	90th centile	00:52:34	00:49:07	01:59:39
	95th centile	01:15:47	01:03:02	00:20:25
	99th centile	02:18:40	01:31:01	00:29:56

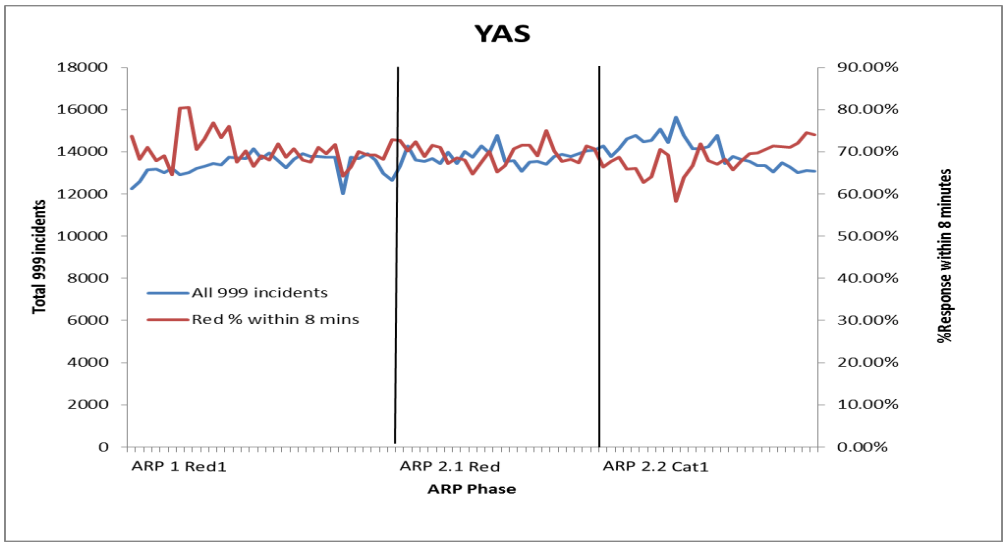
For all categories 2-4 the reported response times are longer in SWAST than other services. This is an important observation as it indicates that, even with a consistent set of call categories across all 3 services, there is individual variation in the ability to deliver timely response which will be dependent on a range of other factors. These include the operating environment (including the mix of urban and rural geography and seasonal factors such as tourism which affect demand); the fleet mix and the implications for implementing the principles of “right response” particularly in respect of providing a conveying resource “first time” to patients that may need transporting to hospital; and the overall amount of resource available that can be provided within the financial provision they receive to provide their service.

The ability to provide a timely response to those patients with the greatest clinical need has been an important focus of the ARP. We have examined the trends in response performance (% attended within 8 minutes) against demand in the three trial services over the different ARP phases for the most urgent category - Red 1 (Phase 1); Red (Phase 2.1) and Category 1 (Phase 2.2). Figure 14 shows the performance for each service over the 3 phases.

Figure 14: Weekly demand and Red 1/Red/Category 1 performance – August 2015 – May 2017







The categories are not exactly equivalent as the Phase 2 categories are higher volume (6-11% of 999 calls compared to 3% categorised as Red 1). There are some differences between the 3 services. In WMAS before Phase 2 there was a trend that showed steady response time performance although as demand increased performance, predictably, started to decline. After Phase 2 there is a clear early decrease in performance not consistent with demand which then stabilises over time. In SWAST the opposite happens and as Phase 2 progresses there is an upward trend in the proportion of the most urgent calls responded to within 8 minutes relative to demand. The step increase in demand at the beginning of Phase 2.1 most likely reflects the summer seasonal effect on demand that is common to this service. In YAS there is a stable picture with a trend towards improving performance as Phase 2 progresses.

The trends in WMAS are most surprising as this service has consistently been the highest performer against the 8 minute target for both Red 1 and Red 2 calls for some time. A reduction in performance for high acuity calls, despite the substantial decrease in the proportion of calls requiring an 8 minute response in Phase 2, is unexpected. However, this finding can be explained by operational changes that have been made in this service to support implementation of ARP Phase 2. WMAS have provided a description of the operational changes they have made which have most likely impacted on this change in performance (Box 3). In all 3 services there is also a trend towards closer alignment of demand and performance, excepting peak times at Christmas and New Year, suggesting a steady and more consistent response to the most urgent calls despite fluctuations in demand.

The response performance emphasises the compromises that have to be made between providing a rapid response to the most urgent patients and an equitable response, in particular minimising long delays for less urgent patients, within a service that is providing a response to a large and heterogeneous population of patients with health problems of variable clinical acuity. It illustrates that changing the response model requires a range of complex operational changes beyond the call assessment and categorisation process and that a system of ongoing review and refinement is needed to optimise delivery as the requirements of the model become clearer over time. The experience at WMAS also highlights the effects of demand and hospital delays and that, even with a new operating model in a high performing service, there comes a point where resources are utilised to a capacity beyond which further gains are not possible and maintaining performance becomes difficult.

Box 3 – Operational changes in WMAS to support ARP

Pre DOD the Trust would actively assign to incidents on pre-alert regardless of the likely categorisation of that patient. This was inefficient and caused a number of operational difficulties as resources were diverted, reassigned and stood down on more occasions than they actually arrived on scene with patients. When DoD was introduced in October 2015 the education that the dispatch teams received was around delayed assignment to incidents until confirmation of code or the NoC suggested a likely high categorisation. Whilst the NoC has been successful in identifying a high proportion of potential Category 1 patients early, there will inevitably be some patients that are only identified as a category 1 later during the call.

Changes to the AQI's are also a contributing factor. During January 2016 the AQI's clarified when a defibrillator would 'stop the clock'. This meant that it was no longer sufficient to just have a defibrillator available should the patient require it, it had to be confirmed as at the patient side before stopping the clock. Whilst the impact on WMAS was not big it resulted in performance reduction of 1%.

In order to improve efficiencies changes to the operating model came into effect from October 2015. These changes were introduced in order to increase capacity to deal with the prevailing demand during winter 2015 and onward into the future. This ensured that each patient has a response that was paramedic led, was capable of assessing and treating a patient and of onward transportation should that be required. As previously evidenced the Trust continues to see the benefits of this model change in that no patients are kept waiting unnecessarily and there are no lengthy delays where an RRV is on scene awaiting back up to transport the patient from a Double crewed ambulance (DCA).

Pre ARP 2 the total percentage of red 1 incidents were approximately 4.5% for the Trust and comprised, in the main of patients in cardiac or respiratory arrest. These patient groups were easy to identify and were respected as requiring a very quick response by crews and responding resources. With the introduction of more patients into this group it is possible that the highest priority has been somewhat diluted.

There are many external factors that have also impacted on category 1 performance, not least the sharp increase in hospital delays that see operational resources tied up with patients in A&E departments well in excess of the 15 minute handover targets. Additionally the reconfiguration of hospital functions has also meant that it is no longer simply a case of transporting a patient to the nearest A&E department. This has led to increased task times which decreases resource availability. Activity growth continues to be seen with no associated commissioner investment to match the demand rises, rather a do more with what you have requirement.

The Trust recognises that in the period moving forward the performance gap in relation to category 1 incidents needs to be closed. The resourcing capacity issues in the system are now about right so once the response standards are finalised and there is a complete understanding of what is required the appropriate modifications will take place to improve category 1 performance. The Trust has already initiated this piece of work, and an action plan has been started where various initiatives are being explored in order to improve performance whilst not jeopardising operational efficiencies.

These descriptive analyses of call volumes within categories and response time performance for individual categories do not account for other factors which may influence performance. We have addressed this in two ways. Firstly, we have examined trends in a range of operational performance times and resource use in each of the 3 trial services over Phase 2.2 adjusted for call volumes, hours

lost at hospital and staff availability. Secondly we have examined a small set of whole service operational performance and resource use indicators, also adjusted for the same factors, to assess any potential benefits to the 999 caller population over 1 year spanning the introduction of the revised call categories.

Trends in operational performance and resource use Phase 2.2

We measured 39 indicators (percentage Category 1 responses within 8 and 19 minutes; median and 95th Percentile response intervals, resource allocation and on scene, conveyance rates for categories C1-C3 and hear and treat & re-contact rates) for each trial service so 117 measures in total. The full results of the statistical analysis are provided in the supplementary file. To summarise, for the majority of measures (93/117; 79.4%) there was no statistically significant trend indicating a change over the Phase 2.2 period. This indicates stable performance across the trial period, which includes winter peaks in demand and hours lost at hospital, suggesting the new operating model helps to mitigate serious decline in performance when services are under substantial pressure. There were statistically significant changes in trend for 24 (20%) of the measured indicators. In summary, these showed;

- An increase in the proportion of category 1 calls receiving a response within 19 minutes in two services (0.19% and 0.07% per week in YAS and WMAS respectively).
- In WMAS a decrease in allocation of core resources in 4/5 call categories (ranging from -0.0012 to -0.0031 per incident per week) and a reduction in core resources on scene for 5/5 categories (range -0.0010 to -0.0022 per week).
- For call connect to time on scene, in YAS there was a reduction in the 95th percentile time for Category 1(-4.40 seconds per week), and median and 95th percentile time for Category 2R (-6.24 and -98.24 seconds per week respectively). In SWAST there was an increase in median time of 4.03 seconds per week for category C2T but no increase in the 95th percentile time.
- For call connect to conveying vehicle leaving scene or patient discharged at scene there was a median increase of 6.48 seconds per week for C1 and a 95th percentile decrease of -46.69 seconds per week for category C2R in YAS. In SWAST there was a median increase of 7.92 seconds per week for category C2T but again this increase had disappeared for the 95th percentile time.
- The proportion of patients transported increased by 0.16% per week for category C2R in SWAST but showed a trend to decreasing in all 3 services for category C2T in all 3 services and this was significant in SWAST (-0.08% per week) and WMAS (-0.09% per week). For category C3T there was a trend to increasing conveyance in YAS (0.11% per week) and WMAS (0.10% per week).

Overall there was a trend towards improving response performance in some services and this was particularly evident in YAS. The single trend in increasing response performance in one category in SWAST was not evident for the 95th percentile time indicating there was no degradation in overall performance in this category for the majority of calls. It is likely that most of the efficiency gains have been made during Phase 1 and 2.1 but this analysis has shown there are still small but significant efficiency gains being made in one service and no loss of efficiency in the other two. The proportion of incidents transported to hospital has shown some interesting findings with this

reducing in all services for the C2 transport category but increasing in two services for the C3 transport category. This may reflect some of the difficulties in distinguishing who will or will not require transport in the higher acuity category 2 and possibly improving discrimination over time for category 3 as more “correct” disposition decisions are made.

Trends in whole service performance

An important consideration of the different phases of ARP and in particular the changes made to the call categories with corresponding changes in expected response time performance is the effect of service provision on the overall 999 population. We have examined trends in response performance and resource utilisation over a 16 month period in the three phase 2 trial services spanning a baseline phase 1 period and the introduction of phases 2.1 and 2.2 for all 999 incidents receiving a response. The full results of the statistical analyses are presented in the supplementary file. A summary of the results comparing changes between Phase 1, Phase 2.1 and Phase 2.2 are presented in Table 14. For simplicity the value of any significant step change (indicating an immediate effect after a change) or slope change (a change in trend) only is reported without confidence intervals. For some measures there were step and slope changes before and after implementation and for these the net effect is reported in the table.

The observed trends for each whole service measure are presented graphically in Figures 15-17.

Table 14: Whole service response performance and resource utilisation over one year – Phase 2 services

Whole service measure	SWAST	WMAS	YAS
Time from call connect to arrival of first core resource on scene Median 95th Percentile	Phase 2.1 Step ↑323.7 seconds Phase 2.2 Step ↓-201.4 seconds Phase 2.2 ↓-324.4 seconds (22 minutes)	Phase2.1 Step ↑123.0 seconds Phase 2.1 Step and slope change. Net effect ↑6.6 seconds per week	Phase 2.1 Step ↑of 102.3 seconds Phase 2.2 Step ↑of 62.5 seconds Phase 2.2 Step ↑880.8 seconds (14.7 minutes)
Time from call connect to arrival at hospital (see and convey) Median 95th Percentile	Phase 2.1 Step and slope change. Net effect ↓-5.1 seconds/ week Phase 2.2 Step ↓-1292.04 seconds (21.5 minutes)	Phase 2.1 Slope change ↓-0.8 seconds/ week No change	No change No change
Time from call connect to leaving scene (see and treat) Median 95th Percentile	Phase 2.1 Step ↑207.9 Phase 2.2 ↓-235.9 seconds Phase 2.2 Step ↓-847.1 seconds (14.1 minutes)	Phase 2.1 step and slope change. Net effect ↓-0.7 seconds/ week Phase 2.1 Slope ↓-3.2 seconds/ week	Phase 2.1 Step ↑of 208.3 seconds No change
Average core resources per incident – all attended incidents	Phase 2.1 Step and slope change. Net effect ↑0.0232 resources per incident Phase 2.2 Slope change ↓-0.0035 resources per incident per week	Phase 2.1 step ↓-0.021 resources per incident	P2.1 ↓-0.0895 resources per incident
Average core resources per incident – all conveyed incidents	Phase 2.1 Step ↓-0.0521 resources per incident	Phase 2.1 Step ↓-0.0399 resources per incident	Phase 2.1 Step ↓ -0.1014 resources per incident
Average core resources per incident – all see & treat incidents	Phase 2.1 Slope ↑0.0001 resources per incident per week	Phase 2.1 Step ↑0.0101 resources per incident	No change

Figure 15: Time from call connect to arrival of first core resource – median and 95th percentile

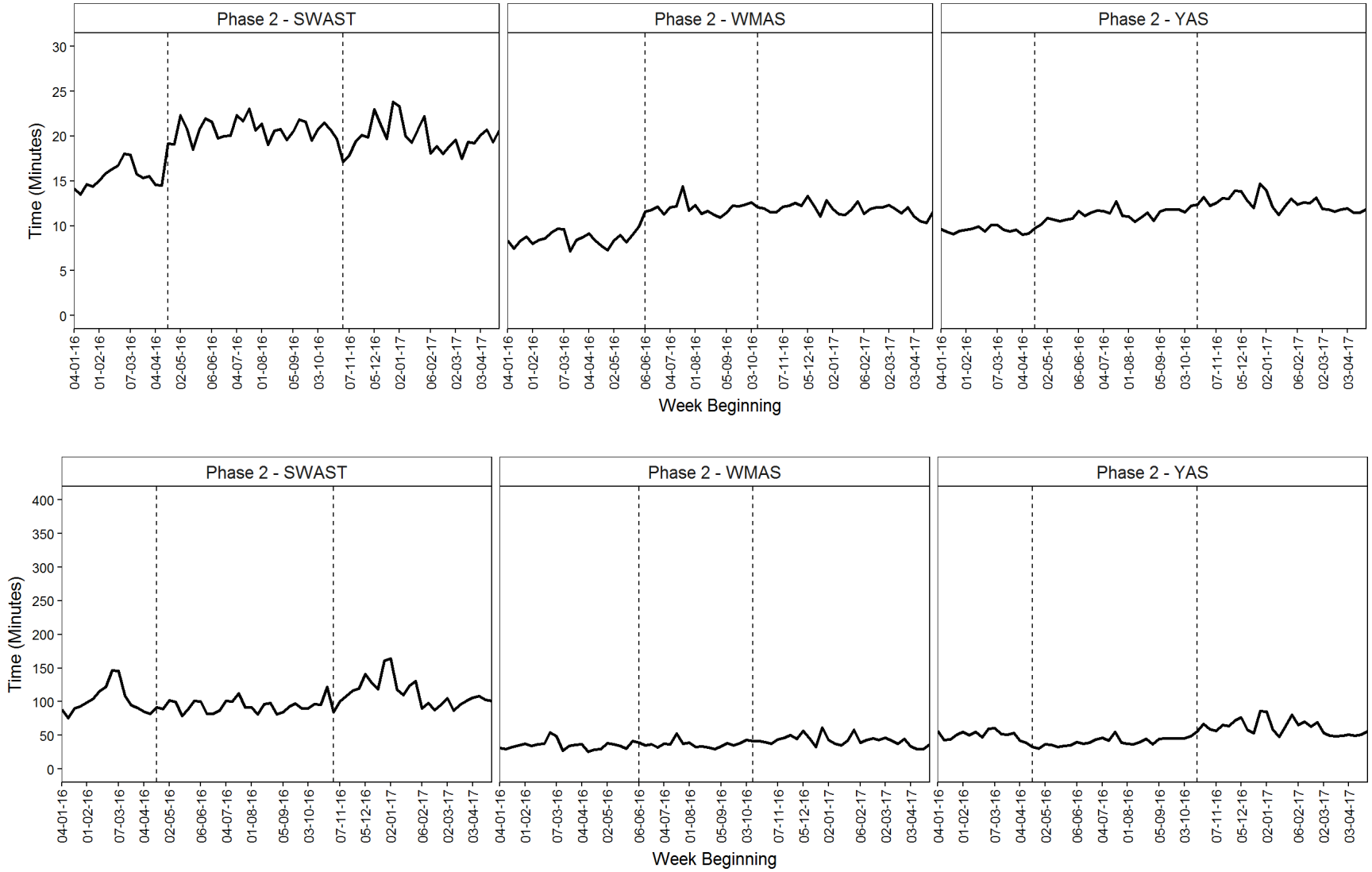


Figure 16: Time from call connect to arrival at hospital - median and 95th percentile

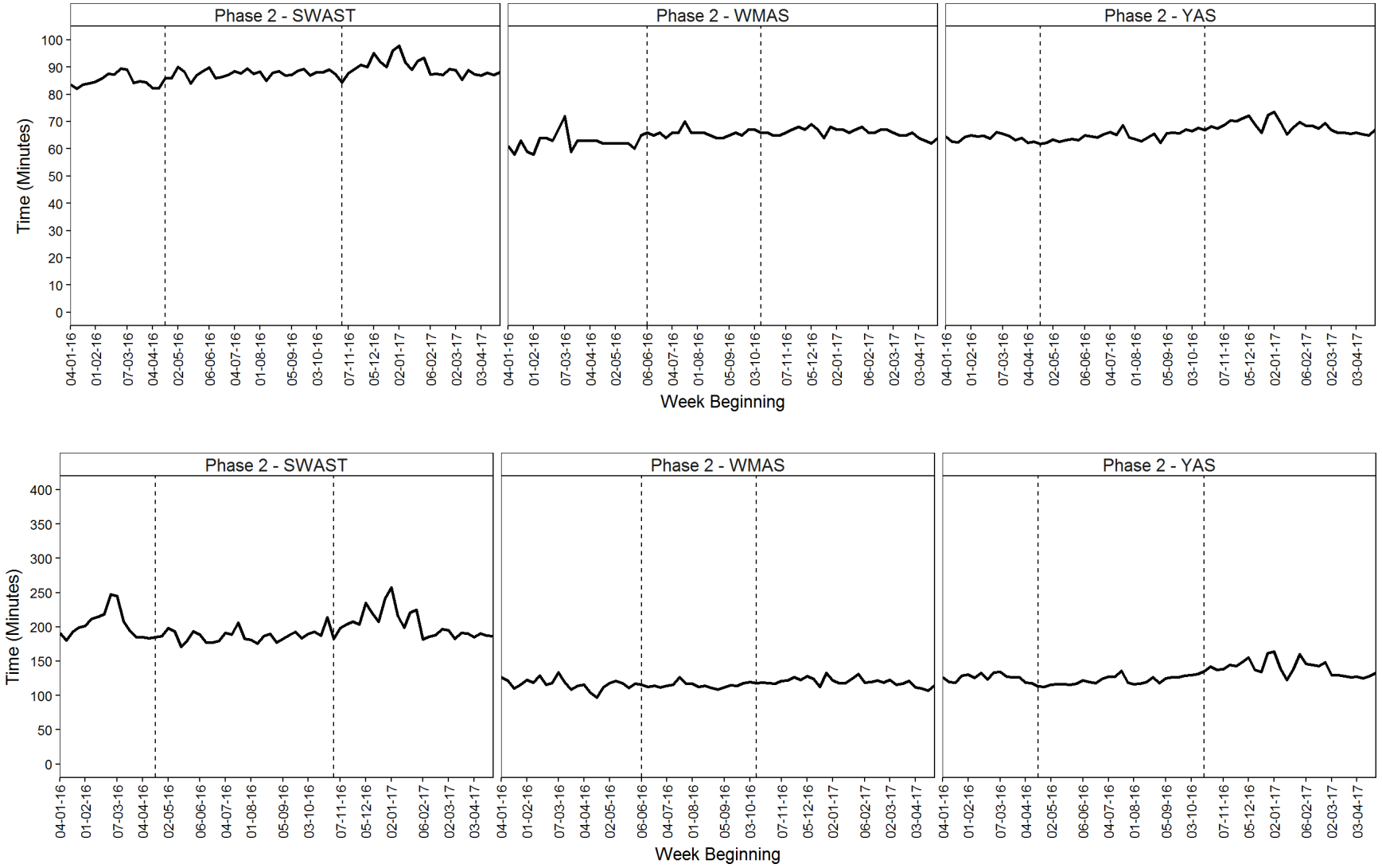
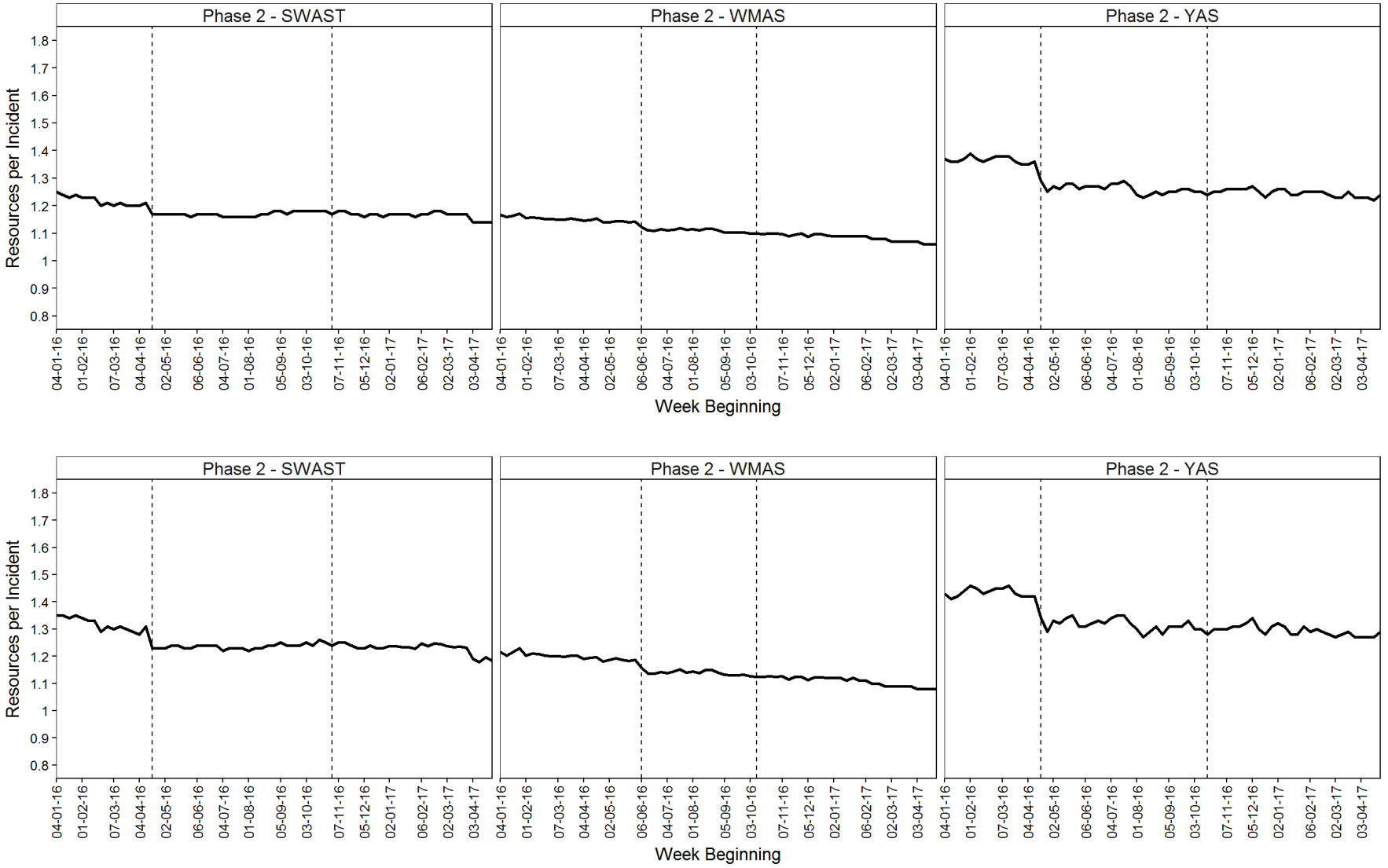


Figure 17: Average care resources per incident – all 999 incidents and all transported incidents



The key findings of the whole service measures analyses are:

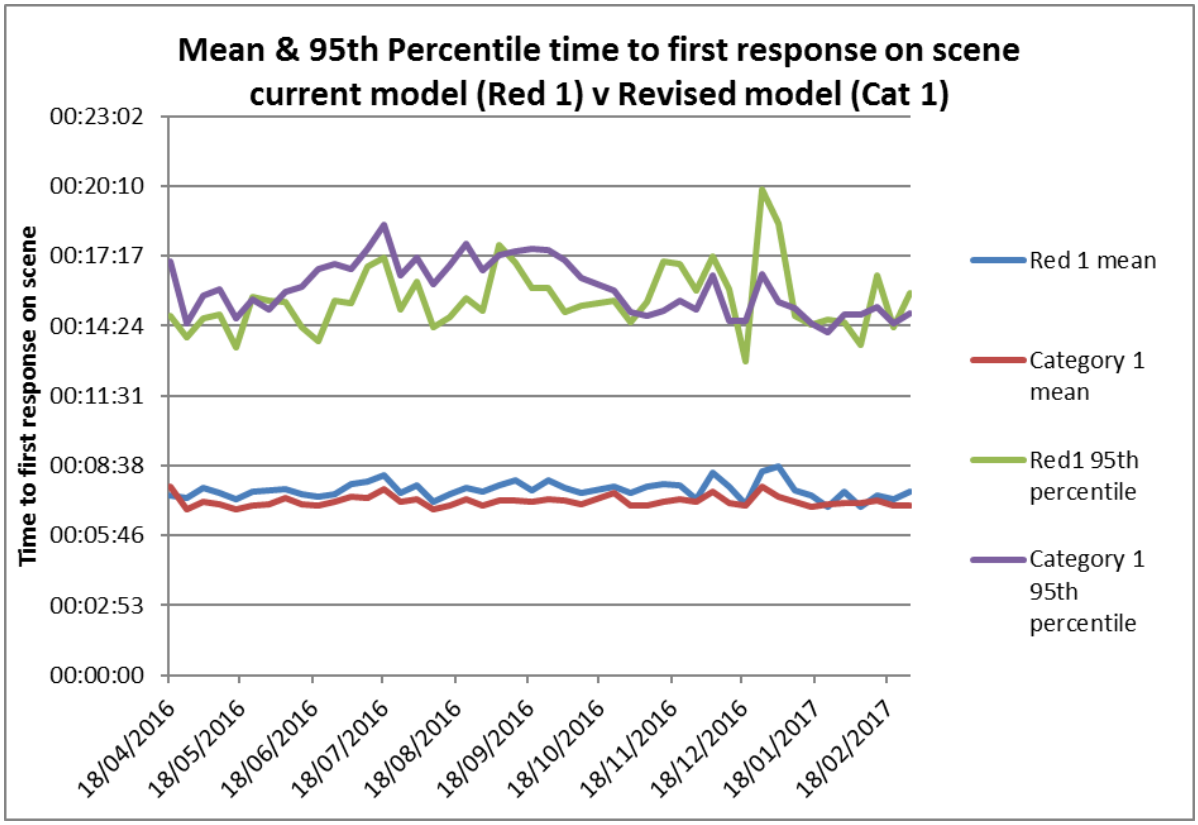
- The median time from call connect to arrival of first resource on scene increased in all 3 services with the implementation of phase 2.1. In phase 2.2 this was reversed to some extent although there was a further increase in YAS. The net effect from changes in both phases was an increase of between 122 and 165 seconds (2-3 minutes). The 95th percentile times are more variable with a substantial reduction in SWAST (22 minutes), a small increase in WMAS (6.6 seconds per week) and a larger increase in YAS (14.7minutes). However these results should be treated with caution as the differences identified in the analysis are not reflected as clear trends in the time series graphs and the values at the extreme ends of the response time curve are subject to much greater variation which will influence results.
- Median time from call connect to arrival at hospital showed small reductions in SWAST and WMAS and no change in YAS. For the 95th percentile there was a substantial reduction in SWAST although the same limitations apply as described above.
- Median time from call connect to leaving scene was increased in all 3 services by 3-4 minutes in Phase 2.1 although this trend began to reverse in one service (SWAST) during Phase 2.2. There were further reductions in 95th percentile times in 2 services.
- There is evidence of further efficiency gains as there was a net reduction in allocation of core resources in all 3 services for both all attended incidents and all conveyed incidents. There were small increases in resource allocation in 2 services for non-conveyed incidents.

The step change increases in median time from call connect to arrival of first resource and leaving scene are unsurprising as, with the introduction of phase 2, the proportion of calls requiring an 8 minute response reduced from around 50% to between 6 and 11% with more flexibility around response interval for other call categories. Given this, the median increases in response time for all 999 calls of 2 – 3minutes and leaving scene of 4 minutes or less are very modest when compared to phase 1 illustrating that although the potential response intervals for the majority of calls can be longer, the service delivered to the whole 999 population only changes by a small amount. The reductions in 2 services in time from call connect to arrival at hospital and no increase compared to phase 1 in the 3rd service show that, despite the greater response interval flexibility the service was improving for patients compared to phase 1 and likely reflects better allocation of the right resource and hence reduced waits for patients where back up conveying vehicles are required. The majority of changes occurred following the introduction of phase 2.1 with performance maintained after the introduction of phase 2.2. The graphs in Figures 15-17 show stable performance following the introduction of phase 2.

The phase 1 trial showed substantial efficiency gains following the introduction of Dispatch on Disposition and it might be assumed that this intervention would produce the biggest gains. However, the whole service analyses have shown that the introduction of phase 2 has produced further efficiency gains, particularly after the introduction of phase 2.1 when the proportion of 999 calls requiring an 8 minute response was considerably reduced. The average reduction in resource allocation across all 3 services for all incidents was 0.038 per incident. Using the same figure for national weekly incidents used in phase 1 (118448) this equates to an estimated 4,501 additional resources available for response per week across England.

Although the adjusted trends measured in the phase 2 trial sites cannot be compared directly to the Phase 1 sites, for comparison Figure 18 shows the mean and 95th percentile time to first response on scene for Red 1 calls in the current phase 1 sites and Red/Category 1 calls in the 3 phase 2 trial sites.

Figure 18: Mean and 95th percentile times for first resource on scene – phase 1 and phase 2 sites.



The category 1 group is over twice the size of Red 1 but shows stable performance with mean Category 1 calls consistently slightly shorter than Red 1 and a clear trend of reducing response times for the 95th percentile, particularly after the introduction of phase 2.2.

The relative whole service performance stability emerging from the combined DoD and revised call category initiatives in the Phase 2 trial sites suggests that the more flexible approach to call assessment, resource dispatch and response intervals may be helping to reduce further deterioration in performance. Longer response time expectations aligned to the revised call categories does not translate in to substantially longer waits for a response form a 999 population perspective.

Comparison of response and call times for urban and rural areas

One criticism of the current, 8 minute target driven operating model is that this may disadvantage patients living in rural areas. Services may concentrate their resources in urban areas where there is highest demand and short distances so they can maximise the number of calls attended in 8 minutes. We have assessed whether ARP has produced any effects on response intervals for

different types of locality in the 3 Phase 2 trial services. Each service provided weekly data for the period January 2016 – April 2017 for 6 time intervals for all attended incidents;

- Mean and 95th percentile call connect to arrival of first core resource
- Mean and 95th percentile call connect to arrival at hospital (see and convey)
- Mean and 95th percentile call connect to leaving scene (see and treat)

This data was also categorised as:

- Predominantly urban (PU)
- Urban with significantly rural (USR)
- Predominantly rural (PR)

For each of the 6 measures we have compared predominantly urban with urban with significant rural and with predominantly rural and repeated this for each of the Phases 1, 2.1 and 2.2 to identify any changes after each phase was introduced. The analyses were adjusted for total number of calls and incidents, hours lost at hospital and seasonality. The full results of the statistical analysis are presented in the supplementary file. Table 15 summarises the difference in times between phase 1 and phase 2.2 only as this model is the likely national model.

Table 15: Urban versus rural analyses - summary of phase 1 and phase 2.2. changes in key response performance measures

Measure	SWAST	WMAS	YAS
Call connect to first resource on scene			
USR v Predominately Urban **Median**			
Phase 1	USR > PU 144 seconds	USR > PU 35 seconds	No difference USRvPU
Phase 2.2.	USR > PU 260 seconds	USR > PU 30 seconds	PU > USR 29 seconds
95th percentile			
Phase 1	PU>USR 1260 seconds	PU>USR 278 seconds	No difference USRvPU
Phase 2.2.	No difference between USR and PU	PU>USR 232 seconds	PU > USR 890 seconds
Predominately Rural v Predominately Urban			
Median			
Phase 1	PR>PU 313 seconds	PR>PU 41 seconds	PR>PU 50 seconds
Phase 2.2.	PR>PU 350 seconds	PR>PU 65 seconds	PR>PU 29 seconds
95th percentile			
Phase 1	PU>PR 1153 seconds	PU>PR 322 seconds	PR>PU 609 seconds
Phase 2.2.	PU>PR 834 seconds	PU>PR 349 seconds	PU>PR285 seconds
Call connect to arrival at hospital			
USR v Predominately Urban **Median**			
Phase 1	No difference USRvPU	USR > PU 322 seconds	No difference USRvPU
Phase 2.2.	USR > PU 204 seconds	USR > PU 353 seconds	PU > USR 84 seconds
95th percentile			
Phase 1	No difference USRvPU	USR > PU 247 seconds	No difference USRvPU
Phase 2.2.	No difference USRvPU	USR > PU 370 seconds	PU > USR 653 seconds
Predominately Rural v Predominately Urban			
Median			
Phase 1	PR>PU 703 seconds	PR>PU 150 seconds	PR>PU 817 seconds
Phase 2.2.	PR>PU 842 seconds	PR>PU 259 seconds	PR>PU 800 seconds
95th percentile			
Phase 1	PR>PU 781 seconds	PR>PU 354 seconds	PR>PU 1474 seconds
Phase 2.2.	No difference PRvPU	PR>PU 481 seconds	PR>PU 782 seconds

Measure		SWAST	WMAS	YAS
Call connect to leaving scene				
USR v Predominately Urban	**Median**			
	Phase 1	PU>USR 538 seconds	USR > PU 142 seconds	PU>USR 182 seconds
	Phase 2.2.	PU>USR 346 seconds	PU > USR 201 seconds	No difference USR vPU
	95th percentile			
	Phase 1	PU>USR 2327 seconds	No difference USRvPU	PU>USR 597 seconds
	Phase 2.2.	PU>USR 1057 seconds	PU > USR 457 seconds	PU>USR 1213 seconds
Predominately Rural v Predominately Urban				
	Median			
	Phase 1	PU>PR 206 seconds	No difference PRvPU	No difference PRvPU
	Phase 2.2.	PU>PR 198 seconds	PU>PR 150 seconds	PR>PU 142 seconds
	95th percentile			
	Phase 1	PU>PR 2469 seconds	No difference PRvPU	No difference PRvPU
	Phase 2.2.	PU>PR 2172 seconds	PU>PR 381 seconds	PU>PR 314 seconds

Figure 19: All Incidents - Call Connect to arrival of first core resource – median time

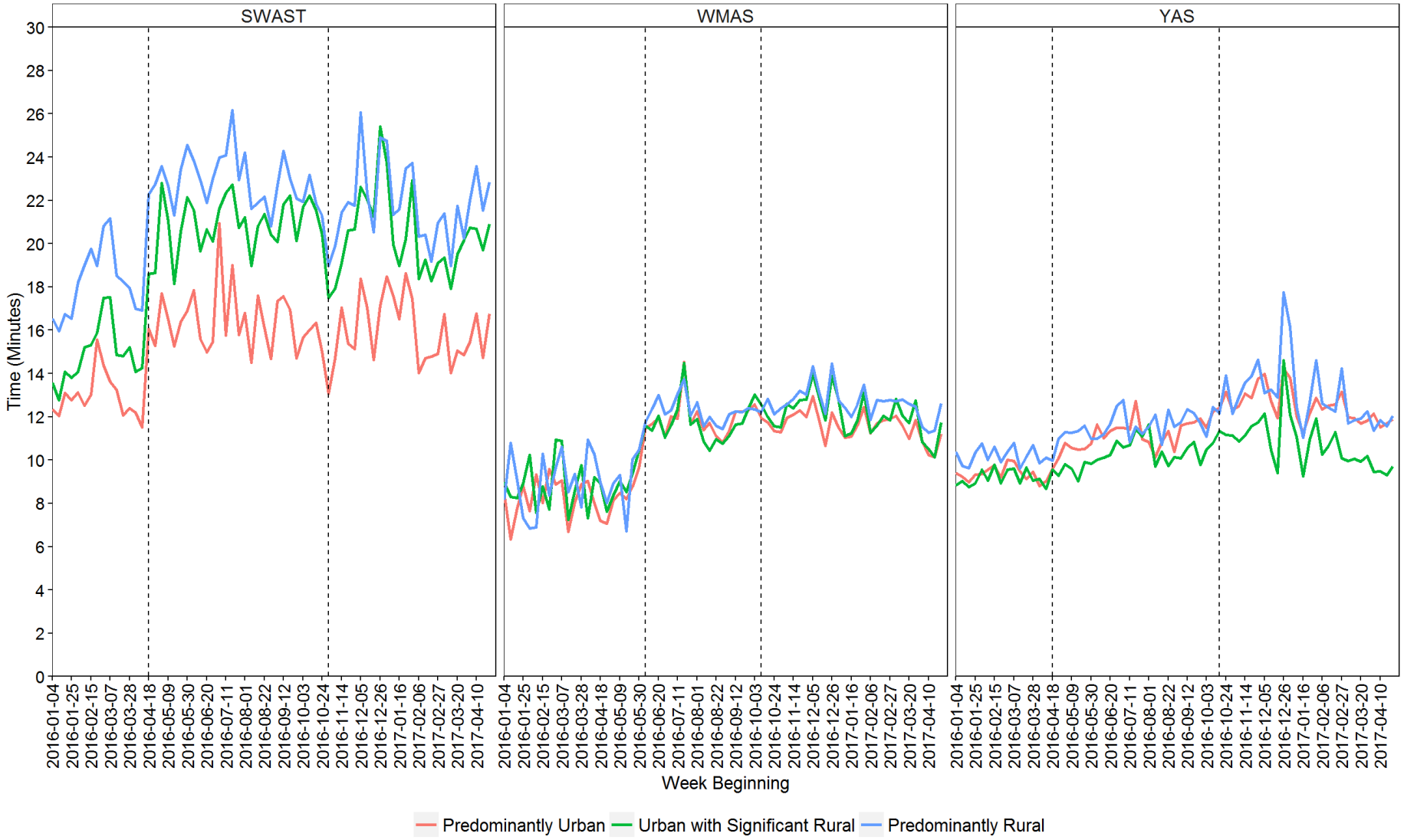


Figure 20: All Incidents - Call Connect to arrival of first core resource – 95th percentile

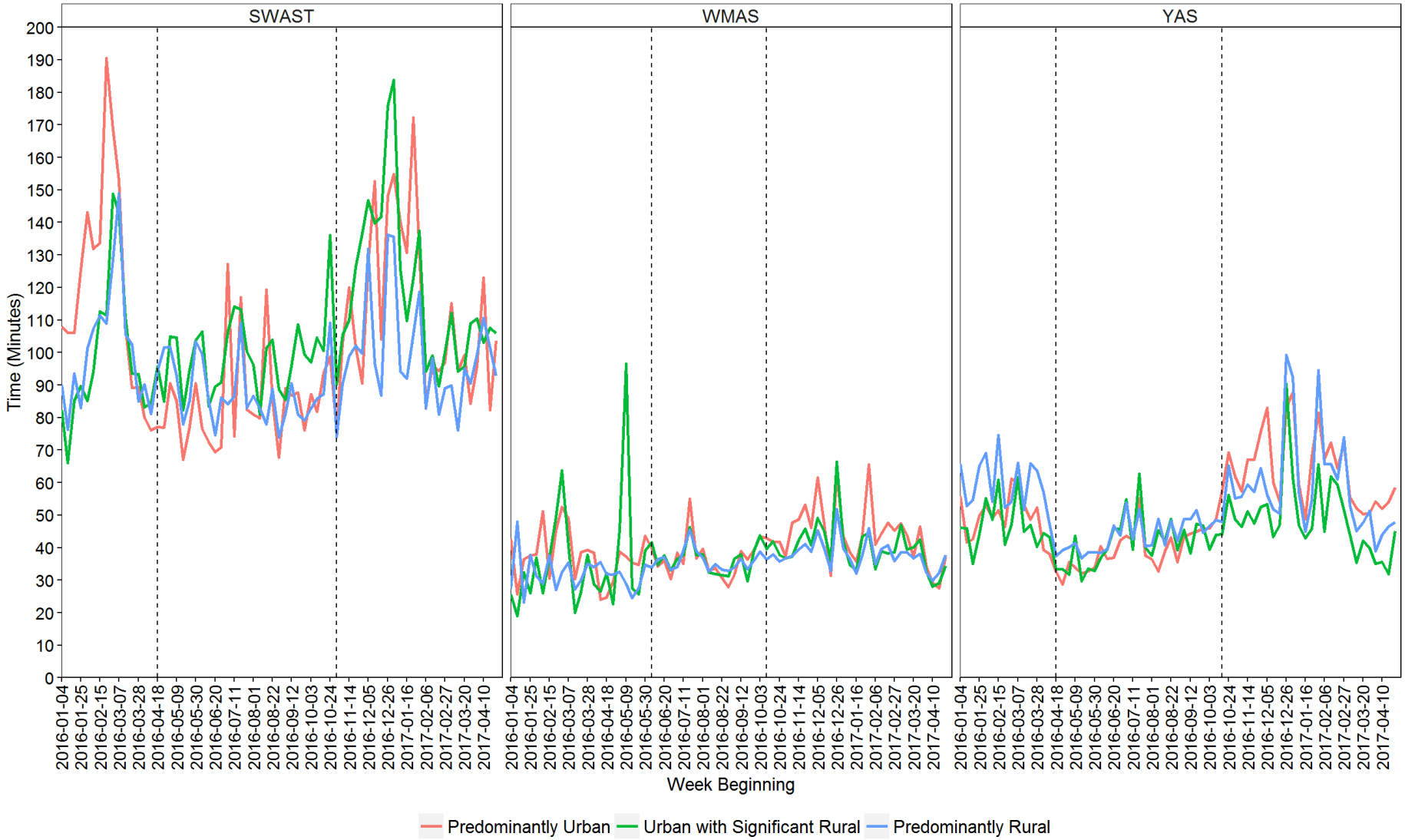


Figure 21: Call Connect to arrival at hospital – median time

Figure 22: Call Connect to arrival at hospital – 95th percentile

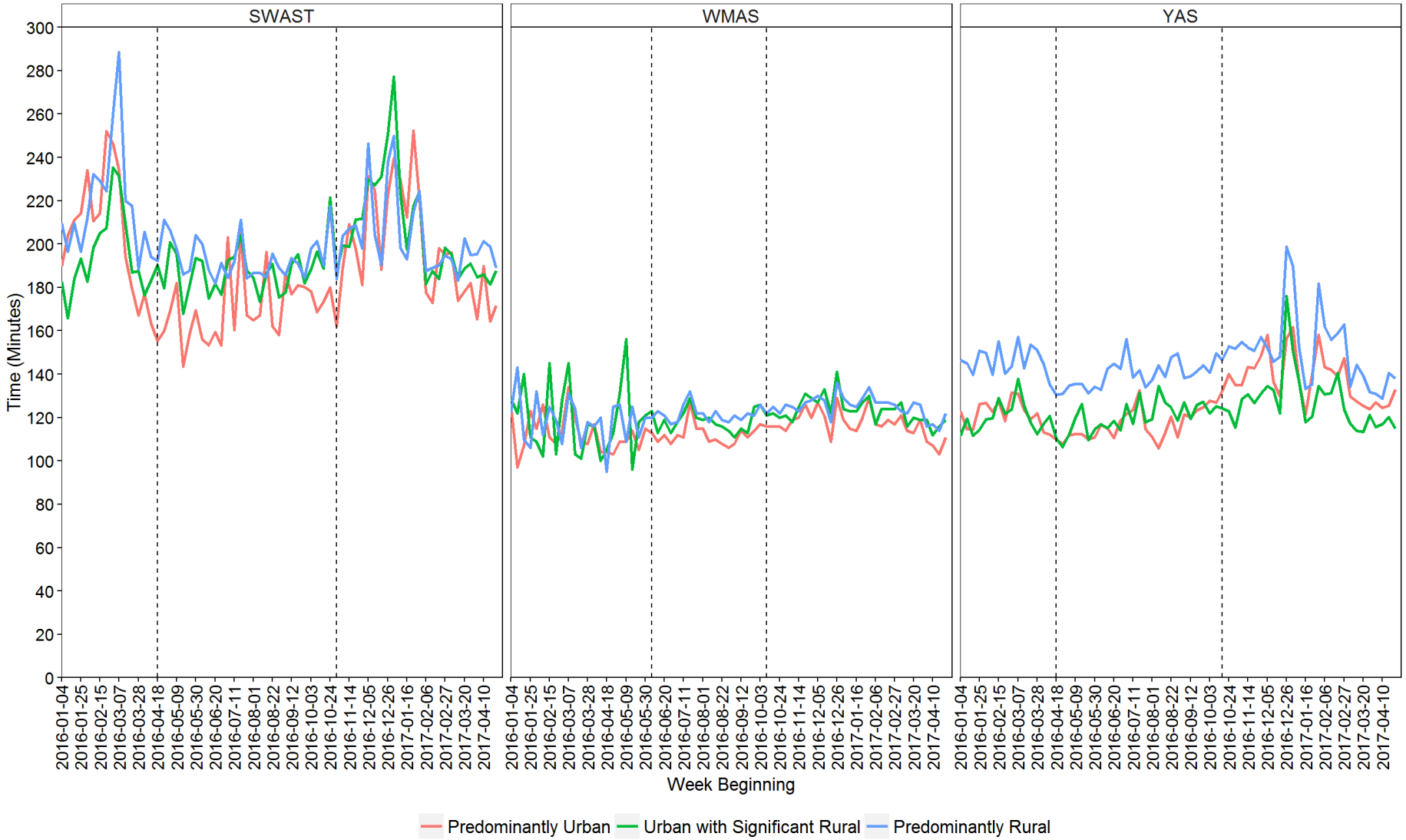
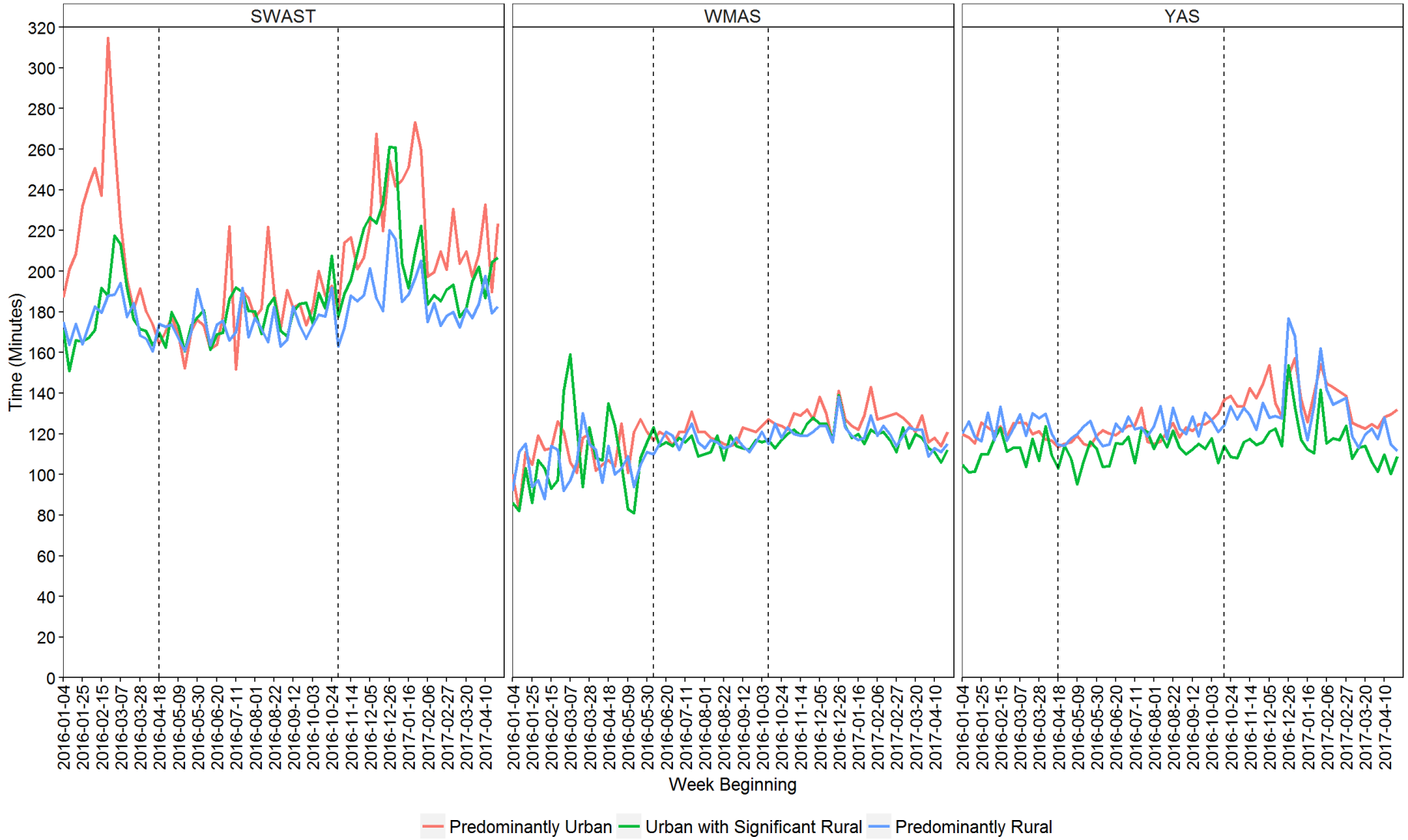


Figure 23: Call Connect to leaving scene– median time

Figure 24: Call Connect to leaving scene– 95th percentile



Call connect to arrival of first core resource

In WMAS and SWAST median time to arrival of first core resource was longer in USR areas than urban areas in both phases. In SWAST the USR time increased by 144 seconds after phase 2.2 was introduced whereas in WMAS this reduced by 5 seconds. For the 95th percentile times both services had longer times in urban areas than USR areas in phase 1 but within phase 2.2 the differences reduced becoming non-significant in SWAST. In contrast, in YAS for both median and 95th percentile times there was no difference between USR and urban times in phase 1 but an increase in time for urban areas by 29 seconds and 14 minutes respectively in Phase2.2.

For the predominantly rural versus predominantly urban comparison, median time was longer in rural areas than urban areas in both phases in all 3 services. Rural times increased in 2 services (by 37 and 23 seconds) and decreased in one service (by 20 seconds) during phase 2.2. For the 95th percentile, times were longer in urban areas than rural areas in both SWAST and WMAS in both phases with this reducing in SWAST by 319 seconds and increasing in WMAS by 26 seconds in phase 2.2. In YAS the 95th percentile time is 607 seconds longer in rural areas in phase 1 but this is reversed and becomes 285 seconds longer in urban areas in phase 2.2.

Time trends are illustrated in Figures 19 & 20.

Call connect to arrival at hospital (see and convey)

In SWAST and YAS there was no difference in times between USR and urban for both median and 95th percentile times in phase 1. In phase 2.2 median USR times increased in SWAST but became non-significant for the 95th percentile time whereas in YAS both median and 95th percentile times became significantly longer (84 and 653 seconds respectively) for urban areas compared to USR. In WMAS USR times were longer than urban times in both phases and increased in phase 2.2 by a median 30 seconds and 123 seconds for the 95th percentile time.

Both median and 95th percentile times were significantly longer in rural areas than urban areas in all 3 services for phase 1. During phase 2.2 in SWAST median rural times increased by 139 seconds but the 95th percentile time showed no significant difference between rural and urban areas, in WMAS median and 95th percentile times in rural areas increased by 109 and 127 seconds respectively whereas in YAS median and 95th percentile times in rural areas reduced by 17 and 692 minutes respectively (Figures 21&22).

Call connect to leaving scene (see and treat)

Urban times were longer than USR for both the median and 95th percentile during phase 1 in SWAST and YAS. For phase 2.2 in SWAST median and 95th percentile urban times reduced by 192 seconds and 1270 seconds (21 minutes) respectively and in YAS although the difference became non-significant for the median time the 95th percentile time showed an increase of 616 seconds in urban areas. In WMAS the median time was longer in USR areas in phase 1 but this reversed in Phase 2.2 and urban times were longer than USR by 201 seconds. Similarly for the 95th percentile a non-significant difference between USR and urban in phase 1 became a significantly longer time in urban areas in phase 2.2 by 457 seconds.

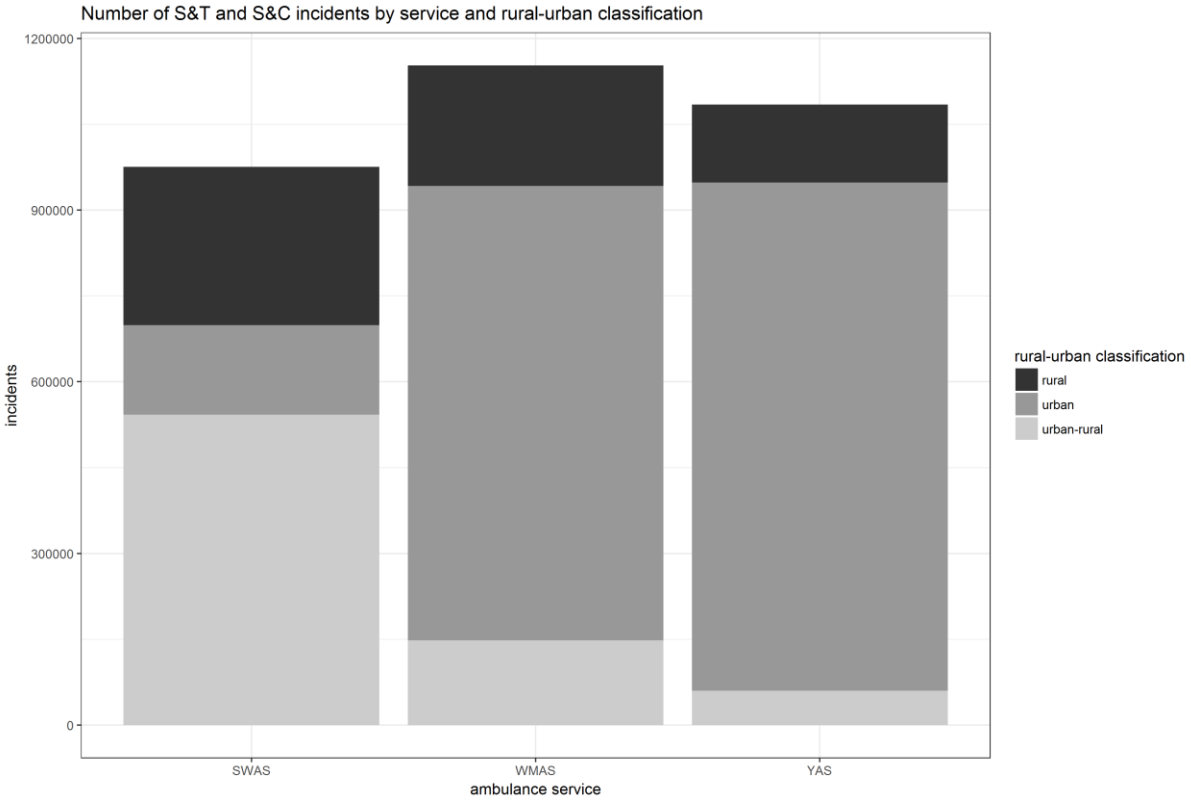
There was no difference between rural and urban areas for both median and 95th percentile times in WMAS and YAS during phase 1. In WMAS median and 95th percentile times increased in urban areas in phase 2.2 (150 and 381 seconds respectively) and in YAS median times were longer in rural areas but this changed to longer urban times (314 seconds) for the 95th percentile. In SWAST times were longer in urban areas than rural areas in both phases but decreased in phase 2.2 by a median 8 seconds and 297 seconds for the 95th percentile (Figures 23&24).

The analysis presents a complex picture both in terms of changes following the introduction of phase 2.2 and differences between services. The main purpose of this analysis was to assess whether there were any reductions in differences between geographical areas in services but this complexity means there are no clear and obvious trends across all 3 services. The most consistent pattern of reduction in differences was in SWAST, particularly for 95th with a closing of differences across all 3 measures. WMAS maintained a consistent performance across phases and although there were some increases in time these were small (typically less than 2 minutes for median and 5 minutes for 95th percentile times) which are small within a framework of flexible response for categories other than Category 1. In YAS there were reductions in rural and USR times for some categories but this was sometimes at the expense of increased times in urban areas. The sometimes substantial changes for 95th percentile times in terms of reducing differences may indicate that, although for the 50% of shortest times there may be some increases, for the bigger population of 999 callers the introduction of the revised call categories may be helping services to better manage the overall 999 population and, in some cases, reducing long waits although this is not consistent across every measure. There was also a reduction in the 95th percentile time to arrival at hospital in 2 services for predominantly rural calls. These calls are likely to have long transfer times which the ambulance service cannot control so, this is a real gain given distances to hospital are fixed. The changes made to call assessment and dispatching processes, better allocation of the “right” resource, more response interval options and a decreased emphasis on the 8 minute target for a substantial proportion of incidents appear to be helping develop a more equitable service.

An interesting feature is that for a substantial number of measures times were longer in predominately urban areas than in mixed or rural areas in all 3 services and this is particularly evident for 95th percentile times. There has been an assumption that response performance in rural areas is consistently longer than in urban or mixed urban and rural areas and this analysis shows this inequity may not be as clear as first thought . Without more detailed investigations it is difficult to understand why these differences occur. It is also possible that other factors come in to play, for example, as rurality increases so population density and associated demand decrease and although distances may be shorter in urban areas increasing demand and congested road networks may erode any perceived advantages in these areas. Performance in different geographical areas may be dependent on the relative proportions of calls deriving from each type of area within individual services. Figure 25 shows the proportions of calls for each type of geographical area within the 3 study services. There are some clear differences between the 3 services. There is a much higher proportion of USR and rural areas within SWAST than the other two services whereas YAS has a substantially larger proportion of incidents originating from urban areas. This may partly explain why there is a consistent pattern of higher times in urban areas in this service as 82% of their workload is generated from urban areas. Relatively small volumes of incidents are generated from USR and rural areas so changes in these areas will be more apparent. Nevertheless in two services there was an overall improvement in performance for calls originating in predominantly rural areas both in ternms

of response and arrival at hospital. What is clear is that, whilst the ARP initiatives may help reduce some inequities, differences in performance in different types of geographical areas are likely to be influenced by a range of other external factors and more detailed investigation is needed to better understand what these factors are and how they influence service delivery.

Figure 25: Proportions of calls generated in Urban, Urban with significant rural and rural areas in 3 phase 2.2 trial services



3.4 Other ARP Phase 2 related measures

For Phase 2.2 we have briefly updated some of the analyses conducted for Phase 1 to ensure that the objectives of ensuring early identification of the most urgent calls and patient safety have been maintained following the introduction of the revised call categories. Phase 2 represents a substantial change in the response model for 999 calls and has only been in operation for a relatively short period of time with two iterations. To supplement the evaluation we have also drawn on the experiences of the three Phase 2 services and utilised the results of small internal analyses they have conducted and their overall views on the success and challenges associated with implementing this change. We have also repeated the staff survey to assess their views on the changes and the impact on the way they work.

Identification of category 1 and cardiac arrest calls

With the introduction of Phase 2 the NoC descriptor list for NHS Pathways sites was revised to reflect the new call categories and include descriptors that are highly likely to require a Category 1 response given the reduction in the proportion of calls that will be assigned a category 1 (8 minute) response. The current phase 2.2 trial has been running for a relatively short period but early data from one service using NHS Pathways (WMAS) shows that the Category 1 capture rate is comparable to the rate recorded in phase 1 (Table 16). Continued use and review will allow further refinement.

Table 16: Category 1 call capture rates using revised NoC call processes – phase 2

Column1	**Number (%)**
Total Incidents (reported)	15478
Cat 1 NoC	
Choking	309
Arrest/Peri Arrest	4359
Fitting Now	4469
Under 5 + priority Symptoms	2201
Operation Consort	0
PLATO	0
Total Cat 1 identified at C1 NoC	**11338 (73.3)**
Cat 1 identified by other Red NoC	3116 (20.1)
Cat 1 identified by other NoC or unclassified	**1024 (6.6)**

For phase 2.2 the relatively short duration of operation means so far only a few weeks data on cardiac arrest capture rate is available for the two services using NHS Pathways. Table 17 shows the cardiac capture rate for these 2 services. SWAS data includes two NoC descriptors (Breathing Problems – not alert/inefficient and Unconscious – not noisy breathing) to generate a Category 1 response that generate a Category 2 response in WMAS and also include the NoC descriptor for "fitting" whereas WMAS only use "fitting now". WMAS also use a NoC "Under 5 – Priority symptoms".

The cardiac arrest capture rates for phase 2.2 Category 1 calls are comparable to those identified in phase 1 of DoD. The Category 1 rate is slightly higher in SWAST but additional NoC descriptors are used and cardiac arrests coded as cardiac arrest is higher in WMAS. The NoC descriptor "unconscious – not noisy breathing" accounted for an additional 10.5% of cardiac arrests in WMAS. The combined category 1, unconscious – not noisy breathing , chest pain and breathing problems descriptors account for 83.3% and 88.9% of all cardiac arrests in SWAS and WMAS respectively demonstrating consistency with the findings of the phase 1 analysis.

Table 17: Cardiac arrest capture rate using NoC Phase 2.2 categories

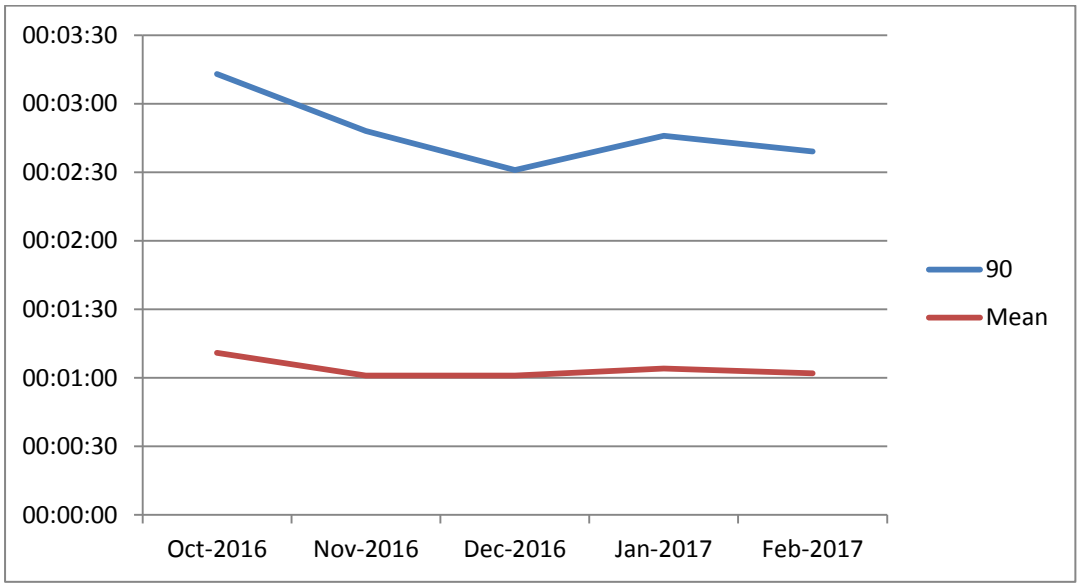
	SWAS (25/10-27/11 2016)	**WMAS (Sept-Oct 2016)**
NoC	**numbers that arrested (% of all arrests)**	**numbers that arrested (% of all arrests)**
Category 1 (C1)		
Arrest/Peri-Arrest	133 (58.6)	384 (67.4)
Choking	2 (0.9)	2 (0.3)
Drowning/Water incident	1 (0.4)	
Fitting now		5 (0.9)
Fitting	2 (0.9)	
Unconscious (NOT Noisy Breathing)	24 (10.6)	
Breathing Probs (NotAlert/Inefficient)	7 (3.1)	2 (0.3)
Under 5 (Priority symptoms)		2 (0.3)
% cardiac arrest identified as C1	**74.5%**	**68.9%**
Cardiac arrest assigned Category 2 NoC	23	97
% cardiac arrest identified as C2	**10.1%**	**16.9%**
Cardiac arrest assigned Category 3 NoC	29	58
% cardiac arrest identified as C3	**12.8%**	**10.2%**
Cardiac arrest assigned Category 4 NoC or other	6	25
% cardiac arrest identified as C4 or other	2.6%	4.4%
Total Cardiac arrests	**227**	**570**

No cardiac arrest data capture rate data is currently available from the service using AMPDS. However YAS have provided some data from their own internal monitoring processes on 1) the time taken to recognise calls with a cardiac arrest AMPDS code during phase 2.2 and 2) a time series of response time data for calls with AMPDS cardiac arrest codes for the period October 2014 (one year prior to the introduction of ARP) to February 2017.

Figure 25 shows the mean and 90th percentile times for T0 (call connect) to T5 (triage completed and AMPDS determinant) for all calls with an AMPDS determinant of cardiac arrest for phase 2.2.

This data is only for a short period of time but shows that during phase 2.2 there was a modest decrease in mean time to establish a cardiac arrest determinant from 1 minute 11 seconds in October 2016 to 1 minute 2 seconds in February 2017. The 90th percentile time shows a more substantial reduction from 3 minutes 13 seconds to 2 minutes 39 seconds over the same time period. YAS have implemented pre-triage questions in to the call assessment process and a programme of EMD support and training which appears to be producing benefits in early recognition of cardiac arrest calls.

Figure 25: Mean & 90th percentile times for T0 to T5 AMPDS triaged cardiac arrest calls -Phase 2.2.



For the response time series we have plotted the weekly mean and 90th percentile response times to calls with cardiac arrest AMPDS determinants for the period October 2014-February 2016 (Figures 26 and 27) and analysed the trend across the whole time period adjusted for total number of incidents, hours lost at hospital and seasonality. The time series graphs suggest an upward trend in response time performance after the introduction of Phase 2. The adjusted analysis shows that there were no step changes or slope (trend) changes in response time performance after Phase 2.1 or 2.2 were introduced indicating that, despite increases in demand and reduced resource availability from increasing hospital delays, there has been no degradation in the timeliness of response to the most urgent calls after the introduction of ARP initiatives in this service.

Figure 26: Mean response time to calls with AMPDS cardiac arrest determinants Oct 2014-February 2017

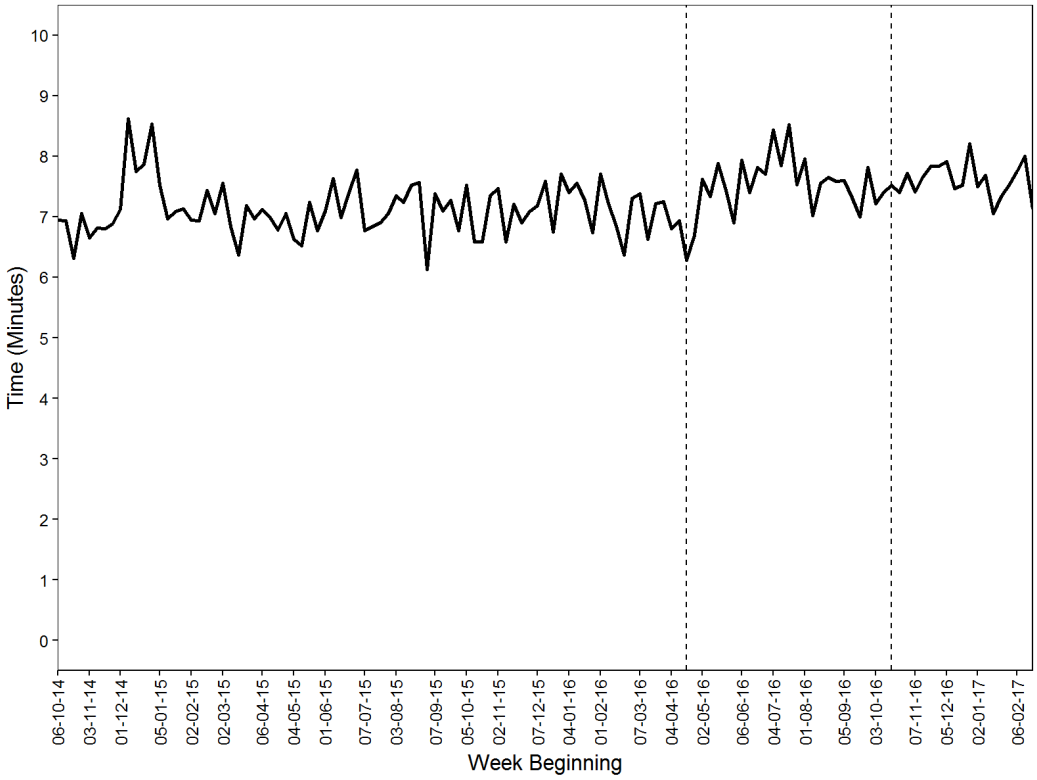
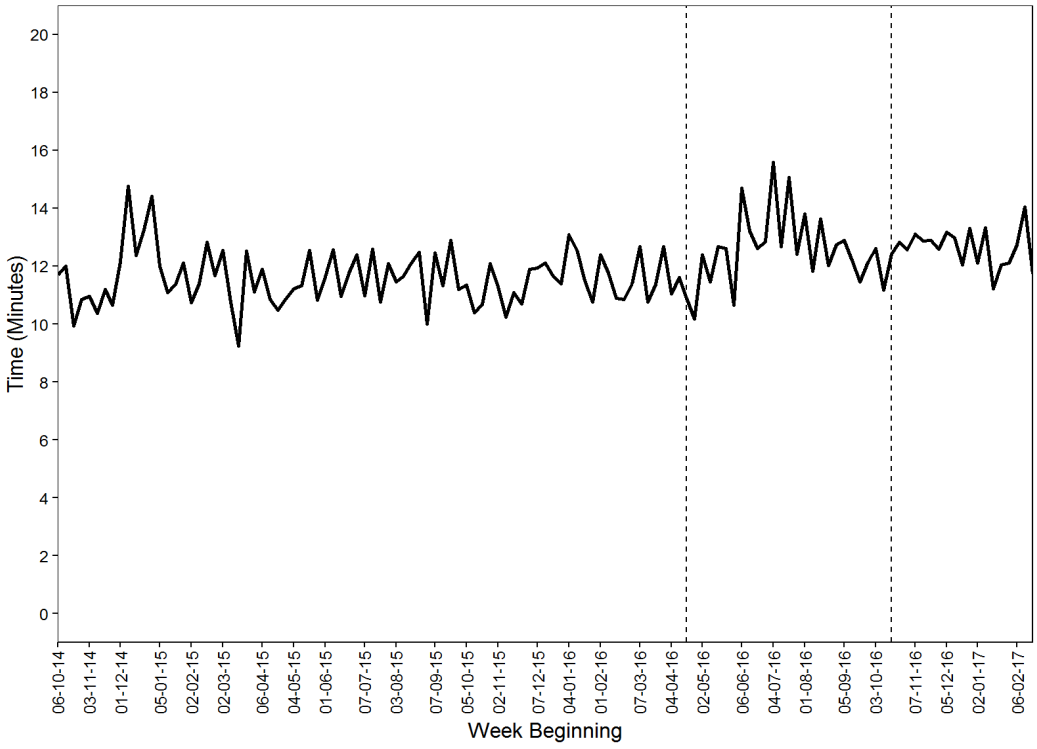


Figure 27: 90th percentile response time to calls with AMPDS cardiac arrest determinants Oct 2014-February 2017



For phase 2 there has been insufficient time to accumulate cardiac arrest outcome data. The most recent information available from SWAST shows an overall ROSC rate of 17.6%. For calls identified as arrest or peri-arrest the ROSC rate is 14.3% which is higher than that shown in the phase 1 analysis. The highest ROSC rates are for NoC descriptors of chest pain and breathing problems however with only 4 weeks information available the numbers of cases within descriptors is too small to make any inferences about changes in ROSC rates after the introduction of the new call categories.

Clinical outcomes

An objective of ARP is to improve clinical outcomes for patients. However the relatively short duration of Phase 2 means that it is difficult to establish impact on specific clinical outcomes. At present 3 clinical quality indicators (CQI) are currently measured (cardiac arrest, stroke and STEMI) and of these only cardiac arrest has a patient outcome (ROSC and survival). Monthly case numbers for these conditions are small and at individual service level there is baseline variability in performance on a month to month basis so changes between phase 1, 2.1 and 2.2 that are a consequence of ARP rather than other factors are difficult to detect within such short time periods. There is also a time lag in the publication of CQI data so only limited data is available for Phase 2.2. Figures 23 – 26 illustrate trends in CQIs for the 3 Phase 2 services and England from the start of the ARP evaluation time period (October 2014) up to the most recent available data (January 2017) https://www.england.nhs.uk/statistics/statistical-work-areas/ambulance-quality-indicators/ambulance-quality-indicators-data-2016-17/

Figure 28: Clinical quality indicator performance October 2014-October 2016 – WMAS

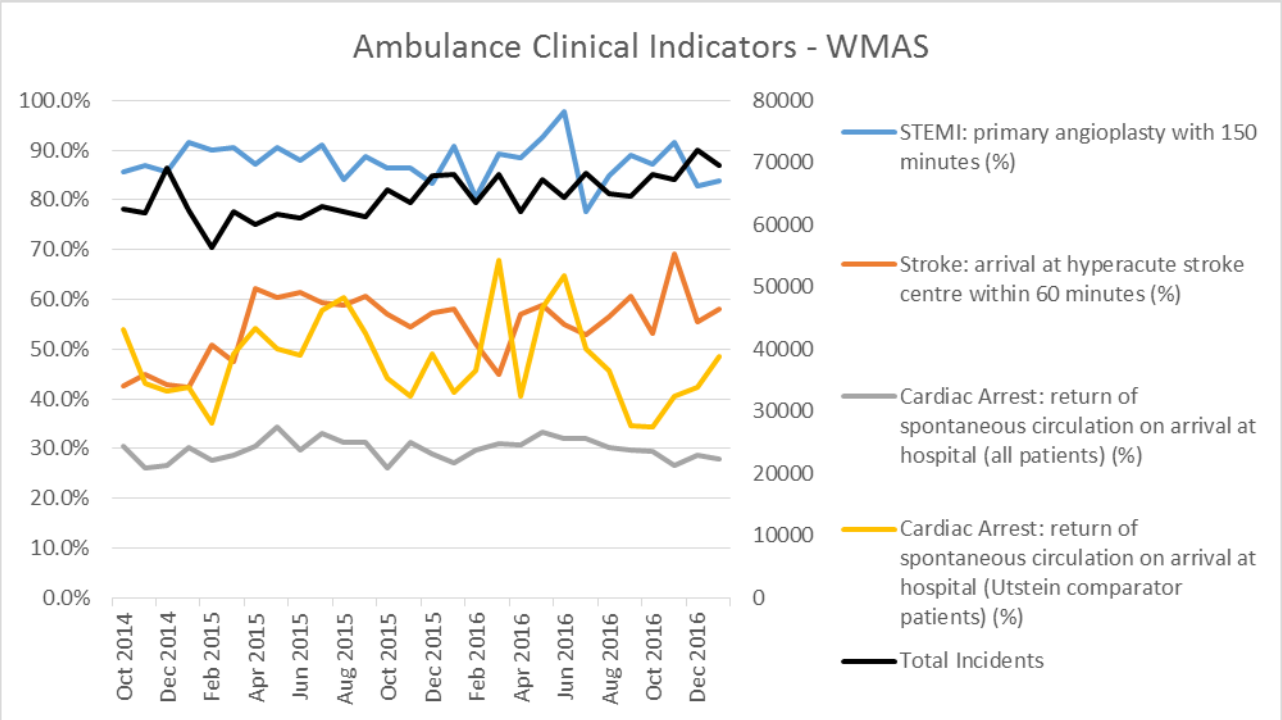


Figure 29: Clinical quality indicator performance October 2014-October 2016 – SWAST

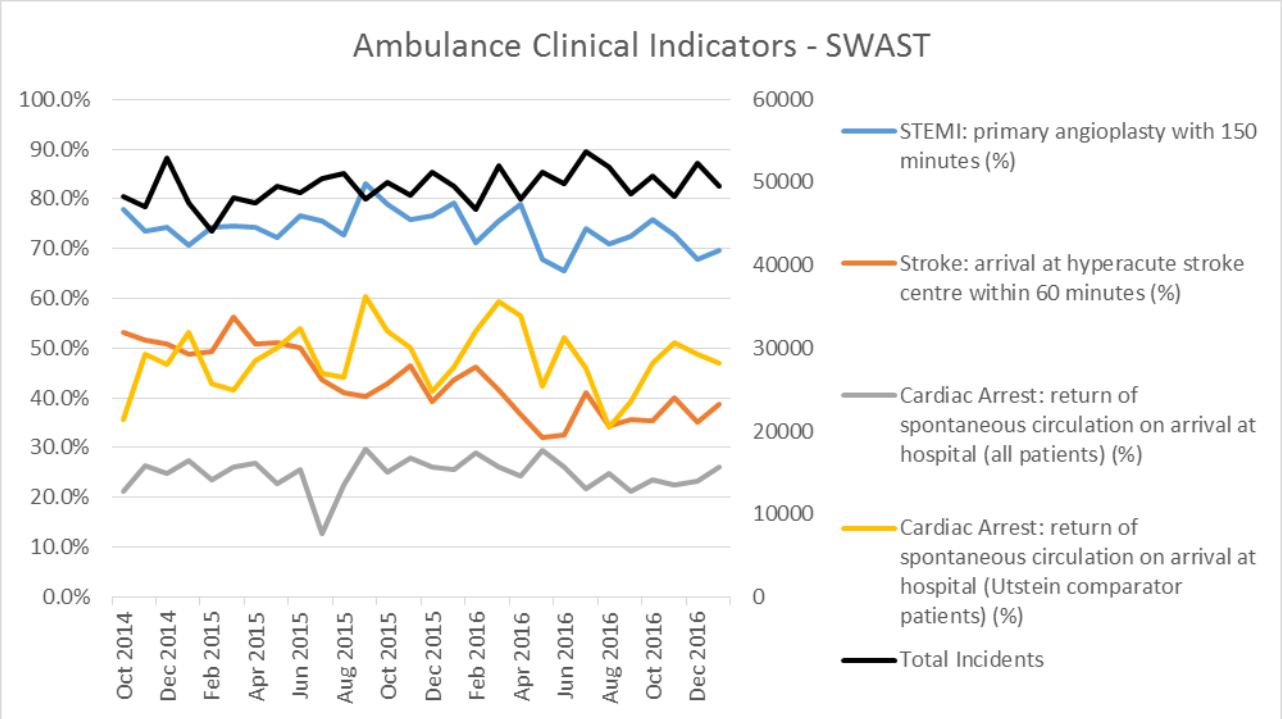


Figure 30: Clinical quality indicator performance October 2014-October 2016 – YAS

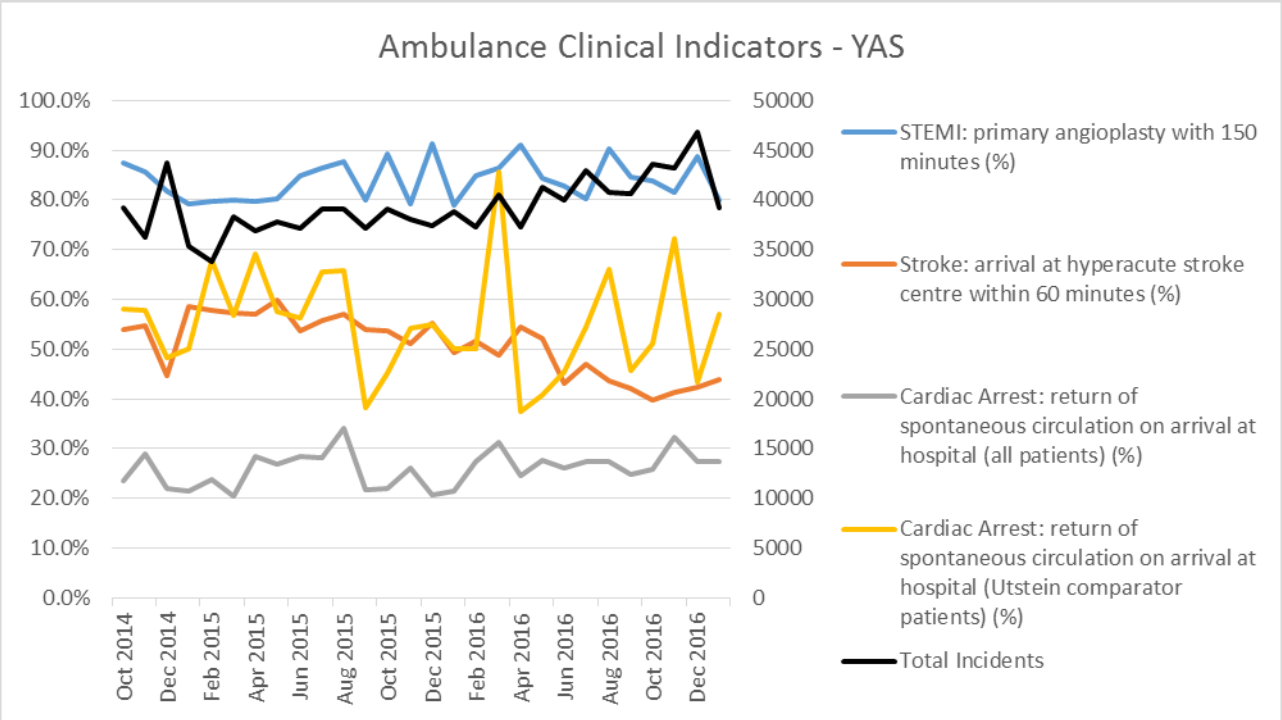
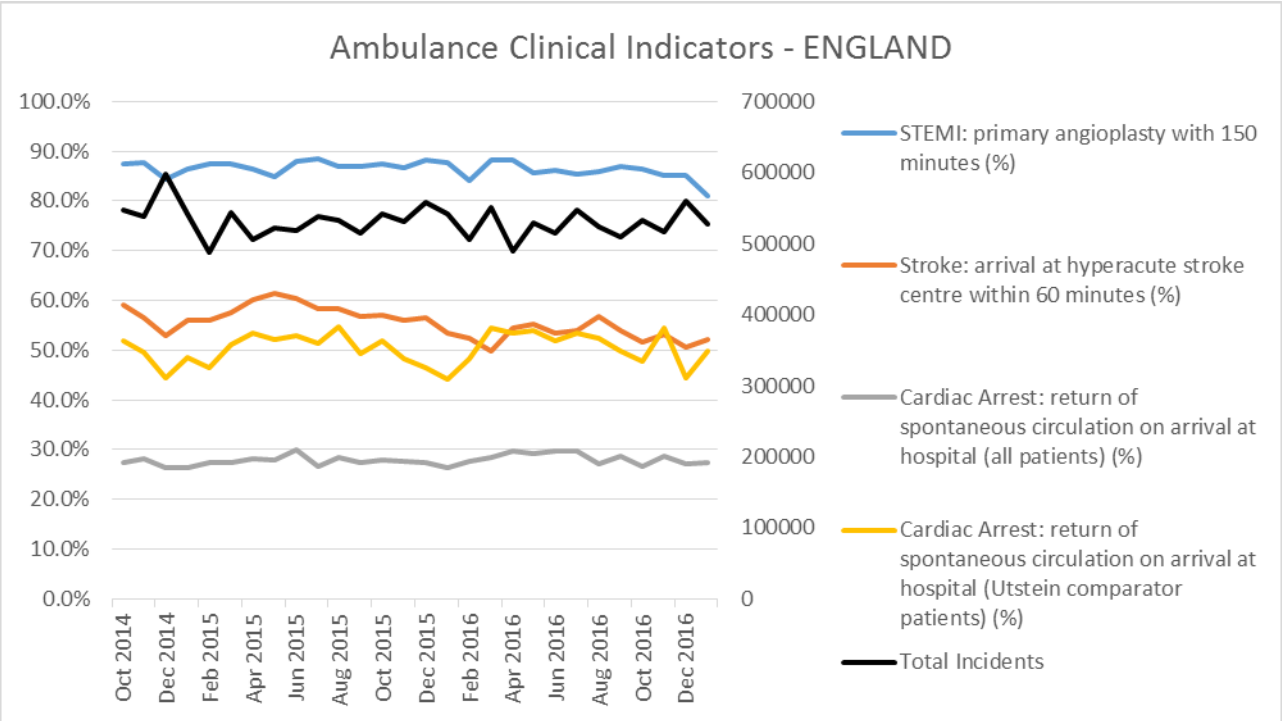


Figure 31: Clinical quality indicator performance October 2014-October 2016 – England



The CQI's show unadjusted performance and the interaction between CQI performance and demand is clear in the graphs. For stroke there is a downward trend in the proportion of patients arriving at a hyper-acute stroke unit (HASU) in 2 services but this is also reflected in the national picture although there is an improving trend in WMAS despite increases in demand. There may also be other factors influencing performance such as changes in HASU provision which are outside the control of ambulance services. For STEMI patients receiving primary angioplasty there is relationship between demand and performance but in YAS and WMAS the trend is for closer alignment between demand and performance suggesting the effects of demand alone are to some extent being closed. ROSC rates for all patients are stable in all services. The changes in ROSC rates for Utstein comparator patients show substantial monthly variation which is probably a consequence of very small numbers of monthly calls at an individual service level which makes random variation more obvious. Given this natural volatility and the small number of months included since the introduction of phase 2 it is not possible to determine any real change in trends for this indicator. A longitudinal study over a much longer period and adjusted for factors such as demand will be needed in order to determine any effects related to the implementation of ARP on clinical outcomes.

Patient Safety and adverse incidents

As in Phase 1, the 3 Phase 2 services have closely monitored patient safety by scrutinising all adverse incidents reported via datix and conducting regular reviews of all calls where there are long waiting times for an ambulance response across all of the revised categories. During both Phase 2.1 and 2.2 there has not been any identified adverse incident or patient safety concerns associated with the ARP changes.

Phase 2.2 Individual service trial feedback

In addition to the information supplied from individual services described previously in this section, some have also provided an overview of the perceived positive effects of ARP for their service and the associated challenges.

South Western Ambulance Service NHS Foundation Trust provided the following feedback on their overall experience of ARP (Box 4)

Box 4: South Western Ambulance Service NHS Foundation Trust

Positives

- The clinical coding has made management of the work load clearer for dispatchers which really helps them prioritise limited resource to the most critical patients
- ARP brings greater focus on specific "life changing" conditions e.g. Stroke, STEMI
- The introduction of Response and Transport subcategories helps improve resource utilisation particularly with RRVs
- Clinical Hub staff in the South West are recognising that the aim of ARP is to be more clinically focussed rather than target/performance focussed. This leads to greater satisfaction that they are doing the right thing for the patient rather than the quickest allocation."

Opportunities

- Increase public awareness of the range of ambulance response times, for example, so that there is not an expectation that a resource will arrive in 8 minutes when it is a Category 4 problem.
- Align the categories between different triage solutions more consistently
- Review clock start AQI metrics for calls which can't identify location, as location is necessary for allocation.

West Midlands Ambulance Service NHS Foundation Trust provided a description of some of the changes they have made to support ARP and the impact they have found as an individual service (Box 4).

Box 4: West Midlands Ambulance Service NHS Foundation Trust

Operational Efficiencies

Before WMAS entered into the ARP trial, the Trust deployed a resource split of 32% Rapid response car (RRV)hours and 68% Ambulance hours average across a month. This was a requirement of needing to meet an 8minute standard for over 45% of the total 999 demand at that time. This situation creates a considerable inefficiency where often an RRV and an Ambulance are required at the scene of each incident, despite WMAS having a strong non-conveyance of around 30%. The average response per incident for a typical month in this mode of operation was 1.3.

Under the ARP model of operation, WMAS switched resourcing to an average 10% RRV hours and 90% Ambulance hours across a month, given 6.5% of incidents require an 8 minute response. This provided a much more efficient mode of operation where the average response per incident fell to below 1.1. The Trust required 4% less overall resource and dealt with an additional 10% of demand.

It is also important to note that whilst the focus on Ambulance provision has been more efficient and reduced waiting for patients (because RRVs aren't awaiting delayed backup), it has also led to an increase non-conveyance. These efficiencies have been demonstrated diagrammatically below:

Operating Model Comparison

	Feb-15		Feb-17			
Ambulance Hours	112156		142746			
RRV Hours	54290	32.6% RRV	16739	10.4% RRV		
Total Hours	166446		159485	Hours	-6961	-4.18%
					Less hours in 2017 overall	
Total Incident Demand	69295		76716	Demand	7421	10.71%
					More demand in 2017	
RPI	1.26		1.10			
Non-conveyance %	38.80%		39.30%	Higher non-conveyance 2017		

3.5 Phase 2 Staff surveys

Survey Process

The staff survey conducted for Phase 1 showed that, overall, staff viewed the changes made with the introduction of Dispatch on Disposition as a positive step. With the introduction of Phase 2 we wanted to explore the views of staff on the effects of changes to call categories. We conducted two further staff surveys after the introduction of Phases 2.1 and 2.2. For these surveys we adapted the questions from the Phase 1 survey to reflect the revised call categories. The surveys were administered electronically with questions requiring a simple tick-box response to multiple choice answers with space for free text comments on perceived advantages and disadvantages. Responses to the survey were entered in to the statistical software package SPSS version 19. Free text comments were tabulated and key themes identified from the narrative responses.

Survey results

For the survey conducted after the introduction of Phase 2.1 one service (YAS) had already completed a recent staff survey when the evaluation survey was launched. Their survey was based on the previous one described above and the results were made available to us so we have included these where the same questions were asked. There are some questions where results can only be given for the other 2 services.
A total of 687 staff responded (378 in SWASFT, 66 in WMAS, 243 in YAS). Of these 568 (85.2%) were operational staff and 119 (14.8%) were EOC/Clinical Hub staff.
There were a much smaller number of participants for survey 3 with only 124 responses (24 in SWASFT, 58 in WMAS and 42 in YAS). 98 (79%) were from operational staff and 26 (21%) EOC/Clinical Hub. This may be because this survey was only a short while after the previous one creating "survey fatigue" and included the busy Christmas and New Year period when staff may have had less time to complete it. Because the response rate for survey 3 was only a fifth of the responses to the previous surveys we have not included this in the quantitative analysis as for some questions, particularly those relevant to specific staff groups such as dispatchers, the number of responses within individual questions were very small. The substantial difference in response rates for surveys 2 and 3 also limits the representativeness of responses and the value of differences in responses between the two survey periods when these are expressed as percentages, as these can be distorted by the much lower number in survey 3. We have therefore only presented the quantitative results for the survey conducted after the introduction of Phase 2.1. However, the 3rd survey after the implementation of Phase 2.2 still provided some useful feedback from staff and so we have included a narrative commentary of the free text questions for this survey.

Impact of ARP phase 2.1 on triage and resource allocation

Table 18 summarises the responses by EOC/Clinical Hub staff to questions on triage and resource allocation. For phase 2.1 only the 2 questions on dispatching of multiple resources and stand downs had responses for all 3 services. There were no equivalent questions in the YAS survey.

Table 18: Survey responses for EOC staff on triage and resource allocation Phase 2.1

	Questions for EOC/Hub staff "**What effect do you think the {item} has had on patient triage and allocation of the right response category?**
	Phase 2.1
Response	**Reduction in the number of calls requiring an 8 minute response n(%)**
Much more effective	21 (35.0)
A little more effective	31 (51.7)
The same as before	4 (6.6)
A little less effective	0
Much less effective	2 (3.3)
Don't know	2 (3.3)
Total	60
	Amber call categories and revised response time standard
Much more effective	9(14.8)
A little more effective	35(57.4)
The same as before	7(11.5)
A little less effective	6(9.8)
Much less effective	2(3.3)
Total	60
	Green call categories and revised response time standard
Much more effective	9(14.5)
A little more effective	24(40.3)
The same as before	19(30.6)
A little less effective	5(8.1)
Much less effective	0
Don't know	3 (4.8)
Total	61
	Has the change in call categories allowed you to reduce the number of vehicles you stand down?
Yes	29(74.3)
No change	2((5.1)
No	8(20.5)
Total	39
	Has the change in call categories allowed you to reduce the number of multiple resource dispatches
Yes	51((71.8)
No change	10(14.0)
No	10(14.0)
Total	71

The responses show that EOC and clinical hub staff considered the changes in call categories had a positive effect on triage and resource allocation, more so for the Red and Amber categories than the Green category. The questions on dispatching multiple resources and reducing stand downs were the same for phase 1 and 2 and comparable to the responses for this same question in the phase 1 survey. In survey 2 we asked EOC/Clinical Hub staff if thought the change to amber and Green categories had made any difference to dispatching the most appropriate resource. 70.7% thought this had improved, 17% there was no change and 12.1% that this had reduced for amber calls. For Green calls 58.5% thought it had improved, 36.5% no change and 4.9% that it had reduced.

For operational staff we asked questions on whether they thought new call categories affected accurate identification of calls. The responses are summarised in Table 19.

Table 19: Survey responses from operational staff on Phase 2.1 call category changes

	Questions for Operational staff "**How do you think the new call categories and response time standard has affected triage and accurate identification of.{Item}?**
	Red calls
Much Better	74(12.7)
A little better	195(33.4)
No change	99(17.0)
A little worse	83(14.2)
Much worse	123(21.1)
Don't know	10(1.7)
Total	584
	Amber calls
Much Better	43(11.6)
A little better	96(25.9)
No change	55(14.9)
A little worse	68(18.4)
Much worse	93(25.1)
Don't know	15(4.1)
Total	370
	Green calls
Much Better	56(15.4)
A little better	92(25.3)
No change	95(26.2)
A little worse	35(9.6)
Much worse	46(12.6)
Don't know	40(11.0)
Total	364
	Have you seen a change in the number of times you have been stood down since the call category trial began?
Decrease in being stood down	223(39.2)
No change	245(43.0)
Increase in being stood down	101(17.7)
	569

For the Phase 2.1 survey the responses by operational staff on triage and accuracy of identification calls showed some changes with a higher proportion reporting they thought this was a little or much worse although there was still a greater proportion who thought it was better or the same. Forty percent of operational staff thought they were stood down less often. This compares to 64.5% reported in phase 1 but it is unsurprising as the gains made by the introduction of DoD will continue in to phase 2 so this most likely reflects additional gains. We also asked a question about timeliness of meal breaks and although there was a small rise in the number reporting this was more timely with each survey (6.7% and 9% respectively), for phase 2.1 there was also an increase in the

responses recording meal breaks as being less timely (30.9%) whereas in phase 1 this was more likely to be reported as no difference (66.5%).

In both surveys we asked all staff two questions on whether they though the ARP helped their service manage demand better and if it had had an impact on their ability to do their job effectively. In survey 1 52.4% of respondents thought ARP had helped their service manage demand a little or much better and 9.6% a little or much worse. For the survey after Phase 2.1 was implemented 2 the responses were 37.1% and 27.6% respectively so there was a shift in the number expressing a view that demand management was not being helped. There is a similar pattern in the responses to the question on impact to do their job effectively with the proportions reporting a little or much more effectively versus a little or much less effectively changing with each survey (Survey 1 30% v 9.7%; Survey 2 29.7% v 22.7%). Again, it is difficult to elicit to what extent this reflects additional changes over and above phase 1 changes or an overall view.

As in survey 1, EOC staff tended to have more positive views than operational staff in the survey after introduction of Phase 2.1. . Some explanations for this change can be identified from the free text comments. One of the main reasons for category changes in phase 2.2 was recognition that the phase 2.1 "Amber" category was too large and not sufficiently discriminating of urgency as it contained a wide range of conditions and levels of clinical acuity. This is borne out by comments in survey 2 from both EOC and operational staff. For EOC staff managing a large "Amber" stack was problematic. Initiatives were developed to try and prioritise the most urgent cases but this added an extra burden. For operational staff a view was expressed that some acute emergencies, particularly STEMI and stroke, were waiting too long for help as they were included within a group that also had calls for problems with much lower acuity. There was also a view that the 3 call type descriptors within the Amber category, (R, T and F), was too complex. The changes made for phase 2.2 tried to address some of these issues by reducing the proportion of calls and splitting categories to just 2 options of response or transport. EOC staff reported this as a positive move but it was seen less positively by operational staff. A number of key themes were evident in the text comments which help explain the issues identified by staff in surveys 2 and 3.

- In both surveys there were multiple and consistent comments made about the effectiveness of identifying life-threatening calls. For both phase 2.1 and 2.2 changes were made to the NoC to ensure that the most serious calls were identified early so dispatch is not delayed. The view of operational staff was that this had become less efficient with phase 2.1 and even more so for phase 2.2 and that the simpler strategy used in phase 1 was more accurate in identifying genuine cardiac arrest cases. Some issues also identified in phase 1 – particularly the "false positives" generated by the questions about noisy breathing for patients who are intoxicated – remained but three other specific examples were highlighted. One was the inclusion of "fitting now or fitting" with operational staff reporting that the majority of these calls were resolved by the time they arrived and most patients not transported. The second was the inclusion of "under 5 with priority symptoms" which, in the view of operational staff, were rarely emergencies and again often did not need hospital assessment or treatment. The third was the inclusion of some call types where bleeding is reported but which was not serious and that they were going many calls as a Red or Category 1 (Purple) call that were for nose bleeds. The general view was that this had become more of an issue with the implementation of Phase 2.2.

- The concern, predominantly of operational staff, was less to do with the actual over-triage itself (although this was seen as a problem) but that there is a group of acutely ill patients within Amber or Category 2 that are being disadvantaged by longer waits when demand is high as resources are sent to the highest priority calls which ultimately are far less urgent than some patients waiting in other categories. There was a strong view that some calls should be in a higher category.
- There was a view that some types of calls have been assigned to the wrong categories with particular concerns about elderly patients with falls that have fractures or patients with other injuries (dislocated shoulder was one example) being assigned to category 4 whilst "cut fingers" are category 3 as are calls from carelines which frequently are not urgent. There were also general comments about excessively long waits in these categories and particularly where patients are in public places and exposed to cold in winter.
- There were some positive comments that the reduction in the proportion of calls requiring an 8 minute response was a positive step forward and that Category 1 and 2 worked well but longer waits in categories 3 and 4 were less welcome.

Other comments were less about the call categories themselves but the effects on working practices that had arisen as the revised call categories had been operationalised. Again these were predominantly from operational staff. The main themes identified were:

- Whilst it was recognised that sending vehicles only to "stop the clock" was a negative practice, the emphasis on providing a conveying resource as the first resource for a substantial number of calls had created significant pressures. Staff reported having to travel much longer distances on lights and sirens and frequently crossing other vehicles that were nearer to an incident, and going out of area to calls that had been waiting for long periods of time. However there was recognition that this was a feature of having the wrong fleet configuration with a higher proportion of single responders and that as this balance is addressed the situation would most likely improve.
- There were a large number of comments that shift over runs had increased, partly because of the long distances but also because within the current system they would only be sent to Red 1 calls when near to the end of a shift but now they are sent to other calls, particularly where calls are out of time are upgraded to emergencies.
- There were comments from all 3 services that single response vehicles were now being less effectively utilised. This was partly that a reduction in the number of practitioners who could be used for "see and treat", for example for elderly fallers, had been reduced when there were calls, particularly in categories 3 and 4, that would be more suitable for this option. Conversely there were also comments that single manned response cars were sent to elderly fallers and would then requires back up to help with lifting so there is no clear agreement on what are the most appropriate calls for single manned response vehicles. There were also comments that cars are held back for Red or Category 1 calls and are under- utilised and that, where there may be delays in getting a DCA, they could be used to provide more timely help and initiate care for some acutely ill Category 2 patients.
- Operational staff reported that the new model is causing friction between staff and patients and that they frequently begin their interaction with complaints from patients

and having to apologise for the time some people have had to wait. One respondent commented that, whilst longer waits were justifiable for many calls, there needs to be more communication with the public about how ambulance service delivery is changing and how long they will wait for a response as there are still clearly expectations that an ambulance sill arrive quickly.

- From an EOC perspective, there was a view that the transport "T" categories cause stacking but that, overall, the revised categories do allow better management of resources. One issue highlighted by staff using AMPDS for call assessment is that there are some calls where the EMD is required to stay on the line. With increased response time windows this had substantially increased the length of some calls meaning staff are unavailable to take new calls. One EOC manager commented that this meant increasing the numbers of EMD's to compensate for these lengthened call times.

To supplement the survey results members of EOC staff at YAS have provided descriptions of their views about the introduction of phase 2 changes. These are presented verbatim below.

ARP has meant that we no longer spend all of our time managing demand and we can invest time in the staff. From the moment we switched to ARP it brought a calmness to the room that we had not seen for a long time, going from months of being at high escalation to now sending the right response for the patient reducing the numbers of resources we need to send to scene.

Duty Manager EOC, Yorkshire Ambulance Service.

ARP has given Dispatchers the autonomy to be able to make decisions on where best to utilise their resources effectively. The pre alert warning means we can allocate resources first to the most critical patients and have a bit of extra thinking time to allocate the most appropriate resources to the lower acuity calls thus saving valuable resources for the next call. As a Team Leader ARP alerts me when a call is more serious and I can support the staff involved from the beginning and ensure any additional specialist resources are allocated. I can see that ARP has definitely brought a more positive vibe into the EOC.

Dispatch Team Leader, Yorkshire Ambulance Service.

Having ARP introduced to the system has helped us as emergency call takers. It helps us to identify where we will go with the call i.e. protocol choice and instructions to be able to help the patient. We are the first point of contact for the patient so it helps identify possible life threatening conditions early and therefore sending the most suitable help to them as soon as possible, whilst we as EMD's are helping the patient over the phone. We have also found that we are working a lot more closely with the clinical hub i.e. identifying purple calls and having the support from them, with these calls. The pre alert questions help with the tone of the call, as in the caller to ourselves feels like they are being listened to, and therefore are more amenable throughout the call.

EMD, Yorkshire Ambulance Service.

From a dispatcher's viewpoint, I find ARP has had a calming effect on the dispatch process, in that there is usually a little more time to think about sending the best, most appropriate resource to each patient. The focus has shifted to accuracy rather than speed, and in general I think the workload is better handled as a result.

Dispatcher, Yorkshire Ambulance Service.

Although these are only the views from one service, they do encapsulate in more detail the particular advantages seen by EOC staff about the impact of implementing the phase 2 call category trial. It is probably fair to say that the impact is more noticeable on EOC staff than operational staff and this is also borne out in the results of the surveys.

Within the written comments in the surveys there was an acknowledgement from staff that many of the issues they identified were a feature of the wider environment rather than the call category review itself. Many of these comments echoed views expressed in survey 1 and clearly continue to exert an influence. As in survey 1, there were multiple comments on calls from NHS 111 and the categories they are assigned to particularly, in survey 3, calls that are allocated as Category 2. There has been a strong and consistent view across all 3 surveys that NHS 111 calls referred for an ambulance response are consistently over-triaged. Similarly the 999 call assessment systems also came in for criticism and it was recognised by some staff that the issues around the over triage of category 1 calls are a feature of the call triage process rather than the actual category itself. Similarly, there was a view that there are still many calls getting a response that do not need an ambulance at all and that more clinicians in EOC could help mitigate this. Of course it is also the case that operational staff have the benefit of a face to face contact with information that is not available in a telephone assessment. It was also recognised that many of the issues, and resultant pressures on staff, are a consequence of demand and insufficient resources to meet that demand in a timely way and that changing call categories will not overcome the gap between demand and resource availability.

Phase 2 staff survey summary

The first staff survey showed that the Dispatch on Disposition element of ARP is generally viewed by both EOC & clinical hub staff and operational staff as a positive development with responses indicating a substantial proportion of staff thought there had been some improvements for all questions. Improvements in the ability to resources dispatch more appropriately were maintained in to Phase 2 and this is supported by the detailed comments from one service. With the introduction of the Phase 2, whilst the shift from an over emphasis on 8 minute response time was seen as a positive step forward, there remained a perception that there is still an element of over-triage to the highest priority categories and suggestions that some calls in Category 2 could be a higher priority. There was also a view that in some cases there are long waiting times for some low priority categories which may be being hampered by the need to discriminate on what type of response is needed at the time of the call. However these have to be viewed in the context of the wider operating environment. As ARP has progressed so the pressures of more demand and the loss of

resources from long waits at Emergency Departments have also increased. For the latter, the attempts of the call category review to move towards better provision of the "right resource, first time" has put more emphasis on allocation of DCA's but it is precisely these resources that are under most pressure from hospital delays. This then translates into challenges for operational performance and staff working environment. What the surveys cannot tell us is, if the revised call categories were able to work as intended and if there were no delays, whether staff would then feel more positive to the changes. There are also some limitations to be considered. The staff surveys have provided a useful and valuable insight in to how the ARP initiatives have translated in to the working practices of frontline staff. However, the responses reflect the views of those staff who took part in the surveys. The services participating in ARP employ many thousands of frontline staff and the surveys reflect the views of a few hundred. What we cannot know is, of those who didn't respond, do they hold the same views or do they not respond because they are generally content with the changes or do not see any difference. We cannot therefore discount the potential effects of response bias in the results presented. Nevertheless, if staff have made the effort to respond and taken time to describe in detail their thoughts and opinions then the views expressed are valid and worth consideration. They provide some useful insights for services themselves and how they operationalise the changes and for the continuing ARP work, both in terms of specific suggestions that can inform ongoing refinement of the call categories themselves, but also by highlighting broader issues such as call assessment processes both in NHS 111 and 999. The detailed views expressed can inform wider considerations around call assessment and the need for systems to evolve so that they are better able to accurately manage what has become a larger population of users that are calling for a more diverse and complex set of health problems than those for which systems were originally designed.

4. Review of Ambulance Service performance measures and quality indicators

Dispatch on Disposition and the call category review have created an alternative operating model for the delivery of ambulance service care with a focus on better alignment of response to clinical need. The third element of the Ambulance Response Programme is a review of the current ambulance system and clinical quality indicators (AQI's) and development of a revised indicator set that can better reflect achievement of an alternative model of service delivery and the impact on patient care. The review has provided an opportunity to not only consider short term changes and amendments to the existing Ambulance Quality Indicators which reflect the operational changes resulting from current ARP work but also consider the longer term requirements needed to support ongoing review and development of more meaningful patient and clinically focussed quality indicators.

4.1 Review process

Performance and clinical quality measures are used by different groups for different purposes – Ambulance Services to monitor and assess how well they are delivering a core NHS service, commissioners to help frame the sort of service they want for local populations, frontline staff to provide feedback on the impact of their care, national bodies to monitor expected standards and patients so they can be assured of services that meet their needs. To ensure that all relevant potential options were considered we used a multi-stakeholder and consensus methods approach to the review.

A working group was convened to conduct the review. Membership included representatives from the following organisations and related stakeholders:

- ARP development group
- National Ambulance Service Medical Directors (NASMED)
- National Ambulance Service Directors of Operations (NDOG)
- National Ambulance Clinical Audit Group
- National Ambulance Research Steering Group
- National Ambulance Commissioners Network
- Association of Ambulance Chief Executives
- College of Paramedics
- Ambulance service staff (EOC, field operations, operational management)
- Patient representatives (Sheffield Emergency Care Forum)
- National Ambulance Information Group (NIAG)
- Call assessment system providers (NHS Pathways and AMPDS)

The review was organised and managed by the evaluation team from ScHARR using a two stage process. The first stage was a 2 day workshop utilising a broad range of stakeholders. The second stage used a small group to review the outputs of the workshop and specifically identify and agree a revised set of AQI's that could be introduced in the short term. The expected outputs were a set of

revised AQI's and a summary of the broader issues related to quality measurement identified during the review together with recommendations on how these could be further developed and managed.

4.2 Consensus workshop

A consensus workshop was held over 2 days in Sheffield on 30th November 2016 with 40 participants from working group organisations outlined above.

Performance and quality measurement for ambulance services have been the focus of considerable debate over many years. Consequently there is a substantial research literature on this subject and existing examples of performance and quality indicators. Prior to the workshop the ScHARR team collated a long list of potential indictors from a range of sources in to a single document to support discussions. These were the existing AQI's; the outputs from systematic reviews and a Delphi survey conducted as part of an existing ScHARR research programme (PhOEBE); a set of AQI's developed for the revised clinical operating model implemented by the Welsh Ambulance Service; indicators identified for a related Delphi survey published in 2016 (Murphy etal)[9] .

We used a modified nominal group approach to the workshop comprising a combination of small group work and open forum discussion with all participants. Briefly, over the two days the following tasks were completed;

1. A framework was agreed that set out the broad components of ambulance service delivery, associated processes and service objectives and expected impact on service delivery and patients. The final agreed framework was a modified version of the 5 step model used to support development of the Wales clinical model (Figure 32) and was used to structure the subsequent activities http://www.wales.nhs.uk/easc/ambulance-quality-indicators

   Figure 32: Framework model for indicator development

2. Small groups then identified potential indicators that could be applied to each component of the framework. They could draw on the collated long list of indicators drawn from existing evidence, a list of the measures included in the current ARP core evaluation datasets and other related activities (for example the National Clinical Audit Group work) to complete the task but also add their own ideas. The outputs from the small groups were discussed by all participants and a set of potential indicators for each component agreed for further review.
3. The small groups then critically assessed each candidate indicator against other criteria using a structured checklist to consider practical implications and comprehensiveness. Criteria

considered included, population applied to; whether it was measurable; relevance to different stakeholders; clarity of definition; acceptability; and an overall assessment of inclusion (yes, no or maybe) and whether it could be used in the short medium or long term.

The workshop considered a large number of potential performance measures/quality indicators and generated a substantial amount of material from both the small group work and broader discussions. Following the workshop all of the checklists completed by the small groups and related notes were examined to identify potential measures/ indicators for each framework. Based on the assessment made by the groups the indicators were further refined by i)excluding any measures where there was consensus that they were unhelpful, irrelevant or too difficult to measure; ii)removing duplicate indicators or combining similar indicators in to a single descriptor; iii)Separating out measures which are purely descriptive (for example call volumes) as these provide context as denominators for subsequent measures/indicators but of themselves do not provide any measure of service delivery or impact, and iv) separated potential indicators in to two groups – those that could be suitable for the short term AQI review and those that have medium or long term potential but need further development . Potential short term indicators were extracted into tables for each of the 5 framework components and these were used as the information source for the next stage of the review. The potential medium and long term indicators were collated and used for the narrative summary on future development.

4.3 Review of current Ambulance Quality Indicators (AQI's)

A smaller group derived from workshop participants and representing each of the stakeholder organisations met in January 2017. This group considered the current list of AQIs – allocated to each of the 5 framework components - and, using the outputs from the workshop, assessed whether each indicator should be retained, modified and/or expanded using the potential indicators identified at the workshop, or removed. The new indicators were constructed to fit with both the 5 component framework and the ARP phase 2.2 call category definitions. Following this meeting a revised set of AQI's were constructed. These were circulated for comment to key stakeholder groups, discussed at the ARP development group and, after further revision an agreed set of indicators produced that would be suitable for early implementation should a decision be made to further extend the revised call category operating model to other ambulance services. The agreed set of indicators is set out in Table 20. Items in blue are for reporting only too provide details of service activity and denominators for the indicators but are not themselves indicators of performance or quality. Items in Red are changes from existing indicators.

Table 20: Revised Ambulance Quality System Indicators

Current indicator	Revised Indicator
Activity (Call volumes) All calls All incidents By category (R1, R2,) Patients not transported Patients transported	**Activity (Call volumes)** All calls All incidents For each category C1; C2R, C2T; C3R, C3T; C4T, C4H **Of all incidents:** Calls not receiving a face to face response Calls receiving a face to face response Patients not transported Patients transported
Call Abandonment rate Call Answering time (seconds median 95th,99th Percentile) **Hear and Treat** % of calls receiving response closed with advice	Call Abandonment rate Call Answering time (seconds median95th,99th percentile) Time to identify C1 Call connect to first of pretriage questions; NoC or T5 – mean; 90th centile **Hear and Treat – All calls with no vehicle response** % closed with advice % referred to alternative service % returned for ambulance response (index call)
Response performance % R1 within 8 minutes R1 Response time (95th centile) % R2 within 8 minutes % All Cat A (R1+R2) within 19 minutes	**Response performance** For each category response time mean and 90th centile C1; C2R, C2T; C3R, C3T; C4T, C4H C1: C2R: C3R: (T categories clock stop is default transporting vehicle) - Time to arrival of transporting vehicle where transported (mean &; 90th centile)
Calls receiving face to face response % Not transported % Transported to type 1or2 ED	Calls receiving face to face response % Not transported % Transported to type 1or2 ED %Transported to other
Recontact rates % recontact with AS within 24 hours – H&T; S&T **Frequent Callers (FC)** % calls from callers with FC procedure in place	Removed – not meaningful in current format as difficult to define & measure properly. Move to CQI Removed – Had a purpose in driving development of FC policies and actions but achieved

The final agreed set reflects some principles agreed by the working group and broader stakeholders;

1. Although response time reporting has been set as mean and 90th percentile, reporting for call answering times has been retained as the 95th and 99th percentile as this measure is for very short times (measured in seconds).
2. There was some discussion on whether transport and non-transport rates should be reported by individual call category or for a whole service. Whilst there are advantages to

reporting by category for individual services this, had to be offset against adding too much complexity to what needs to be a relatively simple set of key indicators reported on a monthly basis that can be understood by different users including the public. It was agreed that a single whole service measure should be used as this provides an overall indication of ambulance service use of alternatives to hospital conveyance.

3. Hear and treat calls should be defined as all incidents where no ambulance vehicle response is sent. All calls defined as hear and treat should be counted not just those in category C4H but reported as C4H and other not by individual category. Hear and treat reporting has been expanded to discriminate between calls resolved by telephone advice only and those that are referred to other services. An additional measure of calls returned for an ambulance response has been added.
4. There was considerable support during the consensus review work to include some measure of delayed transfers of care (handover delays ay hospital) within the AQIs as these have considerable impact on ambulance services ability to deliver a timely service. Following wider consultation it was agreed that this is it is not a measure of ambulance service performance. However, the importance is recognised and there is an expectation that hospitals will be required to report handover delays as part of the ED performance dashboard which will make this issue transparent
5. For transported calls an additional indicator is needed for calls which receive a single person response and then a request for transport so that time to arrival of a transporting vehicle is captured.
6. It was agreed that mean and 90th percentile response times should be reported for all categories C1-C4 so response to all incidents is visible and to avoid the current situation where response to Green calls is hidden. Use of mean and 90th percentile are supported by related work conducted as part of another research study which found these two measures are well understood and meaningful to patients and the public.

The performance and clinical indicator review group has only considered what should be measured as an AQI set. The review did not consider the setting of standards and targets as that is a separate issue although there was broad agreement that expanding the AQI set and reporting times, rather than percentages within targets, is more meaningful and can still provide the leverage for improvement without the perverse incentives the ARP is trying to address. There will potentially be further discussion on whether expected response time intervals, particularly for C3 & C4 should be nationally or locally defined and how these are reported will then need further review.

Consideration was given to whether it would be possible to include a time from call connect to CPR or defibrillation indicator. It was agreed that it would be better to include this within the cardiac arrest Clinical Quality Indicator (CQI) as in order to determine this, the “true” cardiac arrests have to be identified which is a post hoc activity requiring use of the clinical records and cannot be determined using basic CAD or management information data. Time to defibrillation is further complicated by a lack of information on relevant timings where Public Access Defibrillators or other defibrillators (not ambulance service) are used. In the medium term it may be possible to include a routine measure of time to CPR at the time of the call where either it is reported as being in progress or where the EMD/call advisor initiates telephone CPR instructions. This is feasible but would require CAD changes.

The indicators have been incorporated in to a stepwise model based on the 5 framework components used to structure the indicator review which reflects the patient pathway (Figure 33).

Figure 33: Revised Ambulance System Quality Indicators – Patient pathway framework

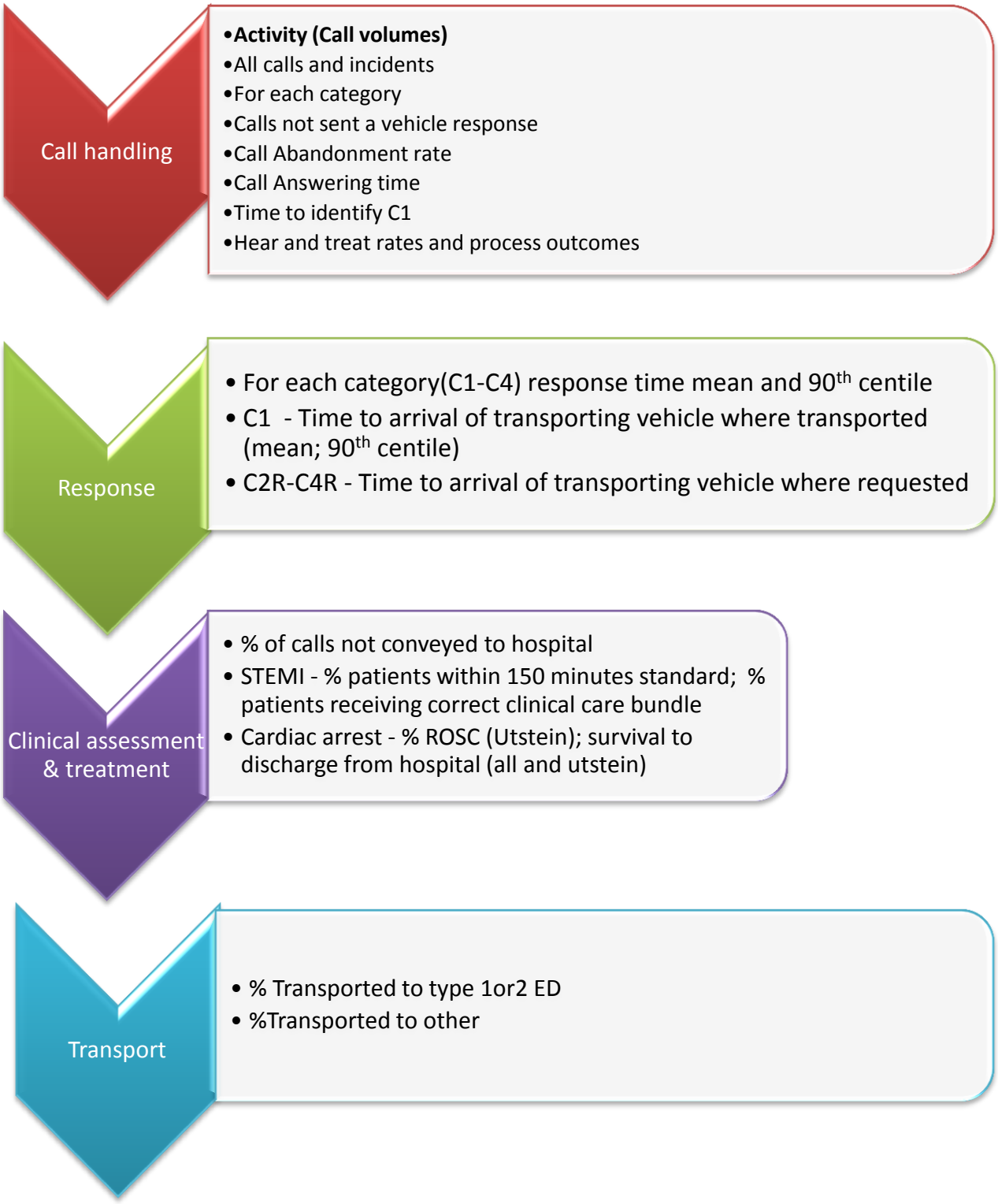


Table 21 provides an example of reported mean and 90th percentile response times for each response category.

Table 21: Mean and 90th percentile response times in 3 services

	WMAS		SWAST		YAS	
	Mean (Call Connect)	Mean (DoD)	Mean (Call Connect)	Mean (DoD)	Mean (Call Connect)	Mean (DoD)
Category 1	00:07:17	00:07:17	00:07:19	00:07:19	00:07:10	00:07:10
Category 2 R	00:10:48	00:08:57	00:19:45	00:18:43	00:17:30	00:16:38
Category 2 T	00:11:44	00:09:53	00:21:13	00:25:36	00:16:21	00:17:29
Category 3 R	00:21:52	00:18:39	00:46:27	00:45:00	00:33:19	00:32:23
Category 3 T	00:22:13	00:19:41	00:50:59	00:53:25	00:41:22	00:40:00
Category 4 T	00:36:00	00:34:15	01:27:23	01:31:01	01:05:28	01:04:19
Category 4 H	00:11:07	00:07:35	00:22:43	00:19:11	00:22:17	00:19:58
	90th centile (Call Connect)	90th centile (DoD)	90th centile (Call Connect)	90th centile (DoD)	90th centile (Call Connect)	90th centile (DoD)
Category 1	00:12:05	00:11:58	00:12:55	00:12:59	00:12:41	00:12:39
Category 2 R	00:17:51	00:16:21	00:41:22	00:39:17	00:35:24	00:34:24
Category 2 T	00:19:48	00:21:37	00:55:45	00:53:52	00:30:46	00:34:07
Category 3 R	00:40:57	00:39:56	01:44:28	01:42:30	01:18:32	01:15:33
Category 3 T	00:42:49	00:45:00	02:03:59	02:02:15	01:38:30	01:33:34
Category 4 T	01:23:47	01:27:12	03:33:06	03:32:30	02:30:12	02:31:10
Category 4 H	00:17:22	00:13:54	00:56:50	missing	00:52:47	00:55:19

The potential revised AQI's were also presented to a Public and Patient Involvement group who provide PPI support to urgent care research (the Sheffield Emergency Care Forum) to gauge their views on relevance and if the indicators were understandable to a lay audience. The group considered the proposed indicators to be useful and an improvement on current indicators, particularly providing response times rather than just proportions within standard.

Clinical Quality Indicators

The review also considered the current CQI's. Measurement of clinical rather than process based performance indicators has always presented particular challenges. The workshop and review process provided a valuable forum for identifying the potential for expanding the scope of clinical indicators, the associated problems which hamper development and potential solutions. These are discussed in the next section.

In the short term we recognise the need to maintain some form of CQI within the national reporting framework. We suggest retaining two of the current 3 CQI's with some refinements

1) Continue using the STEMI CQI in its current form – this is evidence based and processes are well established
2) Continue to use the current Cardiac Arrest CQI but report ROSC only for cases that fulfil Utstein criteria as this outcome is only relevant to this group. Survival to discharge from hospital should continue being reported for all cases and cases fulfilling Utstein criteria. In

the medium term there will be further review of this indicator to explore the addition of additional metrics such as time to CPR and defibrillation.

3) The current CQI's are reported monthly, however these comprise relatively small numbers of cases at individual service level. It is recommended that reporting moves to quarterly or 6 monthly as aggregated numbers over longer periods of time will be more useful in detecting change.

The third current CQI is for stroke. This indicator needs to be reviewed to take account of a)the lack of access to outcome data to identify "true" stroke patients and b) realistic time standards for transport to HASU in the light of ongoing service reconfigurations. In its current form this indicator is not fit for purpose but will be reviewed as part of the NASCQG work-plan to find a better model.

To replace this, a new CQI will be developed and tested during the first year of any changes by accelerating work already in progress. Sepsis is one possible candidate patient group.

4.4 Future developments for system and clinical quality indicators

The workshop generated a substantial number of ideas for both system and quality indicators that could be developed in the future, both in terms of broad topic areas or principles and the ways in which they could be measured. The rich material generated would provide a sound starting point for further development of indicators that could much better reflect the care delivered and the impact of that care for patients than existing time based and response measures. There is not scope to describe these in detail here but examples of topic areas that could be considered and new approaches to measurement included;

- More assessment of the impact of long waiting times on patient outcome and experience and scope to incorporate serious incidents in to a system quality indicator or as part of a broad clinical care bundle for specific conditions or patient groups
- More granular assessment of processes associated with managing patients outside hospital. There is potential for a national indicator if some easy way of identifying and measuring a single component e.g. frequency with which referral to an alternative service is rejected could be found. However, this may be better approached as part of a comprehensive clinical audit cycle for defined groups at individual service level so that local pathways are considered and service gap analysis incorporated as part of the audit cycle
- Support for clinical decision making provided to staff from e.g. clinical hubs. This potentially has an impact on service delivery (e.g. non-conveyance rates), staff confidence and perceptions of support and the contribution of specialist clinician roles within clinical hubs and not just for face to face assessments
- More scrutiny or different types of service user and separating 999 calls from the public and from those made by other agencies such as the police, health care professionals and care-lines to better determine what types of response are required for different types of call.
- Expansion of the current clinical indicators to include a broader group of patients. The current focus on cardiac arrest, stroke and STEMI only represents a very small proportion of patients who access ambulance service care and clinical quality measurement needs to

reflect all types of patients. Potential groups include sepsis, major trauma, mental health problems, and elderly patients who have fallen or groups characterised by ambulance service decisions (long waits, patients left at home, referral and alternative service pathway compliance and availability).

It was recognised that there are differences between identifying useful single indicators that can be easily measured and more detailed and comprehensive clinical audit activities but currently these are not considered together. The consensus work provided a very clear steer that there is considerable scope to make reporting of key routine metrics and clinical audit cycles a much more complementary and joined up process but to be successful this needs to be supported by a comprehensive strategy for a clear, ongoing, performance and quality measurement process. There is a place for simple, easily retrievable metrics that can be measured routinely and frequently but in isolation these often lack context. A comprehensive and complementary AQI and clinical audit strategy that can provide context and more detailed explanation and understanding of processes and outcomes will provide a richer description of how well ambulance services are achieving their intended objectives. In particular, the assessment of potential indicators within the 5 steps has demonstrated the links between the steps and the need to not just consider individual step metrics in isolation, but also to find ways of demonstrating how these fit together as a service pathway.

There is enormous scope to expand the clinical audit process to include a more diverse mix of patients and link this to routine indicator measurement. However the scale of the task cannot be underestimated. For each additional group there is a substantial amount of work involved in designing and testing the set of indicators to be included as part of the care and outcome bundle for that group. The actual audit process itself is then time consuming and labour intensive. This type of comprehensive audit across the patient pathway is much more meaningful and provides better information to support service improvement but they are not suitable for routine and frequent measurement. Instead they need to be incorporated in to a rolling annual cycle so that they can be periodically revisited to assess progress.

A formal, nationally agreed strategy and long term work-plan to support clinical audit cycle development and periodic revision of the AQI's will ensure these important tasks remain relevant and continue to progress. It is recognised that there are significant obstacles to the development of outcome based clinical indicators, not least the difficulties in capturing patient information from other parts of the health care system. A long term strategy will ensure that, as information systems develop, both AQI's and clinical audit care bundles can be refined to incorporate new measures. In addition, the NHS is undergoing substantial changes and re-organisation of services at local levels and a process of ongoing review will be needed to ensure that system and clinical indicators reflect these changes. Finally, any national strategy will need to consider the resources and capability required to support this process. Clinical audit in ambulance services is conducted by dedicated but small teams who already operate at capacity. A formal national programme to expand and develop a more comprehensive process of measuring ambulance service system and clinical quality would need additional financial resources to be successful.

5. Discussion and conclusions

We have evaluated the effects of implementing the two key initiatives developed for the Ambulance Response Programme. The first is Dispatch on Disposition to support more accurate call triage and intelligent dispatching of resources (Phase 1). The second is a trial of a revised set of call categories, with a focus on providing a rapid response to the most urgent calls and a more flexible approach to providing a clinically appropriate response for the diverse range of other calls in 3 services (Phase 2).

Phase 1

The evaluation of Phase 1 shows that Dispatch on Disposition has achieved most of the intended objectives. There has been a modest effect on key performance measures for Red 2 and a reduction in the time to first resource on scene for red 1 and Red 2. In the statistical analysis 23/30 indicators showed some benefit following the introduction of DoD compared to control sites. The most marked effect is on efficiency with a clear and measurable benefit in resource use. We have estimated that nationally there is potential to gain 10243 whole resources per week which would then be available to respond to other incidents. In services using NHS Pathways there has been a reduction in the proportion of calls where a resource arrives before assessment is complete. The additional call triage time does not appear to have had a detrimental effect on time to allocation of resources or to arriving on scene for the most serious (Red) calls with only modest increases of under a minute for Red 2 calls although up to 4 minutes is available for clock start. There are longer times for resource allocation and for arrival of resource for Green 2 incidents which may result from a combination of longer triage times and possible operational changes in dispatch management. As these are the least urgent calls this may simply reflect resource management when there are calls waiting for an available resource and priority is given to more urgent calls. The benefits have been achieved during a time period when there has been a steady and continual increase in demand and the additional constraint on resources resulting from an increasing trend of hours lost at hospital. This pressure is reflected in some of the findings where the benefit in the pilot sites is less a result of reversing the underlying trend in some indicators and more a slowing down the rate of decline in the pilot sites compared to the control sites. In the current environment, the advantage of DoD is most likely that of helping to mitigate deterioration during a period of sustained high demand and there may be scope for further gains if demand begins to reduce.

The one intended objective not achieved is the expected increase in hear and treat rates. There is some indication within individual services that hear and treat has periodically increased – although this is also the case in some control sites – but this is not consistent. It may be that the additional call assessment time alone cannot substantially increase hear and treat rates and that other strategies are needed. In particular, the staff survey revealed that the availability of clinical staff to provide further assessment and advice may limit the number of hear and treat calls that can be appropriately managed, and there may need to be an associated increase in the clinician workforce to fully realise the potential of higher rates of hear and treat. It is also possible that the time period of the study has coincided with a period of higher acuity calls which are not suitable for this type of management , and therefore the opportunity to increase the hear and treat rate has not been present.

The introduction of pre-triage sieve questions and Nature of Call at the beginning of the call assessment process, to ensure time critical events such as cardiac arrest are identified early, appears to be working well in services using NHS Pathways. There are differences between services in how they have adapted use of NoC but a process of monitoring and refinement in the choice of call descriptors that generate an early resource dispatch means that around 75% of Red 1 and specifically cardiac arrest calls can be identified very early, although there may be some associated over triage. Use of NoC is not mandatory for services using AMPDS as to some extent it replicates the early steps in the AMPDS assessment system. Some services are exploring the use of pre-triage sieve questions and key descriptors in AMPDS with some success but these are relatively new developments and a work in progress. An audit in London Ambulance Service has shown potential to reduce the allocation time to cardiac arrest calls by on average 30 seconds. If replicated nationally a cautious estimate is that this could potentially save an additional 250 lives each year. Additional data after a longer period of operation will show if there are any advantages for AMPDS users and if improved outcomes are achieved.

In almost 21 months of operation there have been no reported patient safety incidents attributable to DoD. More calls are being assessed to completion and there is evidence from staff that this is resulting in resources being more appropriately targeted to the right patients. The staff survey has shown that, overall, both EOC/Hub and operational staff view the introduction of Phase 1 as being a positive step in helping to better manage emergency calls in a challenging environment. There is no evidence to suggest that DoD has resulted in a poorer service for patients and there is evidence that it is providing benefits, particularly in terms of efficiency and better use of available resources.

Phase 2

The introduction of Phase 2 has been more challenging to evaluate. For the existing call categories half of 999 calls are allocated to the most urgent Red 1 and Red 2 categories although all available evidence shows that, with the exception of out of hospital cardiac arrest, there are no clinical benefits for responses within the 8 minute timeframe. Cardiac arrest comprises 0.6% of 999 calls. Call prioritisation of 999 calls began in England 20 years ago but during this time demand has more than doubled and the 999 call population comprises patients with a much more diverse range of health problems with variable clinical acuity and urgency. To some extent the existing call categories with a focus on rapid response reflects the historical model of the ambulance service as an "emergency" service rather than the provider of out of hospital health care to a complex range of problems it has become. Of course, providing rapid help to the most acutely ill patients remains the primary objective. This makes it all the more important that, in an environment of increasing demand that shows no sign of diminishing and constrained financial resources, ambulance services can effectively prioritise the most urgent calls and ensure resources are available to respond to them, whilst at the same time providing a clinically appropriate and timely service to other calls. The review and revision of the call categories is an attempt to try and address these issues. It is has not been an exercise in just shifting the goal posts in terms of response time standards – that could be achieved by simply by changing the response time standard for Red 2 and Green calls – but a consideration of the whole 999 population. Revised categories have been defined to try and better reflect the response required by different levels of need. The revised categories therefore represent an entirely new operating model for the management of ambulance resources. This is important from an evaluation perspective as, unlike in Phase 1, we have not been able to conduct a controlled

comparative study as the current and trial call categories are so different direct comparisons cannot be made. Instead we have provided a description of how the operating model has worked in practice using a range of indicators to assess effects over time where possible.

The first iteration of Phase 2 revealed that over 70% of calls were being allocated to a single category (Amber) and that, although there were refinements within the category in terms of type of response, there was too much variation in case mix and acuity within the category. This made it difficult to manage operationally and to ensure the sickest patients were not disadvantaged. The categories were revised to provide more discrimination and the second iteration – Phase 2.2 – implemented in late October 2016.

For phase 2.2 the proportion of calls within categories showed that the number within the highest category (1) requiring an 8 minute response has, as intended, substantially reduced compared to the current model although there were differences between sites in the proportion (6-11%). The highest proportion is within category 2 although this was substantially smaller than the previous phase 2.1 Amber category. There were between service differences in both the number of calls assigned to each category and the within category Response and Transport options. These most likely reflect differences produced by the two call assessment systems, with overall a higher proportion assigned to C1 and C2 in the service using AMPDS. The allocation of AMPDS codes and NHS Pathways SG/SD combinations to the call categories was made using the best available evidence at the time. As phase 2.2 has progressed there is scope to further review the allocation of different types of calls using information accumulated by using the categories in a real operating environment and taking in to account the feedback from staff on possible over-triage in Category 1 and potential re-allocation of some other call types. Overall, there is not a very clear and obvious distinction in conveyance rates between the Response and Transport categories with substantial conveyance rates in each. It may be that the current call assessment systems are not sufficiently discriminatory enough to allow decisions about which incidents may or may not need transport at the time of the call and consideration should be given to whether this distinction is worth retaining.

We assessed response performance for each category for a range of percentiles. Across Phase 2 (including Phase 2.1) the proportion of Red or C1 calls responded to within 8 minutes increased in one service and decreased in another. Evidence from the services shows this is less to do with the categories themselves and more a consequence of other operational changes such as fleet configuration that have occurred to support the new operational model. This is an important point to consider and illustrates that potential benefits may not be immediately apparent, as there will be a process of continual review and readjustment of a whole range of functions that underpin operational delivery before the right balance is found to maximise effectiveness. Proportion of calls responded to within 8 minutes is a crude measure of performance as it provides no picture of what is happening for calls over 8 minutes. The percentile response times show that, for all 3 services, 50% of Category 1 calls received a response in less than 7 minutes, 90% within 13 minutes and 95% within 16 minutes. For other categories the figures are more varied and were noticeably shorter across all categories in one service (WMAS). For the call category trial no standard was set for categories 2-4 in terms of a proportion to be achieved. The percentile response times for each category in each service provide a picture of what can be achieved operationally but not the achievement of standards. These times also reflect provision of the “right” resource as for conveyed patients the “clock stop” is the conveying vehicle so, unlike the current reporting of response times

which use first resource on scene, these figures better reflect the provision of appropriate care as any waiting times for a conveying resource are not hidden. The revision of the AQI's conducted as part of the evaluation will make this more transparent, as the recommendation of the review working group is that mean and 90th percentile times are reported for all categories rather than proportion within a fixed time for only the most urgent calls.

We measured adjusted trends for a range of response intervals and resource utilisation indicators for Phase 2.2. Some gains had been made in phase 2.1 and overall in phase 2.2 there was a trend towards improving response performance in some services and no obvious degradation in performance in any service. Small but significant additional efficiency gains were also made in one service. We analysed changes in key performance measures for the 999 population across the 3 ARP phases and found that, despite the absence of response time standards for the majority of calls, the service delivered to the whole 999 population only changes by a small amount and performance is more consistent and stable than that seen in the current operating model. This reflects probable better allocation of the right resource and hence reduced waits for patients where back up conveying vehicles are required. Further efficiency gains were made during phase 2.1, equivalent to making an additional 4500 resources per week available for response across England.

We also examined differences in response performance between urban and rural areas in the 3 services and found a complex picture that cannot be explained by ARP changes alone. There was evidence of some reduction in inequities between different area types for some measures with a reduction in 95th percentile response times and time to hospital in 2 services for very rural areas. There was also an unexpected finding that operational times are often longer in urban areas than other areas and this warrants more detailed investigation.

There was no indication of reported patient safety issues during Phase 2 and a high proportion of the most urgent and cardiac arrest calls were identified by NoC in the two services using NHS Pathways. In the AMPDS service, on average, calls with cardiac arrest codes are identified in around 60 seconds. The staff surveys showed that the changes were broadly welcomed by EOC staff who considered further improvements had been made in allowing them to manage resources efficiently and effectively. For operational staff there were mixed views so although the reduction in the proportion of calls requiring an 8 minute response was welcomed there were some concerns around whether the right types of calls have been allocated to the right categories. This was more obvious in the third survey but the response rate was very low and it was conducted just after Christmas and new year when demand was at a peak so responses may to some extent reflect timing and another survey at a period when activity is more stable and the Phase 2.2 model has had longer to mature may be more useful.

ARP within the wider health system

Although phase 2.2 has only been in operation for a relatively short time a common feature of all the adjusted time based analyses showed a consistent pattern that, in contrast to trends seen in control sites, phase 2 brought a degree of stability to operational performance during a period of high demand and significant operational challenges. Clearer changes may become more apparent when this new model of service has been operating for a longer period but the results so far suggest that

the more flexible approach to call assessment, resource dispatch and response intervals may be helping to reduce further deterioration in performance and there is no indication that for any group of patients the service is worse. There are still delays and some long response times for some calls in lower urgency categories but this is also true of the current model. The release of resources identified in both phases appears to be helping services better manage demand and call queues and helping provide a consistent service to patients. In an ideal world there would be no long waits for any patient but in order to provide a service that could respond to all patients within an expected timeframe there has to be sufficient resources to allow achievement. The ARP initiatives have shown that there is scope to improve efficiency in resource use which can then be used to improve operational performance for the whole 999 population. However, there will be limits to how much extra performance can be generated from changing operations within the existing resource capability. Changing the response model to one which is more clinically appropriate does not necessarily "save" resources. The experience at West Midlands has shown that, at least in the short term, this change may present its own challenges. Changing the fleet mix towards more conveying vehicles ensures patients who need transport are more likely to get the right resource but it effectively reduces resource capacity as more cars than DCA's can be provided for the same money. WMAS believe this is the right strategy but, as with any new operating model, time is needed to refine the model and allow it to mature before all potential benefits are realised.

Resource capability is also affected by wider issues. In particular, delays in handing over patients at hospital have become acute during the last year, and this has had real repercussions for ambulance services as every vehicle held at hospital is not available to respond to a call. Our data shows over 600,000 ambulance service hours were lost at hospital during 2016 and that is 600,000 patients who could have got a more timely response if vehicles were not delayed. Delayed vehicles are also conveying vehicles and so this produces a serious obstacle to attempts to provide the "right vehicle, first time". A recent report by the National Audit Office also highlighted the impact that hospital delays are having on performance[10] and the need to resolve this issue. NHS England and NHS Improvement are actively addressing handover delays, which is welcome as a substantial reduction would increase resource availability within ambulance service and there may then be scope to gain more benefits from ARP initiatives.

However, as the National Audit Office also found, it is the case that funding for ambulance services has not grown at the same rate as activity. The ARP changes can go some way to improving ambulance service delivery and care by making better use of the resources available. However, as demand continues to increase and financial resources remain flat or even reduce there will come a point where no further gains can be made. The will be an upper limit to how much changing the call assessment process and call categories alone can achieve. This will most likely be compounded by other changes within the NHS and in particular service re-organisations within Sustainability and Transformation Plans. Changes to emergency department provision and specialist services such as hyper-acute stroke services will have effects for ambulance services as they may have to travel further distances to take patients to appropriate care. This in turn will create longer job cycle times and place additional strain on what are already stretched resources. In the absence of additional funds to meet these challenges and further increases in demand, the efficiencies gained through ARP may absorb some of this activity but this may be at the cost of reducing response time performance. ARP has produced some benefits, clearly in efficiency in phase 1 and a clear indication of improved ability to maintain stability in performance during a period of high demand in phase 2.

However, there needs to be realistic expectations about how much ARP can deliver in terms of service improvement, as opposed to maintenance and prevention of performance degradation, in this challenging environment.

Expectation is also an important factor to consider, especially for the public who, for the most part, highly value the ambulance service both for the clinical care they provide but also a quick response. The steep decline in achievement of response time performance for Red 1 and 2 calls has produced many negative media headlines and reports that ambulance services are "failing" to deliver a good service. The continued emphasis on 8 minute response times potentially reinforces the expectation that all 999 calls will, and should, get a rapid response and fuels a perception that the majority of 999 calls are for life-threatening conditions, when in reality these only comprise less than 10% of calls and cardiac arrest only 0.6%. It may also be the case that, whilst patients are familiar with queues and waiting times in other parts of the NHS, including emergency departments, there is a perception that because the ambulance service is an emergency response service (and highly visible on the roads) there will always be a resource to send and that there is no "queue" waiting for an ambulance. The revised call categories have been defined to take in to account the wide range of clinical problems and levels of urgency that present as 999 calls and realistic timeframes for providing a response within a very difficult environment can be set accordingly. However explaining these will be challenging. The ARP work can provide an opportunity to begin to have a much more open discussion with the public, and with policy makers, about the challenges ambulance services face. This includes the nature of the 999 call population and range of response options that are available, and what type of service is needed for the future that balances expectations with what can be realistically achieved for the money available to pay for it.

These difficult conversations have already happened in Wales where a new clinical response model, broadly similar to the ARP, was implemented in October 2015 and after a year of operation and independent evaluation, was recently permanently approved by the Welsh Government http://gov.wales/newsroom/health-and-social-services/2017/170227ambulance/?lang=en

The changes have happened more rapidly in Wales but have been broadly accepted by the public. A similar change is also being trialled in Scotland although no results are available. Very similar changes to those implemented by the ARP have also been made in Victoria, Australia http://ambulance.vic.gov.au/ and an evaluation has highlighted similar findings with faster dispatch to life-threatening calls, early indication pf potential improved outcomes from cardiac arrest and a more appropriate clinical response to other calls.[13] This highlights that the issues faced in the UK are not unique and alternative solutions to providing a safe and clinically appropriate service are being addressed internationally. The National Audit Office, in their report, considered that too much emphasis had been placed on response time performance by commissioners, regulators and providers and recognised the potential for the ARP initiatives to support future development of more clinically focussed ambulance services in England.

A major limitation of this evaluation is that we have not been able to conduct any detailed assessment of the impact of ARP on patient outcomes. This is due to several factors. Firstly, only 3 clinical quality indicators are currently measured (cardiac arrest, stroke and STEMI) and of these only cardiac arrest has a patient outcome (ROSC and survival). Monthly case numbers for cardiac arrest are very small and different changes have been made at different time points over the 18 months of

the ARP programme so it is difficult to accurately measure any change in outcomes over 3 phases of short duration. The other two CQI's measure processes rather than patient outcomes, and again are small numbers and normal performance is very variable on a month to month basis, so changes between phase 1, 2.1 and 2.2 that are a consequence of ARP rather than other factors are difficult to detect within such short time periods. Finally, collection of clinical outcome data is complex and time consuming and there is a minimum 3 month time lag in publishing CQI data. A longitudinal study over a much longer period will be needed in order to determine any effects related to the implementation of ARP. More broadly, clinical outcomes are an important consideration and ideally, in time, measures that reflect clinical care would replace the operational process measures that are currently used. As part of this programme we conducted a consensus workshop on performance and clinical quality measures with a wide range of stakeholders and there was clear agreement that a shift towards better clinical measures is the right way forward. However, it was equally clear that there are no single, easily reported clinical indicators. It is not surprising that time based process measures have dominated ambulance quality indicators as these are easily measured and reported using routine data systems. Measuring clinical indicators and outcomes is a post-hoc activity that is much more time consuming and resource intensive. The indicator review produced many examples of potential clinical indicators but these are most useful when measured as part of a package that combines performance and outcome measures linked across the different steps in the patient pathway. This could be possible for much more varied groups of patients than the current focus on cardiovascular disease. This expanded scope would provide a more meaningful and useful picture of service delivery that takes into account contextual factors, such as variation in local service provision, which in turn is then more relevant to supporting improvement. There is enormous scope to improve the measuring and reporting of ambulance service performance and care, but for this to move forward the functions of clinical audit and AQI's need to be aligned and considered together. They also need a considerable amount of development and an agreed plan on frequency of measurement, reporting and review. This cannot be achieved by small, local clinical audit teams. If progress is to be made these functions need to be part of a co-ordinated national long term programme that can develop, plan and review on an ongoing basis and which is resourced to fulfil these objectives. Without this it is unlikely that any progress from the current position will be made.

Conclusions

The evaluation of phase 1 has provided strong evidence that the introduction of longer call assessment times produces clear benefits for operational efficiency and this is translated in to better response time performance for the most seriously ill patients. Re-analysis of the Phase 1 data since the interim report shows that the benefits have become more evident. There is no evidence to suggest that patients, particularly those with time critical conditions, are disadvantaged by this initiative and the processes put in place to ensure early detection and dispatch to patients with cardiac arrest or potential cardiac arrest are working well. Overall, the evidence supports the phase 1 changes as providing benefits to services that allows them to better manage the resources they have which becomes increasingly important as demand continues to rise. There does not appear to be any reason why Dispatch on Disposition should not continue.

Phase 2, and particularly the most recent iteration of phase 2.2 is a much more substantial change in operating practice and has only been in practice for a relatively short period of time. Because it is

such a significant departure from the current model it has not been possible to conduct a comparative analysis so we cannot say if the new model is "better" or "worse" than the existing model, only that it is different. In practice, the trends in phase 2.2 indicate a period of operational stability in the 3 trial services during a period of high demand and at a time when response time performance continued to deteriorate in services operating the current national model. The relative whole service performance stability emerging from the combined DoD and revised call category initiatives in the Phase 2 trial sites suggests that the more flexible approach to call assessment, resource dispatch and response intervals may be helping to reduce further deterioration in performance. It is also evident from the experiences of the trial services that changing the response model requires a range of complex operational changes beyond the call assessment and categorisation process. A system of ongoing review and refinement is needed to optimise delivery as the requirements of the model become clearer over time. This means that ARP initiatives cannot be seen as a "quick fix" for current poor operational performance. With time and maturity there may be further benefits to be realised but additional review will be needed over a longer time period to see if these materialise. Resolution of the current problems in handover delays will most likely go some way to helping this.

It is worth re-emphasising that modern day ambulance services in England provide a response to a large and heterogeneous population of patients with health problems of variable clinical acuity. These range from life-threatening emergencies to the complex health problems of an ageing population and minor problems that can be resolved over the telephone, but the right solution is expected from a single service that is providing both emergency and urgent care. This poses real challenges as the requirements of an emergency population are quite different to an urgent or minor health problem population and raises questions about whether any single organisation can effectively meet these competing demands. The call category trial initiative has attempted to address this issue and makes clear that, in an environment where demand will continue to increase but financial resources will be constrained, a more realistic but clinically appropriate operational model can be put in place that differentiates response and clinical need. The alternative is to retain the current model but the evidence over the last few years has shown that provision of an 8 minute response to at least half of 999 callers is simply not achievable and creates inefficiencies that are not compatible with providing an equitable or clinically appropriate service. It is too early to say if the phase 2.2 model is the "right model". However, the current evidence shows that it has produced benefits without compromising patient safety and supports provision of a consistent response despite substantial fluctuations in demand and so is a realistic alternative.

References

1. England. High quality care for all, now and for future generations: Transforming urgent and emergency care services in England - Urgent and Emergency Care Review End of Phase 1 Report. 2013.Gateway Reference 00691-00694. http://www.nhs.uk/NHSEngland/keogh-review/Documents/UECR.Ph1Report.FV.pdf

2. Turner J, Coster J, Chambers D, Cantrell A, Phung V-H, Knowles E, et al. What evidence is there on the effectiveness of different models of delivering urgent care? A rapid review. Health Serv Deliv Res 2015;3(43) http://dx.doi.org/10.3310/hsdr03430
3. Cummins R, Ornato J, Thies W, Pepe P. Improving survival from sudden cardiac arrest: The "Chain of Survival" Concept. A statement for health professionals from the advanced life support committee and the emergency cardiac care committee, American Heart Association. Circulation 1991; 83: 1832-47.
4. O'Keeffe C, **Turner J**, Goodacre S, Nicholl J. The role of ambulance response times on survival from out of hospital cardiac arrest. EMJ (2010) doi:10.1136/emj.2009.086363
5. Pons PT, Markovchick VJ. Eight minutes or less: does the ambulance response time guideline impact trauma patients outcome? J. Emerg. Med 2002; 23: 43-48.
6. Pons PT, Hankoo S, Bludworth W, et al. Paramedic response time: does it affect patient survival? Academic Emerg. Med 2005; 12(7): 594-600.
7. Turner J, Nicholl J, O'Keeffe C, Dixon S. The costs and benefits of implementing the new ambulance service response time standards. Final report to the Department of Health. Medical Care Research Unit, University of Sheffield; January 2006
8. Blackwell T, Kline J, Willis J, Monroe Hicks G. Lack of association between pre-hospital response times and patient outcomes. Prehosp Emeg Care 2009; 13(4): 444-450.
9. O'Keeffe C, Nicholl J, Turner J, Goodacre S. Role of ambulance response times in the survival of patients with out-of-hospital cardiac arrest. http://dx.doi.org/10.1136/emj.2009.086363
10. Hawkesa C, Bootha S, Ji C, et al. Epidemiology and outcomes from out-of-hospital cardiac arrests in England. http://dx.doi.org/10.1016/j.resuscitation.2016.10.030
11. Murphy A, Wakai A, Walsh C, Cummins F, O'Sullivan R. Development of key performance indicators for pre-hospital care. EMJ 2016; 33(4): 286-292. http://emj.bmj.com/content/33/4/286
12. National Audit Office. NHS England – NHS Ambulance Services. Report by the Comptroller and Auditor General. NAO, January 2017. https://www.nao.org.uk/report/nhs-ambulance-services/
13. Ambulance Victoria. Revised clinical response evaluation report. Delivering our patients the right care, at the right place, at the right time. Department of Research and Evaluation, Ambulance Victoria, July 2017 http://ambulance.vic.gov.au/about-us/research/our-clinical-response-model/

Acknowledgments

We would like to thank the 10 NHS Ambulance Service Trusts in England for their co-operation in providing the weekly data for this evaluation. Particular thanks go to Tracy Rayment-Bishop (West Midlands Ambulance Service NHS Foundation Trust) for co-ordination and management of the data returns. We would also like to thank the Sheffield Emergency Care Forum for providing valuable and helpful feedback for each Phase of the programme.

Appendix 1 – Nature of Call (NoC) descriptors and cardiac arrest allocation

List of Nature of Call Descriptors

Nature Of Call (NoC)
Red 1
Choking
Drowning/Water incident
Unconscious (Not Red1 scenarios)
Allergic Reaction
Bleeding
Breathing
Chest Pain/Cardiac Prob/Back Pain (Upper)
Collapse
Diabetic Probs
Fall Injuries Unknown
Fitting
Major Incident Declared
Maternity
PLATO
RTC Roll Over
Running Red
Stroke/Neurological
Trauma Major
Medical
Electrocution/Shock
Abdominal/Flank Pain
Air Incident
Alcohol Related
Bomb Threat
CBRN
Concern For Welfare
Death expected <18
Death unexpected all ages
Explosions
Fire Persons Reported
Fireams
Flooding
HAZCHEM
Headache
Heat/Cold Exposure
Marine Incident on Coast
Multiple Casualty Event
Overdose
Rail Incident

Nature Of Call (NoC)
RTC
Section 136
Stabbing
Suicide
Assault
Marine Incident at Sea
Mental Health
Burns
Trauma
Back Pain (Lower)
Death expected >18
Eye Problems
Fall Non-Injury
Fire Request To Standby
HCP
Major Incident Standby
Medical Minor
Social
Telecare
Pathways Overridden
111 / Unknown

Nature of Call descriptors assigned to cardiac arrest in 4 services Oct 15-March 16

	Total of NoC	Total arrests	% arrests	% of all arrests
Red 1	20622	3353	16.3%	68.2%
Choking	1661	22	1.3%	0.4%
Drowning/Water incident	208	9	4.3%	0.2%
Unconscious (Not Red1 scenarios)	32425	419	1.3%	8.5%
Allergic Reaction	4400	0	0.0%	0.0%
Bleeding	36538	18	0.0%	0.4%
Breathing	91136	240	0.3%	4.9%
Chest Pain/Cardiac Prob/Back Pain (Upper)	91823	124	0.1%	2.5%
Collapse	16222	64	0.4%	1.3%
Diabetic Probs	7864	5	0.1%	0.1%
Fall Injuries Unknown	26398	23	0.1%	0.5%
Fitting	25039	48	0.2%	1.0%
Major Incident Declared	3	0	0.0%	0.0%
Maternity	5021	3	0.1%	0.1%
PLATO	0	0	0	0
RTC Roll Over	1377	6	0.4%	0.1%
Running Red	4160	3	0.1%	0.1%
Stroke/Neurological	30040	48	0.2%	1.0%
Trauma Major	3771	12	0.3%	0.2%
Medical	65247	62	0.1%	1.3%
Electrocution/Shock	515	0	0.0%	0.0%
Abdominal/Flank Pain	37157	12	0.0%	0.2%
Air Incident	173	0	0.0%	0.0%
Alcohol Related	5362	1	0.0%	0.0%
Bomb Threat	29	0	0.0%	0.0%
CBRN	13	1	7.7%	0.0%
Concern For Welfare	17066	48	0.3%	1.0%
Death expected <18	13	5	38.5%	0.1%
Death unexpected all ages	588	201	34.2%	4.1%
Explosions	22	0	0.0%	0.0%
Fire Persons Reported	614	1	0.2%	0.0%
Fireams	681	0	0.0%	0.0%
Flooding	6	0	0.0%	0.0%
HAZCHEM	254	0	0.0%	0.0%
Headache	5005	0	0.0%	0.0%
Heat/Cold Exposure	108	0	0.0%	0.0%
Marine Incident on Coast	207	0	0.0%	0.0%

	Total of NoC	Total arrests	% arrests	% of all arrests
Multiple Casualty Event	419	0	0.0%	0.0%
Overdose	16486	14	0.1%	0.3%
Rail Incident	15	0	0.0%	0.0%
RTC	11322	21	0.2%	0.4%
Section 136	468	0	0.0%	0.0%
Stabbing	793	4	0.5%	0.1%
Suicide	4568	13	0.3%	0.3%
Assault	7855	2	0.0%	0.0%
Marine Incident at Sea	801	0	0.0%	0.0%
Mental Health	12495	1	0.0%	0.0%
Burns	1316	0	0.0%	0.0%
Trauma	86619	19	0.0%	0.4%
Back Pain (Lower)	9003	1	0.0%	0.0%
Death expected >18	214	18	8.4%	0.4%
Eye Problems	523	0	0.0%	0.0%
Fall Non-Injury	20659	12	0.1%	0.2%
Fire Request To Standby	191	0	0.0%	0.0%
HCP	21225	5	0.0%	0.1%
Major Incident Standby	23	0	0.0%	0.0%
Medical Minor	54726	21	0.0%	0.4%
Social	806	1	0.1%	0.0%
Telecare				
Pathways Overridden	49064	0	0.0%	0.0%
111 / Unknown	13227	50	0.4%	1.0%
Total	845350	4913	0.6%	100.0%

Appendix 2 – Staff survey questionnaire Phase 1

Appendix 3_Survey 1.pdf

Appendix 3 – Call category review process

Ambulance Response Programme

Emergency call review

Janette Turner

ScHARR, University of Sheffield

On behalf of the emergency call review group

Helen Daly (NHS England)

Peter Fox (NHS Pathways)

Anil Gil (NHS Pathways)

Mona Johnson (NHS Pathways)

David Macklin (Yorkshire Ambulance Service NHS Trust)

Tracy Rayment-Bishop (West Midlands Ambulance NHS FT)

Susan Tuckett (South West Ambulance Service NHS FT)

Darren Worwood (NHS Pathways)

Fenella Wrigley (London Ambulance Service NHS Trust)

Sue Watkins (London Ambulance Service NHS Trust)

1. Background

The current time-based ambulance response standards have been effective in driving improvements and maintaining response times to the most critically ill and injured patients.
However, efforts to comply with these standards in the face of steadily rising demand have led to a range of operational behaviours that appear increasingly inefficient, and which have the potential to create a system with unevenly distributed clinical risk.
The NHS England Ambulance Response programme (ARP) has been established to review ambulance response performance standards and explore strategies that can reduce operational inefficiencies whilst focussing on clinical need to maintain a very rapid response to the most seriously ill patients, reduce overall clinical risk in the ambulance system and improve the quality of care (effectiveness, safety, experience) for patients, their relatives and carers. One of these strategies is to revisit the current ambulance response categories.

The majority of patients currently coded as Red 2 do not derive clinical benefit from the arrival of an ambulance resource within 8 minutes, however the Red 2 target leads to a range of behaviours that undermine the efficiency of the ambulance service with substantial variation in patient care and unevenly distributed clinical risk. As part of the Ambulance Response Programme (ARP) a commitment was made to review the Red 2 code set to determine whether the current system can be improved and trial possible changes during spring 2016 within the current ARP evaluation plan.

Early discussion with the ARP Expert Reference Group (ERG) concluded that in order to effectively review the current Red 2 code this could not be done in isolation and instead all ambulance dispositions within both AMPDS and NHS Pathways should be reviewed. We have therefore started with a “blank sheet of paper” and developed a call review process that, where possible, takes an evidence-based approach byutilising existing data, collecting new data and applying and analysing this data in a transparent and consistent way.

A clinical sub-group was formed to conduct the call review on behalf of the ARP with input and oversight from the Programme Expert Reference Group. The call review process has been conducted using the following steps:

1. Creation a new set of definitions for emergency calls based on condition/symptom type, clinical response needed (assessment; treatment/management; transport) and operational response required.

2. A review all call descriptors (AMPDS code or final NHS Pathways symptom group and symptom discriminator[SG/SD] combination) and a category from the revised definition framework set out in(1) allocated to each descriptor.

3. A validity check of the assigned categories using available data to establish, as far as possible, the match between clinical acuity and need for the call descriptor and assigned category using available descriptive call data. Where there is insufficient data to confidently assign a category to a call descriptor a higher response category has been allocated until sufficient data becomes available.

4. A check of the proportions of calls within each type of call assessment/prioritisation system assigned to each of the new defined categories to ensure that the same type of patient is assigned to the same category regardless of the system used.

5. Agreement of the final call descriptor & category assignment with the ARP Expert Reference Group.

2. Call review methods

2.1 Development of new call category definitions

Definitions for a new set of call categories in terms of clinical and operational response required were developed collaboratively by the Expert reference Group. We started with the premise that clinically and operationally it makes sense to divide 999 patients into the following categories:

Category X: Life threatening. The patient needs immediate treatment at the scene to treat or preserve life where life can be saved. Mortality rates are high; a difference of one minute in response time is likely to affect outcome and there is evidence to support the fastest possible response. Example: cardiac arrest.

Category Y: Emergency. The patient needs an emergency response. Mortality rates are lower, however a difference of fifteen minutes is likely to affect outcome and there is evidence to support early dispatch. This group can be divided into:

Y1: Patient treatment (at scene) is a priority, and they may or may not need subsequent transportation. Example: hypoglycaemia.

Y2: Patient transportation is a priority because they require the services of a specialist facility. Example: stroke.

Y3: Patient assessment is a priority with treatment and referral where appropriate so that transport to ED is not needed.

Category Z: Urgent. The patient needs an urgent response. Mortality rates are very low or zero, however a difference of one hour or more might affect outcome. There is evidence to support alternative models of care. This group can be divided into:

Z1: See and transfer (ideally to a non A&E destination)

Z2: See and treat (discharge at scene +/- referral to other services)

Z3: Hear and treat (resolve via telephone +/- referral to other services)

These initial definitions provided the framework used to guide the subsequent call review processes. The final agreed version of the categories and their definitions (having taken account of modifications made during the call review process) is provided in Appendix 1.

2.2 Mapping call descriptors to revised category definitions and supporting data

Ideally the process of allocating revised category types to call descriptors would be conducted using patient outcome information but access to relevant data is limited. Instead we have utilised a range of proxy indicators that may reflect clinical acuity to support the decision making process on allocating call types to category.

The London Ambulance Service has previously used a similar approach to reviewing call categories and we have used the same set of indicators for this review. The indicators used to assist allocation of a revised categoryfor each call descriptor (AMPDS code or final NHS Pathways SD/SG)are:

- Total number of incidents and proportion of total call volume
- Total percentage of patients conveyed to hospital i.e. 'see and convey'
- Total percentage of patients not-conveyed i.e. 'see and treat'
- Total percentage of patients managed through ' 'hear and treat'' i.e. no ambulance vehicle response (as an indicator of low acuity)
- Total number of cardiac arrests and as a percentage for the specific descriptor (as an indicator of high acuity)
- Total number of patients requiring transport to hospital using a pre-alert or blue lights (as an indicator of high acuity)

Sensible use of the indicators required thresholds to be set and weighted, principally in terms of clinically significant proportions of any given indicator that could be used to support decision making about allocation of a revised code to each call descriptor. The principles used were:

Category X: Dispositions where the patient has a $\geq$ A% chance of either being in cardiac arrest on ambulance arrival, or sustains an EMS witnessed cardiac arrest

Category Y1: Dispositions where the patient has a $\geq$ B% but < A% chance of being in cardiac arrest on ambulance arrival (or sustains an EMS witnessed cardiac arrest) and requires the delivery of drugs and/or paramedic interventions at scene, but is conveyed to hospital in < C% of cases.

Category Y2: Dispositions where the patient has a $\geq$B% but <A% chance of being in cardiac arrest on ambulance arrival (or sustains an EMS witnessed cardiac arrest) and is conveyed to hospital in $\geq$C% of cases with lights and sirens used > D% of the time.

Category Y3: Dispositions where the patient has a $\geq$B% but <A% chance of being in cardiac arrest on ambulance arrival (or sustains an EMS witnessed cardiac arrest) and is conveyed to hospital in $\leq$C% of cases with lights and sirens used > D% of the time.

Category Z1: Dispositions where the patient has a <B% chance of being in cardiac arrest on ambulance arrival (or sustains an EMS witnessed cardiac arrest) and is conveyed to hospital in > E% of cases with lights and sirens used < F% of the time.

Category Z2: Dispositions where the patient has a <B% chance of being in cardiac arrest on ambulance arrival (or sustains an EMS witnessed cardiac arrest) and requires the delivery of drugs and/or paramedic interventions at scenein G% of cases, but is conveyed to hospital in < E% of cases with lights and sirens used < H% of the time.

Category Z3: Dispositions where the patient has a <J% chance of being in cardiac arrest on ambulance arrival (or sustains an EMS witnessed cardiac arrest) and requires the delivery of drugs and/or paramedic interventions at scene in < K% of cases, and is conveyed to hospital in < L% of cases.

The previous work conducted by LAS had set thresholds for some of these indicators and these, together with input from the ARP Expert Reference Group, were used to set initial thresholds for each indicator within each revised category. The initial thresholds and revised thresholds arrived at during the call review process (indicated in blue) are provided in Appendix 2.

As part of the current ARP project data for the set of indicators set out above (call volumes; conveyance; hear and treat; cardiac arrest and pre-alert/ blue light rates) were requested from all 10 ambulance services taking part in the programme. However, given the very short timescales services were limited in how much data they could retrieve and although data was submitted services were not able to provide this at the discriminatory level needed (AMPDS data had to be aggregated by card number and main disposition category Echo-Omega, NHS Pathways data could only be retrieved by Symptom group but not symptom discriminator). This data was therefore used as a "sense check" where needed particularly for example, for overall conveyance rates within a defined clinical population. We did have available a range of different existing datasets from research studies and projects conducted within individual ambulance services to support the call review. The datasets used for each call assessment system are described in more detail in the next section.

2.3 Call Review process

The initial call review process was undertaken in November & December 2015. Two teams independently reviewed all call types and allocated a revised category code to each individual AMPDS Code and SG/SD combination (for NHS Pathways). Decisions on category allocation were made using a range of criteria;

- available evidence and the initial thresholds set for call indicators described in Appendix 2
- Coroners prevention of future death reports
- Consideration of operational factors such as the incidence of indicators within specific call descriptors (where rare conditions that rarely cause cardiac arrest could be allocated to a higher category with minimal operational impact, whereas very common dispositions allocated to a higher category will have an impact on all patients using the service)
- Some calls, regardless of clinical need, an ambulance presence may be expected and needed for reputational reasons, e.g. building fires.
- Potential contextual considerations in addition to condition specific criteria described within individual system codes which may have an impact on the type of response required. For example age groups (responses may be different for the same condition if it is a child or very elderly patient) or conditions where the threat to life may be low but the nature of the condition may warrant a faster response to provide good clinical care (such as pain relief for fractures)

NHS Pathways

Review team – Anil Gill, Peter Fox, Mona Johnson, Darren Worwood (NHS Pathways); Susan Tuckett, (SWASFT); Tracy Rayment-Bishop, (WMAS); Janette Turner, (UoS); Helen Daly (NHSE).

Information sources

- Aggregated Symptom Group (SG) Group data from multiple sites with see and convey, see and treat and hear and treat for each group (April to October 2015). Some additional data was also available for blue in and cardiac arrest numbers
- Symptom group and Symptom descriptor (SD) combinations and numbers of calls assigned to each of current categories (Red 1, Red 2, Green 2, other ambulance, ED or Primary care supplied by SECAMB (April-July 2015) and NEAS (May – November 2015) and
- SWASFT data with see and convey, see and treat and hear and treat blue light and cardiac arrest numbers for each combination
- Aggregated call data from NHS Pathways analytics department

The team from NHS Pathways reviewed the raw data collected from SWASFT and other services Because of the mode of reporting there was some difficulty deriving the SG/SD combinations. Further data was sought from the NHS Pathways data analytics department and they generated a table from all data submitted from NHS Pathways 999 user sites to date. NHS Pathways then examined all the SG/SD/Dx combinations generated and allocate them to a code. This was cross-checked against the data produced by SWAST to ensure that the allocation met the criteria in (appendix 2).

AMPDS

Review team – Dave Macklin, (YAS); Fenella Wrigley, Sue Watkins, (LAS); Janette Turner, Helen Daly

Information sources

- Aggregated data from multiple sites of AMPDS codes (truncated by card number and response level Echo – Omega)and for each code conveyance, hear and treat blue light and cardiac arrest numbers
- LAS data providing individual AMPDS codes and conveyance, hear and treat blue light and cardiac arrest numbers
- Sheffield data (6 months 2013) of AMPDS code and initial clinical impression on scene from ePRF record

Each card was worked through sequentially by number and new categories allocated to individual AMPDS codes using the criteria listed previously.

During the course of both reviews some clear additions to the initial revised categories emerged;

- The category Z3 was originally envisaged as encompassing “hear and treat” but this definition was found to be too restrictive. The NHS Pathways team discriminated between calls that could be closed by advice only (Z3) and calls that required referral on to other providers such as primary care which they defined as P(Z3). Both teams also recognised

there are situations where local arrangements for dealing with problem are in place (e.g. expected death) and an additional Z category was added (Z4) where hear and treat is not applied and no ambulance response is sent. Other types of local arrangement were identified by the AMPDS team and allocated as a new category L. Following subsequent discussion with the Expert Reference Group the consensus decision was made to retain the Z3 category only (rather than add additional categories) but expand the definition to include all calls where an ambulance vehicle response is not needed and reflect that within this category there are a range of alternative options available some of which can be locally determined.

- An additional category S was added – where there is a standard response for specific situations (e.g. fire, explosion, chemical incident) and special response (e.g.HART) is deployed or where for reputational reasons and ambulance response will always be sent.
- The majority of codes were heterogeneous and few were clearly very high or very low conveyance or hear and treat rates (as set out in thresholds table) but the thresholds were useful particularly for cardiac arrest rates.
- For some conditions/symptoms allocation of a new category was difficult without some guidance on expected timescales (e.g. there are possible shifts between X & Y and Y & Z ambulance responses depending on wait times).

A subsequent expert group meeting allocated time frames to the revised categories and these were used in later stages of the review process.

Reconciliation of between category variation

A key issue within the call review was to check the proportions of calls within each type of call assessment/prioritisation system (AMPDS and NHS Pathways) allocated to each of the revised categories to ensure that the same type of patient is assigned to the same category regardless of the system used. The two systems have very different architectures underpinning the assessment function with NHS Pathways having a more complex and less linear questioning approach than AMPDS presenting potential differences in the degree of interrogation available to assess clinical need.

To address this we have assessed what the likely call volumes will be within each category as this will reflect both equitable distribution between systems and the potential operational impact for services.

As a first step we calculated the likely call volumes for each category for AMPDS and NHS Pathways. For AMPDS we used two sources of data – 1)call volume data for each code recorded in the data supplied by LAS and used to assign new category definitions to each code and 2) a supplementary dataset available at Sheffield from a current project examining variation in non-conveyance rates and comprising one month's call data from 4 other services. For NHS Pathways call volumes for each category were calculated using aggregated data from a number of services supplied by the NHS Pathways central team.

We calculated and compared the estimated call volumes for each category for the two systems using these data sources for the initial round of call review allocations. This initial comparison revealed a substantial difference in allocations between the two systems, particularly the highest level of urgency (X). To address this issue two additional pieces of work were undertaken.

1. Each of the review teams independently re-assessed the first iteration of category assignment and adjusted allocations where it was considered to be clinically appropriate and taking account the additional guidance provided by allocation of response time intervals for each category agreed by the Expert Reference Group. The AMPDS code list was also checked for alignment against the revised clinical category groupings developed as part of the development of the new clinical operational model in the Wales Ambulance service Trust (WAST). Recalculated call volumes for each category showed a substantial reduction in variation between systems with much closer agreement for the X category.
2. In order to ensure that both proportions of calls within categories were equitable between systems but also that responses allocated were clinically comparable, a final meeting of both review teams was held to establish clinical consensus on allocation of codes to the revised category definitions.
   The combined review team were;
   NHS Pathways - Dr's Anil Gil, Mona Johnson, Darren Worwood, Peter Fox (NHS Pathways), Dr Andy Smith and Sue Tuckett (SWAST)
   AMPDS – Dr Dave Macklin (YAS) & Sue Watkins (LAS)
   Janette Turner (ScHARR, Observer)

   Particular emphasis was placed on assessing and reviewing that the right response category had been assigned to high acuity and therefore high risk conditions that required the highest level (X) response. The conditions reviewed were:

- Workable cardiac arrest
- Peri arrest
- Unconscious diabetics
- Unconscious overdose
- Fitting now
- Choking
- Sepsis
- Children
- Bleeding
- Breathing problems
- Maternity

Following discussion and agreement adjustments were made to the allocated category for specific codes within these broad clinical categories. For NHS Pathways, volumes within each group could be defined either using the final coding arrived at by the assessment system or the final response made (taking in to account calls where the response type was higher than the system decision). As a starting point we have used the final response decision as this reflects more closely current practice but there remains an opportunity during the subsequent trial period to review the response allocations.

The output from the joint meeting and subsequent revision was complete lists of call codes for each system with every individual code allocated to one of the revised categories.
Consideration was also given to the descriptive "labels" to be attached to the revised call categories (the X, Y, Z labels only being used as placeholders during the review process). In particular review group members with operational expertise felt the 1-3 distinctions within each category may cause some confusion for EOC/clinical hub staff as these could be misinterpreted as levels of urgency (i.e. that Y1 takes precedence over Y3 when the urgency is the same and the difference is the required clinical response). This was discussed at the subsequent expert reference group and a decision made to a) rename the category headings as Red, Amber and Green as these are easily understood, are currently used and align with the WAST model, and b) replace numbers with letter suffixes (R – Clinical & possible conveying response; T – Transport only; F – Face to Face assessment only primary response; H – Hear & treat or no ambulance).
These revised descriptors are provided in Appendix 1.

Call volumes for each category have been calculated again following the final review meeting using the previous data (LAS, other AMPDS services and NHS Pathways national data) with an additional 1 year of call data from Yorkshire Ambulance Service (YAS) (Table 1).

Table 1 - Proportions of call volumes within revised categories

New category	AMPDS sites			
	LAS	**YAS**	**Other**	**NHS Pathways (national)**
All X(RED)	**56631 (5.7)**	**32223 (6.2)**	**10536 (6.3)**	**20000 (6.6)**
Y1(AR)	415080 (41.9)	159533 (25.3)	62790 (37.5)	65962 (21.8)
Y2(AT)	173545 (17.5)	90043 (17.4)	30144 (18)	13711 (4.57)
Y3(AF)	77749 (7.9)	33878 (6.6)	15084 (9)	40442 (13.47)
Y?				99464 (33.1)
All Y(Amber)	**666374 (67.3)**	**283454 (54.7)**	**108018 (64.6)**	**219579 (73.12)**
Z1(GF)	101640 (10.3)	36740 (7)	19078 (11.4)	675 (0.22
Z2(GT)	30795 (3.1)	81030 (14.4)	7532 (4.5)	36560 (12.2)
All Z1/Z2	**132435 (13.4)**	**117770 (22.7)**	**26610 (15.9)**	**37235 (12.4)**
Z3 (including L) (GH)	132359 (13.4)	47704 (9.2)	21648 (12.9)	23400 (7.79)
All Z(GREEN)	**264793 (26.8)**	**165474 (31.9)**	**48258 (28.8)**	**60635 (20.9)**
S	1688 (0.2)	508 (0.09)	461 (0.3)	74 (0.2)
Unallocated		26952 (5.2)		
Total	989487	518546	167273	300288

LAS data includes calls from NHS 111 allocated a revised category code based on the DX code passed by 111 (e.g. Red 1 response has been allocated to X or RED). Data from other AMPDS sites does not include these calls.

Proportional allocations for current call categories derived from the aggregated data from 10 ambulance services submitted October – December 2015 as part of the ARP are given in Table 2.

Table 2 – Proportions of calls by category for AMPDS & NHS Pathways

Column1	AMPDS	NHSP
Red1	2%	3%
Red 2	46%	43%
Green1	6%	1%
Green2	30%	43%
Green3	4%	1%
Green4	12%	10%

3. Next steps and approval processes

The final agreed lists of codes assigned to revised categories for each system will be presented for discussion and approval to relevant external committee (ECPAG) and for information and checking to NASMED, and NDOG before proceeding to live trial in two services.

Alongside the call review process a number of other issues have emerged that will need consideration and guidance for services in order for them to develop their operational implementation plans. These are;

- Where clinically a lower level response is warranted (e.g. hear and treat – Z3) but the patient is in a public place, a higher level ambulance response may need to be allocated.
- Some calls may need a higher level of response because of environmental factors (geographical location, weather e.g. a patient outside during very cold weather)
- Not all calls have a full assessment and therefore a dispatch code (e.g. 3rd party calls or requests from other emergency services). Common agreement will be needed on how these calls are managed and which category they should be allocated to.
- Currently calls where a resource arrives on scene before an assessment is complete are coded as Red 2. Agreement will be needed on how these calls will be managed within the new category framework.

Discussion is needed on how this should be managed, i.e. are a set of "rules" needed for EOC about circumstances which would need an automatic upgrade to a different type of response and what that response should be that can be used across all services or should this be locally determined.Specific guidance will be produced to support implementation of the new call categories prior to trial in two test sites.

Response time intervals have been agreed by the ERG together with suggested performance targets (Appendix1).

Finally, all services receive calls passed from NHS 111 which are not re-triaged and are currently sent as requiring an ambulance within one of the existing categories (Red1, Red2, Green1). Calls from NHS 111 requiring an ambulance response will need to be mapped to one of the revised category definitions. The NHS Pathways author team will, as part of their system revision, ensure

that revised DX codes created to reflect the new call categories can be mapped between NHS 111 and Ambulance Service EOC/Clinical Hub dispatch.

Revised call category definitions (Phase 2.1)

<table>
<tr><th>Call type definition</th><th>Response</th><th>Resource</th></tr>
<tr><td>Red– Life-threatening(X)
Time critical life-threatening event needing immediate intervention and/or resuscitation e.g. cardiac or respiratory arrest; airway obstruction; ineffective breathing; unconscious with abnormal or noisy breathing; hanging.
Mortality rates high; a difference of one minute in response time is likely to affect outcome and there is evidence to support the fastest response.
Time interval & performance target– 75% within 8 minutes, Ambulance response within 19 minutes</td><td>Defibrillator
Person trained to use defibrillator
Ambulance clinician who can assess and deliver advanced life support</td><td>Operational response plan to deliver fastest suitable resource</td></tr>
<tr><td rowspan="4">Amber– Emergency(Y)
Potentially serious conditions (ABCD problem) that may require rapid assessment, urgent on-scene intervention and/or urgent transport.
Mortality rates are lower; a difference of an extra 15 minutes response time is likely to affect outcome and there is evidence to support early dispatch.
Time interval within 19 minutes
(Call that need conveying clock stop is by the vehicle that actually conveys)</td><td colspan="2">All categories need face to face assessment by a suitably qualified clinician plus</td></tr>
<tr><td>AR(Y1)Assess; Treat; Transport
e.g. Probable MI, serious injury</td><td>Suitably qualified clinician who can assess and treat and vehicle that can transport</td></tr>
<tr><td>AT(Y2)Assess; Transport
e.g. Stroke</td><td>Vehicle that can transport</td></tr>
<tr><td>AF(Y3) Assess; Treat
e.g. Fits; diabetic hyper/hypoglycaemia; overdose; unconscious with normal breathing</td><td>Nearest available resource (any type) with suitably qualified clinician who can assess and treat</td></tr>
<tr><td rowspan="3">Green– Urgent(Z)
Urgent problem (not immediately life-threatening) that needs transport within a clinically appropriate timeframe or a further face to face or telephone assessment and management.
Mortality rates are very low or zero; a difference of one hour or more might affect outcome and there is evidence to support alternative pathways of care.
Time interval – 60 minutes (unless otherwise specified by HCP)</td><td>GF(Z1)Face to face assessment and management which may include transport</td><td>a)Suitably qualified clinician who can assess & manage
b)Transporting vehicle where needed</td></tr>
<tr><td>GT(Z2) Transport only required</td><td>Transporting vehicle with suitable HCP (within specified timeframe)</td></tr>
<tr><td>GH(Z3) Calls which do not require an ambulance responsebut do require onward referral or attendance of non-ambulance provider in line with locally agreed plans or dispositions, or can be closed with advice (Hear & Treat)</td><td>Suitably qualified clinician in EOC who can assess & manage</td></tr>
<tr><td>Type S – Specialist response
Incidents requiring specialist response i.e. hazardous materials; specialist rescue; mass casualty</td><td>Locally agreed plans apply</td><td></td></tr>
</table>

Thresholds for indicators used to support category allocation

New category	Seen and transported N(%for determinant)	Seen & not transported N(%for determinant)	Hear & treat N(%for determinant)	Cardiac arrest N(% for determinant)	ROSC (% for determinant)	Pre-alert (blue lights) of patients conveyed	Pre-alert (blue lights) all calls for determinant
X – Life-threatening	>90%(70) for live patients			>5%	>5% where resus attempted	>50%	
Y1 Assess, treat, transport	>80%(60) for live patients			>1% - <5%		15 %	
Y2 Assess, transport	>80%(60) for live patients			0.1-0.5%		15%	
Y3 Assess, treat		>60%		0.1 -0.5%			<10%
Z1 Assess, treat, transport (longer timeframe)	>60% (50)			<0.1%		<2%	
Z2 Transport only (longer timeframe)	>80% (65)			<0.1%			< 1%
Z3 Closed by Hear & treat	<10%	<10%	>80% (65)	<0.1%			<1%

Phase 2.2 – Additional call category review

A trial of the new categories in operational practice began in two ambulance services in April (SWAST & YAS) and a third in June (WMAS). During this early trial period comprehensive weekly monitoring of performance against the expected time standards and a small number of accelerated clinical outcomes has been carried out. The purpose of monitoring was to ensure clinical safety, assess the operational feasibility and assess call volumes within a real environment so that adjustments could be made as needed.

The early results from the trial sites have shown, as expected, a substantial reduction in the proportion of calls requiring an 8 minute (Red) response (6-8%) when compared to the previous Red 1 and Red 2 proportions of 50% or more. However the group of "Amber" calls have posed considerable operational challenges as these comprise a large proportion of calls (around 70%) which require a 19 minute response (compared to the previous 50-55%) and also the "correct" response in terms of whether a conveying vehicle is needed and therefore stops the clock. Management of a volume of calls of this magnitude, with inevitable variability in clinical acuity and need, is problematic in terms of prioritising dispatch of available resources as there is insufficient clinical discrimination to help dispatchers.

The clinical coding subgroup has begun to address these issues. Review of the call categories has been approached using a number of basic assumptions:

- The underlying premise used for the initial call category review that response should be based on clinical urgency and type of response needed still holds true
- The current assignment of call codes (AMPDS or NHS Pathways SG/SD combination) to the Red category is correct and does not need further review
- The current assignment of call codes to the Amber categories is too large and does not sufficiently discriminate between calls for emergency conditions which require a response within 19 minutes or less and urgent conditions which require a response that could be longer than 19 minutes but need a shorter timeframe than the Green 60 minute target
- The current separation of Amber calls into 3 categories based on need for treatment and transport, transport or face to face assessment is too complex (or at least too difficult to discriminate with a high degree of specificity at the time of the call) and could be improved by a simpler transport or assessment split
- There are some calls where a longer response (for assessment and onward management decisions) would be appropriate
- The current 7 categories have probably introduced a level of complexity that is not helpful in terms of managing call stacks and appropriate allocation of resources in response to clinical need
- To be successful the right balance has to be found between call volumes within categories and the associated time targets and the capability to deliver the expected performance in the current operating environment.

Taking these factors in to account the subgroup revisited the first iteration of call category definitions and created a set of revised categories that may better reflect the response required for different types of conditions. The suggested revisions have incorporated a further differentiation in

time standards to enable better discrimination between different types of calls and a reduction in the number of categories to support operational implementation but retaining the principle of allocating the right type of response rather than any response. The suggested revised categories were discussed and agreed by the ARP Expert Reference Group. The clinical coding subgroup then repeated the process described for the first iteration to allocate AMPDS codes and NHSP DX codes to the revised categories. In summary, the steps taken were:

- Trial sites provided data from the existing coding trial calls to describe the distribution of calls to each of the initial category descriptors and conveyance rates for each AMPDS code and SG/SD combination. This provided the baseline evidence on the need for further revision and assisted in the reallocation of codes to revised categories
- AMPDS and NHS Pathways clinical teams re-allocated codes to the revised categories.
- Estimates were be made on likely proportions of calls within each category using data from trial services on current numbers of calls assigned to AMPDS codes or NHS Pathways DX codes
- Once agreed appropriate changes to CAD systems will need to be made and any additional training for staff carried out
- Estimated date for live implementation of revised categories in trial sites of 1st October
- For the NHS Pathways system the codes that were allocated to "Red" were reallocated to Category 1. There was no further evidence to uplift any of the codes allocated to "Amber" to Category 1 at that time. Codes commencing with Dx011 in the NHS Pathways system were allocated to Category 2. The trial to date had demonstrated that these cases could be acceptably managed within a 19 minute response framework. Work to consider the type of response had already demonstrated whether transport was required in these cases and the codes were appropriately allocated to the subcategories – the majority of codes requiring transportation to a higher level of care. Evidence from ambulance services suggested that calls that were coded in the Dx012-family had safely waited for up to 45 minutes. This supported the decision allocate these codes to Category 3. These codes in general had a higher proportion of "treat on scene" outcomes and were mapped to the relevant Category 3 subcategory. A small number of other call types for specific for conditions that require a response within 1 hour were uplifted from "Green" to Category 3. Previously reported evidence showed that the majority of other Green codes were being responded to in longer timeframes than 1-2 hours and these were mapped across to the relevant Category 4 response. During the time of the trial, additional codes were created to allow fractionation within the code subcategories and movement within these according to the responses required. For AMPDS a similar process using data from phase 2.1 on conveyance rates and waiting times was used but, as in phase 2.2, individual AMPDS codes were mapped to relevant categories rather than the broad groups of calls that the NHS Pathways DX codes allow.

As in Phase 2.1, the call types assigned to each revised category were reviewed by NASMED and NDOG and after adjustment the final agreed sets of codes and allocated categories were approved by ECPAG in September 2016 so the Phase 2.2 trial could commence in October 2016.

Revised call category definitions (Phase 2.2)

Call type definition	Response and Resource
Category 1 -Life-threatening Time critical life-threatening event needing immediate intervention and/or resuscitation e.g. cardiac or respiratory arrest; airway obstruction; ineffective breathing; unconscious with abnormal or noisy breathing; hanging. Mortality rates high; a difference of one minute in response time is likely to affect outcome and there is evidence to support the fastest response. **Time interval & performance target– 75% within 8 minutes, Ambulance response within 19 minutes**	Defibrillator Person trained to use defibrillator Ambulance clinician who can assess and deliver advanced life support Transporting vehicle where transport required Operational response plan to deliver fastest suitable resource
Category 2 - Emergency Potentially serious conditions (ABCD problem) that may require rapid assessment, urgent on-scene intervention and/or urgent transport. Mortality rates are lower; a difference of an extra 15 minutes response time is likely to affect outcome and there is evidence to support early dispatch. **Time interval 19 minutes** **(Calls that need conveying clock stop is by the vehicle that actually conveys)**	**C2T** Assess; Treat; Transport e.g. Probable MI, serious injury, stroke, sepsis, major burns Suitably qualified clinician who can assess and treat **and** vehicle that transports where needed
	C2R Assess; Treat e.g. Fits; unconscious with normal breathing Nearest available resource (any type) with suitably qualified clinician who can assess and treat
Category 3 – Urgent Urgent problem (not immediately life-threatening) that needs treatment to relieve suffering (e.g pain control) and transport or assessment and management at scene with referral where needed within a clinically appropriate timeframe. Mortality rates are very low or zero; a difference of one hour or more might affect outcome and there is evidence to support alternative pathways of care. **Time interval 40 minutes (Calls that need conveying clock stop is by the vehicle that actually conveys)**	**C3T** Assess; Treat; Transport e.g. serious injury modalities without systemic compromise; burns (not major); non-emergency late pregnancy/childbirth problems.
	C3R Assess; Treat Calls within scope of advanced clinical practice and suitable for treat and leave. E.g. uncomplicated diabetic hyper/hypoglycaemia; not immediately at risk drug overdoses; non-emergency injuries; abdominal pain.
Category 4 – non-urgent Problems that are not urgent but need assessment (face to face or telephone) and possibly transport within a clinically appropriate timeframe. **Time interval 90 minutes to complete assessment in person or by telephone. Onward management is locally agreed including transport times for HCP calls**	**C4T** Assess; Treat; Transport • 999 or 111 calls that may require a face to face ambulance clinician assessment • Requests for transport by health care professionals
	C4H Non-ambulance response Calls which do not require an ambulance response but do require onward referral or attendance of non-ambulance provider in line with locally agreed plans or dispositions, or can be closed with advice (Hear & Treat)
Type S – Specialist response Specialist response incidents i.e. hazardous materials; specialist rescue; mass casualty	**Locally agreed plans apply**