Ambulance Response Programme Review

NHS England Gateway Publications Reference: 08296

<Blank Page>

Equality and Health Inequalities Statement

Promoting equality and addressing health inequalities are at the heart of NHS England's values. Throughout the development of the policies and processes cited in this document, we have:

- Given due regard to the need to eliminate discrimination, harassment and victimisation, to advance equality of opportunity, and to foster good relations between people who share a relevant protected characteristic (as cited under the Equality Act 2010) and those who do not share it; and
- Given regard to the need to reduce inequalities between patients in access to, and outcomes from healthcare services and to ensure services are provided in an integrated way where this might reduce health inequalities

Introduction

On 13 July 2017 NHS England announced a new set of performance targets for ambulance services in England which saw standards applied to every 999 call for the first time. This move was made as a result of the findings of the Ambulance Response Programme (ARP); the largest study of ambulance services in the world.

The Ambulance Response Programme (ARP) was established in 2015 as an integral part of the Review. The ARP aimed to increase operational efficiency whilst maintaining a clear focus on the clinical need of patients.

Throughout the ARP we have held three objectives in view:

I. Prioritising the sickest patients, to ensure they receive the fastest response;

II. Driving clinically and operationally efficient behaviours, so the patient gets the response they need first time and in a clinically appropriate timeframe;

III. Putting an end to unacceptably long waits by ensuring that resources are distributed more equitably amongst all patients.

The new targets aim to save lives and remove “hidden” and long waits previously endured by millions of patients, including reducing lengthy waits for the frail and elderly. The new system was backed by the Association of Ambulance Chief Executives, the Royal College of Emergency Medicine, the Stroke Association and the British Heart Foundation amongst many others.

The review of the old system followed calls from paramedics for the modernisation of an approach to ambulance response that was developed and introduced in 1974, as well as criticism from the National Audit Office and Health Select Committee.

The scale of this national change cannot be underestimated. This has been the biggest substantial change in ambulance operating practice in England for 40 years and has required enormous effort from ambulance services to operationalise the required changes. This has involved not only the complex technical challenges required to support new call triage and dispatch processes but also the wider organisational challenges of new working practices for staff, wholesale review of fleet configurations and staff rostering.

To ensure the successful implementation of the Ambulance Response Programme, a group of clinical and operational experts were tasked to monitor and review the logistical, practical and operational issues associated with national roll-out. This group also discussed and reviewed weekly and monthly reporting data for monitoring and safety purposes and were responsible for overseeing the continued evaluation and further development of the programme. They commissioned and developed the ARP Review in conjunction with Sheffield University’s School of Health and Related Research and the Association of Ambulance Chief Executives.

The aims of the ARP Review were to:

- Undertake a review of the implementation of the various initiatives within the Ambulance Response Programme;
- Provide recommendations relating to further development of the programme;
- Provide oversight, analysis and monitoring of safety and performance of ambulance services operating the Ambulance Response Programme;
- Provide clinical expertise as to the recommended time and type of response that is appropriate for specific conditions;
- Ensure the programme delivers coherent outcomes and benefits to patients, ambulance services and the wider Urgent and Emergency Care system.

The ARP Review followed 13 key lines of enquiry. This review report shows that the ARP has been successfully implemented across all ambulance services, and at a time when ambulance trusts were under extreme winter pressure. As before the ARP, there is variation across services in terms of the achievement of expected response standards, and for a small number of ambulance services performance remains a significant challenge. For others performance for Category 1 and 2 calls has been maintained despite high demand and substantial hospital handover delays. For Category 3 and 4 calls, performance remains outside the expected standard for some services, and there is more work to be done to improve performance for patients in these lower acuity categories.

Whole service analysis of all 999 calls has shown that from a population perspective, the substantial revision of response time standards has had minimal impact on the overall service provided to the majority of 999 callers in both the time to arrival of an ambulance, and time to arrival at hospital.

There remain some differences in the proportion of calls assigned to each call category with higher than expected volumes in Category 2. Work is ongoing to better understand this in detail, and a review of the alignment of call types to categories to reduce variation between AMPDS and NHS Pathways triage systems is recommended.

There is some evidence that the new response time standards have increased call volumes from duplicate calls. Services have managed this by introducing call scripts to improve patient awareness and manage expectations. A standardised approach to this public messaging is recommended.

The Ambulance Quality Indicators introduced as part of the ARP have been reviewed and for the most part retained with the addition of a small number of amendments including the addition of a mean response indicator for Category 3. Patient safety remains a core commitment and an additional measure of the number of calls with long waits in each category will be added, initially to the weekly monitoring reporting. A measure of mean and 90th response times for those mental health patients requiring Section 136 conveyance will be reported nationally for the first time. Services will also put in place a formal and standardised process for reporting and reviewing serious incidents.

Analysis of differences in key performance indicators between urban and rural areas shows a complex picture, however for the main 999 population there is no overall disadvantage to predominantly rural areas in terms of the arrival of the first ambulance response.

Following on from the evidence of the ARP Review the ARP Development Group recommends that the following be adopted nationally:

- Reporting of the Category 3 mean against an indicator of 60 mins.
- Reporting of the mean and 90th percentile response times for Section 136 calls.
- Removal of all defibrillator clock stops.
- Mandated use of a Nature of Call/Pre-Triage Sieve question/key words process for all 999 calls in England with associated nationally reported measures.
- Revision of the Clinical Quality Indicators (CQIs) for Stroke, heart attack (STEMI) and out of hospital cardiac arrest (OHCA) to monitor patient care across the full clinical pathway.
- Development of CQIs for sepsis, falls (where the patient remains on the floor) and patients with mental health needs.
- Pilot of a revised and standardised respnse framework to requests for ambulance transport from healthcare professional with additional guidance to follow in autumn, and new standardised guidance for inter-facility transfer requests.
- Standardised use of call handler scripts to support caller expectation of ambulance response times.
- Re-constitution of the Emergency Code Prioritisation Ambulance Group (ECPAG).
- Clarification of the legal position regarding the use of blue lights and sirens.
- Use of an ambulance balanced scorecard to support system wide understanding of ambulance performance.

The revised AQI standards and accompanying guidance for ambulance services can be found at: www.england.nhs.uk/statistics/statistical-work-areas/ambulance-quality-indicators

An overview of performance standards is outlined below:

Category	Headline Description	Sub description	Average Response Target	90th Percentile Response Target
1	Life Threatening	A time critical life-threatening event requiring immediate intervention or resuscitation.	7 minutes	15 minutes
2	Emergency	Potentially serious conditions that may require rapid assessment and urgent on-scene intervention and/or urgent transport.	18 minutes	40 minutes
3	Urgent	An urgent problem (not immediately life threatening) that needs treatment to relieve suffering and transport or assessment and management at the scene with referral where needed within a clinically appropriate timeframe.	None (Mean indicator of 60 minutes)	2 hours
4	Less-Urgent	Problems that are less urgent but require assessment and possibly transport within a clinically appropriate timeframe.	None	3 hours

Ambulance Response Programme Review

Report of ARP phase 2.3 implementation

Janette Turner

Richard Jacques

Centre for Urgent & Emergency Care Research

University of Sheffield

May 2018

Executive summary

The aim of the Ambulance Response Programme (ARP) has been to explore strategies that can support operational efficiency and performance and improve the delivery of high quality care for patients. Over an 18 month period from October 2015 to April 2017 two major changes were made to ambulance service operations – The introduction of additional time to triage 999 calls to enable better dispatch of an appropriate response (Dispatch on Disposition) and a revision of call categories to support provision of responses that are a better fit between urgency, clinical need and appropriate response. After careful piloting and evaluation the decision was made to implement ARP across all ambulance services in England in July 2017 and this was achieved by November 2017.

This has been a substantial change in service delivery and as such these changes have been monitored and reviewed by the ARP development group to assess and identify areas of further work. The ARP Review brings together the outputs from these activities and provides an overview of progress since ARP changes were implemented nationally.

The independent evaluation team from the School of Health and Related Research at the University of Sheffield have conducted a number of key tasks:

1. Performed a descriptive analysis of reported performance against the revised AQI standards.
2. Assessed the proportion of calls assigned to each category in line with those estimated at the time of the call category review.
3. Explored some additional measures and potential unintended consequences based on feedback reported from services
4. Repeated some of the statistical analyses conducted for phase 1 and 2 (whole service performance, and differences in urban and rural performance) with additional services.
5. Reviewed safety issues and monitoring.

The main findings are:

- ARP has been successfully implemented across all trusts and at a time when the services were under extreme winter pressure. As before ARP, there is variation across services in terms of the achievement of expected response standards and for a small number of services performance remains a significant challenge. For others performance for Category 1 and 2 calls has been maintained despite high demand and substantial hospital delays and more recent data shows that, as winter pressures ease, the majority of services are achieving expected performance. For Category 3 and 4 calls, even with longer response standards, actual performance is outside the expected standard in a small number of services and there is more work to be done to improve performance for these categories.
- Proportions of calls in categories 1 and 2 are higher than expected and this is more evident in services using the AMPDS triage system. More detailed work is ongoing to better align call types to the right response category and reduce the volume of Category 2 calls.
- The main unintended consequence has been a possible increase in call volumes from repeat callers where patients or callers are requesting updates on when an ambulance is likely to arrive. This has a potentially detrimental effect on call answering times for new calls. There has been some success in developing call handling scripts to better manage patient

expectations and provide helpful information. There is now a process in place to introduce standardised call handling scripts across all services.

- Whole service analysis of all 999 calls has shown that, from a population perspective, the substantial revision of response time standards has had minimal impact on the overall service provided to the majority of 999 callers with only very small changes in both the time to arrival of a resource and time to arrival at hospital. There is also evidence that further efficiency gains have been achieved by reducing average allocation of resources per call.
- Analysis of differences in key performance indicators between urban and rural areas shows a complex picture but some clear findings that, overall, for the main 999 population there is no disadvantage to predominantly rural areas in terms of arrival of first response and any disadvantage is manifested as increases in predominately urban areas where there is highest demand. Time from call to arrival at hospital is longer in urban with significant rural and predominately rural areas - reflecting longer distances to hospital - but changes in differences after the introduction of ARP are modest. There has been an assumption that patients in rural areas receive a substandard service compared to those in urban areas. The analyses of differences in performance between urban and rural areas showed that this does not hold as a general rule. There are longer times in rural areas in some services but it is equally if not more likely, overall, that times can be longer in urban areas and that the overall proportion of calls originating in urban areas is a bigger driver for differences in response time performance than geographical area.
- Patient safety remains a core commitment and no serious incidents attributable to ARP changes have been reported. Long waits for some patients remain under scrutiny and methods to provide continuous reporting and measurement of the effects of long waits will continue to be developed as ARP moves in to business as usual.
- The Ambulance Quality Indicators introduced with ARP 2.3 have been reviewed and for the most part retained with the addition of a small number of amendments. These will continue to be reported monthly but in addition a balanced score card has also been developed to support ongoing monitoring as ARP becomes “business as usual” so that any major issues can be swiftly identified. This scorecard comprises weekly reporting of call category response standards – with the addition of mean response indicators for Category 3; call answering times; time to arrival of a transporting vehicle for Category 1 calls; proportions of calls managed by different call closure types (hear, treat, convey) and number of calls breaching 2 and 3 times the 90th percentile target; and the number and nature of serious untoward incidents, using a standardised reporting methodology
- The ARP changes have been successfully implemented across all ambulance services in England. It is testament to the hard work and enthusiasm of all services that this has been achieved in a short space of time and in an environment of substantial pressure. The review of implementation shows that for the most part this has been a positive step forward but challenges remain. Some of these may be tackled by further refinement and maturing of phase 2.3 initiatives. Services have reported significant problems over the winter period with queueing calls and no resources available to send. Demand and performance are closely linked, and there will come a point where services have little or no capacity to maintain performance as demand increases. ARP initiatives will have helped to mitigate this to some extent however demand and supply problems need solutions outside those that ARP can deliver if services are to be expected to deliver against the expected standards.

Introduction

The Ambulance Response Programme (ARP) is an initiative established by NHS England. The aim of the programme has been to explore strategies that can support operational efficiency and performance, maintain a rapid response to the most seriously ill patients, reduce clinical risk in the ambulance system and improve quality of care for patients. Over an 18 month period from October 2015 to April 2017 two major changes were made to ambulance service operations – The introduction of additional time to triage 999 calls to enable better dispatch of an appropriate response (Dispatch on Disposition) and a revision of call categories to support provision of responses that are a better fit between urgency, clinical need and appropriate response. Alongside these changes a consensus approach was used to review and refine the existing Ambulance Quality Indicators.

The operational changes were carefully piloted and independently evaluated. The evaluation found the new operating processes resulted in substantial efficiency gains with better use of available resources, stable performance using the new call categories and no identified serious patient safety issues. In July 2017 the Secretary of State for Health approved the ARP interventions for national implementation. Between July and November 2017 all ambulance services in England introduced the revised call categories and began reporting performance against a set of revised quality indicators.

As with any new way of operating a complex service, there was an expectation that processes would be reviewed and revised in the light of experiences as the number of services changing to new ways of working increased. As such, the ARP development group and associated organisations have continued to provide oversight and a comprehensive programme of work to monitor progress and refine guidance for specific activities. The ARP Review brings together the outputs from these activities and provides an overview of progress since ARP changes were implemented nationally.

For this report the independent evaluation team have:

1. Conducted a descriptive analysis of reported performance against the revised AQI standards.
2. Assessed the proportion of calls assigned to each category and compared this to the expected proportions estimated at the time of the call category review.
3. Explored some additional measures and potential unintended consequences based on feedback reported from services
4. Repeated some of the statistical analyses conducted for phase 1 and 2 (whole service performance, and differences in urban and rural performance) with additional services.
5. Reviewed safety issues and monitoring.

Information sources

All 10 services in mainland England have continued to submit weekly data for a range of resource and performance indicators. This data has been supplemented with specific one-off data requests (e.g. rural and urban performance) and the national AQI data for items not included in the weekly data. Individual services have also provided supplementary information to support a description of initiatives undertaken to operationalise ARP changes.

Operational performance following the introduction of revised call categories and Ambulance Quality System Indicators

We have examined response performance in each of the 10 services in relation to a range of time measures and the revised AQI standards. Services implemented the new response model at different times and so results reflect different numbers of weeks in individual services ranging from 35 to 18 weeks spanning the period August 2017 to March 2018. The new standards are:

	Standard (hour:min:sec)
Category 1 Mean 90th centile	 00:07:00 00:15:00
Category 2 Mean 90th centile	 00:18:00 00:40:00
Category 3 90th centile	 02:00:00
Category 4T 90th centile	 03:00:00

Table 1 provides a summary of response times for arrival of first resource for the different call categories. Response times are calculated from call connect for all categories and so include call triage time before clock start (up to 30 seconds for Category 1 and up to 240 seconds for categories 2-4) and any dispatch delays resulting from queueing calls awaiting a resource to become available. From a patient perspective these times reflect the waiting time from their call being made to arrival of help. The results show substantial variation across services for all call categories. For Category 1 calls mean response time from call connect ranges from 4 minutes 18 seconds to 10 minutes 9 seconds. Taking in to account the average 27 seconds to clock start (for AQI reporting) 3/10 services provided a response within the mean 7 minute standard and in 8/10 services 95% of calls receive a response within 20 minutes.

For Category 2 calls mean response time ranges from 7 minutes 52 seconds to 34 minutes 52 seconds. 3/10 services provided a response within a mean time of 19 minutes. Even accounting for the average 2 minutes to clock start for Category 2 calls this only increases to 5/10 services and the same number provide a response within an hour in 95% of incidents.

For Category 3 calls 4/10 services provide a mean response within an hour (taking in to account the average 2 minutes 18 seconds to clock start) and although 9/10 services provide a response of substantially less than an hour for half of calls (median) at the 95th percentile the range shows wide variation from 1 hour 4 minutes to 6 hours 25 minutes. There is a similar picture for Category 4 calls with 8/10 services providing a response in less than 90 minutes for 50% of calls but wide variation from 1 hours 52 minutes to 7 hours 22 minutes at the 95th percentile.

The revised AQI's require services to report mean and 90th centile performance using the clock start times specified following the introduction dispatch on disposition. As such these reported times will more accurately reflect performance against expected standards than times from call connect presented in table 1. The total time from clock start to clock stop (arrival of a resource or transfer for

telephone advice) for mean and 90th percentile standards are summarised for each service in Table 2. For Category 1 calls this shows a smaller range of mean times from 6 minutes 38 seconds to 9 minutes 43 seconds and a 90th centile range of 11 minutes 20 seconds to 17 minutes 22 seconds. 2/10 services had an average mean response time of 7 minutes or less and 5/10 within 8 minutes. 6/10 services provided a 90th centile response with 15 minutes and all within 18 minutes.

For Category 2 – 4 calls response times reflect arrival of a transporting vehicle for patients taken to hospital (in contrast to first resource on scene in table 1). For Category 2, 3/10 services reported response times within the mean and 90th centile standard times, Category 3, 1/10 met the 90th centile standard and Category 4, 2/10 met the 90th centile standard.

The tables reflect a number of features. Firstly, there is substantial variation in performance across different services using the same call categorisation model indicating that changing the call categories alone is not sufficient to instigate service improvement and wider factors are contributing. Secondly they reflect services under pressure that are struggling, even in comparatively "well performing" services to meet expected response standards and 95th and 99th centile times show there remain long waits for some patients. It is important to recognise that for many services the figures presented here reflect average weekly performance over a short period of time covering a period of the year when services are under most pressure from a combination of increased demand, continued problems with hospital delays reducing resource availability and holiday periods that reduce availability of some urgent care services. There is general recognition that recent months have been extremely difficult across the whole NHS and emergency and urgent care system in particular. For all services these events will skew performance so the figures here cannot be regarded as typical. Only at least a full year of data showing operational performance will be able to take account of this seasonal variation and provide a more accurate overview of overall performance against the revised quality standards. Figure 1 shows performance for a more recent week with more "normal" demand and illustrates a much improved picture with the majority of services performing within or close to the expected standards although there are still some long waits in a small number of services, particularly for Category 3 and 4 calls.

The summary response times presented in tables 1 and 2 are also a reminder of some of the cautions set out in the 2017 evaluation report with regard to the limits of effects that efficiency gains created by better call triage and dispatch may have. Services have reported significant problems over the winter periods with queueing calls and no resources available to send. Demand and performance are closely linked and there will come a point where services have little or no capacity to maintain performance as demand increases. ARP initiatives will have helped to mitigate this to some extent but there will come a point beyond which all efficiencies are exhausted and demand and supply problems need solutions outside those that ARP can deliver if services are to be expected to deliver against the expected standards.

Table 1: Summary of average weekly response times for revised call categories in England August 2017 – January 2018

Time from call connect to resource on scene		EMAS	EoE	LAS	NEAS	NWAS	SCAS	SECAMB	SWAS	WMAS	YAS
Cat1	Mean	00:08:53	00:09:16	00:07:46	00:07:21	00:10:06	00:05:05	00:08:39	00:10:09	00:04:18	00:08:12
	50th centile	00:07:37	00:07:51	00:07:00	00:06:32	00:08:20	00:04:16	00:07:24	00:08:26	00:03:49	00:07:16
	95th centile	00:19:25	00:20:03	00:14:39	00:15:01	00:21:23	00:11:38	00:18:21	00:22:20	00:08:48	00:17:09
	99th centile	00:27:51	00:29:33	00:21:57	00:22:35	00:41:49	00:16:43	00:25:27	00:34:09	00:13:03	00:25:15
Cat2	Mean	00:25:53	00:24:33	00:22:34	00:26:39	00:29:50	00:11:33	00:17:57	00:34:52	00:07:52	00:19:09
	50th centile	00:18:14	00:17:28	00:16:15	00:19:35	00:18:39	00:08:15	00:15:03	00:24:49	00:06:53	00:15:04
	95th centile	01:13:05	01:03:17	01:00:06	01:05:41	01:31:34	00:31:31	00:38:57	01:34:55	00:16:21	00:53:21
	99th centile	02:00:05	02:00:00	01:45:38	01:37:33	02:42:07	00:57:46	01:02:18	02:52:06	00:23:56	01:27:37
Cat3	Mean	01:16:29	01:28:05	01:08:57	02:10:20	00:59:54	00:41:58	01:27:00	01:17:06	00:20:25	00:52:29
	50th centile	00:43:48	00:51:12	00:41:12	01:27:30	00:34:26	00:26:23	00:58:36	00:46:55	00:13:48	00:31:19
	95th centile	04:19:40	04:51:26	03:44:41	06:25:35	03:14:18	02:11:16	04:20:04	04:10:11	01:00:04	02:45:29
	99th centile	07:05:59	07:45:22	06:19:37	08:49:10	05:39:00	03:47:15	06:46:24	07:02:21	01:31:03	04:58:37
Cat4	Mean	01:45:46	01:49:21	01:54:59	01:28:20	01:33:24	01:01:23	02:14:43	02:08:27	00:33:14	02:05:43
	50th centile	01:07:27	01:07:49	01:23:43	00:57:35	01:14:27	00:41:55	01:30:39	01:34:08	00:19:07	01:27:12
	95th centile	05:06:46	05:43:06	05:12:13	04:46:37	03:53:14	03:01:58	06:26:54	06:00:44	01:52:35	07:21:58
	99th centile	06:20:15	09:09:39	07:45:20	07:40:41	06:41:31	04:20:17	09:42:40	09:12:13	02:48:16	12:04:05
Cat4 H	Mean	00:24:27	02:32:50	01:44:56	01:11:12	00:45:00	00:00:00	00:53:37	00:24:46	00:07:19	00:38:25
	50th centile	00:12:26	01:58:06	01:17:40	00:59:21	00:30:07	00:00:00	00:21:34	00:18:43	00:05:35	00:23:46
	95th centile	01:23:35	05:57:42	04:50:44	03:14:04	02:14:53	00:00:00	02:51:55	01:09:19	00:17:36	02:01:01
	99th centile	02:11:04	08:27:32	07:40:04	03:25:53	03:50:26	00:00:00	04:41:07	01:48:30	00:27:25	03:55:03

Table 2: Summary of average weekly clock start to clock stop response times for revised call categories in England August 2017 – January 2018

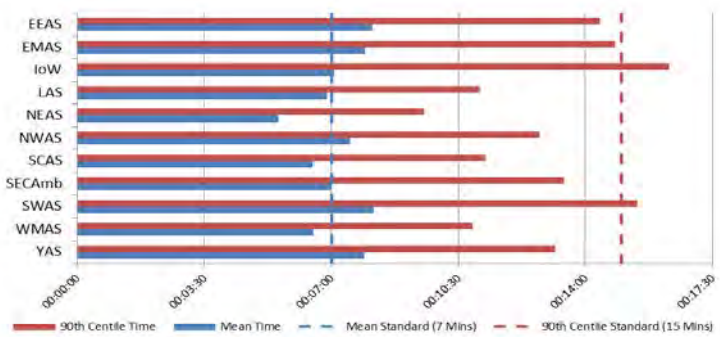
Time from clock start to clock stop		EMAS	EoE	LAS	NEAS	NWAS	SCAS	SECAMB	SWAS	WMAS	YAS
Cat1	Mean	00:08:56	00:08:45	00:07:17	00:06:38	00:09:43	00:07:21	00:08:12	00:09:41	00:06:53	00:07:50
	90th centile	00:15:53	00:15:47	00:11:48	00:11:20	00:16:09	00:13:27	00:14:48	00:17:32	00:11:52	00:13:58
Cat2	Mean	00:34:21	00:27:40	00:22:01	00:23:06	00:31:59	00:17:30	00:17:59	00:32:59	00:12:47	00:24:20
	90th centile	01:14:30	00:56:34	00:45:47	00:47:42	01:12:46	00:35:28	00:33:34	01:08:41	00:23:20	00:53:58
Cat3	Mean	01:27:52	01:24:12	01:08:31	01:37:01	01:04:34	01:04:09	01:27:05	01:15:13	00:36:45	00:57:19
	90th centile	03:29:49	03:31:58	02:42:00	03:50:02	02:32:09	02:30:04	03:18:43	02:58:42	01:24:26	02:12:50
Cat4	Mean	01:28:40	01:38:40	01:15:03	01:24:04	01:19:53	01:34:36	02:15:22	02:06:44	00:57:17	02:09:51
	90th centile	03:23:08	03:58:46	02:32:20	03:21:18	03:01:34	03:34:24	05:14:32	04:44:07	02:26:49	05:09:10

Figure 1: Response time performance for 1 week (11 services)

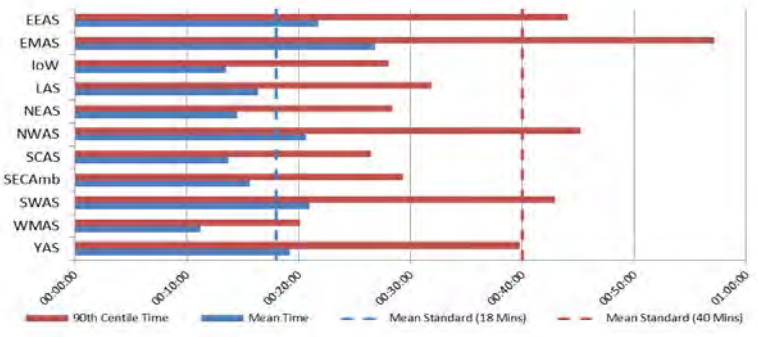
Ambulance Response Programme

Response Times Summary w/c 23rd April 2018

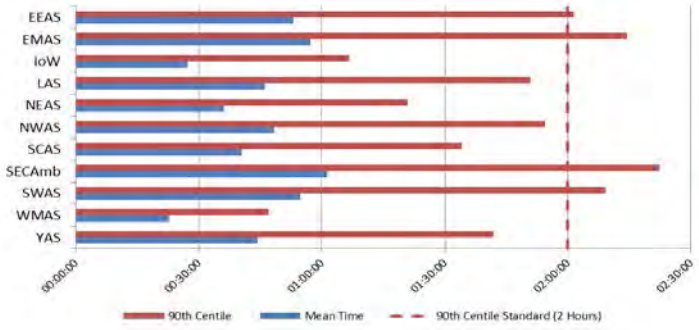
Category 1



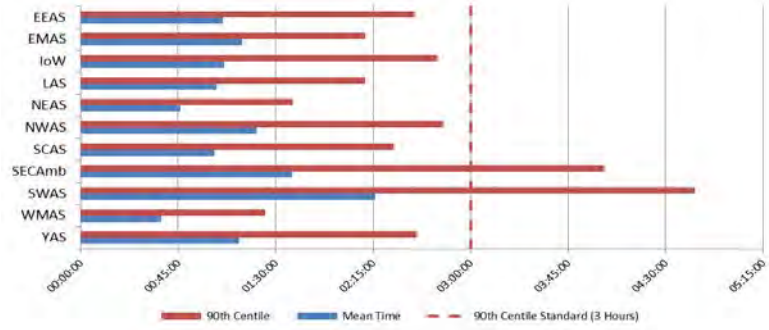
Category 2



Category 3



Category 4



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. In the previous evaluation report we examined trends in response performance and resource utilisation over a 16 month period up to April 2017 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. For the ARP Review we have repeated these analyses using a longer phase 2 period in the 3 original trial sites (up to December 2017) and adding in two additional sites comparing phase 1 and phase 2 performance for the year 2017. These two additional sites have been included because they moved to phase 2 in the middle of 2017 providing comparable before and after periods. The full results of the statistical analyses are available on request. A summary of the results comparing changes between Phase 1 and phase 2 are presented in Table 3. For the original trial sites (SWAST, WMAS, YAS) the results are split by phase 2.1 and 2.2 although phase 2.2 also includes some weeks of 2.3 but this is a relatively short period so we have not specified this as an additional phase as it is likely not long enough and spans the complex winter period to make any useful distinction. The two additional sites (EMAS, NWAS) moved to phase 2.3 reporting in July and August 2017 providing a longer consistent “after” comparison period. Essentially, for all services the overall comparison is between the previous call categories (Red 1, Red 2, Green) and the revised categories. 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 2-10

Table3: Whole service response performance and resource utilisation January 2016 – December 2017 spanning the introduction of phase 2 call categories

Whole service measure	SWAST	WMAS	YAS	EMAS	NWAS
Time from call connect to arrival of first core resource on scene Median	Phase 2.1 Step ↑359.6 seconds Phase 2.2 Slope ↑8.0 seconds per week	Phase2.1 Step ↑70.1 seconds	Phase 2.1 Step ↑ 78.2 seconds	Phase 2 Step ↑179.1 seconds and slope ↑8.1 seconds per week	Phase 2 Step ↑168.2 seconds
95th Percentile	Phase 2.2 ↑53.8 seconds	Phase 2.2 Slope change ↑32.7 seconds per week	Phase 2.1 Slope change ↑28.7 seconds per week		No change
Time from call connect to arrival at hospital (see and convey) Median	Phase 2.1 Step ↑318.0 seconds Phase 2.2 Step and slope change. Net effect Step ↑7.1 seconds/ week	Phase 2.1 Slope change ↓-11.1 seconds/ week Phase 2.2 Slope change ↑4.4 seconds per week	No change No change	No change	No change
95th Percentile	Phase 2.2 Step and slope change. Net effect Step ↑37.5 seconds/ week	Phase 2.2 Slope change ↑20.5 seconds per week	Phase 2.1 Slope change ↑20.5 seconds per week		No change
Time from call connect to leaving scene (see and treat) Median	Phase 2.1 Step and slope change. Net effect Step ↑13.3 seconds/ week	Phase 2.1 Slope ↓-9.5 seconds/ week Phase 2.2 Slope ↑7.9 seconds/ week	Phase 2.2 Step ↓-280.5 seconds	No change	Phase 2 Step ↑821.9 seconds (13.7 minutes)
95th Percentile	Phase 2.2 Step and slope change. Net effect Step ↑62.3 seconds/ week	Phase 2.1 Slope ↓-65.3 seconds/ week Phase 2.2 Slope ↑39.9	Phase 2.2 Slope change ↑21.4 seconds per week		Phase 2 Slope ↑68.6 seconds per week

Whole service measure	SWAST	WMAS	YAS	EMAS	NWAS
		seconds/ week			
Average core resources per incident – all attended incidents	Phase 2.1 No change Phase 2.2 Slope change ↓-0.0011 resources per incident per week	Phase 2.1 step ↓-0.0272 resources per incident Phase 2.2 step ↓-0.0101 resources per incident	P2.1 ↓-0.1150 resources per incident	No change (but this is influenced by substantially variable values before phase 2)	Phase 2 Step ↓-0.026 per incident
Average core resources per incident – all conveyed incidents	Phase 2.1 Step ↓-0.060 resources per incident Phase 2.2 Slope ↓-0.001 resources per incident per week	Phase 2.1 Step ↓-0.043 resources per incident Phase 2.2 Step ↓-0.012 resources per incident	Phase 2.1 Step ↓ -0.1054 resources per incident	Slope ↑0.0022 per incident per week	Phase 2 Step ↓ -0.0390 resources per incident
Average core resources per incident – all see & treat incidents	Phase 2.1 Slope ↓-0.0003 resources per incident per week	Phase 2.2 Step ↓-0.0053 resources per incident	Phase 2.1 Step ↓ -0.027 resources per incident Phase 2.2 Slope ↑0.0009 resources per incident per week	No change	No change

Figure 2: Time from call connect to arrival of first core resource – median

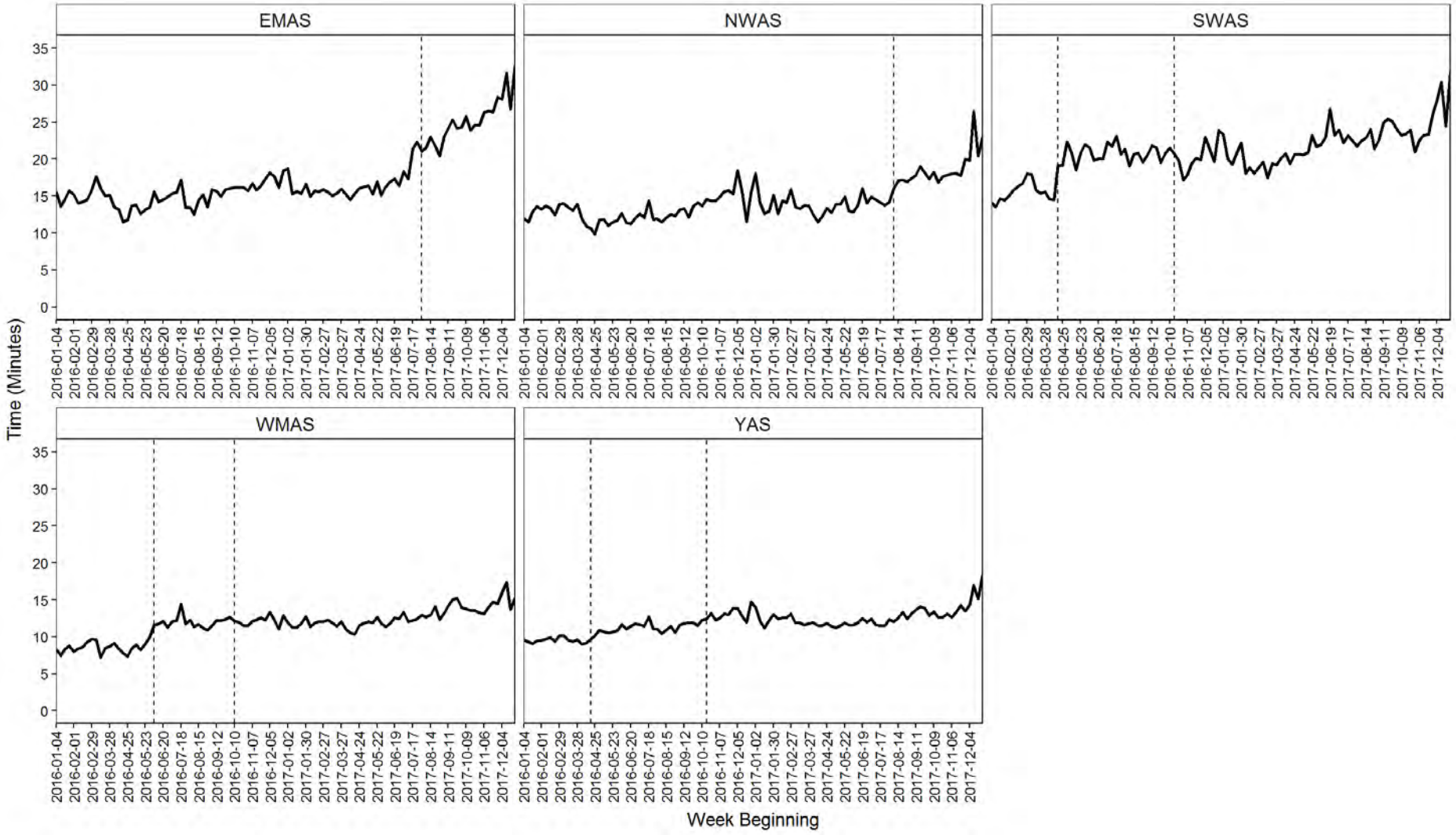


Figure 3: Time from call connect to arrival of first core resource – 95th percentile

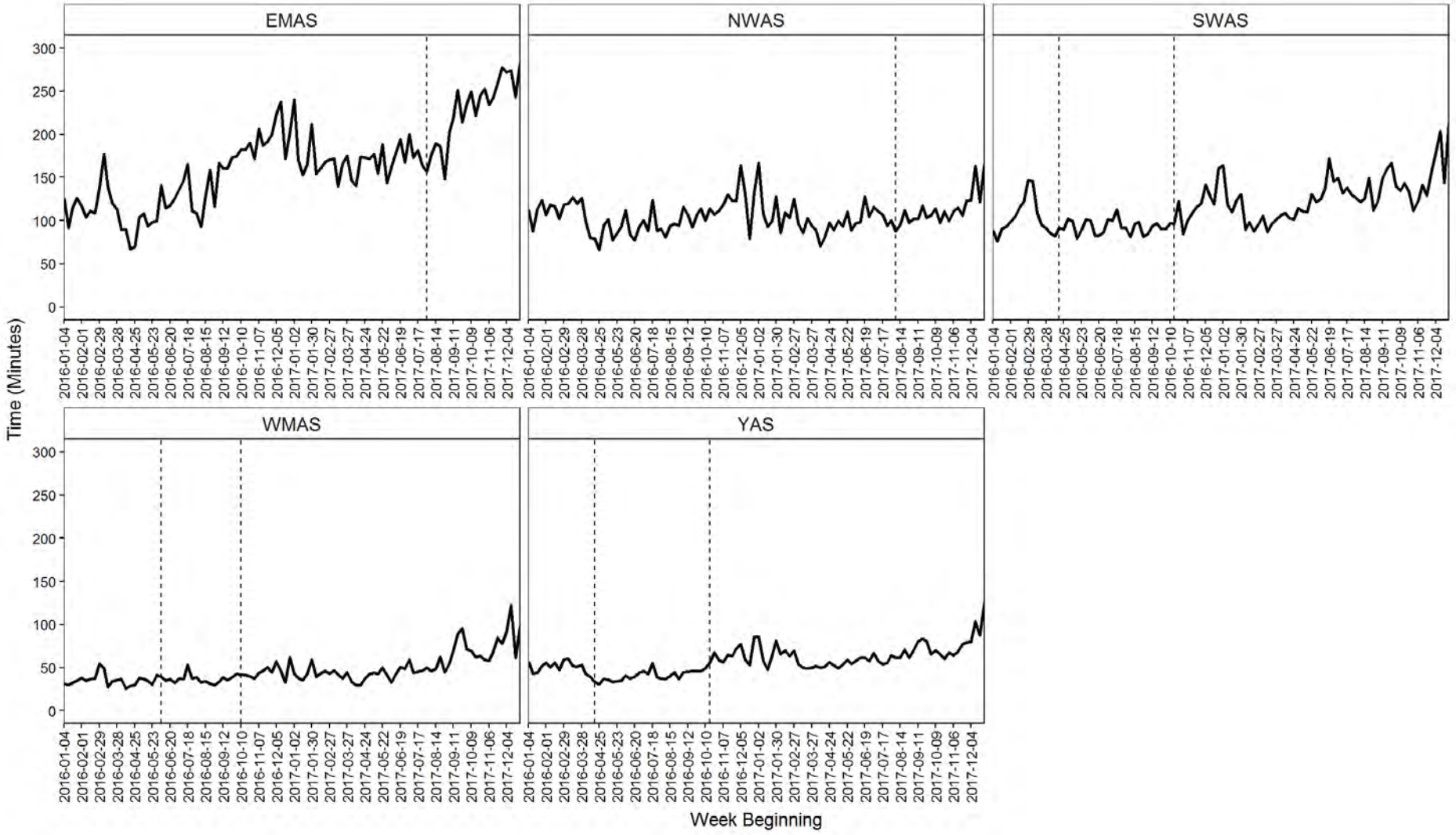


Figure 4: Average Core Resources per Incident – All Attended Incidents

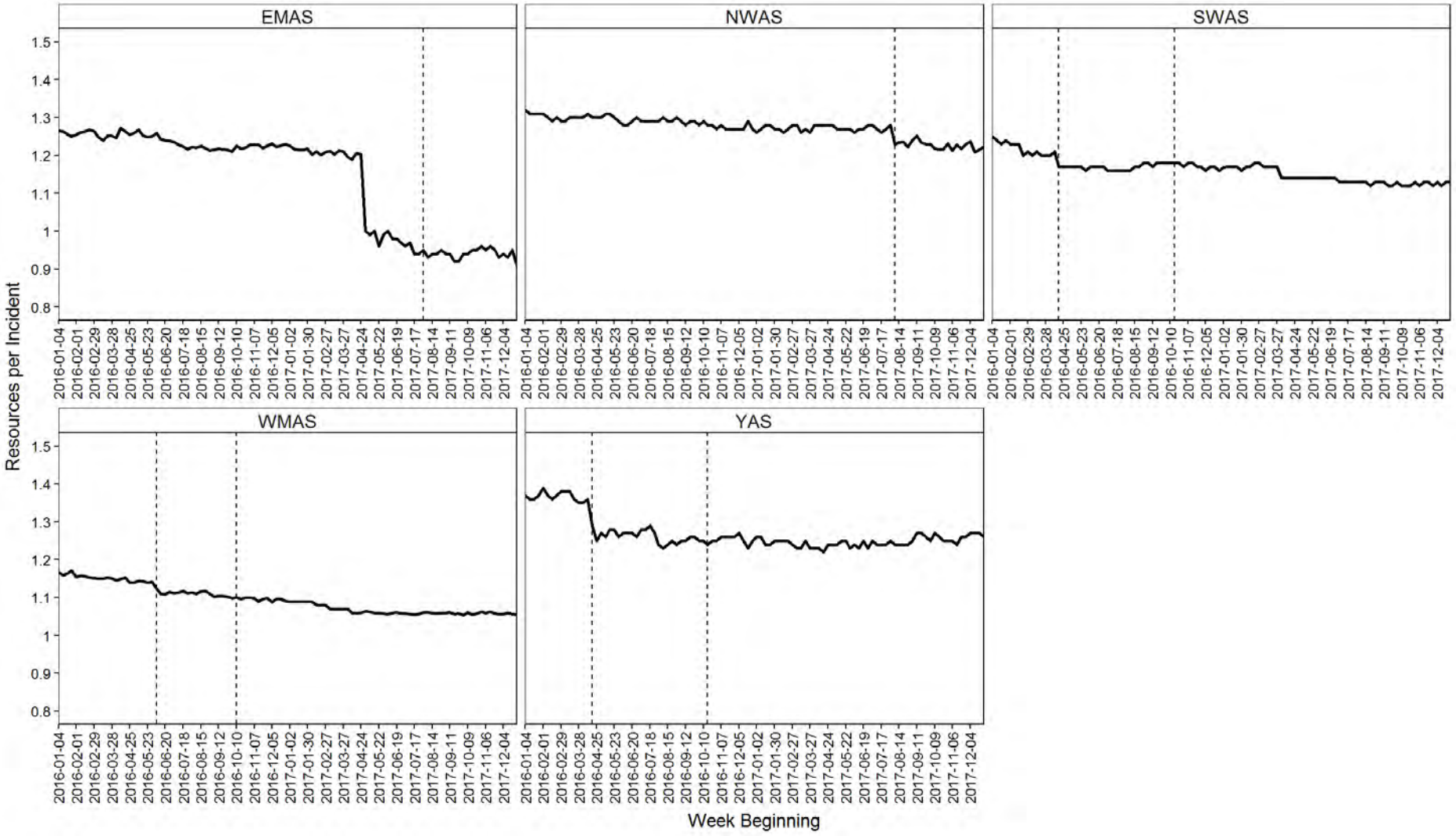


Figure 5: Time from call connect to arrival at hospital – median

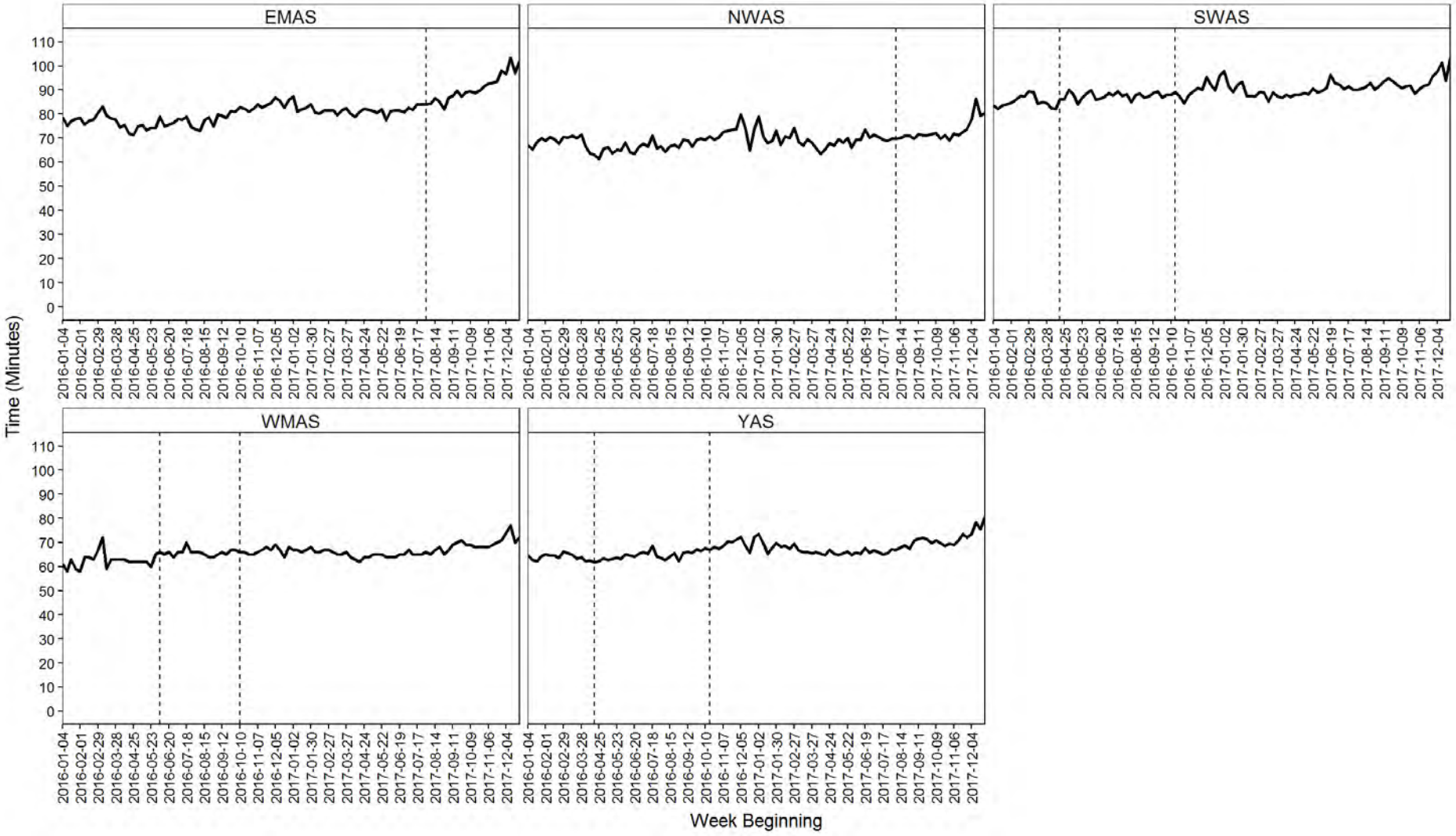


Figure 6: Time from call connect to arrival at hospital – 95th percentile

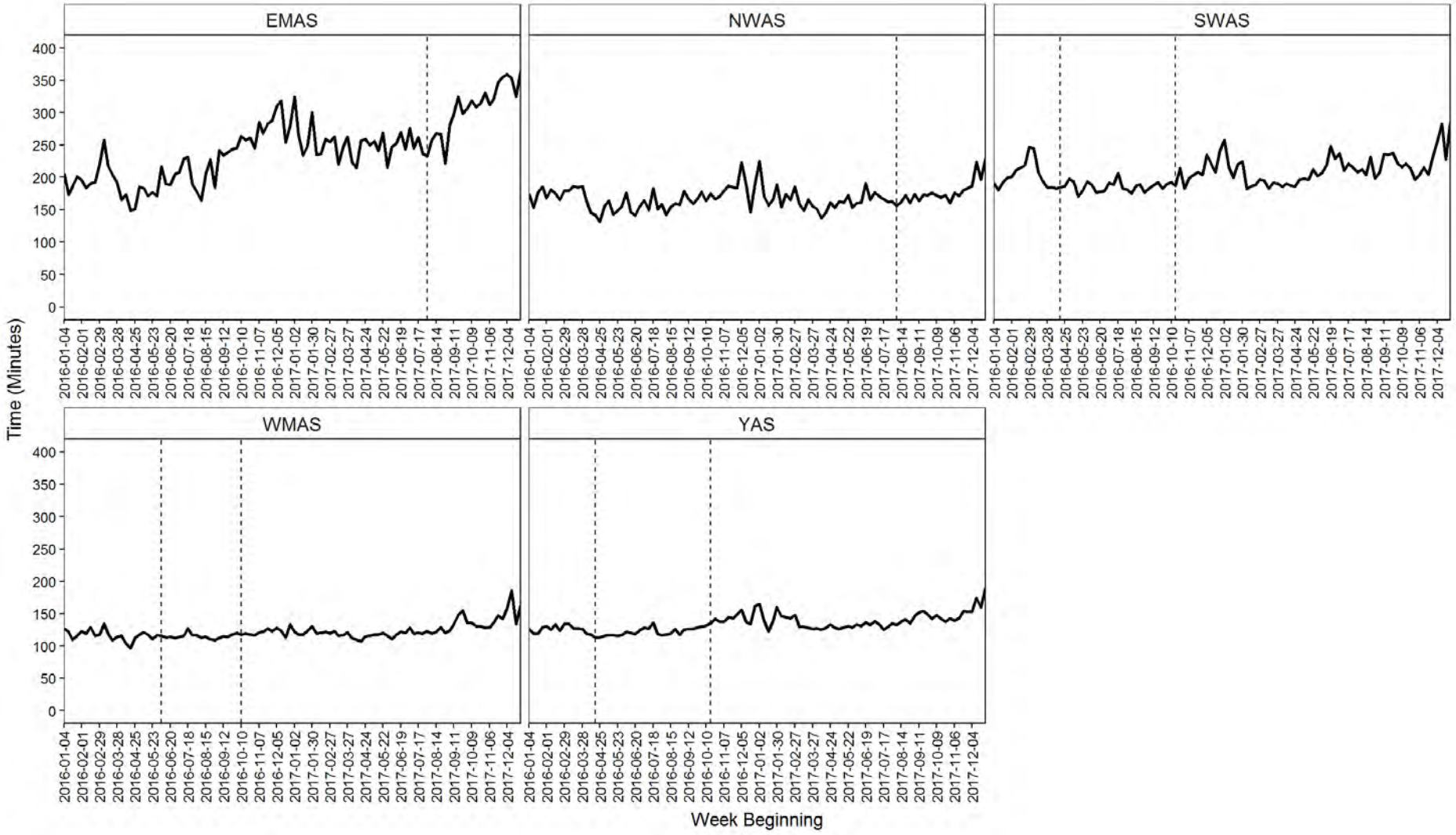


Figure 7: Average Core Resources per Incident – All Conveyed Incidents

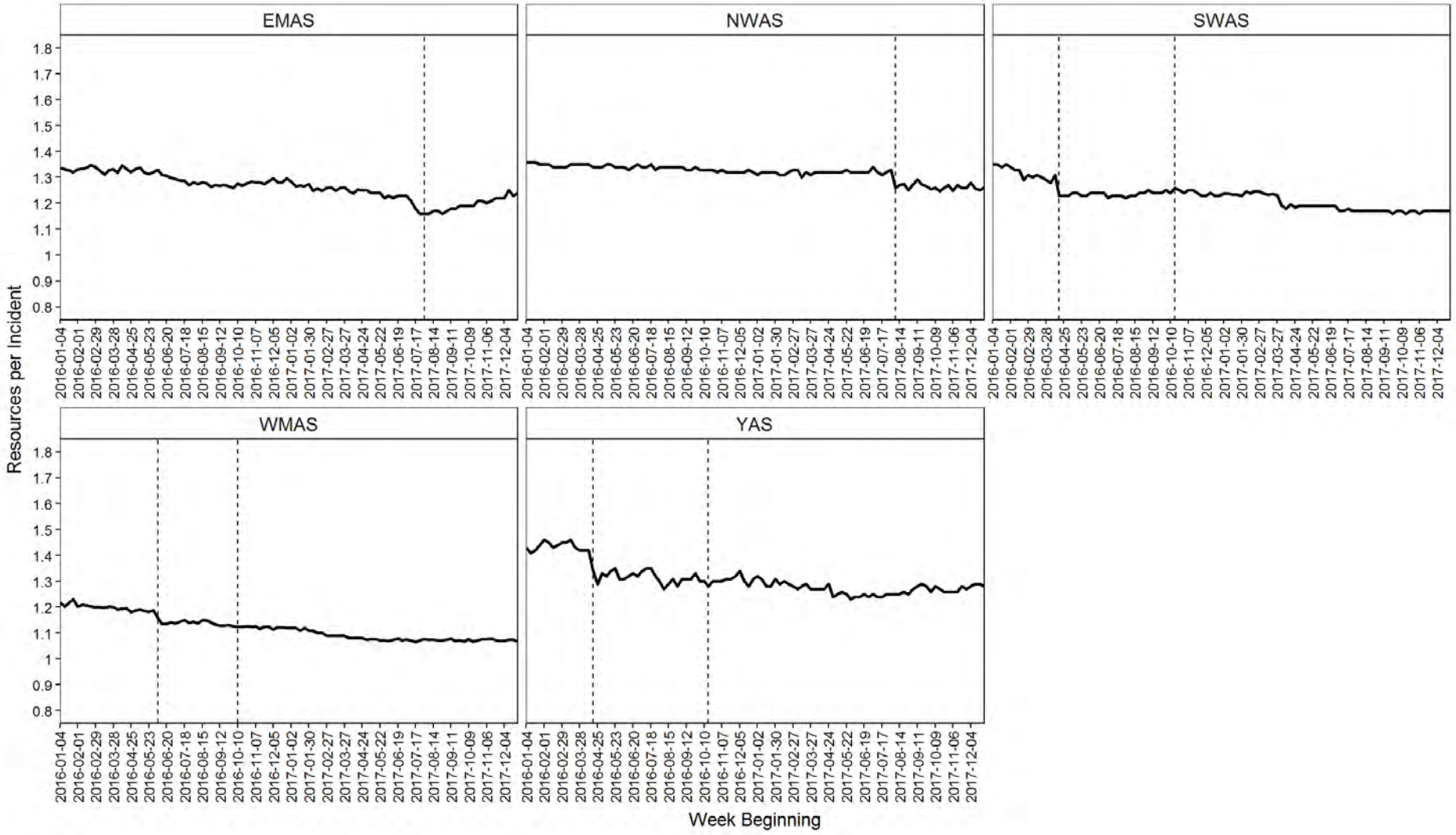


Figure 8: Time from call connect to leaving scene – see and treat – median

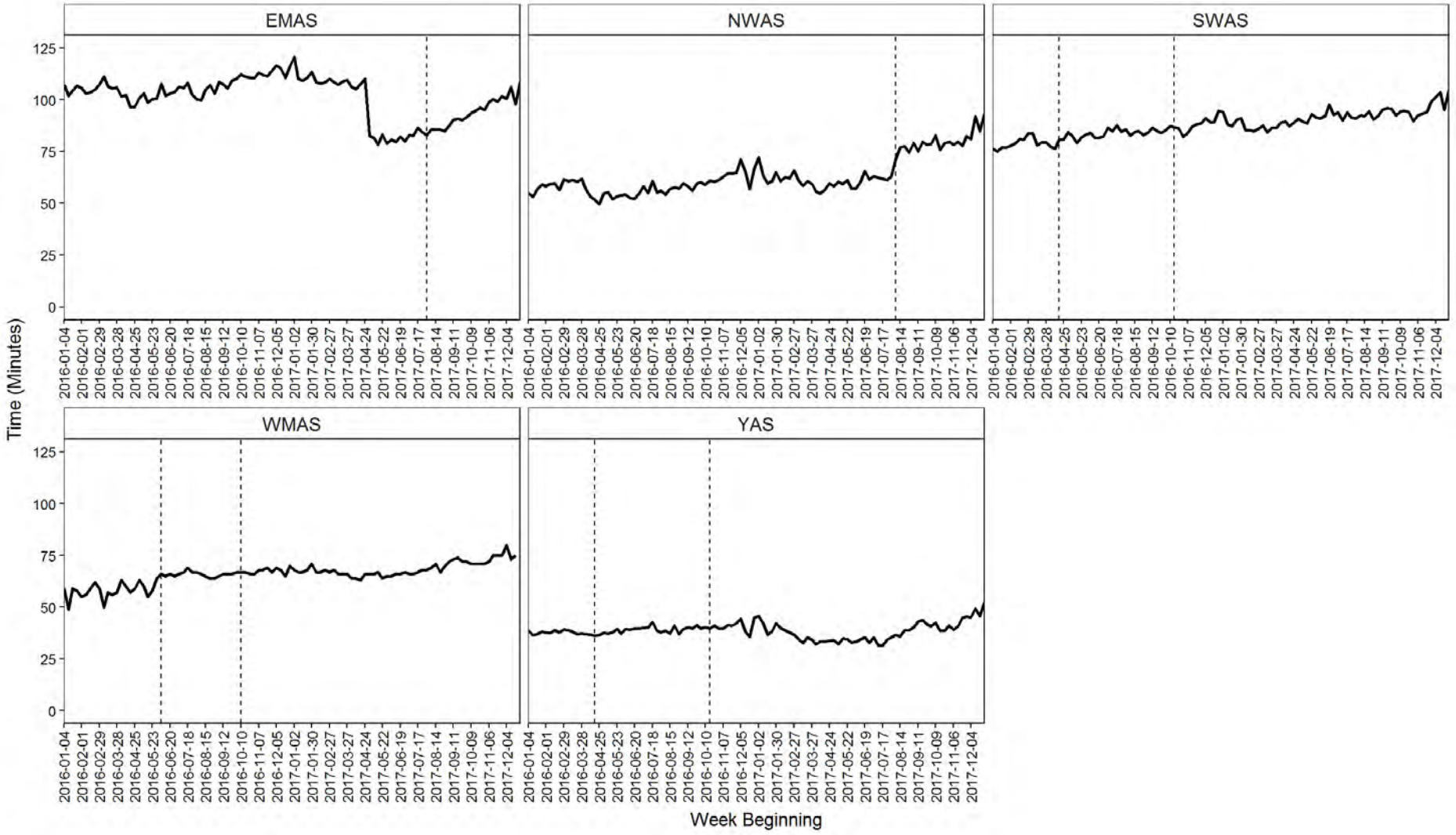


Figure 9: Time from call connect to leaving scene – see and treat – 95th percentile

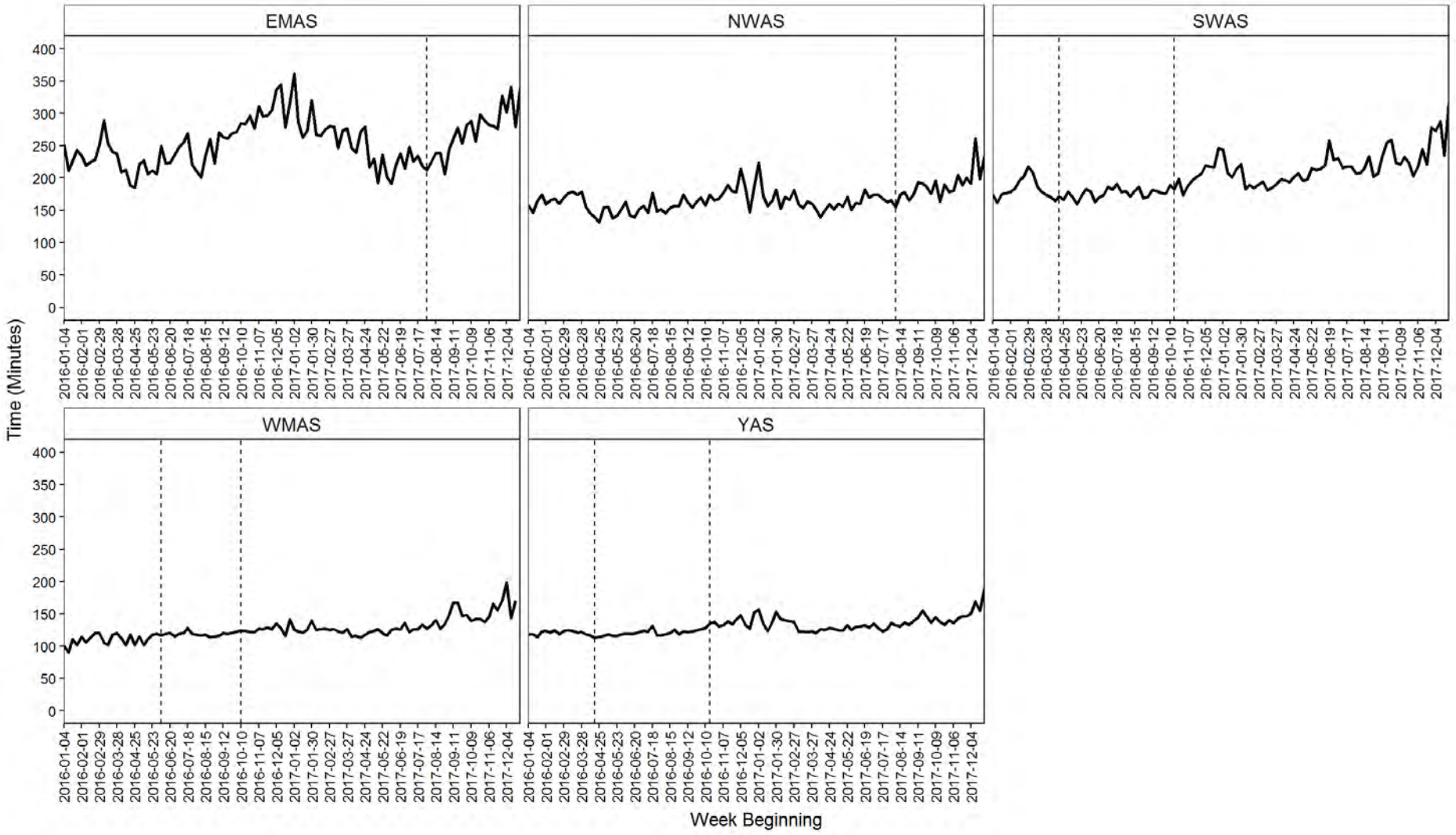
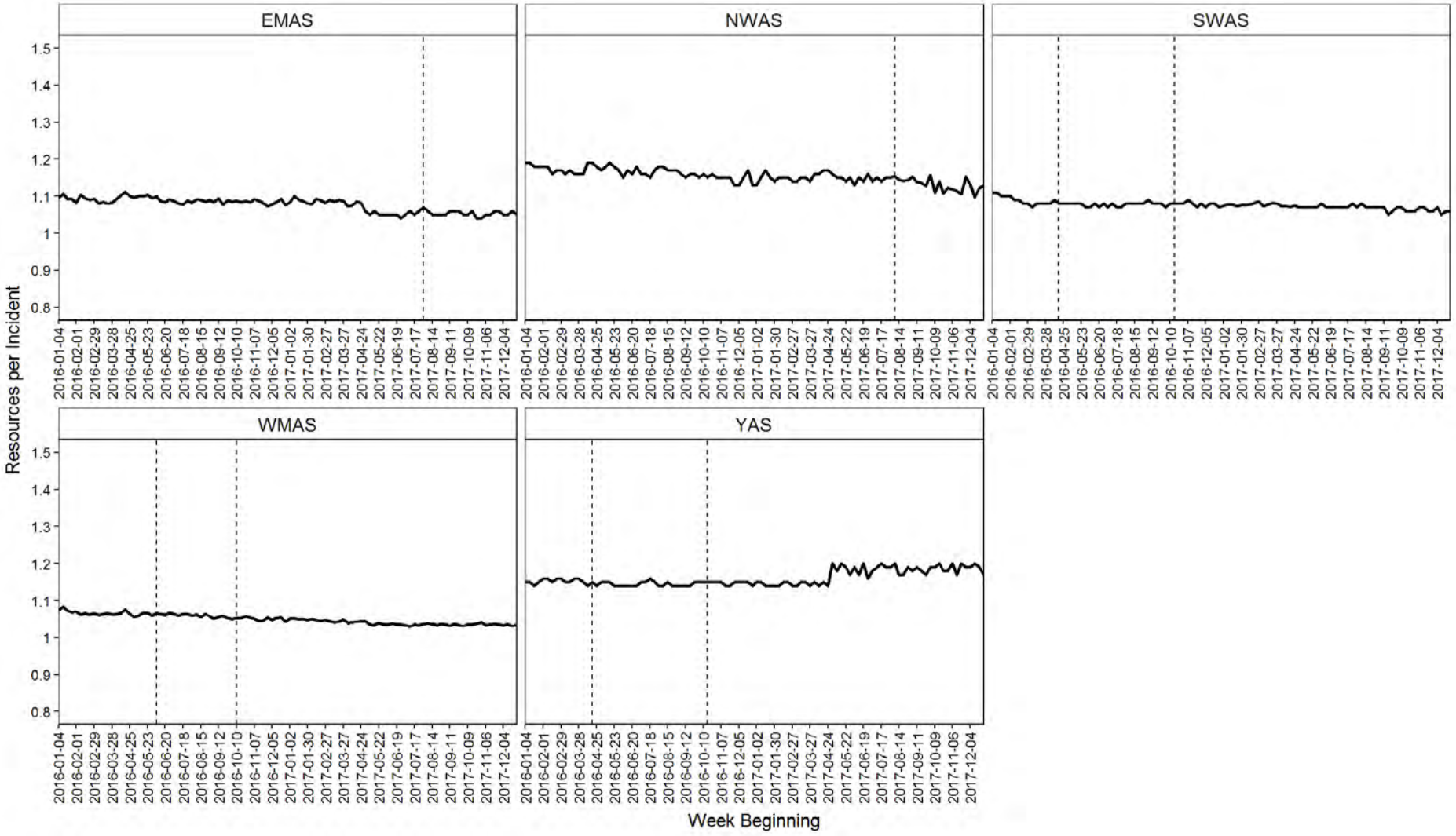


Figure 10: Average Core Resources per Incident – See and Treat



Summary of findings

- There was a step change increase in median time from call connect to arrival of first resource on scene in all 5 services after the introduction of phase 2 (range 70 – 360 seconds). In the 3 early trial sites this occurred immediately after the introduction of the new call categories in phase 2.1 although there was an additional slope change in 1 service after the introduction of phase 2.2. The 95th percentile times show an increase in phase 2.1 in one early trial site and phase 2.2 for the other 2 sites (54-29 seconds). Of the 2 additional sites there was no change in one site. We have not reported the result for EMAS as the statistical model is inconsistent with the performance shown in Figure 3 and is not therefore reliable.
- Median time from call connect to arrival at hospital showed no change in 3 services. In one early trial service there were increases in both phase 2.1 and phase 2.2 and in another there was a step reduction in phase 2.1 but followed by a slope (weekly) increase in phase 2.2. For the 95th percentile the pattern shown for median was repeated with increases in the 3 early trial sites no change in one later site. The EMAS result is not displayed for the same reason given above. In the 3 early trial sites the increases were small (37-20 seconds) suggesting only modest increases in time to hospital despite substantially greater response time standards.
- Median time from call connect to leaving scene was much more variable. In early services this increased overall in one service in phase 2.1 but with no subsequent effect, in one service decreased in phase 2.1 but then gradually increased for phase 2.2 and decreased substantially in one service (by 280 seconds) in one service. Of the two more recent services there was no change in one and a large increase (13 minutes) in the other. For the 95th percentile there was also an increase in this service although smaller (69 seconds). The EMAS result is not displayed. In the more established trial sites there were increases in phase 2.2 in 2 sites (61 – 21 seconds/week) and an early reduction in followed by an increase in phase 2.2. in one site.
- Overall there is evidence of further efficiency gains or no change across most sites as there was a net reduction in allocation of core resources 4/5 services for all attended incidents and all conveyed incidents and 3/5 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%. For phase 2.1 and 2.2 there was more flexibility around response interval for other call categories and with the introduction of phase 2.3 (most pertinent to the EMAS and NWAS figures as they have only reported against phase 2.3 standards) a further expanded set of response time standards with the introduction of 2 hour and 3 hour 90th centile standards for Category 3 and 4 calls respectively. Given this, the median increases in response time for all 999 calls of between 1 and 6 minutes and 95th centile increases of under 1 minute suggests that, from a population perspective, the substantial revision of response time standards has had minimal impact on the overall service provided to the majority of 999 callers. Similarly, there was no change in median time to hospital in 3 services, a modest change in one and a bigger increase (5.5 minutes) another. The 95th percentile increases of under 1 minute indicate that longer response time standards is not increasing overall transport times to hospital. Time to

discharge at scene is much more variable with a large increase in one service and 95th centile increases in most services. Given these are larger than in other categories they may reflect other changes unrelated to the call categories, for example if services are aiming to increase their see and treat rates a consequence may be longer scene times to enable safe discharge. Of course we cannot distinguish these individual effects in this aggregated data.

Overall the introduction of phase 2 has resulted in either no or very modest increases in response time performance for the 999 population comparative to the substantially increased time standards for the revised call categories suggesting broadening the required response time does not translate in to an equivalent increase in overall performance.

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 although smaller in scale but these will be contributing to maintaining the overall service for the population.

The results also reaffirm a feature identified in the Phase 1 and 2 evaluation, which is that the introduction of the revised call categories goes some way to supporting stable performance. The median and 95th centile response performance graphs show that in the more established sites the seasonal peaks expected are less marked than in in the other two sites – the Christmas and New Year weeks for 2016 provide a good example.

One interesting feature of this analysis is the extended phase 2 period in the 3 early trial sites which shows that increases in response performance, time to hospital and time to leaving scene are more likely (though not exclusively) likely to be slope rather than step changes. This means that, rather than immediate effect that is then sustained, trends are gradual but increasing (in some instances reversing an early reduction) and this may be an indication of the margin of gain being reached and services reaching capacity in terms of their ability to respond within existing resources as demand continues to increase.

There are some limitations to these analyses. Problems with data and the ability to establish reliable trends and hence statistical models in one service have already been highlighted. The trend graphs also showed some anomalies that were not always consistent with the statistical findings, for example extreme variations in values or clear step changes at times not consistent with ARP changes. Where appropriate we have conducted sensitivity analyses recalculated removing outliers but these have not changed the overall findings. Step changes illustrate the nature of ARP changes in that they are not simply about revising call categories but require significant operational changes to implement the new model in practice. All services have been initiating and implementing operational reviews of fleet configurations, staff rotas, hospital handover delays etc. as well as broader initiatives around providing more care closer to home which will all have an impact on the performance and resource use measurements we have made. This means that it does become increasingly difficult to distinguish ARP effects from other effects in our statistical models as we cannot account for all the potential changes that may be having an effect.

Comparison of response and call times for urban and rural areas

A criticism of the previous, 8 minute target driven operating model was 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. In the evaluation of ARP Phase 1 & 2 we reported an analysis of response performance in urban and rural areas before and after the implementation of revised ARP call categories in the 3 early trial sites. We found a complex picture of differences which were not always consistent across the 3 sites although there was some evidence that there may be a reduction in inequities for some indicators. There was also an unexpected finding that operational times are sometimes longer in urban areas than other areas. We have replicated these analyses using an extended period of operation for the 3 early trial sites and have added two additional sites. The 3 early sites provided weekly data for the period January 2016 – December 2017 and the 2 additional sites for the year January – December 2017 spanning the transition from phase 1 to phase 2.3. We have measured 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 and 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 available on request. Tables 4 - 6 summarise the difference in times between phase 1 and phase 2. For the 2 new sites (EMAS & NWAS) phase 2 reflects performance using the current (phase 2.3) response time standards. For the 3 earlier sites phase 2 reflects both phase 2.2 (when categories 1-4 were introduced) and phase 2.3 so there is some variation in the expected performance standards for the whole period. We have not treated phase 2.3 separately as these services only began reporting against these standards from September onwards.

Table 4: Comparison of urban and rural call connect to first resource on scene

Call connect to first resource on scene	EMAS	NWAS	SWAST	WMAS	YAS
USR v Predominately Urban					
Median					
Phase 1	USR > PU 136 seconds	No difference	USR > PU 118 seconds	USR > PU 35 seconds	No difference
Phase 2	USR > PU 194 seconds	No difference	USR > PU 257 seconds	USR > PU 34 seconds	PU > USR 106 seconds
95th percentile					
Phase 1	No difference	PU>USR 762 seconds	No difference	No difference	No difference
Phase 2	No difference	No difference	No difference	PU>USR 572 seconds	PU > USR 1115 seconds
Predominately Rural v Predominately Urban					
Median					
Phase 1	PR>PU 488 seconds	No difference	PR>PU 310 seconds	PR>PU 41 seconds	PR>PU 50 seconds
Phase 2	PR>PU 569 seconds	PU>PR 222 seconds	PR>PU 394 seconds	PR>PU 81 seconds	PR>PU 28 seconds
95th percentile					
Phase 1	PR>PU 2022 seconds	PU>PR 1781 seconds	No difference	No difference	PR>PU 604 seconds
Phase 2.	PR>PU 2039 seconds	PU>PR 2929 seconds	No difference	PU>PR 585 seconds	PU>PR 552 seconds

Figure 11: Time from call connect to arrival of first resource - Median

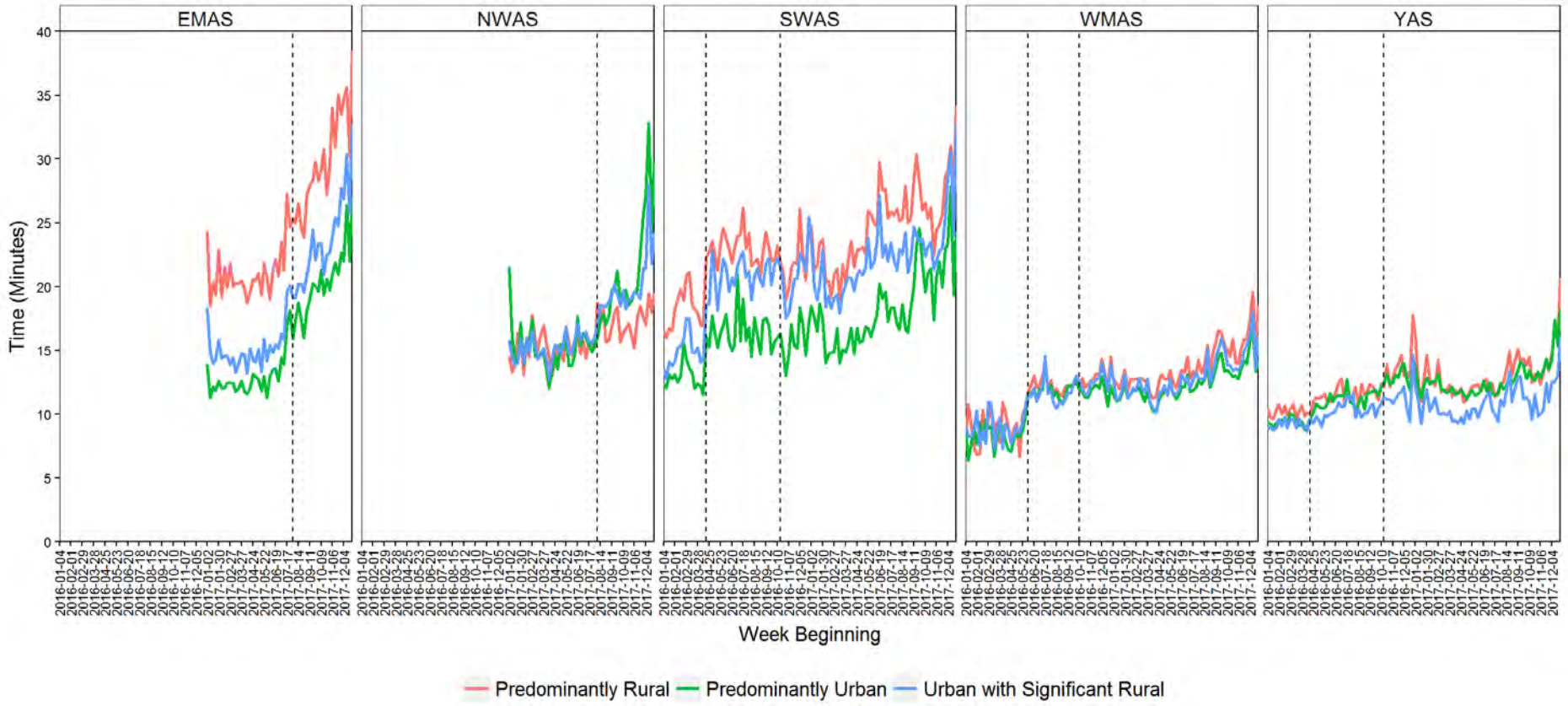


Figure 12: Time from call connect to arrival of first resource - 95th percentile

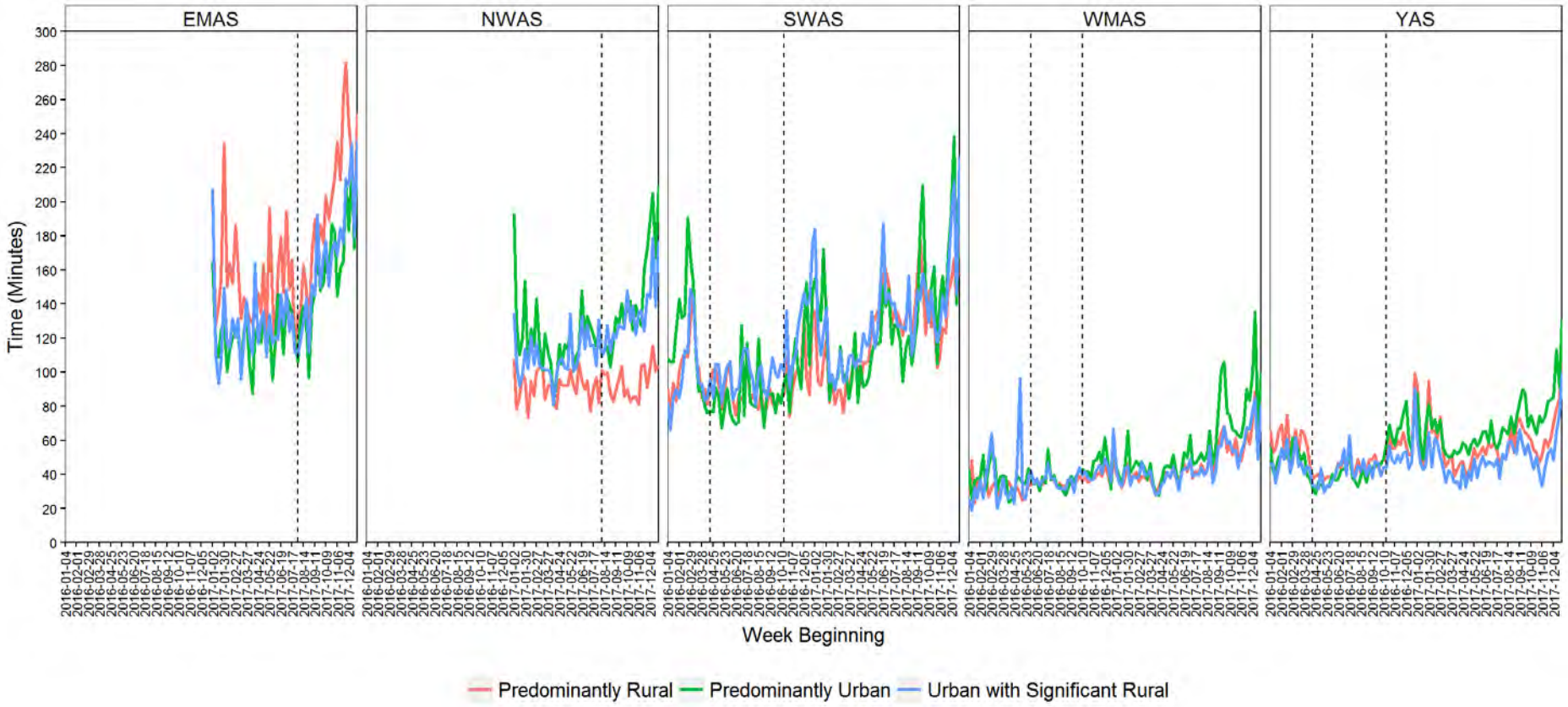


Table 5: Comparison of urban and rural call connect to leaving scene

Call connect to leaving scene – See & Treat	EMAS	NWAS	SWAST	WMAS	YAS
USR v Predominately Urban					
Median					
Phase 1	USR > PU 168 seconds	No difference	PU > USR 533 seconds	USR > PU 143 seconds	PU > USR 184 seconds
Phase 2	USR > PU 259 seconds	No difference	PU > USR 193 seconds	PU > USR 243 seconds	No difference
95th percentile					
Phase 1	No difference	PU>USR 1260 seconds	PU>USR 2293 seconds	No difference	PU > USR 601 seconds
Phase 2	No difference	PU>USR 1896 seconds	PU>USR 762 seconds	PU>PR 778 seconds	PU > USR 1431 seconds
Predominately Rural v Predominately Urban					
Median					
Phase 1	PR>PU 588 seconds	PU>PR 552 seconds	PU>PR 201 seconds	No difference	No difference
Phase 2	PR>PU 755 seconds	PU>PR 843 seconds	No difference	PU>PR 200 seconds	PR>PU 82 seconds
95th percentile					
Phase 1	PR>PU 1331 seconds	PU>PR 2837 seconds	PU>PR 2439 seconds	No difference	No difference
Phase 2	PR>PU 2041 seconds	PU>PR 4106 seconds	PU>PR 1475 seconds	PU>PR 704 seconds	PU>PR 598 seconds

Figure 13: Time from call connect to leaving scene - Median

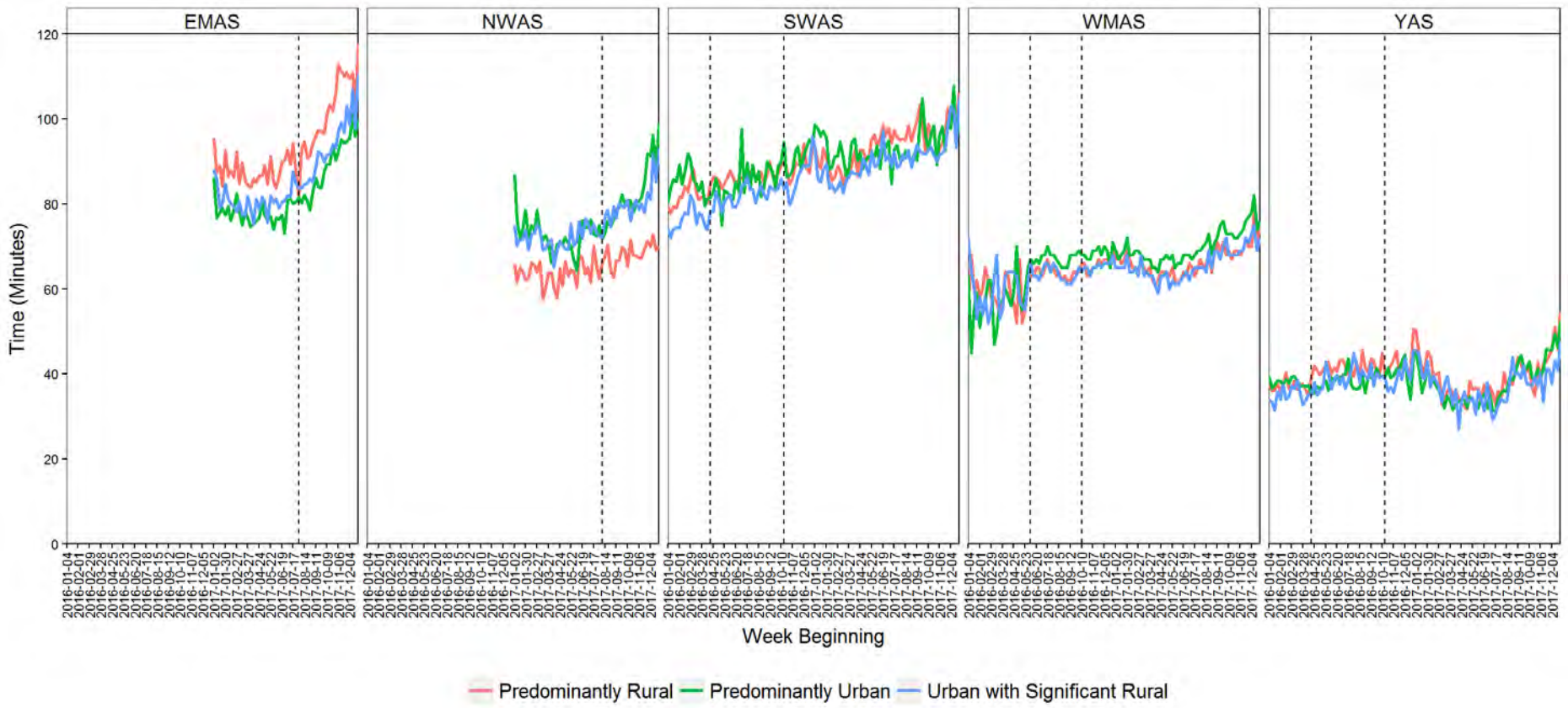


Figure 14: Time from call connect to leaving scene – 95th percentile

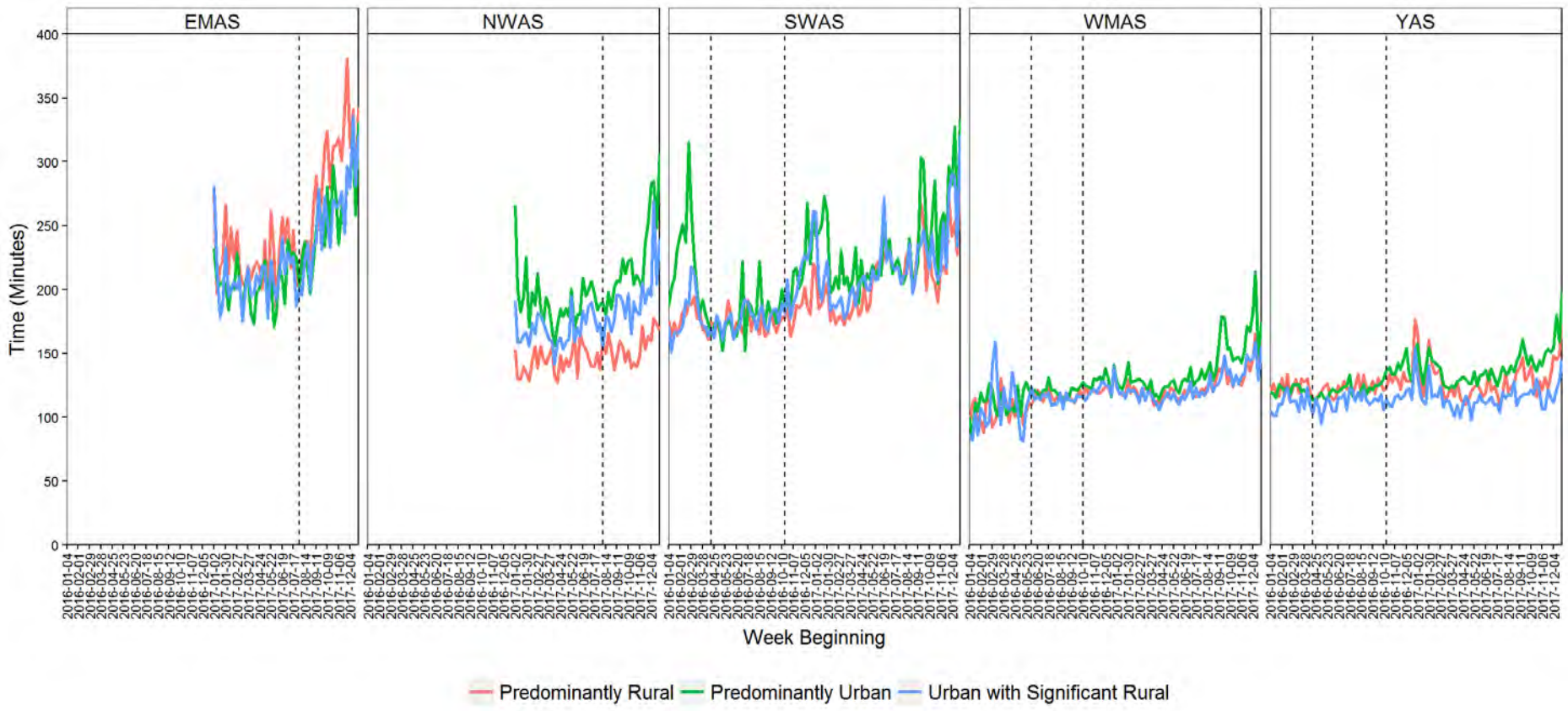


Table 5: Comparison of urban and rural call connect to arrival at hospital

Call connect to arrival at hospital – See & Convey	EMAS	NWAS	SWAST	WMAS	YAS
USR v Predominately Urban					
Median					
Phase 1	USR > PU 606 seconds	PU > USR 134 seconds	No difference	USR > PU 330 seconds	No difference
Phase 2	USR > PU 651 seconds	No difference	USR > PU 290 seconds	USR > PU 343 seconds	PU > USR 164 seconds
95th percentile					
Phase 1	USR > PU 1095 seconds	No difference	No difference	USR > PU 362 seconds	No difference
Phase 2	USR > PU 1378 seconds	No difference	USR > PU 747 seconds	No difference	PU > USR 880 seconds
Predominately Rural v Predominately Urban					
Median					
Phase 1	PR>PU 1420 seconds	No difference	PR>PU 701 seconds	PR>PU 247 seconds	PR>PU 817 seconds
Phase 2	PR>PU 1531 seconds	PU>PR 258 seconds	PR>PU 1007 seconds	PU>PR 253 seconds	PR>PU 747 seconds
95th percentile					
Phase 1	PR>PU 3552 seconds	No difference	No difference	PR>PU 339 seconds	PR>PU 1472 seconds
Phase 2	PR>PU 3462 seconds	PU>PR 1930 seconds	PR>PU 1048 seconds	PR>PU 254 seconds	PR>PU 489 seconds

Figure 15: Time from call connect to arrival at hospital - Median

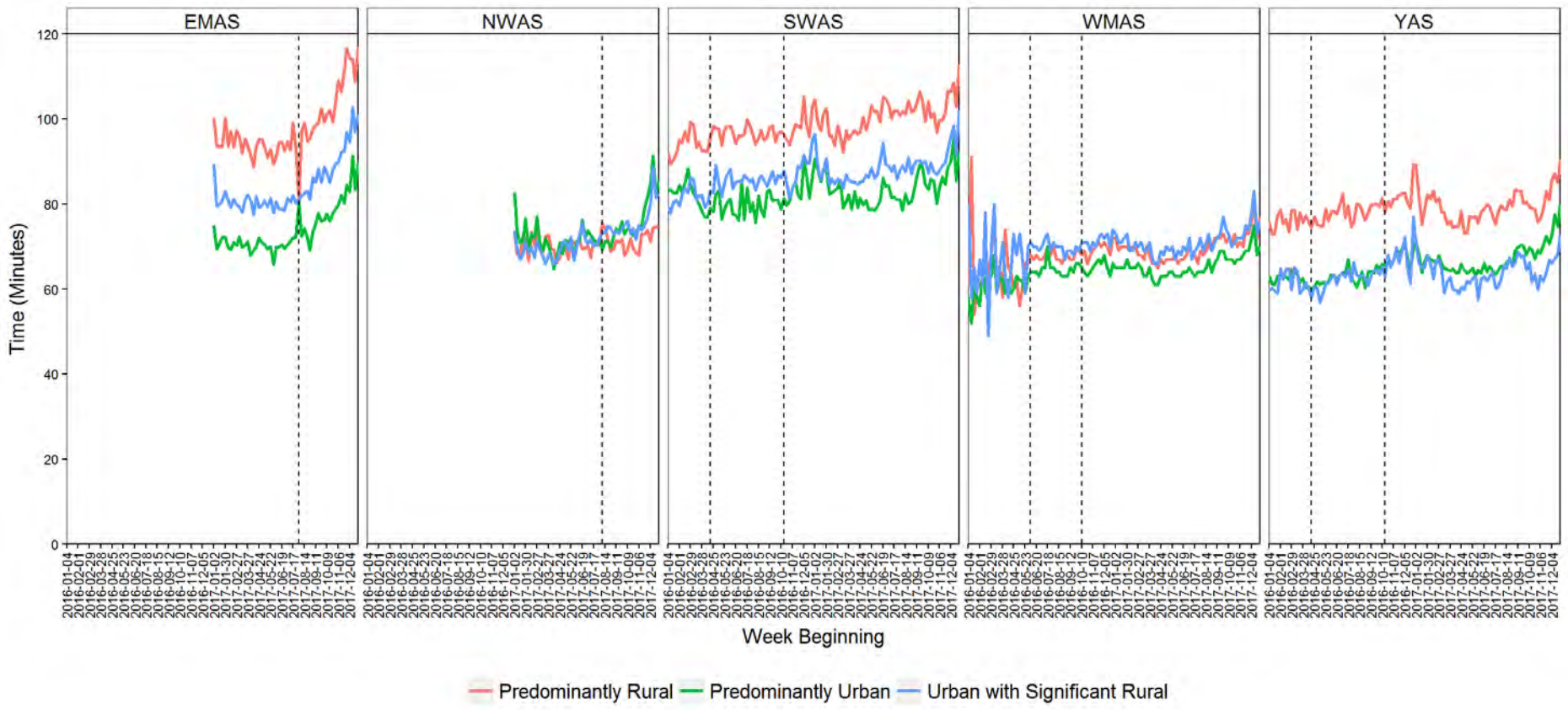
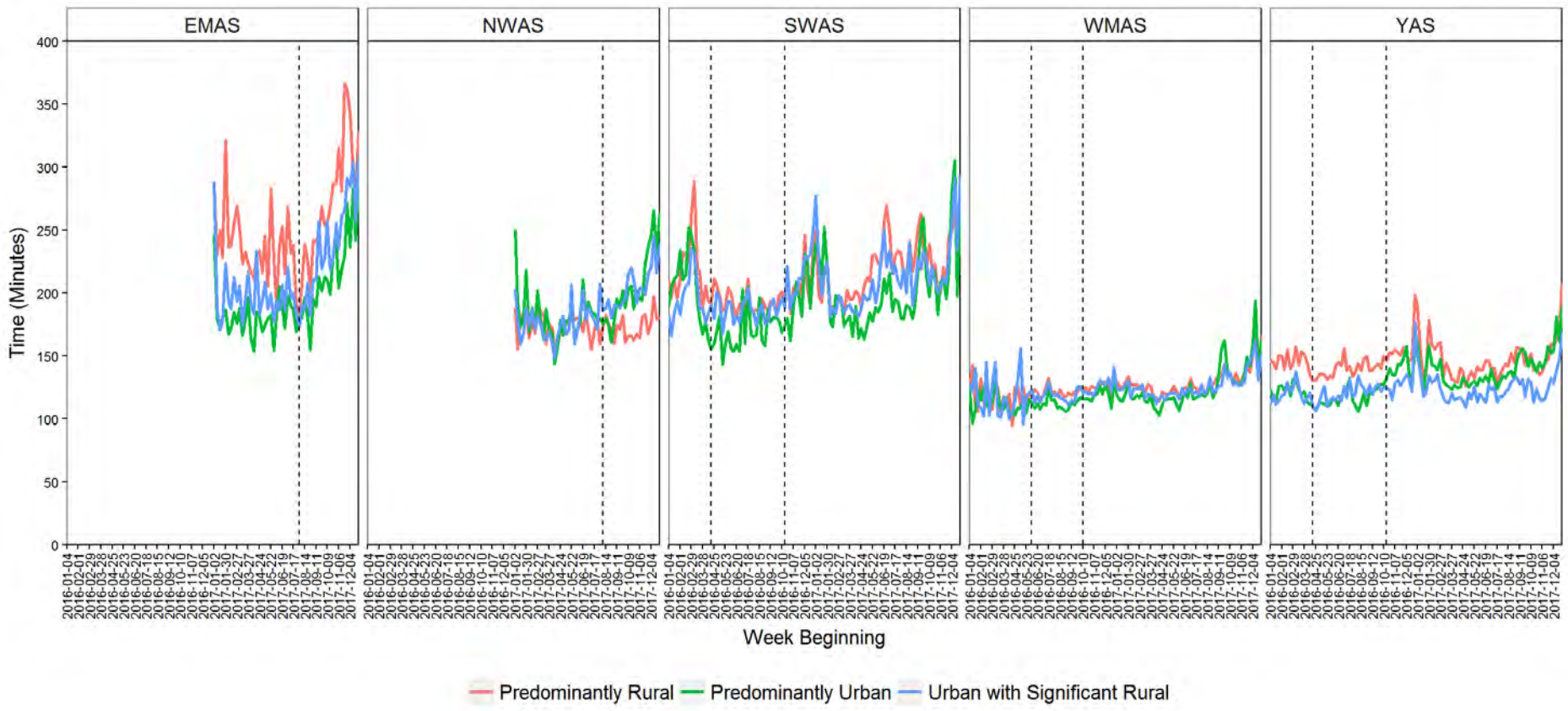


Figure 16: Time from call connect to arrival at hospital - 95th percentile



Time to first resource on scene

Median Response times were longer in urban with significant rural (USR) areas than predominantly urban (PU) areas in 3 services both before and after the introduction of phase 2 with an increase in phase 2 of 1-132 seconds. In one service there was no difference between USR and PU before and after and in the other PU times were greater than USR in phase 2. For the 95th percentile times there were no differences between USR and PU before or after the introduction of phase 2 in two services, in 2 services there was no difference before but PU times were longer after and in 1 service times were longer in urban areas before with no difference after. In all 5 services the 95th percentile response time showed no disadvantage for USR areas compared to urban areas.

For the comparison of predominantly rural (PR) areas with predominantly urban median times, in 3 services PR times were greater than PU both before and after with times increasing after by 40 and 84 seconds and reducing in one service by 22 seconds. In the other two services there were no differences between PR and PU before or after in one service and PU was longer than PR after in one service. For 95th percentile times PR was longer than PU before and after in one service but in 3 services PU was longer than PR after the introduction of phase 2. In one service there was no difference in either phase. Although there are individual differences, overall the results show that for the majority of the 999 population (using 95th percentile times) there are no disadvantages in USR and PR areas and any disadvantage is for urban areas where times are more consistently longer.

Time to leaving scene

The median time was longer in USR compared to PU both before and after in one service. In 2 services times were longer in urban areas after and in 2 services there was no difference during phase2. This is more marked for the 95th percentile times where this is longer in urban areas in 4/5 services after the introduction of phase 2.

For predominantly rural areas there was a consistent pattern of PR times longer than PU times before and after for both median and 95th percentile times in one service. In one service median times were longer in PR areas after the introduction of phase 2 but for the other 3 services there was no difference or times were longer in urban areas. This is repeated for 95th percentile times where in 4/5 services times were longer in urban areas than rural areas.

Time to arrival at hospital

In one service times were longer in USR areas both before and after and for median and 95th percentile times with increases after of 45 seconds and 283 seconds respectively. In one service median time was longer for USR but at the 95th percentile time there was no difference. USR times were greater than urban times during phase 2 only in one service for both median and 95th percentile times and conversely in another service both median and 95th percentile times were greater in urban areas than USR.

Median times were longer in PR areas than PU areas both before and after in 3 services. In one service PU time was longer in phase 2 and in the other longer PR time was reversed to longer PU time in phase 2. At the 95th percentile PR time was longer than PU in 3 services both before and after

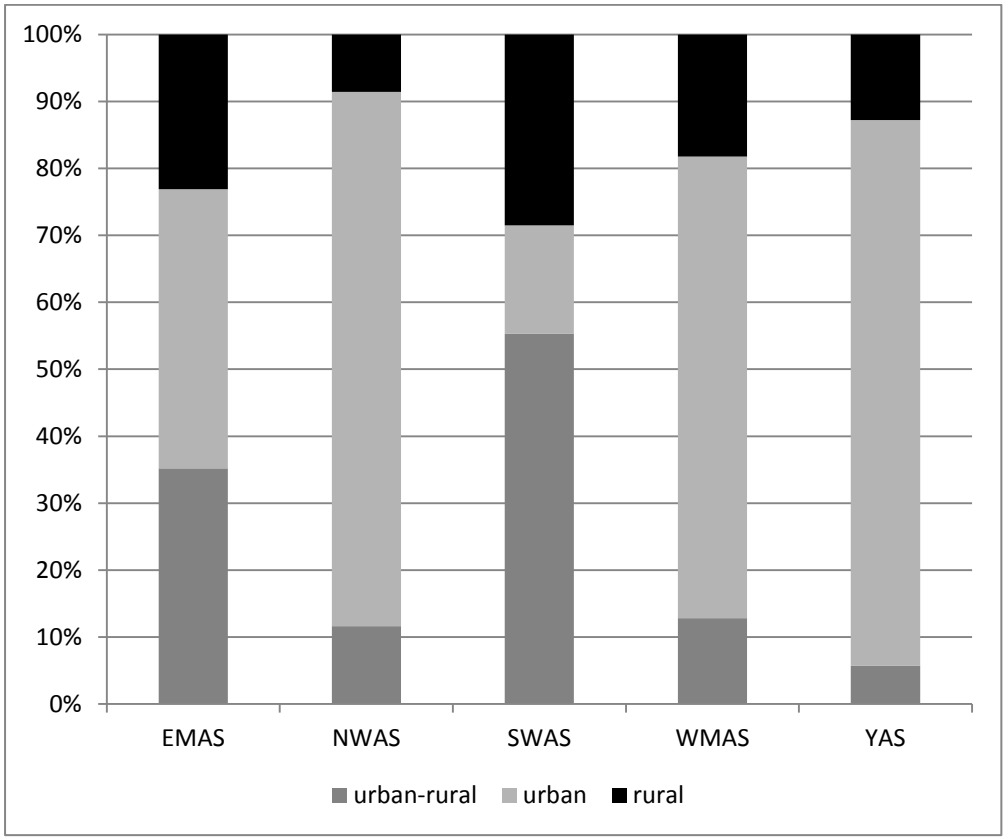
but in all three the time difference reduced in phase 2 (by 85, 90 and 983 seconds). In one service PU times were longer in phase 2 and another PR times longer in phase 2 compared to no difference in phase 1.

Summary of findings

Our first analyses in the previous report showed a mixed picture and this remains the case although the addition of more services and longer time period is beginning to show some patterns of consistency. Of the 5 services the results for EMAS show a consistent pattern of longer times in urban with significant rural and predominantly rural areas across most of the time intervals measured. Conversely the results for NWAS report no measure where times were longer for USR or PR and where there are differences times are always longer in urban areas. For the 3 early sites the results of this analysis remain consistent with the previous one other than that there is an increase in the likelihood of any difference detected being an increase in times for urban areas when compared to rural or mixed urban and rural and some of the increases are quite substantial. The largest differences between urban and rural areas are evident in the time to arrival at hospital which is not surprising as distances to hospital will generally be longer in rural areas and this time component will be fixed. There is some evidence that small reductions are being made in the difference between rural and urban times to hospital. The biggest effect services can have will be time to scene and leaving scene as this is within ambulance service control. The overall finding show that, particularly for 95th percentile times, the likelihood is that times are longer in urban areas than rural areas. This holds true for call to leaving scene in patients who are not taken to hospital.

There are a number of possible explanations for these findings. Firstly, the relative proportions of calls within each geographical type may be having an influence on performance in the different areas. Figure 17 shows the proportions of 999 calls classified as Urban, Rural and Urban with Significant Rural in each of the 5 services. This clearly illustrates differences which logically fit with some of the findings. About half of demand in EMAS comes from rural or urban with significant areas and it is this service that shows the most consistent pattern of higher times across all three measures in these areas when compared to urban areas. This greater variation will produce challenges where a substantial proportion of calls are generated in localities where population demand is lower but distances longer. In contrast, 80% of demand in NWAS arises in predominately urban areas and less than 10% in predominantly rural areas. Most pressure will therefore be in urban areas and this is reflected in the findings that where there are differences times are longer in urban areas. YAS has a similarly large proportion of demand within urban areas and, whilst not as marked as NWAS, for time to arrival of first resource where there are differences this is more often longer in PU areas. In WMAS urban areas also account for most demand although there is a larger proportion from other areas compared to NWAS and YAS and results from this service are more mixed although for 95th percentile times any difference is predominantly longer times in urban areas. SWAST shows some interesting results as this service has by far the largest proportion of calls generated in PR and USR areas and, whilst median times are longer in USR and PR for time to arrival of first resource there is no difference for 95th percentile times and time to leaving scene is likely to be longer in urban areas. Times to arrival at hospital are consistently longer in rural and mixed urban rural areas in this service which most likely reflects longer travel times to hospital.

Figure 17: Proportions of 999 calls classified as Urban, Rural and Urban with Significant Rural



This variability in proportions of demand generated in different geographical areas will present operational challenges around where to allocate resources. In services with a high proportion of demand generated in urban areas and low volumes of calls in rural areas then most resources will be concentrated in urban areas as that is where they are needed. Low demand in rural areas may be managed by some ring-fencing of resources and a higher likelihood of resources not being diverted once allocated and driving to a more rural area. Where large proportions of a service area are rural and mixed urban and rural (as in SWAST) this may allow more even distribution of resources as, relatively, more demand will be originating in these areas. It may be that where urban and rural areas are evenly split (as in EMAS) the positioning of resources is actually harder to balance in terms of geographical spread and variation in demand.

At times of peak demand then, relatively, there will be larger increases in call volumes in urban areas as this is where the highest number of calls are generated. If resources are low and/or calls are queuing it may be more likely that resources will be diverted from a low priority call to a high priority call in urban areas with consequent longer response times for some calls. This effect can be seen in the figures where times of peak demand (for example winter and particularly the Christmas & New Year periods) tend to show clearer gaps between each area type and urban response times increase. During more stable periods the times for each of the 3 types are much more tightly bunched together showing little variation between types.

One other feature worth highlighting is the impact of moving from the previous operating model where almost 50% of calls were classified as needing an 8 minute response to the new call categories where only around 10% of calls need the fastest response and for categories 3 and 4 the response standards have increased substantially. It would be reasonable therefore to expect that, for the whole 999 population of calls, overall times will increase. This has not happened across the board and for some measures in some services no difference in times has been detected after the introduction of phase 2. More generally, where differences between urban and rural areas has been measured, any increases in the difference during phase 2 have been modest and are measured in seconds or minutes indicating that increasing response time standards for lower urgency calls has not resulted in a substantial increase in overall response times for the 999 population. There was also a reduction in the 95th percentile time to arrival at hospital in 4/5 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.

These results show a comparison of differences in response performance between urban and rural areas before and after the introduction of the revised ARP call categories for all 999 calls. What the results don't show is if there is any variation for different types of calls. Clearly we cannot compare different call types before and after the introduction of phase 2 as the call categories are not comparable. We have plotted the median and 95th percentile times from call connect to arrival of first resource, leaving scene and arrival at hospital for the duration of phase 2 for each call category together with average resource allocation for each call type. The full results can be provided on request. Here the 95th percentile response performance for call connect to arrival of first resource is presented to illustrate PU, USR and PR performance for the 4 call categories for each service (Figures 18-21).

Figure 18: Category 1 – Call connect to arrival of first resource 95th percentile

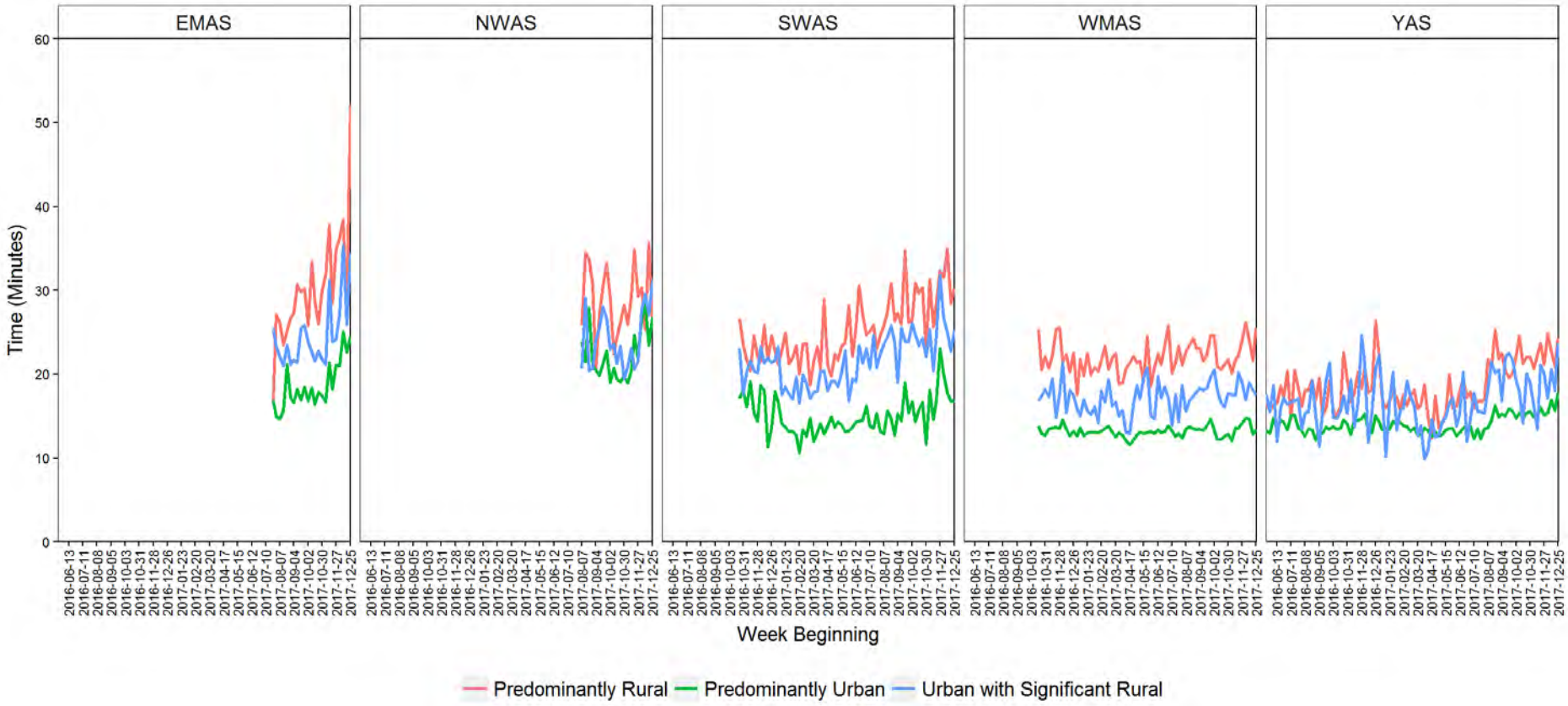


Figure 19: Category 2 – Call connect to arrival of first resource 95th percentile

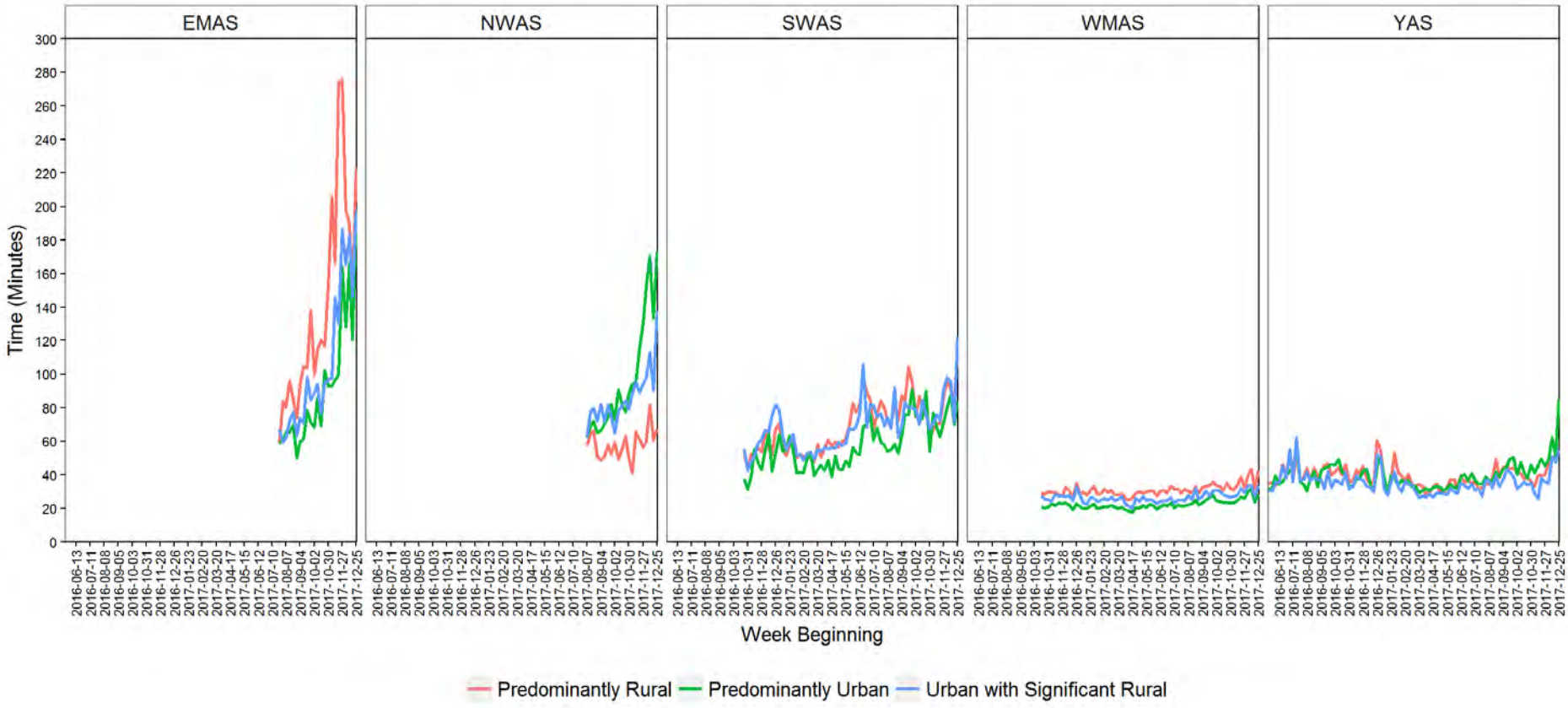


Figure 20: Category 3 – Call connect to arrival of first resource 95th percentile

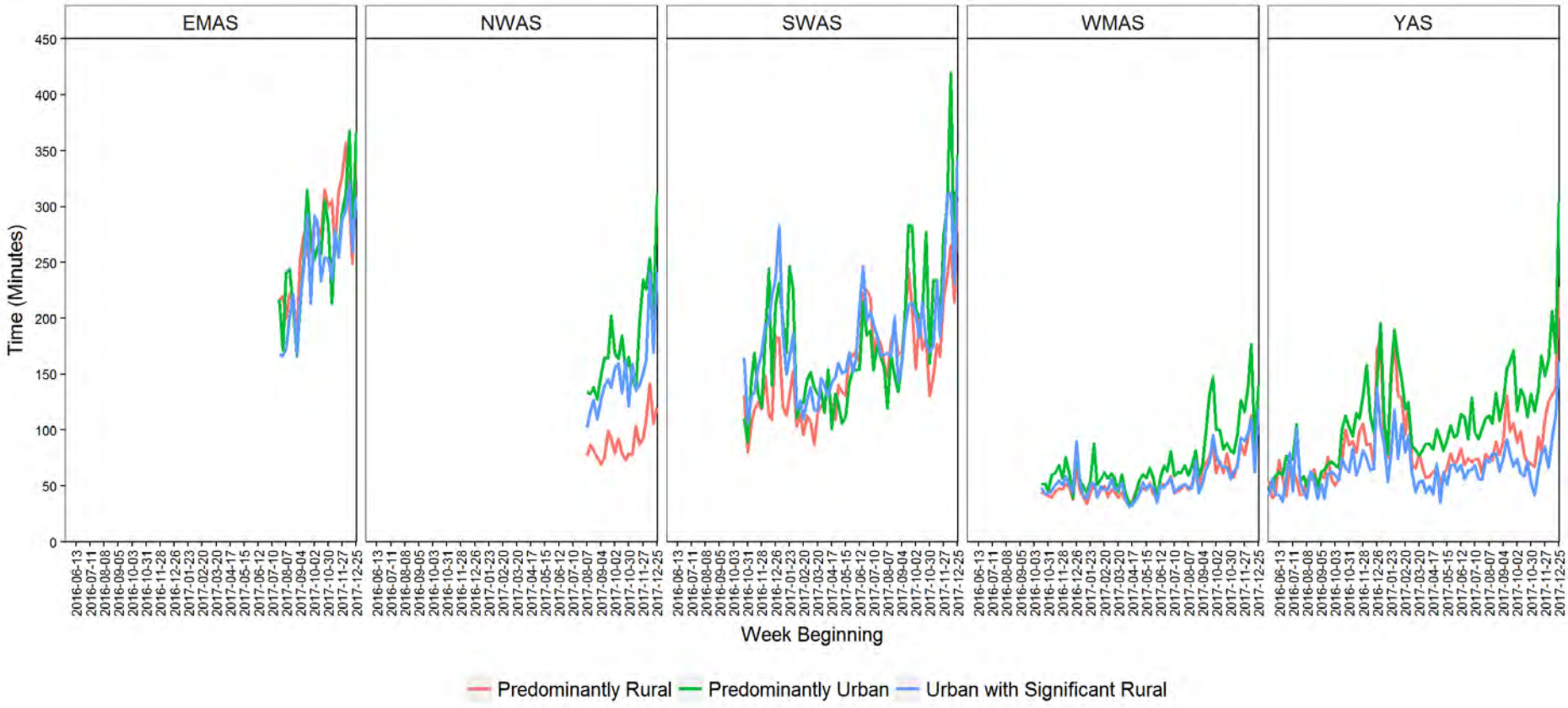
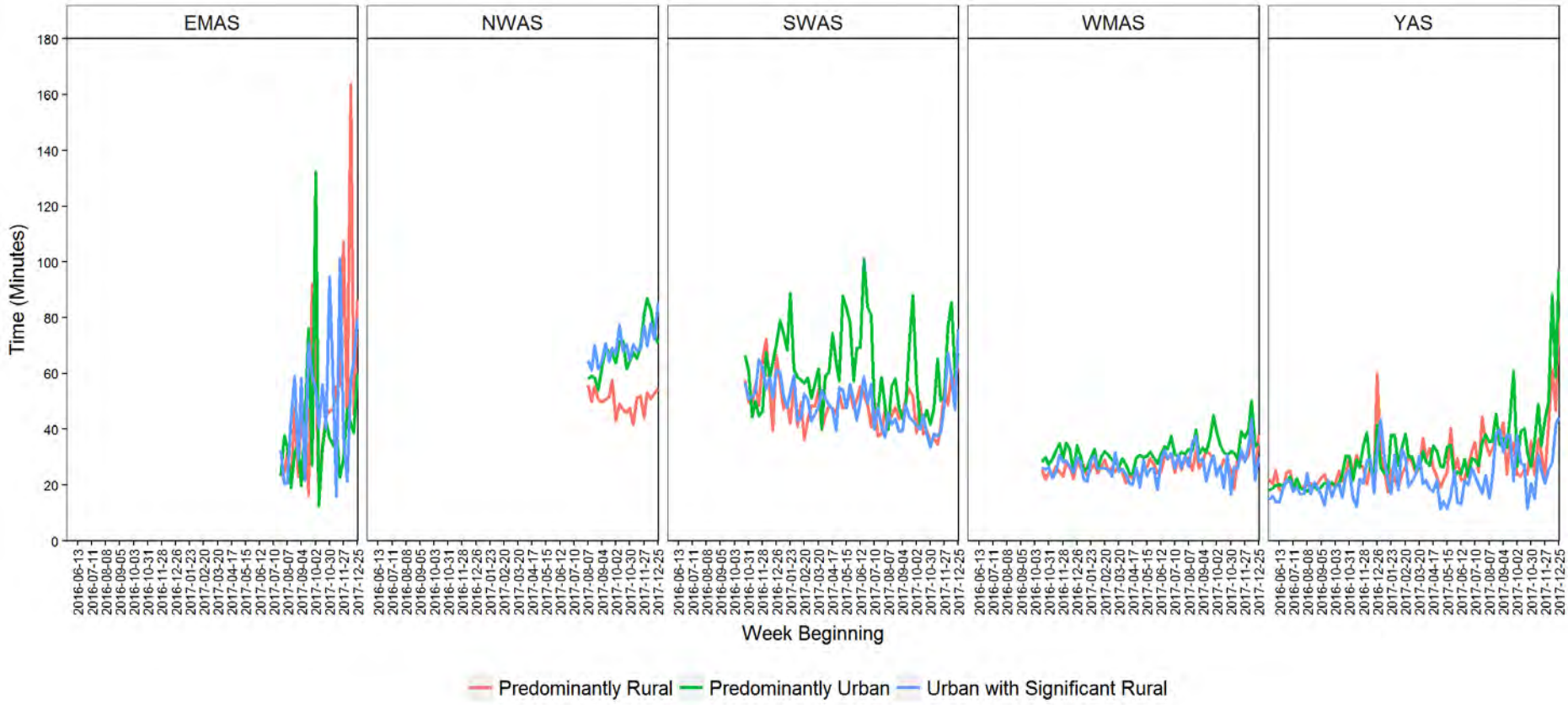


Figure 21: Category 4 – Call connect to arrival of first resource 95th percentile



Unsurprisingly, response time performance is longer in predominantly rural areas across all services for Category 1 calls. Response time is related to distance and for these calls requiring the fastest response the longer the distance, the less likely it will be that a response will be achieved within a short time frame. The only way to reduce this disparity is to have more resources available within rural areas but this creates inefficiency as demand is lower so utilisation will be correspondingly low. In an environment of high demand and limited resources it is difficult for services to justify holding resources is areas where there may be only a small number or no calls in any given day whilst elsewhere in the same service where demand is higher patients are waiting with no resources available to send. Equity comes at a cost. These differences are much less apparent, overall, for Category 2 and particularly in the 3 more established sites where there is little clear difference between different area types suggesting for this group there is an equitable service where slightly longer response time standards have enabled services to provide the same level of service across all areas. For Categories 3 and 4 some separation does begin to appear and in the majority of services the area with longer times is predominantly urban. This may go some way to explain why longer response times were just as likely, and in some services with a high proportion of calls originating in urban areas more likely, to be identified in urban areas when compared to rural areas in the all calls analysis. Category 3 and 4 calls have a 90^{th} percentile response standard of 2 and 3 hours respectively. In urban areas with high demand and therefore most resources, when these start to become scarce there may be a greater likelihood that resources are diverted to higher category calls or these lower category calls held for longer in the call stack to ensure a timely response to category 1 and 2 calls.

In summary, the results have shown that assumptions of response performance in rural areas being consistently longer than in urban or mixed urban and rural areas are unfounded. This is true in some cases where there is an equivalent mix of urban and rural areas which may make resource positioning difficult, but there is a clear picture that any inequities in response performance are driven by the overall workload in urban areas and maintaining a prompt response in rural areas may come at a cost to increasing response times in urban areas.

Proportions of call volumes in each revised call category

The call category review process during phase 2 of ARP developed new definitions of call categories based on urgency and clinical need. The previous call categories of Red 1 and Red 2 required the fastest response (75% within 8 minutes) for half of all calls although evidence has shown that the proportion of calls with life-threatening, time dependent conditions is substantially lower. Revision of the categories aimed to both more accurately classify life-threatening emergencies so that the fastest response could be made to this patient group and assign other calls to categories with response time standards that take account of reasonable time frames with which to provide an appropriate response. Two call triage systems are in use and during the review a detailed process was undertaken to assign each individual call type (AMPDS determinant and NHS Pathways symptom group and symptom descriptor combination) to the revised categories and the expected proportions of calls within each category were estimated.

There have been a number of iterations of the Category 1-4 call allocations including the removal of the distinction between response and transport options in categories 2 and 3 and different response time standards so estimates of proportions of calls within each category have varied but in summary the initial call volumes within each category reported in the 3 trial sites were 7% for C1, 41% for C2, 43% for C3 and 9% for C4 in NHSP sites and 11% C1, 53% C2, 23% C3 and 12 % C4 in AMPDS sites.

The move of all services to the current version of call categories and new response time standards during 2017 (phase 2.3) allows an assessment to be made of the proportions of incidents allocated to each category in operational practice and identify any between service and between triage system differences. Table 6 shows the average proportion of incidents allocated to each category for each service. Weeks included range from 10 to 27 weeks depending on service implementation. Figure 22 shows the proportions of calls assigned to each category for the most recent week of the ARP weekly monitoring dashboard.

There are differences across all services in terms of the proportion of incidents assigned to each category irrespective of triage system used but overall a larger proportion of calls are assigned to Category 1 and 2 in AMPDS services than in NHS Pathways services (range 65-71%) compared to 50-66% in NHS Pathways sites. One AMPDS site has a particularly high proportion of calls assigned to Category 1 and overall AMPDS sites have a Cat 1 proportion 2-3% points higher than NHS Pathways sites.

The proportions of calls assigned to Category 1 and 2 are higher than the 50% reported during the early trial phase of the revised call categories for NHS Pathways sites although two services have maintained this proportion. In the early trial a single AMPDS site reported a combined Cat 1 & 2 proportion of 65 % and current figures show this has remained similar or higher.

There are a number of reasons that could explain the variance in calls assigned to different categories. One obvious one is differences in the two triage systems and the way they function. AMPDS is primarily a sorting system to prioritise calls by urgency with emphasis on the identification of priority symptoms. This means more determinants are likely to be assigned to higher priority categories. The introduction of a new version of AMPDS since the last category review may have exacerbated this for a small number of determinants and there are ongoing discussions with the

National Academies of Emergency Dispatch to review the questioning, particularly for calls where the patient is described as “not alert” which, if refined, should reduce the proportion of calls assigned to Category 1.

Other possible issues are the time period and a possible increase in acuity of illness requiring higher level response over winter and, as services have struggled with high demand, some upgrading of calls, particularly from Cat 3 to Cat 2 if calls have been waiting for a long time for a response. These are illustrated in Figure 23. The introduction of the revised Ambulance Quality Indicators included for the first time the 90th centile standards of 2 hours and 3 hours for Categories 3 and 4 respectively. There is some anecdotal evidence from services that this may be influencing call taking and dispatch behaviour as staff view Category 2 as a 19 minute response (the mean standard) and compare this to the 90th centile 2 hour standard for Category 3 calls which then leads them to upgrade calls as this difference in time seems large. It is difficult to unpick which factors may be having an influence although case-mix, demand and perceptions around time standards would be expected to have an influence across all services but 3 out of 4 services using NHS Pathways show a consistently lower proportion of calls assigned to Categories 1 and 2 compared to AMPDS sites.

There is currently a review in place to revisit the assignment of AMPDS determinants to each call category and additional data has been requested to try and identify if there are groups of calls that are being consistently upgraded. Determinant assignment will be reviewed to try and bring the proportions of calls in each category in closer alignment to the expected proportions and closer to those reported in most NHS Pathways sites. It is also worth considering whether a mean or median response time standard should be introduced for categories 3 and 4 to bring them in line with Categories 1 and 2 and reduce the perceived substantial difference in expected response performance between categories 2 and 3. It is difficult to identify what a mean standard for these categories should be using the current 2.3 data as these reflect a period of high and sustained operational pressure. The more recent data shown earlier in Figure 1 shows all services currently report an average weekly mean response time performance of 60 minutes or less for Category 3 and 8/10 report a mean time of 90 minutes or less for Category 4.

Table 6: Average % of incidents assigned to call categories 1-4 by service – August 2017 – March 2018

	AMPDS Sites						NHS Pathways sites			
% incidents per category	**EMAS**	**EoE**	**LAS**	**NWAS**	**YAS**	**SWAS**	**NEAS**	**SCAS**	**SECAMB**	**WMAS**
Cat1	8.9%	10.1%	8.6%	10.0%	13.4%	8.1%	8.2%	5.9%	5.7%	7.1%
Cat2	59.3%	61.2%	56.1%	57.8%	56.5%	57.6%	57.8%	44.4%	49.5%	45.8%
All Cat 1&2	**68.2%**	**71.3%**	**64.7%**	**67.8%**	**69.9%**	**65.7%**	**66.0%**	**50.3%**	**55.2%**	**52.9%**
Cat3	25.8%	20.1%	23.2%	24.5%	20.8%	28.6%	25.2%	32.3%	37.6%	40.8%
Cat4	0.5%	5.2%	3.0%	5.5%	4.9%	1.9%	1.5%	3.3%	1.9%	2.6%
Cat4 H	5.4%	3.3%	9.0%	2.3%	4.4%	2.8%	7.3%	6.4%	5.4%	3.7%
All Cat 3&4	**31.7%**	**28.6%**	**35.2%**	**32.3%**	**30.1%**	**33.3%**	**34.0%**	**42.5%**	**43.9%**	**45.9%**

Figure 22: Proportion of calls assigned to each call category week beginning 23rd April 2018

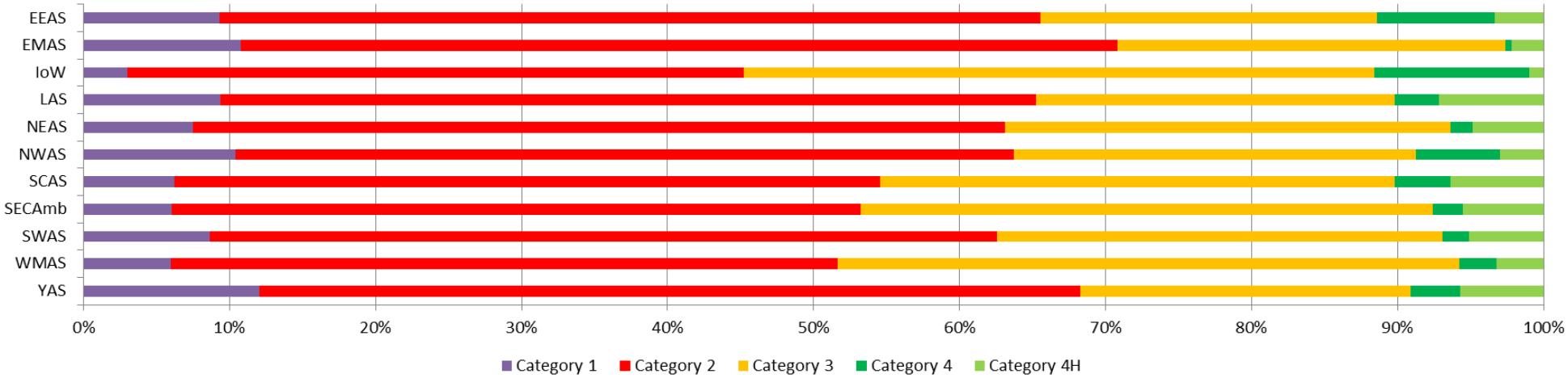
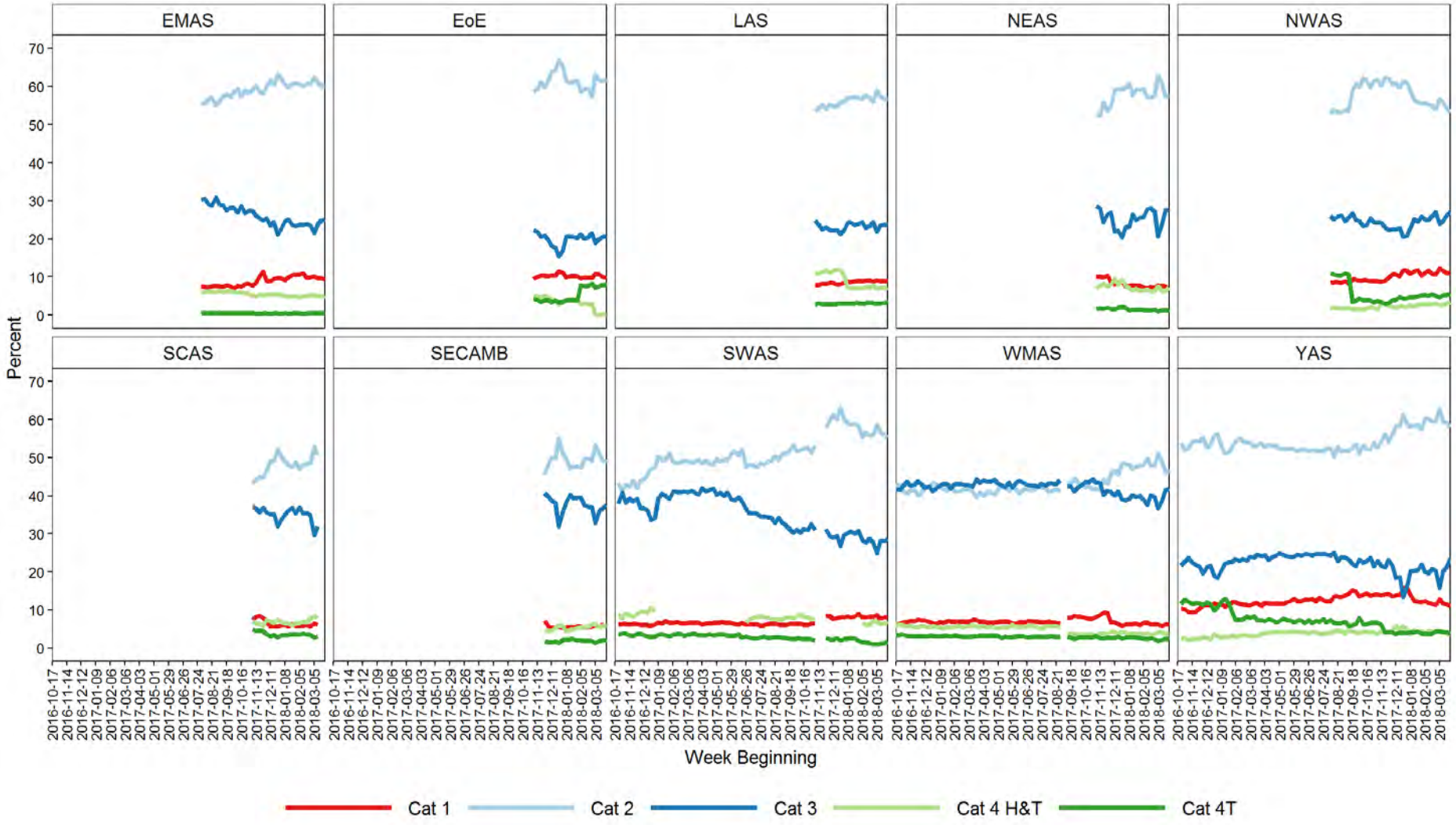


Figure 23: Percentage of incidents by category


EMAS
EoE
LAS
NEAS
NWAS
SCAS
SECAMB
SWAS
WMAS
YAS
Percent
Week Beginning
Cat 1
Cat 2
Cat 3
Cat 4 H&T
Cat 4T

Unintended consequences – call volumes and answering times

The revised response time performance standards are a significant departure from those they replaced with a much smaller proportion allocated the fastest response and Category 3 and 4 calls having a longer response time frame compared to the previous green categories. Following national introduction of phase 2.3 some services have reported an increase in the number of duplicate calls and in particular repeat calls from patients/callers requesting information about where their ambulance is or how long it will be. These additional calls put further pressure on a call answering system that is already coping with increased demand, ties up call takers and can consequently increase call answering times for new 999 calls. We have explored if there is any observable increase in duplicate calls using the ARP weekly data returns by calculating the difference between calls and incidents and plotting if this changes over time. Not all of the difference between calls and incidents will be due to duplicate calls but an increase in these calls should be observable. Table 7 shows the mean difference between calls and incidents before and after the introduction of phase 2.3 for the year January 2017 – March 2018

Figure 24 shows the weekly percentage difference in calls and incidents for 6 services who moved from the previous call category system to 2.3 for the year 2017 so spanning a before and after period. There are some gaps in weekly data during the transition from phase 1 to phase 2.3 hence breaks in the time series plots.

Figure 25 shows a longer, continuous period in the 3 call category trial sites which span phases 1, 2.1, 2.2 and 2.3.

Table 7: Mean difference in proportions of calls and incidents January 2017 – March 2018

Service	Mean difference between calls and incidents (%)				
	phase 1	Phase 2.1 & 2.2	phase 2.3	% change phase 1 to 2.3	Weeks in phase 2.3
EoE	35.0		34.4	-0.6	23
LAS	18.6		30.1	11.5	21
NEAS	30.0		37.0	7.0	21
SECAMB	30.4		31.1	0.7	19
SCAS	29.0		17.6	-11.4	18
EMAS	18.5		21.1	2.6	35
NWAS	22.6		28.6	6.0	34
SWAS	25.8	34.6	32.5	6.7	19
YAS	18.4	7.4	9.1	-9.3	29
WMAS	12.3	13.1	18.6	6.3	30

Table 7 shows there has been an increase in the percentage difference between calls and incidents (as a proxy for duplicate calls) in all services although this is very variable ranging from less than 1 to 11 percent. The data from YAS needs validating as the call volumes appear uncharacteristically low compared to historical values.

Overall 6/10 services show an overall increase. Although percentage changes may appear small, from an ambulance service call volume perspective even small increases results in substantial extra workload. Ambulance services in England manage a minimum of 10,000 and up to 30,000 calls per week. A 1 % increase is an extra 100 – 300 calls per week, 5% increase is 500 – 1500 calls per week and 10% increase is 1000-3000 calls per week. Across England there are 11 million calls per year so 1 and 10% increases in duplicate calls will increase call volumes by 110,000 – 1,100,000 per year.

The baseline difference is also very variable across services. It should also be borne in mind that the "after" period ranges from 6 months to 18 weeks and the majority of services only moved to 2.3 or reporting against 2.3 standards during autumn and winter. December and January have seen peak demand and difficulties in providing a timely response so it is not possible to discriminate between an increase in duplicate calls that is associated with the introduction of 2 and 3 hour time standards for Category 3 and 4 calls (and a 40 minute 90th centile time for Category 2) and an increase in calls asking about response where delays are the result of insufficient resources or indeed an increase in calls for other reasons. A longer time series is needed to better understand changes in duplicate calls when services are providing responses within expected standards. Figure 24 shows a step change in services following the introduction of phase 2.3 but given most services changed in October/November some of this will likely be winter effects. In the two services operating 2.3 for the longest period (NWAS and EMAS) there does appear to be an increase in the difference between calls and incidents after the introduction on phase 2.3 and is in the opposite direction to observed decreases in other services after a summer peak in June suggesting there is some effect with the introduction of the new standards. Similarly in Figure 25 there is a trend of increasing difference between calls and incidents in all 3 services. The big step change in SWAST is unexplained and may be a data issue. There is a clear step change in WMAS following the move to phase 2.3 reporting in September 2017. The difference between calls and incidents narrows in YAS which is an anomaly that so far YAS have been unable to explain.

Figure 24: Weekly % difference between calls and incidents January 2017 - March 2018

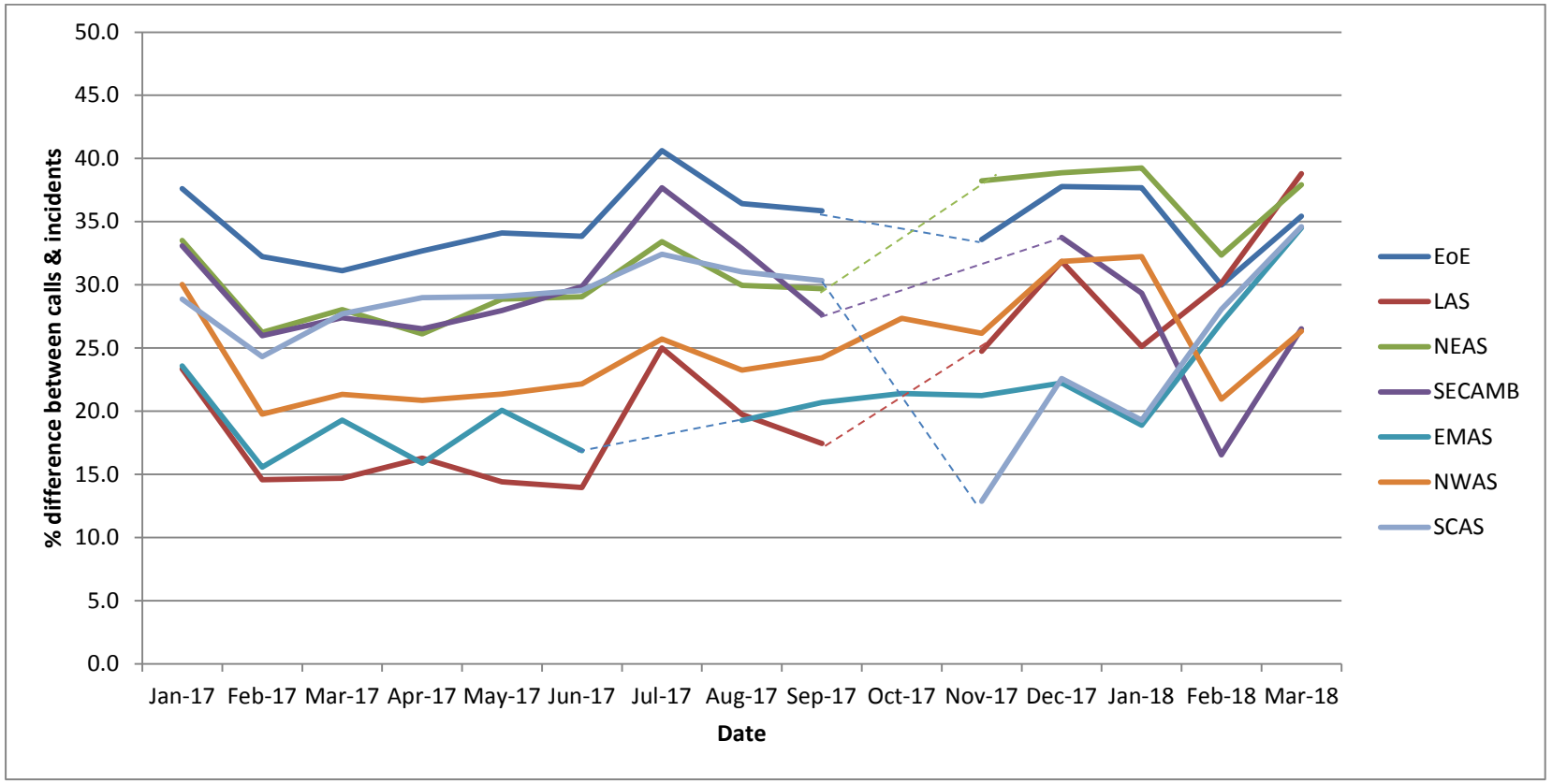
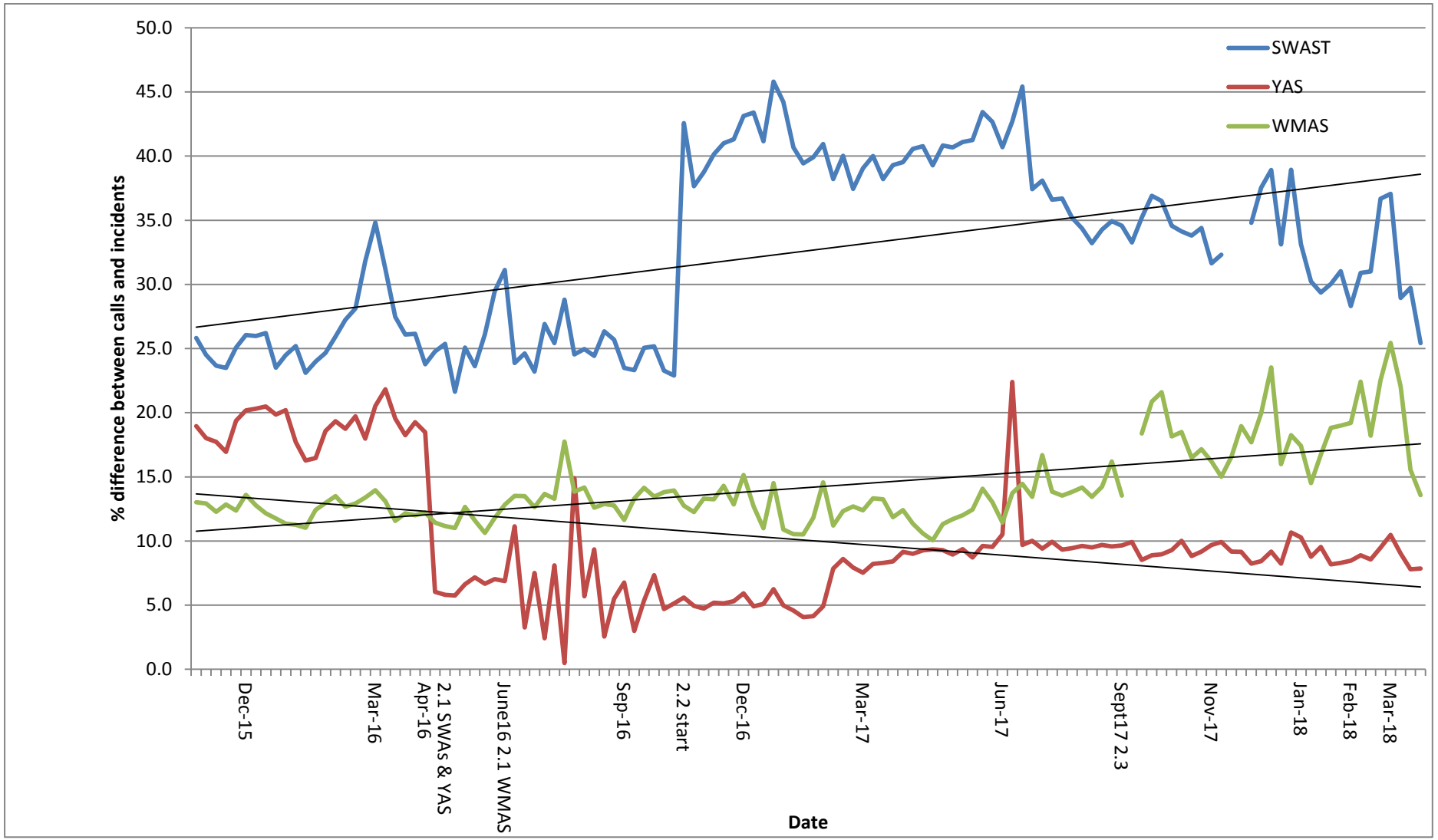


Figure 25: Weekly % difference between calls and incidents September 2015 - March 2018 in three services



We have also examined call answering times and whether there is any change following the introduction of phase 2.3. For this we have used the monthly NHSE AQI data for call answering times for the period April 2016 to February 2018. Call answering times were removed from the initial revised AQIs then reintroduced so there are some months when no data is recorded hence some values displayed as zero in services that began reporting phase 2.3 indicators earlier in 2017. Figure 26 shows the 95th centile call answering time for this period in each service.

Figure 26: Call answering time April 2016 to February 2018 - 95th percentile

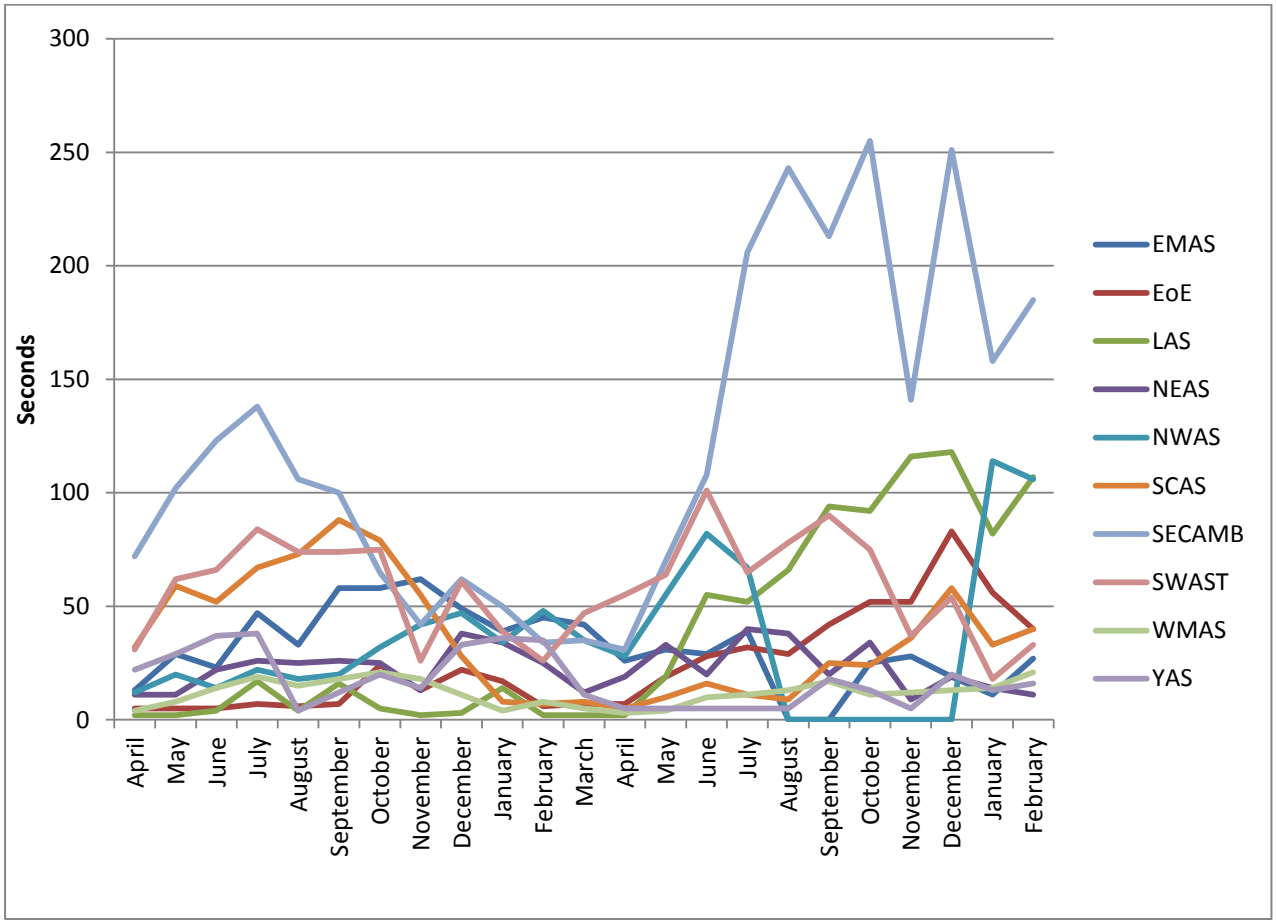


Figure 26 shows there is service variation in call answering times. One service historically has higher times as it has suffered significant recruitment problems of call takers within Emergency Operations although this is a feature in some other services. The graph also shows there were peaks and troughs in call answering times prior to the introduction of phase 2.3. There is a clear increase in some services for the period after phase 2.3 was introduced (LAS, EMAS, SCAS, EOE) but in some cases there was an upward trend prior to this and all services show a peak in call answering times in December although this is bigger in some services than others. With only a small number of months data and for a period of high demand it is not possible to identify if there is a clear relationship between an increase in duplicate calls and call answering times associated specifically with the introduction of Phase 2.3 as data is confounded by the effects of winter demand. It would be worth retaining call answering times as an AQI in order to monitor over time is there is a persistent step change in call answering times.

Recognition by services that there is an increase in duplicate calls associated with questions about response arrival has resulted in some changes to call handling scripts so that 999 call takers provide additional information to callers to manage expectations. North East Ambulance Service developed a set of call handling scripts at the time of implementation which, for calls identified as Category 2 – 4 included a time frame explained as "we aim to get to you within…(e.g. 2 hours for Category 3)" and asking callers not to call back unless condition changes. An internal audit by NEAS found staff and patients found the call scripts of value, maintained the number of duplicate calls about expected time of arrival at the same level as before the introduction of Phase 2.3 and increased the number of calls where people called back to cancel the ambulance by 6%.

All services have adopted some changes to their call handling scripts since the introduction of phase 2.3 although these vary in content and use. Some services use them for all calls and others when there is high demand. Some provide expected timeframes and others don't but emphasise where a call is in terms of the urgency (i.e. distinctions between emergency and urgent) and provide reassurance that a resource will be sent as soon as one is available. Some services also include a warning that when demand is high a vehicle may be diverted to a more urgent call. A common feature in all services is a request to callers to not call back unless there has been a change in patient condition. These changes to call scripts have helped services manage patient expectations both in terms of the new timeframes for response and also potential long waits when they are queueing calls. Taking this forward, the ARP development group have decided that some standardisation is needed so that callers get the same service wherever they are. This principally applies to Category 3 and 4 calls but also Category 2 so that expected response is made explicit for patients. This both contributes to meeting patient expectations and will hopefully reduce the number of repeat calls. NHS Pathways is adding a standardised call handling script to the triage system and this will be consistent in NHS 111 for calls allocated to an ambulance disposition. The same standardised script will be used for AMPDS sites although it is recognised there needs to be some flexibility in how the scripts are utilised in response to demand fluctuations.

Revised Ambulance Quality Indicators

The implementation of the nationally agreed new call categories and associated changes to Ambulance Quality Indicators has introduced revised indicators for measuring response time performance and some additional measures. Here we consider progress so far and the usefulness of the current set of indicators and whether additional measures should be considered.

1. Response time performance

The principle change to measuring response time performance has been the shift from reporting a percentage with a specified time frame (75% within 8 minutes for the previous Red 1 and 2 categories) to reporting mean and 90th centile times. For Category 2-4 calls there has also been a change in that reported times for transported patients reflect provision of a transporting vehicle not just first response. The objective of these changes was to make times for all calls more transparent including any waits for transporting vehicles.

Response time performance has been presented and discussed in detail in the earlier section. Reporting of mean and 90th centile response times is providing a clearer picture of actual performance for all calls and all categories which is an improvement on previous reporting. Review of a range of indicators in tables 1 and 2 presented earlier shows there are some potential subtle differences in response performance that are not reflected in the current two main indicators.

There were sound reasons for using mean response times rather than median as this reflects long waits which is important but it does have disadvantages. Mean does not refer to any specific volume of calls; it is a general measure of all calls with no indication of what proportion of patients actually receive a response within the reported time frame. Although mean times do reflect outlying values (in this case long waits) they are sensitive to very small numbers of cases with exceptionally long times which can results in substantial fluctuations. Mean is also influenced by simple data errors that may be the real cause of outlier values.

One way to manage this would be to report trimmed mean rather than mean as this will still reflect, over time, upward and downward trends and includes most calls but does mitigate the substantial effects that a small number of exceptional cases or data errors. Any increases or decreases in long waits would still be reflected in the 90th centile time. London Ambulance Service has reported a trimmed mean as well as mean in their weekly data returns and Table 8 shows the difference between mean and trimmed mean for each call category over the 13 week period they have been reporting Phase 2.3 data.

The times reported in this table are from call connect so, taking in to account additional average time to clock start (30 seconds for C1, 2minutes for C2, 2 minutes 18 seconds for C3 and 3 minutes for C4) the data shows that in 12/13 weeks the trimmed mean time is within the 7 minute standard for Category 1 calls and 19 minute standard for Category 2 calls compared to 5/13 weeks within the mean standard for both categories. Although there is no mean standard for Category 3 and 4 calls for C3 in 12/13 weeks a trimmed mean response was provided within an hour and for Category 4 in 6/13 weeks there was a trimmed mean response time of 90 minutes or less. The average difference between mean and trimmed mean is around 30 seconds less for C1, 3.5 minutes for C2 and 15 minutes for C3 and 4.

Trimmed mean times could potentially provide a more stable measure of response performance that is less influenced by extreme values resulting from external factors such as increased demand, reduced resources caused by hospital handover delays and call queues which are often outside the control of the ambulance service. This option was discussed by the ARP Development group and, whilst the possible advantages are understood, there was assurance from the recent audit of all services on their data reporting processes that the likelihood of data errors is small. The original decision to use the mean rather than median or a trimmed mean was made to ensure that the impact of long waiting times is transparent in the AQI reporting. It also provides a signal of service pressure as mean values vary by a bigger magnitude when there is high demand or lost resources and this is an important feature of quality monitoring. For these reasons the decision has been made to retain the mean response time interval as the AQI measure.

Table 8: Mean and trimmed mean response time performance from call connect to arrival of first resource in one service

		06/11/17	13/11/17	20/11/17	27/11/17	04/12/17	11/12/17	18/12/17	25/12/17	01/01/18	08/01/18	15/01/18	22/01/18	29/01/2018
Category 1	Mean	00:07:18	00:07:20	00:07:50	00:07:50	00:08:08	00:07:56	00:07:28	00:07:54	00:07:30	00:07:44	00:07:26	00:07:44	00:07:47
	Trimmed mean	00:06:44	00:06:54	00:07:18	00:07:19	00:07:35	00:07:24	00:06:59	00:07:23	00:07:01	00:07:16	00:07:03	00:07:12	00:07:11
Category 2	Mean	00:17:50	00:19:03	00:21:52	00:20:58	00:23:32	00:25:05	00:23:03	00:27:12	00:21:50	00:20:40	00:19:39	00:22:21	00:21:29
	Trimmed mean	00:14:54	00:15:54	00:18:00	00:17:10	00:19:35	00:20:39	00:18:50	00:21:28	00:17:48	00:17:24	00:16:43	00:18:11	00:17:52
Category 3	Mean	00:50:17	00:57:33	01:10:16	01:01:21	01:13:00	01:15:21	01:10:13	01:21:39	01:03:22	01:02:51	00:57:03	01:07:25	01:04:31
	Trimmed mean	00:39:56	00:45:05	00:55:24	00:46:39	00:58:33	01:00:46	00:54:29	01:04:04	00:47:57	00:48:06	00:44:53	00:51:25	00:49:06
Category 4T	Mean	01:44:37	01:44:09	02:09:21	01:57:06	01:59:13	02:10:20	02:01:04	01:58:27	01:55:17	01:39:52	01:32:17	01:48:19	01:48:10
	Trimmed mean	01:30:54	01:29:15	01:51:01	01:38:33	01:42:21	01:53:21	01:44:55	01:42:27	01:38:49	01:26:37	01:19:40	01:32:39	01:32:50

Table 9: Comparison of mean and 90th centile response time for first response (C1) and transporting vehicle response (C1T)

Response time	**EMAS**	**EoE**	**LAS**	**NEAS**	**NWAS**	**SCAS**	**SECAMB**	**SWAST**	**WMAS**	**YAS**
Mean C1	00:08:44	00:08:46	00:07:12	00:06:41	00:10:01	00:07:19	00:08:10	00:10:12	00:06:49	00:07:42
Mean C1T	00:21:29	00:11:56	00:13:17	00:09:11	00:15:37	00:11:32	00:11:15	00:14:14	00:08:19	00:10:55
90th centile C1	00:15:32	00:15:51	00:11:45	00:11:33	00:16:33	00:13:21	00:14:41	00:18:17	00:11:43	00:13:54
90th centile C1T	00:50:39	00:21:37	00:22:49	00:16:13	00:28:36	00:21:47	00:20:12	00:25:43	00:14:42	00:20:17

Secondly, there is an issue around reporting first response or right response. The change to reporting Category 2-4 calls requiring transport to hospital with a clock stop of transporting rather than first vehicle does provide a much clearer picture of response with the right resource and there is no suggestion that this should change as long waits for a transporting vehicle are not hidden from the performance figures. Services have already made or are in the process of making substantive changes to their fleet configurations to better support the principle of sending the right response first time by moving away from the model that had a high proportion of single responder cars that could "stop the clock (with consequent long waits for a back up transporting vehicle) and increasing the number of double crewed transporting ambulances. This has had an effect on reducing long waits for transporting vehicles evidenced in the first evaluation report and tables 1 and 2 in this report show there are only small differences in mean response times for first resource on scene (Table 1) and response times where transporting vehicle stops the clock (Table 2).

However, there is still some debate about the optimum mix of resources. The move to a higher proportion of double crewed transporting vehicles is an advantage for patients needing transport and also reduces crew anxieties where single responders have historically had to spend long periods of time waiting for a transporting vehicle. As ARP changes have become embedded it has also become clear that it is these transporting resources that are taken out of the system when they are queueing at hospitals which then itself results in delays for other patients who need a transporting vehicle and indeed may result in some patients getting a delayed response of any help if there are no resources to send. There is an argument that retaining at least a pool of single responder cars means that, in highly pressurised operating environments as is currently the case, then some resource may be better than waiting a very long time for the right resource. From a patient perspective it may be that getting timely help and reassurance is as important as the "right" resource. In an ideal environment where providing the right resource first time is operationally viable this distinction may not be important but when a system is under pressure it may be useful to separate out performance in terms of providing help and the ideal response.

The current revised AQIs do not make this sort of distinction. For Category 1 calls the clock stop is the first resource on scene given the high risk of a life-threatening condition that may need immediate intervention but this does not reflect any wait for a transporting vehicle and a substantial proportion of these cases will need to go to hospital. Conversely, for categories 2-4 reported response time indicators do reflect provision of a transporting vehicle for patients who need to go to hospital but as mean and 90th centile times are reported for all calls in each category it is not possible to distinguish response performance times separately for transported and non-transported patients. If they were, and if the principle of right resource first time is being achieved then it would be expected that mean and 90th centile times would be broadly similar. If the system is under pressure and there are long waits for transporting vehicles if they are held at hospitals then these may be hidden. It will also not be clear whether, although there may be delays in providing a transporting vehicle (and these will have an impact on overall C2-C4 performance) patients may have received initial help from an alternative resource. If this were demonstrated this may provide some public reassurance.

There are two possible changes to AQIs that would address these issues. Firstly, by reporting mean and 90th centile times for arrival of a transporting vehicle for Category 1 calls. As an interim measure the revised AQIs have included an indicator for Category 1 response time for a transporting vehicle

in patients needing transport to hospital (C1T). Table 9 provides a comparison of C1 (first response) and C1T performance for all services following implementation of phase 2.3.

The table shows that mean C1T time ranges from 90 seconds to 13 minutes greater than mean first response with most services showing a difference of around 4 minutes. For the 90th centile C1T time ranges from 3-25 minutes greater than first response and there is more variance between services with 6/10 services showing a difference of 7 minutes or less. There is clear variability between services in waiting times for a transporting vehicle for Category 1 calls. Some of this may be due to the time period assessed when there is a known problem with handover delays at hospital which reduces the pool of available transporting vehicles and some may be because fleet re-configuration is still in progress. It is worthwhile continuing to measure C1 transporting vehicle times to see if differences reduce over time as the operating environment changes and services make further adjustment to fleet models.

The issue of whether any response should be reflected in the indicators as well as provision of a transporting vehicle only for patients who need to go to hospital could be solved in two ways. Option 1 would be to report C2-C4 response performance separately for transported and non - transported patients. If the ARP model is working well and the majority of transported patients receive the right response first time then response times for each group would be expected to be broadly similar. If a service has insufficient transporting vehicles within its fleet, or if it does but resources are unavailable because of hospital delays, then there may be a difference in performance between the two groups. Small differences may not matter as a principle of ARP is that it is better to delay dispatch, within reason, to send the right vehicle and avoid unnecessary dual responses with substantial delays waiting for back up transport. However if the system is under pressure then increasing variance in response time performance between transported and non-transported patients may be a helpful signal of the size of delays in providing a transporting resource. AQIs already report conveyance rates for different call categories. These are shown in table 10 using the weekly ARP data.

There is some variation in the conveyance rates between services and categories with categories 1 and 2 having the highest rates (around 70%). Rates are lower in categories 3 and 4. Although transported patients outnumber non-transported patients, 30% of Category 1 and 2 patients do not go to hospital and 40-50% of Category 2 and 4 patients are not conveyed so a substantial proportion of the patient population do not necessarily need a transporting vehicle.

Separate reporting of response time performance for transported and non-transported patients would allow better discrimination of parity of response for these two groups. Increasing response times in the transported group would also provide a useful signal of system pressure and delays in providing an appropriate response. What this wouldn't show is how soon patients who need transporting wait to get any ambulance service help – increasing response time may indicate they are waiting with no help or that they are just waiting for the right resource. One solution could be to add a median time to first resource. This would serve two functions, firstly to indicate how quickly at least half of callers receive ambulance service help of some sort and allow some reassurance that when there is system pressure longer waits for a transporting resource does not, in all cases, mean wait with no help.

The ARP Development Group have considered these multiple options for additional measures and although it is agreed that further additions would add richness to the overall picture of ambulance service performance for different types of response this would be at the cost of adding an extra burden in terms of reporting for services. In addition, the purpose of the AQI's is to provide a simple overview of key aspects of performance. Adding more measures would overcomplicate the picture and in particular adding further time measures would be counter to the principle set out at the beginning of ARP that there needed to be a move away from the focus on response times and more emphasis of clinical service delivery and the impact on patient care and the current initiatives to find a "balanced score card" approach to reporting ambulance service delivery across a range of domains. For this reason it was agreed that adding additional time measures would be counterproductive and efforts should instead be directed to further developing indicators that reflect other important service attributes. The monitoring of the Category 3 mean response times as an indicator will bridge the current gap between the monitoring of the category 2 and 3 90th centiles. This however will not be incorporated as a formal standard. An indicator of the 95th percentile call answering time will be added to the balanced score card pending development of a better measure after more detailed investigation of call answering time distributions. This more detailed work is scheduled to be completed by autumn 2018.

Additional quality indicators

The introduction of phase 2.3 and the revised AQIs has included a small number of new indicators. These new measures have only been reported for a relatively short period of time so it is too early to assess their long term usefulness as monitoring tools but here we have provided a summary of the cumulative reporting of these indicators during their first few months of use.

One new introduction is separate reporting of calls from Health Care Professionals (HCP calls) for an ambulance response as these make up a substantial proportion of ambulance activity which is in addition to 999 calls and calls transferred from the NHS 111 service. Table11 shows the AQI reported response performance for HCP calls for all services. For some service the returns are only for a small number of recent months so total incident numbers are low.

The table shows that, overall, most services provide a 90th centile response which is substantially in excess of the requested time frame although a small number provide responses more closely aligned with the expected timeframe. Until there is a full year of data it is impossible to determine if this response performance is a feature of the current operating pressures or a long term trend. Current data suggests, with some exceptions, that services are struggling to provide a timely response within HCP requested timescales, particularly for 1 and 2 hour responses.

Table 10: Proportion of calls transported to hospital for each category – phase 2.3

Volume of Transported Incidents	EMAS	EoE	LAS	NEAS	NWAS	SCAS	SECAMB	SWAS	WMAS	YAS	All
Cat1	69.1%	66.8%	72.2%	69.4%	74.5%	63.4%	62.0%	65.8%	64.3%	76.1%	68.4%
Cat2	73.0%	70.1%	75.8%	76.0%	78.5%	70.3%	70.6%	68.9%	59.4%	78.6%	72.1%
Cat3	59.8%	58.6%	58.9%	63.7%	60.9%	53.0%	54.3%	57.1%	45.9%	68.0%	58.0%
Cat4	59.8%	44.0%	11.3%	43.2%	72.0%	39.2%	43.2%	41.2%	17.4%	53.3%	46.3%

Table 11: AQI reported cumulative response performance for HCP requests August 2017 – February 2018

	1 hour response			2 hour response			3 hour response			4 hour response		
Service	Incidents	Mean h:min:sec	90th centile h:min:sec	Incidents	Mean h:min:sec	90th centile h:min:sec	Incidents	Mean h:min:sec	90th centile h:min:sec	Incidents	Mean h:min:sec	90th centile h:min:sec
England	50,748	1:37:34	3:31:02	46,970	1:58:07	4:08:17	1,850	2:23:41	4:55:46	176,853	2:54:26	6:04:55
EMAS	3,306	1:52:39	3:42:29	3,142	2:20:19	4:04:10	25	2:31:48	3:40:43	12,862	4:01:09	7:11:28
EoE	3,998	2:25:01	5:43:37	3,593	2:58:55	6:48:01	303	3:40:34	8:06:34	4,750	4:19:08	9:55:55
LAS	3,960	1:56:43	4:11:23	6,102	2:10:53	4:28:34	238	2:34:17	4:21:08	5,744	2:33:13	4:43:12
NEAS	2,423	2:16:28	5:24:27	872	2:58:23	6:59:42	-	-	-	169	3:34:15	8:33:25
NWAS	18,130	1:26:58	3:01:01	9,318	1:34:48	3:21:01	757	2:02:28	4:12:12	7,049	1:59:37	4:29:37
SCAS	7,003	0:53:14	1:33:01	5,457	1:36:00	2:49:30	-	-	-	1,504	2:23:18	4:39:59
SECAMB	563	2:09:50	4:51:16	3,986	2:50:02	6:15:17	-	-	-	1,223	3:47:50	8:37:25
SWAST	3,256	1:37:34	3:22:20	2,989	2:01:26	4:10:11	107	3:22:38	8:11:51	3,559	2:24:11	5:06:16
WMAS	6	0:55:27	1:34:34	8,832	1:30:11	3:23:02	-	-	-	8,333	2:02:29	4:40:04
YAS	8,103	1:46:49	4:14:26	2,679	1:39:12	3:39:48	420	1:44:57	3:30:48	15,641	2:39:17	6:06:12

Recognition of Category 1 calls and commencement of CPR in cardiac arrest cases

Two new indicators focus on the most time critical calls and have clinical relevance as this group is the one where time of response and intervention is most likely to have an impact on patient outcome.

A key feature of the early ARP work was a recognition that, if additional time was to be allowed for call triage and dispatch of an appropriate response then a process was needed to ensure early identification and immediate dispatch of a resource for potentially life threatening calls. The use of pre-triage questions and Nature of Call was developed and implemented to support this early identification. A new AQI has been introduced to report the rate of rapid identification of Category 1 calls and the mean and 90th centile times to C1 identification. Table 12 provides a summary of cumulative data reported so far. Call volumes vary by service depending on how long phase 2.3 has been implemented.

Table 12: Category 1 calls identified with Nature of Call (NoC) or Pre-Triage Questions (PTQ)

				Time to identify C1	
Service	Incidents identified by NoC/PTQ	Total Cat 1 calls	% Cat 1 identified by NoC/PTQ	Mean h:min:sec	90th centile h:min:sec
England	138,837	289,802	47.9%	0:49	1:31
EMAS	13,217	37,987	34.8%	0:34	1:22
EoE	24,077	25,790	93.3%	0:40	1:16
LAS	14,225	32,450	43.8%	1:15	2:45
NEAS	3,602	10,707	33.6%	0:32	1:03
NWAS	16,041	58,482	27.4%	1:08	2:30
SCAS	8,802	11,142	79%	0:40	1:23
SECAMB	6,770	9,929	68.2%	1:03	2:28
SWAST	-	17,113		-	-
WMAS	26,113	35,398	73.8%	0:58	1:02
YAS	25,990	50,804	51.1%	0:30	0:52

Overall, almost 50% of Category 1 calls are identified using either PTQ or NoC. There is considerable variation between services in the proportion of C1 calls identified by PTQ and NoC. The end of Phase 1 and 2 evaluation report looked in detail at this issue and found that services using NHS Pathways reported a consistently higher proportion of calls identified by PTQ and NoC than AMPDS sites. This pattern is repeated here with 3 NHS Pathways sites (SCAS, SECAMB and WMAS all reporting close to or over 70% of C1 calls using PTQ and NoC. One other NHS Pathways service (NEAS) reports a much lower rate. There is more variation in the remaining AMPDS sites where the C1 identification rate ranges from 27.4% - 93%. It has been recognised that implementing PTQ and NoC alongside AMPDS is more problematic as there is some replication of questioning. A separate work group has been developing and refining the PTQ and NoC process for AMPDS and is reporting separately. This indicator will be useful for identifying if any changes recommended by the working group results in C1 identification rates increasing. A low PTQ/NoC identification rate does not necessarily translate in

to a longer time in identifying C1 calls as, for all calls the mean time to identification is 49 seconds (range 30-75 seconds) and most services achieve this within a mean 60 seconds. 90% of C1 calls are identified within 91 seconds (range 52 seconds – 2 minutes 45 seconds). Further refinement of PTQ/NoC processes may reduce this time in some services.

The second additional clinically relevant indicator added to the AQIs is time to commencement of bystander CPR in cases of out of hospital cardiac arrest. This is an important measure as early CPR is associated with improved survival. Table 13 summarises the data reported so far but not all services have a reporting mechanism for this indicator yet so there is limited information.

Table 13: Incidents where Cardio-Pulmonary Resuscitation (CPR) is started by a bystander

		Time to start of CPR	
Service	Incidents	Mean h:min:sec	90th centile h:min:sec
England	4,623	4:45	7:56
EMAS	1,814	3:50	6:38
EoE	-	-	-
LAS	-	-	-
NEAS	416	6:47	10:47
NWAS	-	-	-
SCAS	905	5:11	8:47
SECAMB	783	5:11	8:24
SWAST	-	-	-
WMAS	705	4:51	7:59
YAS	-	-	-

The data reported to date shows an overall mean time of 4 minutes 45 seconds and 90th centile time of almost 8 minutes to start of CPR. These times are shorter than the mean and 90th centile ambulance response times indicating that resuscitation begins more quickly in cases where bystander CPR is initiated which has the potential to improve patient outcome. What is not clear from the indicator is the proportion of all cardiac arrest cases that receive bystander CPR. In the longer term there is potential for this indicator to provide a focus for services to identify strategies they could introduce to further reduce the time to CPR although this may be limited as some time is needed to triage the call and identify the problem and then, if CPR is not already taking place, provide instructions to a caller before CPR can start.

Patient Safety

The need to maintain patient safety has been a core component of ARP and throughout the different ARP phases there has been a continuous monitoring process to review any potential adverse incidents related to ARP changes. Services also assess long waits to distinguish between calls where this happens as a consequence of resource issues from those that may be related to ARP, for example calls that potentially may have been assigned to the wrong category. Patient safety monitoring is the responsibility of individual services as this is a regular process aligned to safe

operational delivery where potential problems need to be identified quickly. It is not a one off exercise conducted by the evaluation team after a period of implementation. Every service has in place a robust process for monitoring patient safety and throughout the ARP changes there has been a mechanism for early reporting of potential adverse incidents to the ARP programme team so that independent clinical scrutiny can be provided and each service provided a patient safety report at the monthly ARP delivery group meetings. The National Ambulance Medical Directors Group (NASMED) have provided oversight to ensure processes are in place to monitor patient safety related issues and report any identified problems and have provided a statement of assurance that no ARP related adverse incidents have been identified during the implementation of phase 2.3.

"The Medical Directors of the English ambulance services have confirmed that all English ambulance services have processes in place which monitor for response-related adverse incidents and confirm that no incidents have been reported that are directly attributable to the introduction of the revised operating model in the Ambulance Response Programme."

For monitoring long waits there is no standard process but each service has put a strategy in place. As an example, South Central Ambulance Service (SCAS) has a well-defined process for monitoring long waits as part of their broader patient safety monitoring which reports to their Quality and Safety Committee. The long waits group investigates initial causes for long waits utilising other resources (such as datix reports) to determine whether there are any patient risk or poor experience issues that need further action. As part of this process ARP related changes are considered when identifying causes of long waits.

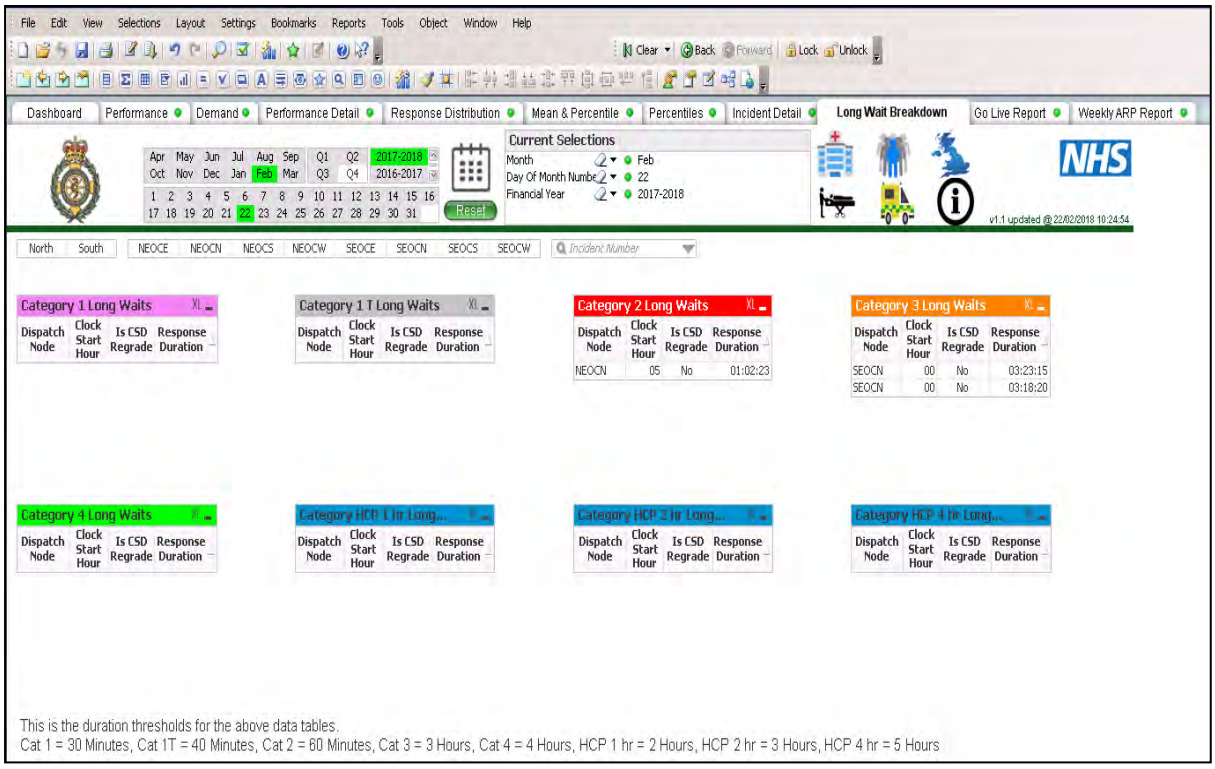
An important feature of monitoring long waits is defining what a long wait is and the criteria for identifying which calls to review. SCAS have determined a time threshold at the 99th centile and calls which exceed this threshold trigger a review. The time thresholds have been revised in line with the Category 1-4 response standards and the current threshold are described in Table 14.

Table 14: Long wait threshold times – South Central Ambulance Service

SCAS ARP Long Wait Thresholds		
Category	**National or local HCP Standard**	**Long Wait Threshold / 99th centile**
CAT 1	7 mins mean; 15 mins 90th	**30 mins**
CAT 1 T	Internal Transport measure – 18 mins	**40 mins**
CAT 2	18 mins mean; 40 mins 90th	**60 mins**
CAT 3	120 mins 90th	**3 hrs**
CAT 4	180 mins 90th	**4 hrs**
HCP 1 hr	60 mins	**2 hrs**
HCP 2 hr	120 mins	**3 hrs**
HCP 4 hr	240 mins	**5 hrs**

These threshold mean that, for example, if the 99th centile response time for C1 calls is 30 minutes or longer, all calls with a response time of 30 minutes or more will be reviewed. A similar strategy is used by other services. The advantage of using pre-defined time thresholds is that there is a systematic and consistent approach to identifying calls for review. It also allows the development of automated systems which can flag up calls with long waits in real time and collate them for review. Figure 27 illustrates the automated process within SCAS used for rapid identification of calls with long waits.

Figure 27: Automated long wait call identification system - SCAS



A potential disadvantage of using a pre-determined threshold is that the volume of calls that require review will fluctuate depending on the operating environment and external pressures. So, when a service is performing well if the 99th centile response time does not reach the threshold there will be no long waits to review. If the 99th centile threshold is reached then 1% of calls will be reviewed. If there is high pressure and falling performance the threshold may be reached before the 99th percentile – for example this may be at the 95th percentile generating a bigger proportion of calls potentially eligible for review. If this happens then either the longest 1 % of calls can be reviewed (those over the 99th centile only) or all calls exceeding the threshold – 6% - could be reviewed but this may create an unmanageable workload.

Other services take a different approach and as an alternative to using pre-determined thresholds linked to centiles, the volumes of which will fluctuate, set time only thresholds but concentrate on what could be termed very long waits as it is these cases that are of most clinical concern and where there is highest risk of patient harm. To illustrate, if for Category 2 there is a threshold of 60 minutes and the 95th percentile time is 61 minutes then 6% of calls would exceed the threshold and

potentially require review. However, within this 6% there will be calls that only just exceed the threshold – 65, 70 or 80 minutes, and others are the end of the distribution which may be more than 2 hours. Reviewing all 6% would be a major task but not all will be a major risk. This could be mitigated by only reviewing those over the 99th centile but there may then be cases with a response time just below this where a problem is missed.

An alternative is to set a threshold for very long waits that are proportionally much longer than the expected standard. One strategy would be to identify all calls that are 2 or 3 times greater than the 90th centile time for a call category. South Central Ambulance Service has provided an example of the likely call volumes of long waits for each call category if this strategy is utilised (Table 15). The example shows that the numbers of calls in categories 1 and 2 that exceed response standards by this magnitude are very small but there are potentially large numbers for categories 3 and 4 when demand is high and services are under pressure.

Of course not all long waits will result in harm (though inconvenience and not meeting patient expectations). Where harm does occur this may be for many reasons, not just a delayed response. This means setting thresholds and criteria for identifying calls where patients wait for long times outside the expected response standards is complex and there is no obvious "right" way to assess which option is best. However, as part of the ongoing monitoring process as ARP moves to "business as usual" some oversight of long waits is agreed to be necessary which needs a standardised process that all services could use to ensure the process is robust and provide consistency. This process needs to be useful in terms of providing a "signal" where long waits are becoming a short term or persistent problem but not over burdensome in requiring large numbers of complex case reviews. As a first step it has been agreed that reporting of numbers of calls exceeding the 90th percentile response standard by 2 and 3 fold for each call category will be added to the balanced scorecard. This will provide the signal. The next step will be to develop and introduce a standardised process for reviewing serious incidents where long waits may have been a contributory cause, so that the effects of long waits are transparent and lessons from these events can be learned.

Table 15: Number of potential cases for long waits review - SCAS

Using National Standard for 90th centile

		Cat1			Cat2			Cat3			Cat4		
Week Start	**Measure**	**90th centile**	**x2 90th centile**	**x3 90th centile**	**90th centile**	**x2 90th centile**	**x3 90th centile**	**90th centile**	**x2 90th centile**	**x3 90th centile**	**90th centile**	**x2 90th centile**	**x3 90th centile**
13/11/2017	**Response Time**	00:15:00	00:30:00	00:45:00	00:40:00	01:20:00	02:00:00	02:00:00	04:00:00	06:00:00	03:00:00	06:00:00	09:00:00
	No of Incidents	52	1	1	233	38	8	351	34	6	50	1	0
	Total incidents	830			4509			3733			454		
25/12/2017	**Response Time**	00:15:00	00:30:00	00:45:00	00:40:00	01:20:00	02:00:00	02:00:00	04:00:00	08:00:00	03:00:00	06:00:00	09:00:00
	No of Incidents	74	4	0	808	236	98	868	340	87	82	24	9
	Total incidents	626			5740			3519			310		
29/01/2017	**Response Time**	00:15:00	00:30:00	00:45:00	00:40:00	01:20:00	02:00:00	02:00:00	04:00:00	08:00:00	03:00:00	06:00:00	09:00:00
	No of Incidents	41	0	0	229	23	7	351	60	0	28	2	0
	Total incidents	557			4687			3402			339		

Summary of ARP Phase 2.3 progress

Implementation of the recommendations of the evaluation of ARP phases 1 and 2 was mandated in July 2017 and since that time all 11 ambulance services in England, having already implemented dispatch on disposition, have now moved to the clinical response model using revised call categories with corresponding new response performance standards (the Isle of Wight has not yet completed this transition). The scale of national change cannot be underestimated. This has been the biggest substantial change in ambulance operating practice in England for 40 years and has required enormous effort from ambulance services to operationalise these changes. This has involved the complex technical challenges required to support new call triage and dispatch processes but also the wider organisational challenges of new working practices for staff, wholesale review of fleet configurations and rostering and explaining and persuading wider audiences of trust boards, governors and the public of the value of a change that could, by many, be interpreted as a "downgrading" of services given the prominence that response times have always held in determining judgements of quality and setting expectations. The overarching aim of the Ambulance Response Programme was to develop a clinical operational model that could adapt to an environment of continued increases in demand and limited resources but ensure the fastest response to those patients who need it whilst maintaining a safe and clinically appropriate service to every other patient who requests ambulance service help. Overall, this has been achieved and, for half of the services, the necessary changes have been made during the busiest and most challenging time of the year. The fact that between July and November 2017 five services had successfully transitioned to the new operating model and that within the space of 2 years two substantial changes have taken place across 11 ambulance services serving a population of 65 million people should be recognised as a huge achievement.

This does not mean the transition has been easy. Analysis of response performance has shown that, over winter, many services were still struggling to achieve the expected response standards and there were some exceptionally long waits for lower priority categories. This could be interpreted as a failure but, of course, setting any standard does not guarantee achievement. There have always been long waits for some patients, despite an expectation of an 8 minute response to 50% of calls, but previously these waits were hidden. Now they are transparent and any changes obvious. The move to reporting mean and 90th percentile times for each call category has been a real step forward and provides a much more truthful reflection of the type of service the public are receiving. More recent data as the pressures of winter ease show a much more encouraging picture with the majority of services meeting or nearly meeting the expected standards and in some cases performing well above what is expected. This demonstrates that the performance standards set for the revised categories are not unrealistic. There do remain a small number of Ambulance Trusts where performance against expected standards remains a challenge and more work needs to be done but this emphasises the fact that delivering ambulance services is a complex process that requires much more than a set of standards to optimise service delivery. The collaborative working across services throughout ARP does mean that services will continue to learn from each other as they seek solutions to specific challenges.

A feature of the new call categories is the substantial increases in the expected response times for different categories of calls. It could be assumed that this would result in equivalent changes in service delivery but our analyses of whole service performance show this is not the case. Overall the introduction of phase 2.3 has resulted in either no or very modest increases in response time performance for the 999 population comparative to the increased time standards for some of the revised call categories. They also show some additional efficiency gains in terms of further reductions in average resource allocations per call. The results also reaffirm a feature identified in the Phase 1 and 2 evaluation which is that the introduction of the revised call categories goes some way to supporting stable performance across the whole population.

There has also often been an assumption that patients in rural areas receive a substandard service compared to those in urban areas. The analyses of differences in performance between urban and rural areas showed that this does not hold as a general rule. There are longer times in rural areas in some services but it is equally if not more likely, overall, that times can be longer in urban areas and that the overall proportion of calls originating in urban areas is a bigger driver for differences in response time performance than geographical area.

The introduction of ARP phase 2.3 has revealed some unintended consequences. One is a possible increase in the number of duplicate calls generated by callers or patients calling back to check when a response can be expected. This in turn has a detrimental effect on call answering times for new calls as call handlers are managing a greater volume of calls. Some strategies have already been developed including scripts to inform callers of the likely wait and have seen some success. There is now a process in place to standardise the use of call scripts across all services to mitigate this effect. It also illustrates that there is probably more work needed to inform the public about the revised call categories and adjust expectations about how quickly they can expect a response, particularly when their problem is not a clinical emergency.

Alongside the implementation of the revised call categories a new set of ambulance quality indicators has been introduced. These have, as previously described, made response performance much more explicit and transparent across all categories. They have also seen the introduction of some new, more clinically focussed indicators such as time to CPR for patients in cardiac arrest. To support the revised system indicators the ambulance clinical quality indicators have been revised and for the first time will be able to link to national registries to report more accurate outcome data for the key conditions of cardiac arrest, stroke and heart attacks. New indicators are also being developed for sepsis and falls. This development is a work in progress but in the short term the ARP recognises that the changes implemented are still new and will require ongoing, consistent and frequent monitoring. To support this a balanced score card has been developed that will report, on a weekly basis, the key response performance standards but also some additional indicators to provide a broad overview including response time for a transporting vehicle for Category 1 calls; call answering times (which will be further developed); proportions of calls with different types of call closure (hear and treat, see and treat and see and convey) and the number of calls with long waits exceeding twice and three times the 90th centile time in each call category as a patient safety measure, followed by the consistent reporting of serious untoward incidents (SUIs).

The ARP changes have been successfully implemented across all ambulance services in England. It is testament to the hard work and enthusiasm of all services that this has been achieved in a short

space of time and in an environment of substantial pressure. The review of implementation shows that for the most part this has been a positive step forward but challenges remain. Some of these may be tackled by further refinement and maturing of phase 2.3 initiatives. Services have reported significant problems over the 2017/18winter period with queueing calls and no resources available to send. Demand and performance are closely linked and there will come a point where services have little or no capacity to maintain performance as demand increases. ARP initiatives will have helped to mitigate this to some extent but there will come a point beyond which all efficiencies are exhausted and demand and supply problems need solutions outside those that ARP can deliver if services are to be expected to deliver against the expected standard.

Clinical Quality Indicators

Ambulance Clinical Quality Indicators (CQIs) have been in place since 2011 to measure and monitor the impact of ambulance services on patient outcomes, and in particular to provide an overview of the clinical quality achieved by ambulance services based on four CQI topic areas:

- Return of Spontaneous Circulation (ROSC) at hospital following Cardiac Arrest;
- Out of Hospital Cardiac Arrest (OHCA) Survival to Hospital Discharge;
- Management of ST Elevation Myocardial Infarction (STEMI); and
- Management of Stroke.

As one component of the Ambulance Response Programme (ARP) NHS England has reviewed the current CQIs to determine if they could be improved to better reflect clinical care. We have undertaken a process of stakeholder engagement, facilitated by Sheffield University, to ask ambulance leaders, clinicians, commissioners and patient representatives to identify the areas of clinical care on which we should focus.

Following the engagement exercise, and after discussion with the Secretary of State for Health and Social Care, the following focus areas were agreed:

- STEMI: 999 call to angiography (mean & 90th percentile)
- Stroke: 999 call to CT scan, and 999 call to thrombolysis (mean & 90th percentile)
- OHCA: Survival to hospital discharge following out of hospital cardiac arrest (Utstein group)

Development of these areas will be ongoing as we work to introduce reporting of data across the patient pathway and as ambulance Trusts begin to link into the national outcome databases for Stroke, STEMI and Out of Hospital Cardiac Arrest. Phase 1 of the Stroke and STEMI work will report:

- Call to Door and Door to Balloon (angioplasty) times for STEMI; and
- Call to Door (arrival at stroke centre), Door to Scan and Door to Reperfusion times for Stroke.

Phase 2 work will be ongoing with the end position reporting of:

- Proportion of confirmed STEMI patients that receive definitive treatment within 150 minutes of making a 999 call; and
- Proportion of patients that complete their pathway of care (call to reperfusion) within 180 minutes of making a 999 call.

We expect 90% of patients will meet this standard by 2022.

NHS England and partners are continuing to monitor clinical care bundles for STEMI, have now more accurately described the Stroke care bundle as a diagnostic bundle and will implement a care bundle for those patients who have a sustained ROSC before arrival at hospital. These measures will further support the delivery of care alongside the focus on time based measures.

NHS England is also working with the National Ambulance Service Medical Directors group (NASMeD), ambulance service colleagues and commissioners to develop meaningful clinical indicators for sepsis, patients who have fallen and are still on the floor, and those with mental health needs.

The first set of revised CQIs was published in April 2018 reporting data from November 2017. This time lag is due to the preparatory work required for the new indicators. We will continue to work to reduce the time gap between the reporting period and publication dates.

We believe that ambulance trusts will require a ‘grace period’ until September 2018, in line with the 2018/19 refreshed planning guidance, to enable the new clinical standards to bed in, and to allow time for the required system changes to be made. Thus commissioners are expected to work with their providers in the meantime to monitor performance against the new clinical standards and to discuss progress, but not to formally hold providers to account.

For the time being, ambulance commissioners and providers should continue to submit all existing monthly data collections. We are working with ambulance colleagues and national clinical audit database suppliers to link data across the patient pathway, and will be specifying the data collection methods in due course.

It is our intention that the revised ambulance Clinical Quality Indicators and reporting requirements will support a clear focus on clinical care, patient outcomes and experience and these will be included in the new ambulance balanced scorecard.

Use of Nature of Call (NOC) in MPDS Trusts

AACE review & recommendations for NHS England ARP Development Group

March 2018

Prepared by - The Association of Ambulance Chief Executives
Daniel Gore – Mike Boyne – Dr David Macklin

Contents

Document history	
20180119	Initial draft
20180130	Appendix 2 added
20180209	Further drafting DG / MB
20180216	Further comments DM/MB – submission in draft to ARP DG
20180223	Minor changes post feedback DG
20180307	Updated x 1sfn chart following Trust correction (NWAS / LAS)
20180308	Final release to NHS England
2180316	Amended chart page 9

1. Summary

The evaluation of the Ambulance Response Programme (ARP) conducted by the Sheffield University School of Health and Related Research (ScHARR) outlined the findings of extensive trials of a range of new operational practices within Ambulance Trusts.

The evaluation of ARP Phase 1 focussed on the process known as Dispatch on Disposition (DoD) which allows additional triage time for all calls other than those which are immediately life-threatening. The evaluation concluded that there was

"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."

The intended benefits of DoD are to:

- provide a more clinically appropriate response by targeting the right resource (clinician skills and vehicle type) to the right patient
- reduce 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)
- increase "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.

In order to safely allow additional triage time for DoD to be conducted by call handlers, additional processes were introduced to identify patients with immediately-life threatening conditions at the earliest opportunity. As explained in the evaluation report:

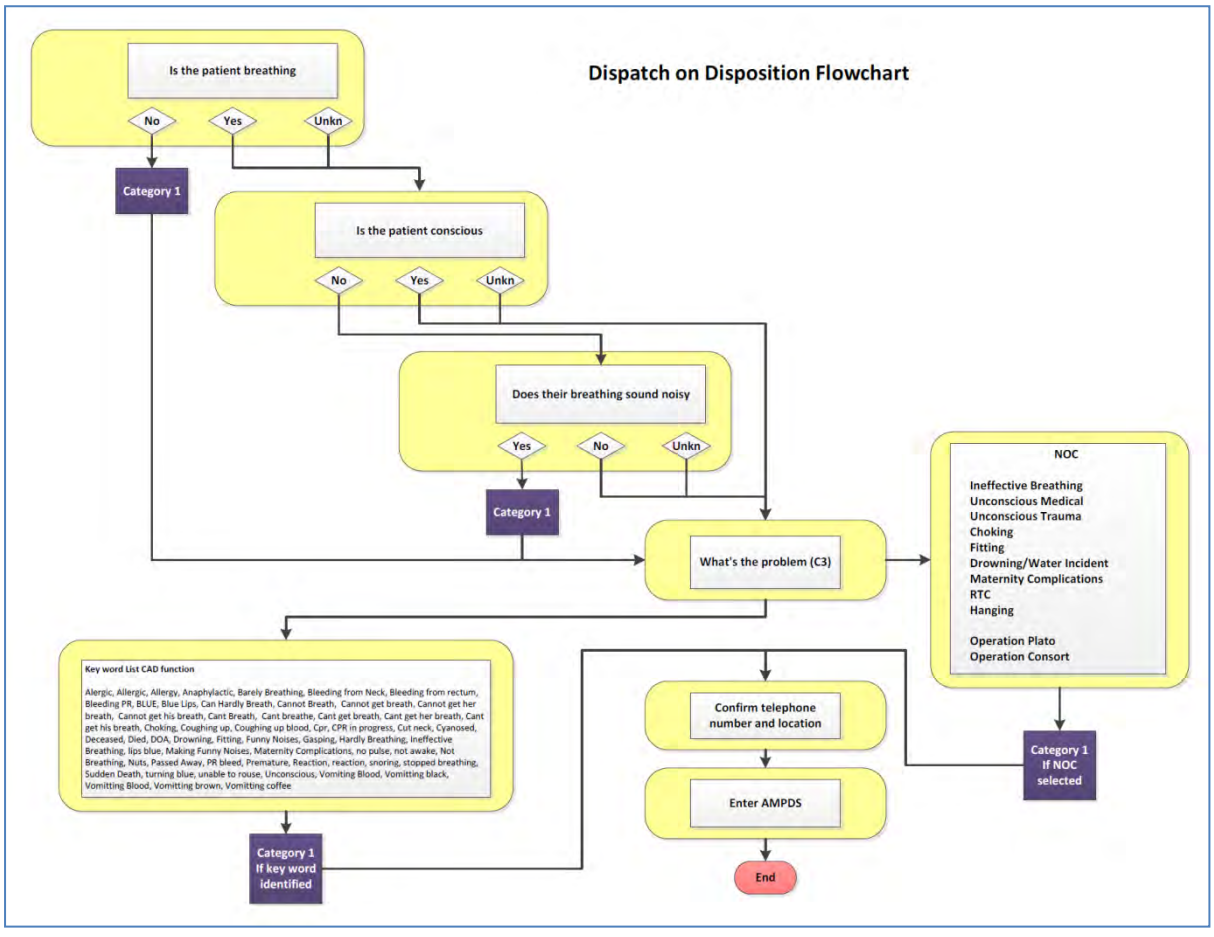
"A component of the Dispatch on Disposition initiative and subsequent call category trial has been the introduction of three pre-triage sieve (PTS) questions and nature of call identification using a predefined list of problems (collectively NoC) at the beginning of the call management process. The purpose 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."

As a consequence of these findings, the use of the three PTS questions was mandated for all Ambulance Trusts on joining the ARP in order to provide the "safety net" to ensure that a reasonable proportion of immediately life-threatening patients were identified at the earliest opportunity.

The PTS questions were developed to enable rapid identification of patients in cardiac arrest or at immediate risk of cardiac arrest. Patients who are not breathing or those who are unconscious with noisy breathing are a cohort of patients who require the fastest response from the ambulance service in order to improve their chance of survival. It is therefore vital that these patients are identified ahead of other less critically ill patients.

The evaluation found strong evidence that the addition of NoC added further benefit for trusts using NHS Pathways and hence the use of NoC was mandated for those trusts. But the evidence of benefit from the NoC process for trusts using MPDS was less conclusive and therefore the use of the process was not mandated for trusts using MPDS

A useful example summary of the PTS/NOC process appears below:



NHS England, through the ARP Delivery Group, commissioned the Association of Ambulance Chief Executives (AACE) to form a task and finish group in order to establish a definitive position on the question of whether NoC added benefit for trusts using MPDS and whether it should be mandated in the same manner as for trusts using NHS Pathways.

This report summarises the output of the task and finish group and will present evidence to support the following findings:

1. That Cat 1 patients are identified much sooner in the process where early predict arrangements are in place when compared to the time taken to reach an MPDS code for these calls.
2. That NoC has been found to increase the proportion of Cat 1 incidents that can be early predicted.
3. That the combined use of the three “building blocks” of PTS, NoC and CAD based Keywords produces the highest level of early prediction.
4. That correct use of the mandated PTS process and recommended NoC options increases the accuracy of early prediction.
5. That there remains scope to improve the accuracy, speed and efficiency of early prediction still further and that a group should be established to continually review and make recommendations about the future design of pre-triage processes.

2. Objectives

The objectives of this piece of work were to establish a definitive position on the question of whether NoC adds benefit for trusts using MPDS and whether it should be mandated in the same manner as it is for trusts using NHS Pathways.

It was agreed that AACE on behalf of NHS England would lead a task and finish group with the following core objectives:

- Assess each Trusts current call flow process to gain understanding of their approach to Category 1 early identification.
- Agree key measures for data collation relevant to Category 1 process.
- Consider the data provided and agree the relevant measures required to inform the recommendation to NHSE.
- Analysis of the key data measures in the context of the success of a NOC providing early predict of Category 1 patients.
- Provide a recommendation to NHSE as to whether a NOC has a place in MPDS Trusts to support early identification of Category 1 patients and if it should be mandated.

3. Approach to the work stream

In order to fulfil the objectives set out above, AACE convened a task and finish group with representation from each of the six MPDS Trusts.

London Ambulance Service NHS T (LAS)
East of England Ambulance Service NHS T (EoE)
East Midlands Ambulance Service NHS T (EMAS)
North West Ambulance Service NHS T (NWAS)
Yorkshire Ambulance Service NHS T (YAS)
South West Ambulance Service NHS FT (SWAST)

The process that each Trust was using at the start of the monitoring period was shared with the group so that the part this plays in Category 1 early predict was fully understood. Of the Trusts, three were using a NOC already (LAS, YAS and EEAST). Helpfully the NOC codes that were being used were broadly the same in each of the

three Trusts. It is worthy of note that in YAS and EoE, the PTS and NOC process is imbedded within the CAD, currently in LAS this is more of a manual process.

There were differences in the order in which Trusts were undertaking the call flow and associates questions, for example the point where location was verified, however the five Trusts used were all undertaking the mandated pre-triage sieve as their first process step. These process step variances will impact on the data.

It should also be noted that during the period Trusts were working through various Category 1 improvement initiatives internally which, in some cases, changed their process. AACE have worked closely with all the Trusts to support and understand these initiatives and any impact they may have had on the data and recommendation have been factored in.

It was agreed at the outset that it was important to wait until each Trust was fully live on ARP and able to provide three months data for the required measures. This meant that data for November, December 2017 and January 2018 was used.

SWAST data has not been included in the review as the trust was in the process of transitioning from use of NHS Pathways to MPDS, a number of process changes were made during the data gathering period and the trust has since completely revised their approach to early prediction.

In total, there were 39 data measures agreed for Trusts to submit relating to their Category 1 process. Not all of the measures relate specifically to PTS/NOC processes, however the wider set of measures will be beneficial in terms of other shared learning.

Following the first review of all of the data measures, the group agreed that those measures which should be specifically considered for the NOC recommendation were:

Identification
% of Cat 1 incidents early identified
% of Cat 1 incidents early identified from PTS
% of Cat 1 incidents early identified from NOC (where in place)
% of Cat 1 incidents early identified from key words (where in place)
% of Cat 1 incidents **not** early identified
Length of time from call pick up to make prediction
Average time call pick up to early predict achieved (for Cat 1 calls with an early prediction)
Average time call pick up to early predict achieved PTS
Average time call pick up to early predict achieved NOC
Average time call pick up to early predict achieved Key words
Average time call pick up to MPDS determinant where **no** early predict
False predict rate
% Cat 1 incidents **not** finally coded as Cat 1
% Cat 1 incidents positive predict finally coded as Cat 2
% Cat 1 incidents positive predict finally coded as Cat 3
% Cat 1 incidents positive predict finally coded as Cat 4
% Cat 1 incidents positive predict finally coded as Cat 1

These measures were chosen by the group as it was agreed that many of the other measures were significantly impacted by other confounding factors outside of the call taking process. For example, if the response time standards were used as a measure, these are influenced by issues such as such as allocation times and availability of resources. If the speed of allocation was used as a measure of success, this can be impacted by other factors such as adherence to system status plans or individual dispatcher allocating speed.

Using the measures above allows us to assess from the point of call pick up to the first point at which a potential Category 1 patient can be alerted to dispatch for onward processing. These measures insulate the call handling process, and in particular pre-triage processes, from most confounding factors resulting in a data set that provides reliable measures upon which recommendations may be made.

4. Data findings summary

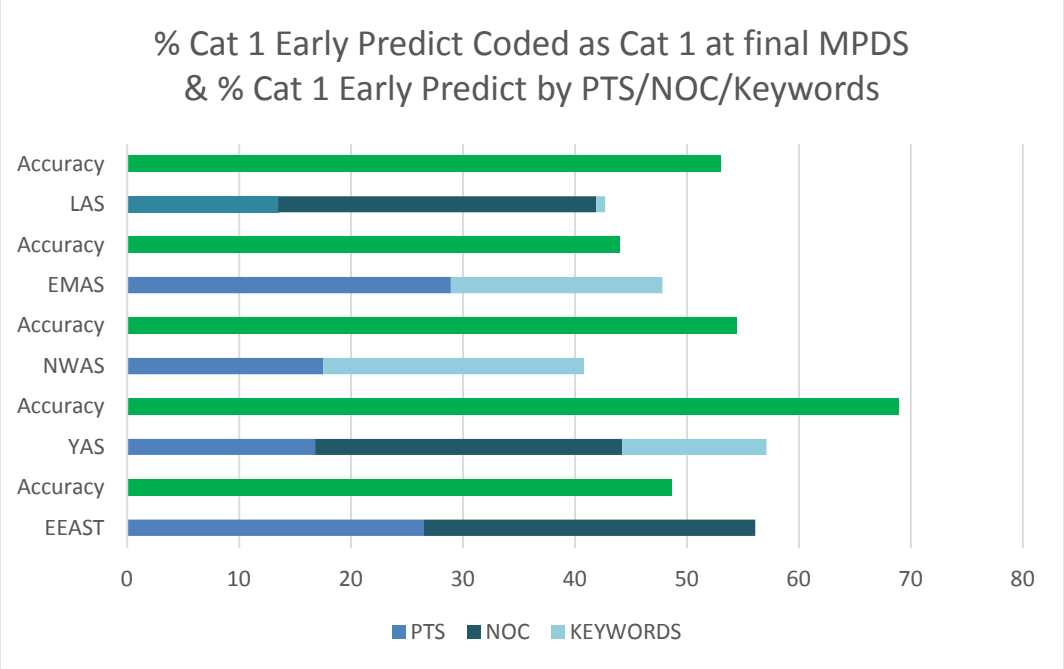
The 15 measures outlined in section 3 above focussed on the three factors judged to be most important in determining whether it is appropriate to mandate the use of NoC for trusts using MPDS:

1. **Cat 1 early predict capture rate:** Does the use of NoC improve the percentage of Cat 1 incidents that are early predicted over and above the use of MPDS alone, PTS and/or MPDS plus PTS?

2. **Time to early predict:** Does the use of NoC add time to the call handling process that would lead to undue delay in reaching a dispatch decision for Cat 1 incidents?

3. **Accuracy:** Does the use of NoC have an impact in improving the accuracy of early prediction of Cat 1 incidents?

A detailed analysis of the data collected is outlined in section 5 of this report. A summary of the key elements is outlined in the tables below:

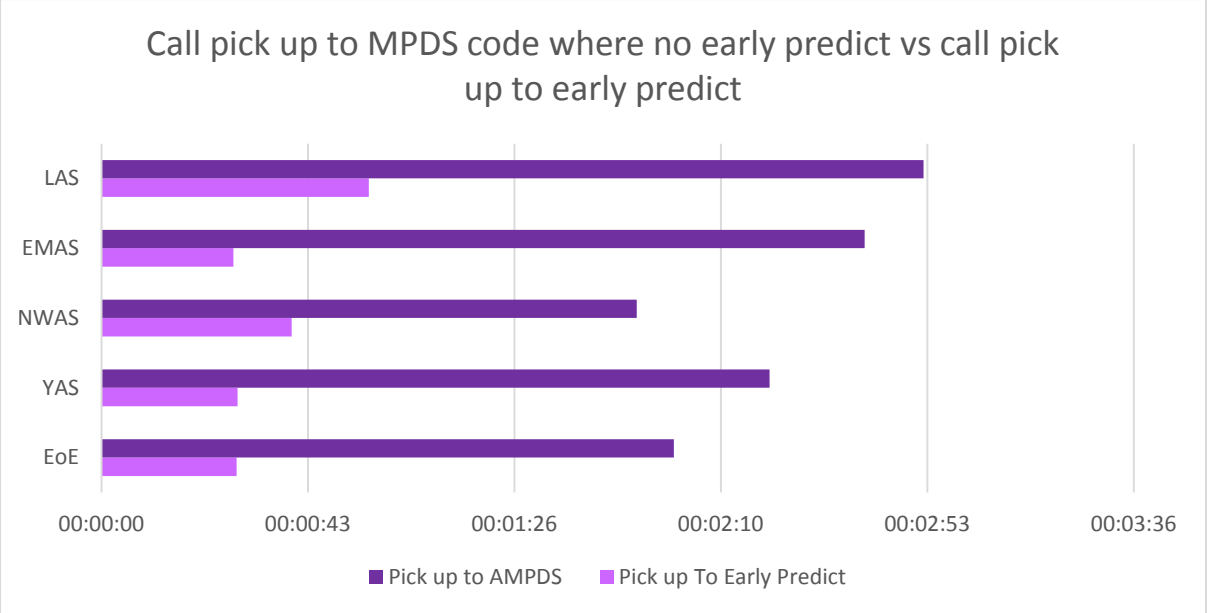


The chart above shows the average percentage of Cat 1 incidents that were early predicted by each trust over the three-month data gathering period. The chart shows the proportional contribution made by PTS, NoC and Keywords in each trust. Although there is some variation between trusts in their application of the three early predict components, the data indicates that there are clear improvements in the early predict capture rate where NoC is in use and embedded properly within the CAD.

Note: During the data collection period the PTS and NoC in LAS was not embedded within the CAD which will have negatively impacted on its effectiveness.

The chart also shows the average percentage of early predicted Cat 1 incidents that were subsequently coded as Cat 1 when arriving at a final MPDS code. It is a good measure of the accuracy of early prediction.

The data indicates that the highest Cat 1 early predict capture rate and the highest level of early predict accuracy were both achieved in YAS through the use of PTS, NoC and Keywords.



The chart above shows the average time from call pick up to the time to reach a final MPDS code where **no** early prediction has been made. It has been included as a proxy measure to indicate the time taken in MPDS to reach a definitive coding outcome if no early predict tools to be used. The chart also shows the average time taken from call pick up to the time that an early prediction of Cat 1 was made.

The data gives strong evidence to support the position that early predict tools lead to more rapid identification of Cat 1 incidents than the use of MPDS alone. YAS, EEAST and EMAS all achieved an early prediction at 28 seconds on average over the three-month data collection period.

The data indicates that YAS, through a combination of PTS, NoC and Keywords are capturing the highest proportion of Cat 1 early predictions (57%), with the highest degree of accuracy (69%) at a speed (28 seconds) that matches early predict times of trusts which are only using two of the three pre-triage components (EMAS and EEAST).

EEAST achieved a similar level of Cat 1 early prediction (56%) and with the same speed as YAS (28 seconds). They had the second highest level of accuracy (49%) but that is markedly lower accuracy than YAS (20% points lower). EMAS (who were predicting all unconscious patients as Cat 1 during the PTS process) achieved a lower level of Cat 1 early prediction than YAS (48%) with the same speed as YAS (28 seconds) but with the lowest degree of accuracy (44%).

It should be noted that the data collected relates to the various early predict systems that were in use by trusts during the data collection period. From this we are confident that we have identified the most effective combination of processes that are currently in use by trusts and that we can make recommendations about the best practice that could form the basis of a mandated process. On the strength of the existing data combined with operational observation in all trusts, we recommend the YAS approach as current best practice (combined use of PTS, NOC, Keywords).

However, it is also possible, likely even, that efficiency can be improved still further through continual development of the precise design of PTS and NoC processes. Therefore, our findings give a reliable indication of the best solution currently known to us but it is unlikely to be the definitive, optimal pre-triage solution.

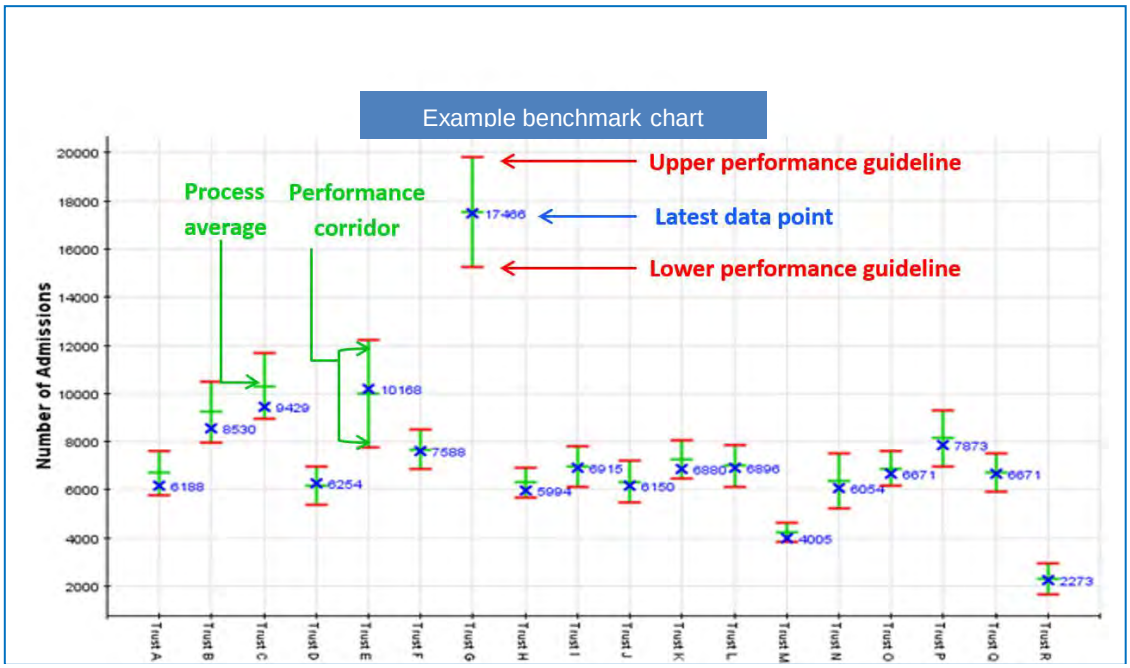
5. Detailed data analysis

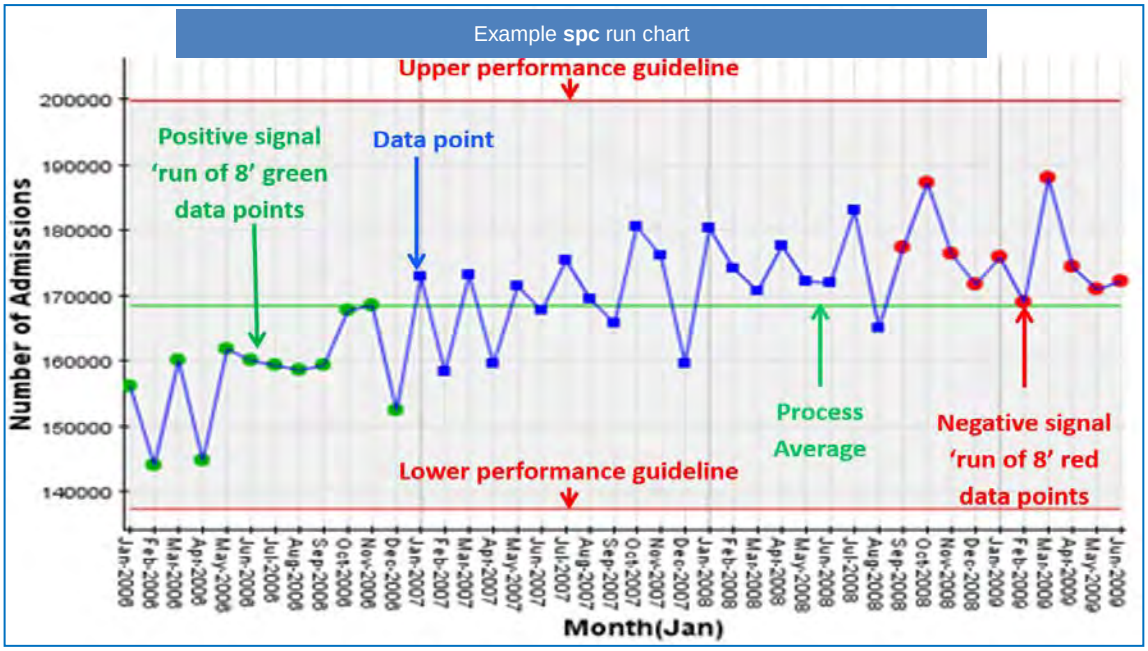
This section sets out the more detailed data analysis for the specific measure used as set out in section 3. The charts used are **spc** and benchmarks charts which have been supplied within the Lightfoot Solutions data analytics platform who have kindly supported AACE and the five Trusts with this work stream.

Where there are slight differences in the figures quoted in section 4 above, from those shown below, this is because the benchmarking tool shows the latest performance where appropriate with process breaks, rather than the average of the whole data period, to highlight both positive and negative changes. Moving forwards, AACE will use the tool to support Trusts in highlighting and sharing best practice.

Lightfoot ***sfn*** charts explained

The two charts below are provided to help readers understand the types of chart being used.





For ease, green text boxes have been added in the benchmark charts below showing the mean for each measure and Trust.

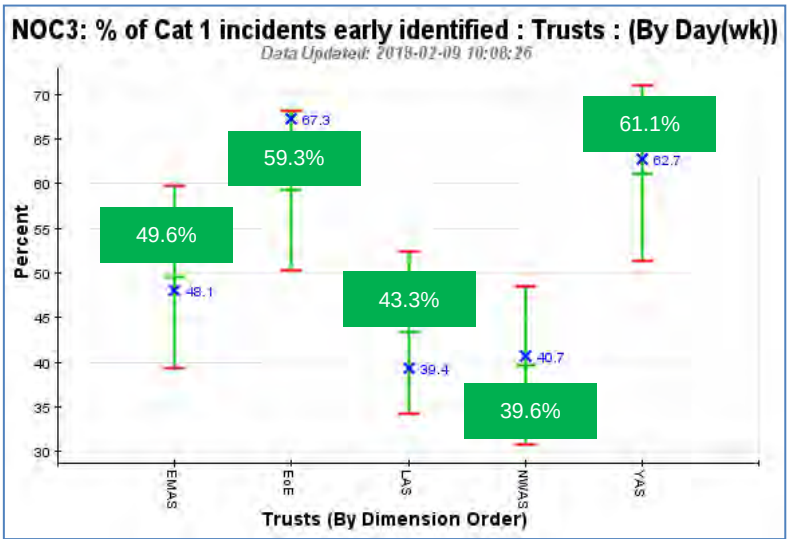
Cat 1 Patient early identification

Percentage of Cat 1 patients early identified overall

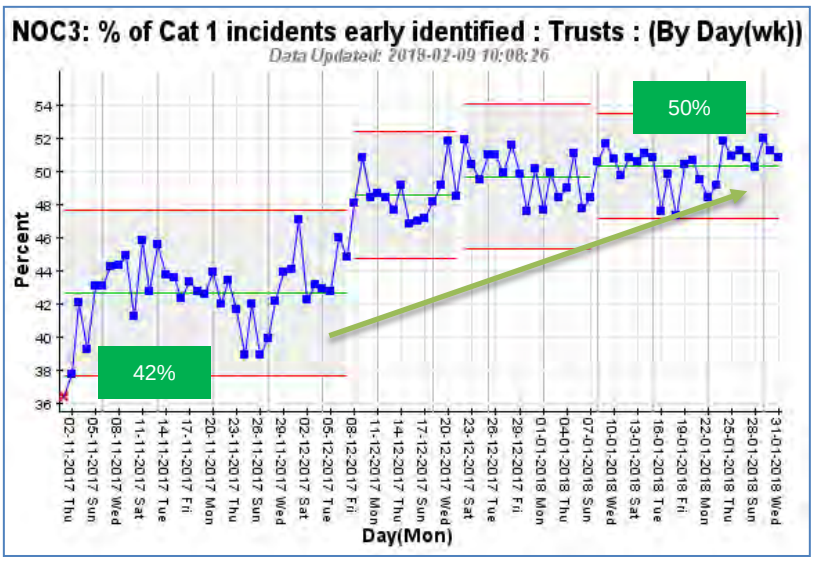
As described within this report, it is essential that control rooms are able to identify potential Cat 1 patients as quickly as possible in the call cycle and present them to dispatch for allocation.

The chart below sets out for each Trust the mean percentage of Cat 1 incidents early identified.

YAS and EoE predict the highest percentages of Cat 1 patients early with their process which includes use of an MPDS NOC

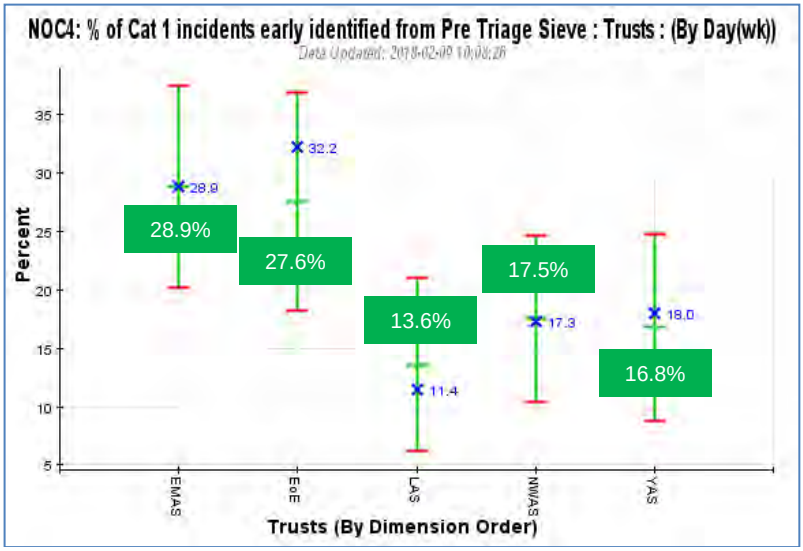


As can be seen below, throughout the data collation period, the volume of patients early identified at a “five Trust” view has increased as Trusts work through their improvement initiatives.



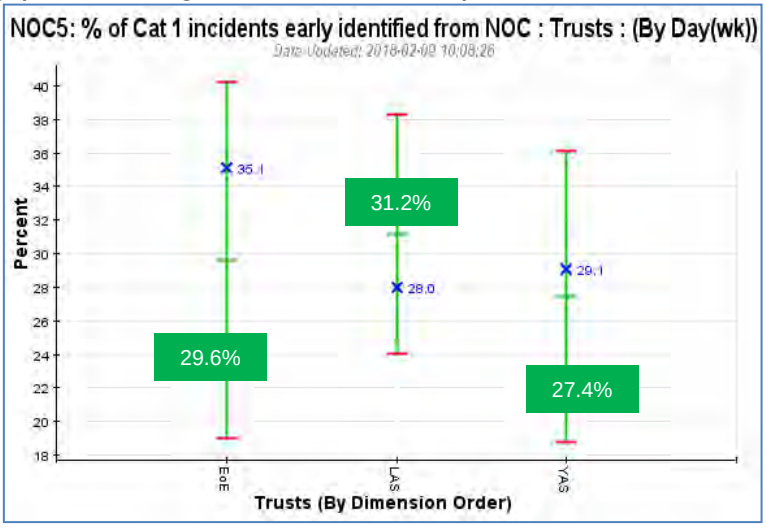
Cat 1 Patient early identification Pre-triage sieve

The chart below shows the percentage of Cat 1 patients identified through the pre-triage sieve. As can be seen there is variation across the Trusts. In EMAS we believe it may be their current process of early predicting all patients declared as unconscious in the PTS as potential Cat 1, which is causing the higher level of predict at the PTS.
For EEAST it is thought this is associated with some further work undertaken on identifying potential Cardiac Arrest. In summary it would seem reasonable to suggest that on average around 17% of Cat 1 patients are identified through the three pre-triage questions where those not breathing or unconscious with noisy breathing are selected.



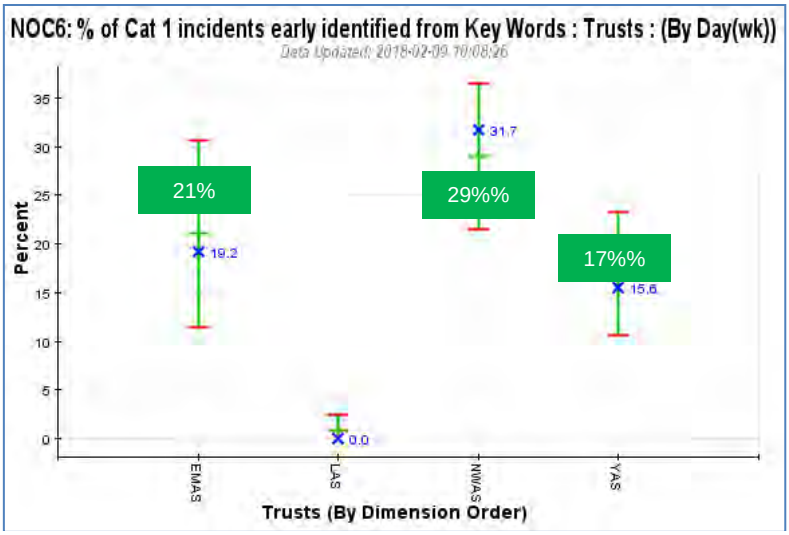
Cat 1 Patient early identification NOC

There is some variation in terms of the percentages of patients identified through an MPDS NOC. This could be factors such as slight variation in the NOC types used, or call taker understanding and application, however the chart below sets out that it is possible to early predict a high number of Cat 1 patients via an MPDS NOC.



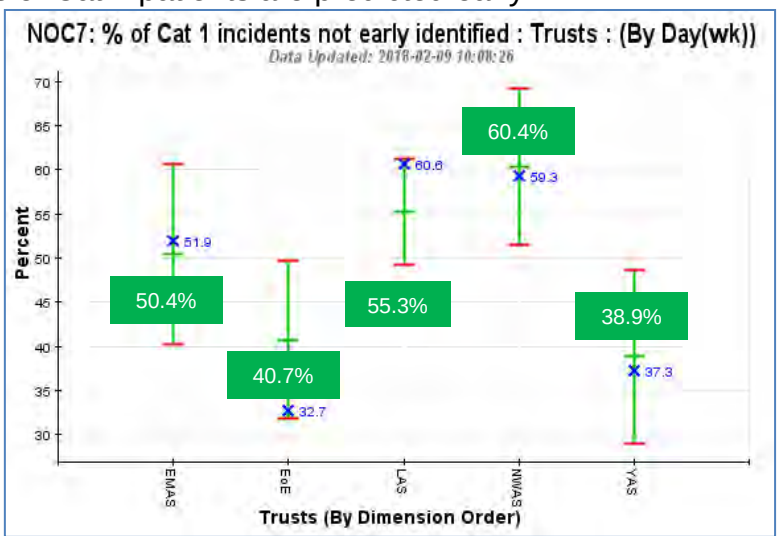
Cat 1 Patient early identification Key Words

For those Trusts who have CAD systems which facilitate the use of key words or phrases, the chart below shows that it is possible to early predict Cat 1 patients in this way. It is difficult to say if these patients could be captured in the NOC and individual call taker training and understanding will have a bearing on this. For example, we can see that overall EEAST predict a similar number of Cat 1 incidents as YAS without the use of keywords. It would however appear useful that where CAD systems have this functionality, Trust's use it as a further enhancement to their early predict processes.



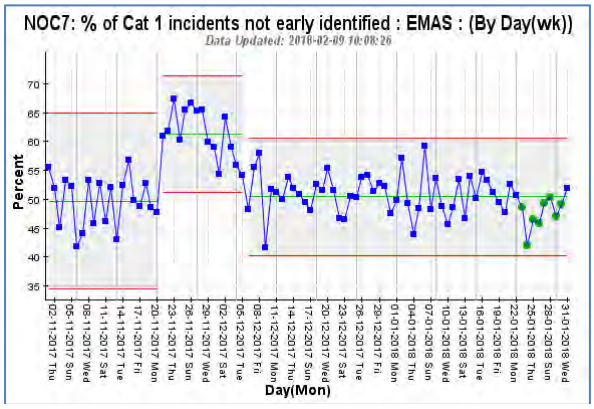
Cat 1 patients not early identified

The chart below shows that there is significant variation amongst Trusts in terms of the percentage of Category 1 patients not early identified. As EEAST and YAS continue to refine their early predict processes, we can see that they have the lowest level of "non-Cat 1 predict". This further re-enforces that where a NOC is in use, higher numbers of Cat 1 patients are predicted early.

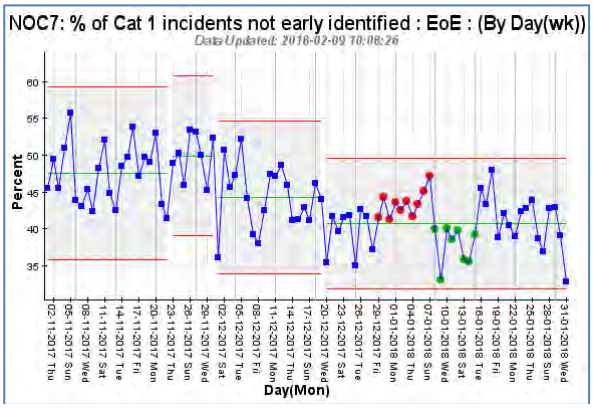


The five charts below set out the trended **spc** view for each Trust in terms of the percentage of Cat 1 patients not early predicted. As can be seen there is a downward trend in all Trusts as they work through improvement initiatives which is encouraging. Further sharing of best practice for early predict arrangements will improve this further and reduce variation.

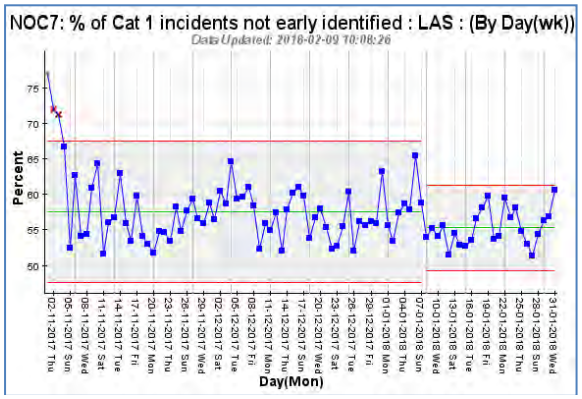
EMAS



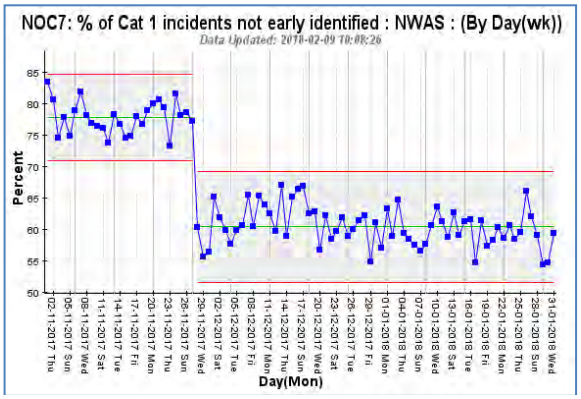
EEAST



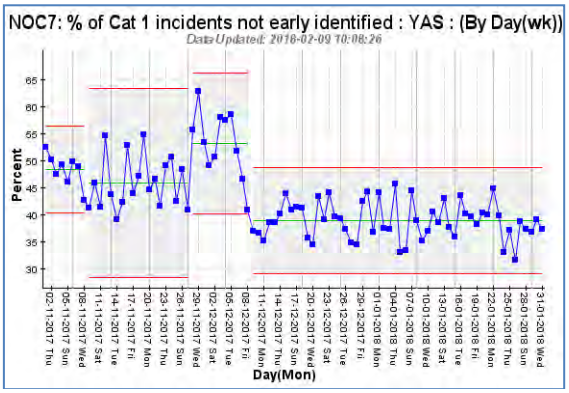
LAS



NWAS



YAS

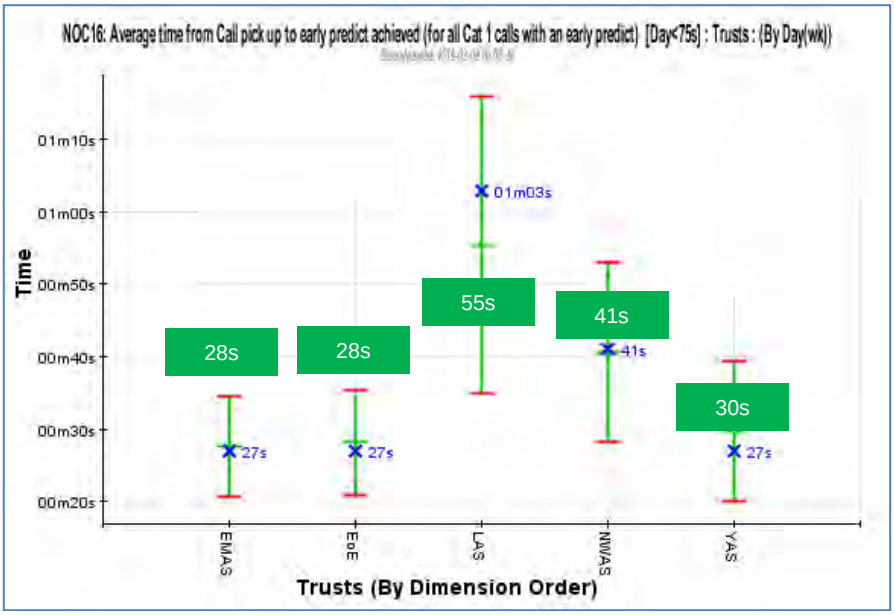


Time to make early prediction

As described in section 3 above, the group agreed that the most appropriate “time based” measures were those measured from call pick up not call connect. Principally this was to factor out any variation in call pick up times between Trusts.

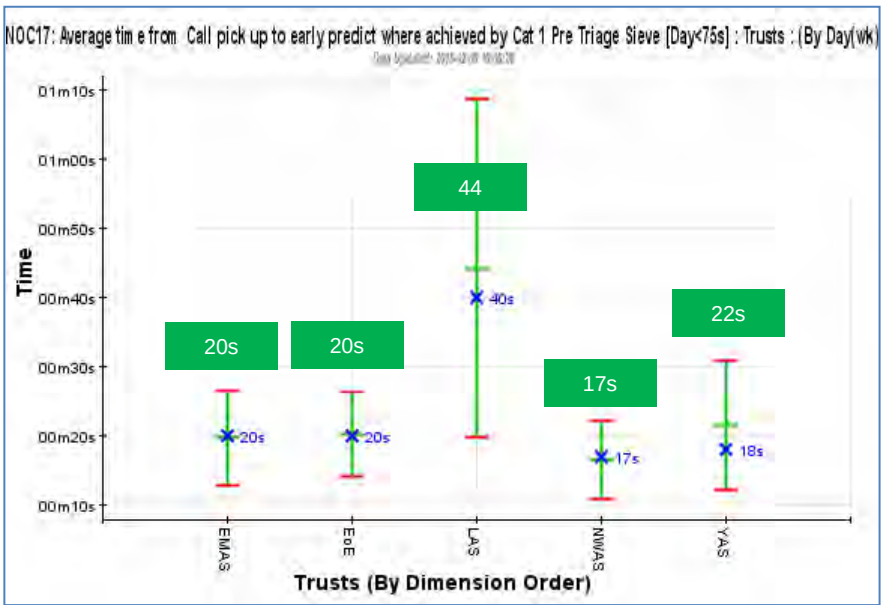
Call pick up to early predict Cat 1 overall

The chart below sets out the average time from call pick up to early predict for those Cat 1 calls with an early predict. As can be seen LAS and NWAS take longer to early identify Cat 1 patients than the other the Trusts. This is acknowledged by LAS and NWAS who are working on initiatives supported by AACE to reduce their time to predict Cat 1 patients. This shows that it is possible to elicit an early predict within an average of 30s for Cat 1 patients.



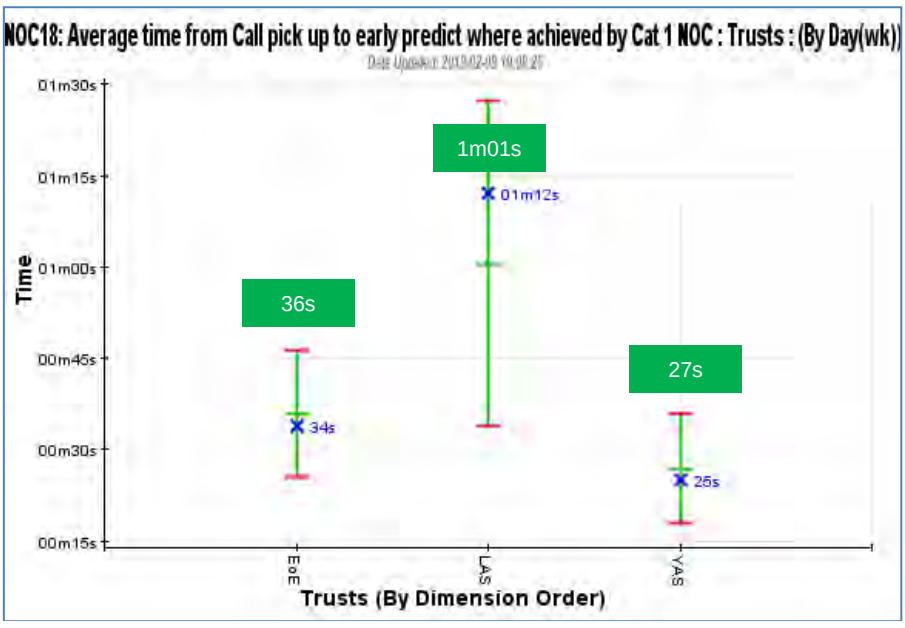
Call pick up to early predict Cat 1 identified in PTS

As would be expected, the pre-triage sieve is fastest from call pick up to early predict, principally because it occurs first in the process and has three very structured questions. As previously described, LAS do not currently have the PTS embedded within their CAD which most likely accounts for the higher time and large variation.



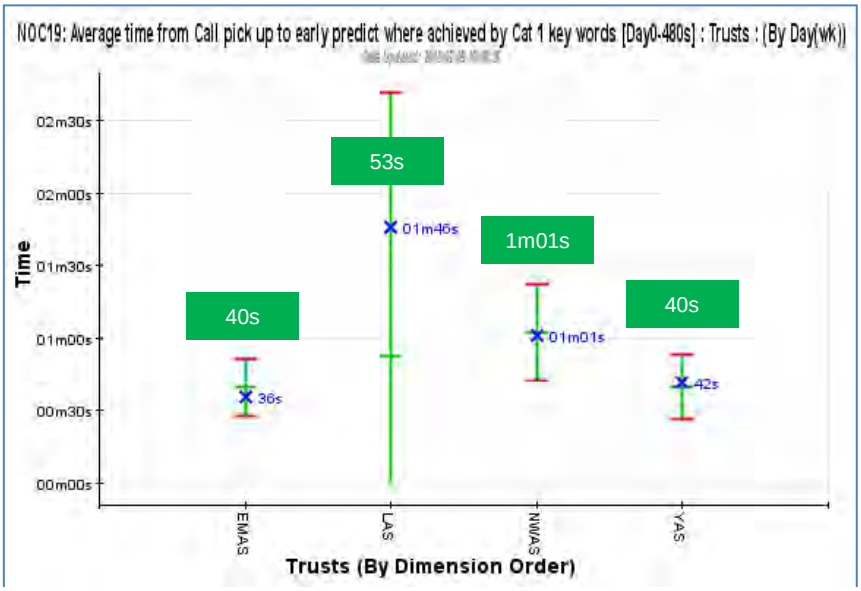
Call pick up to early predict Cat 1 identified in NOC

Of the three Trusts operating the NOC, the times from call pick up to early predict are shown below. As expected these times are longer than for the PTS given this happens later in the process. In EEAST and YAS, we can see that the time from call pick up to early identify in the NOC is still very quick. LAS are working on the required changes to reduce their time to NOC supported by AACE.



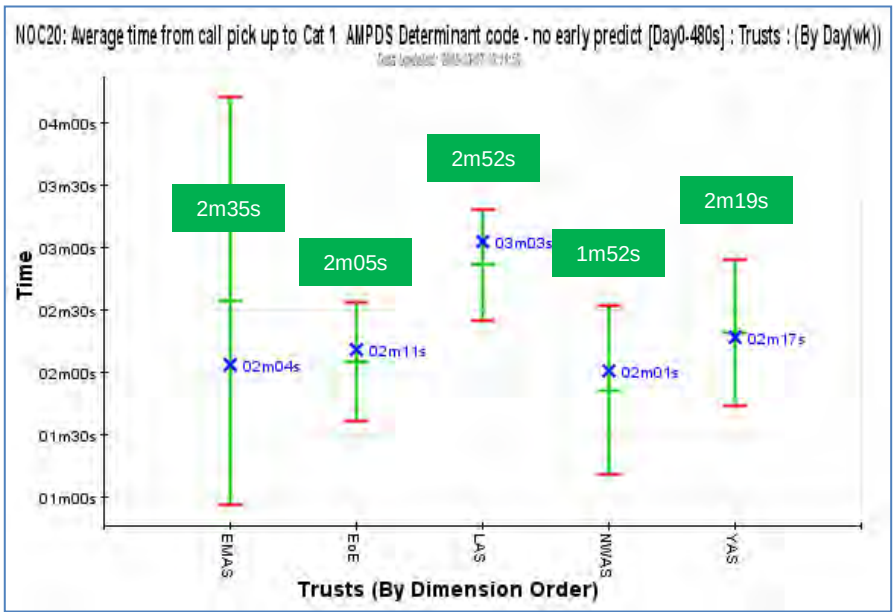
Call pick up to early predict Cat 1 identified in Key words

The average time to elicit an early predict with key words is set out below. As can be seen there is more variation here. LAS do not really use key words, rather this data is derived from a former CAD process and is extremely low numbers. In NWAS the additional time to key word predict is associated with the current order of the call taking process which AACE are working with NWAS to improve.



Call pick up to early predict Cat 1 where **no** early predict

The chart below shows the time from call pick up to Cat 1 MPDS code where there has been **no** early predict. As can be seen this is far longer than the time from call pick up to early predict set out above.



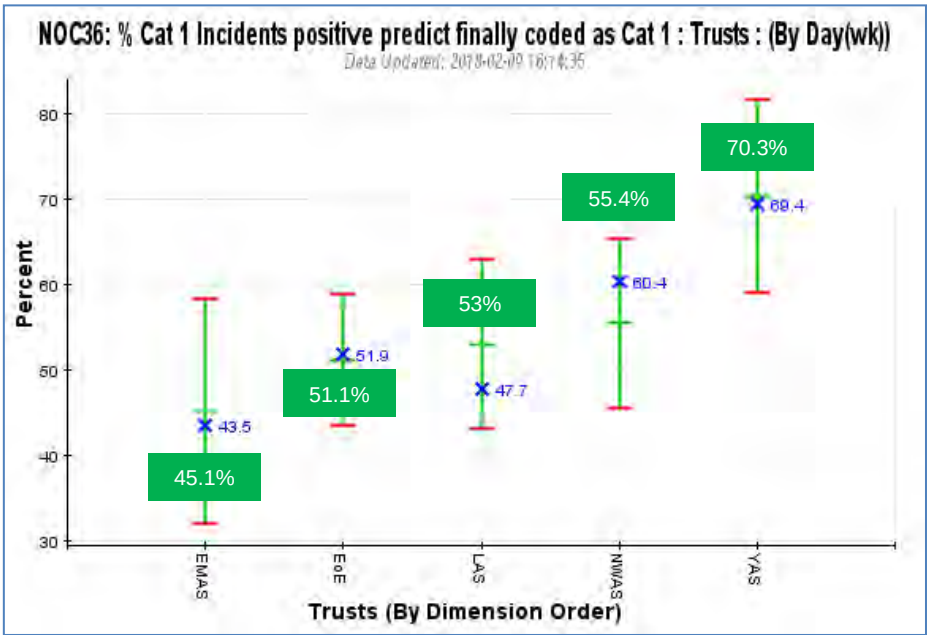
In summary as can be seen, Category 1 patients can be identified far quicker in the call cycle through PTS, NOC and keywords than awaiting the final MPDS determinant.

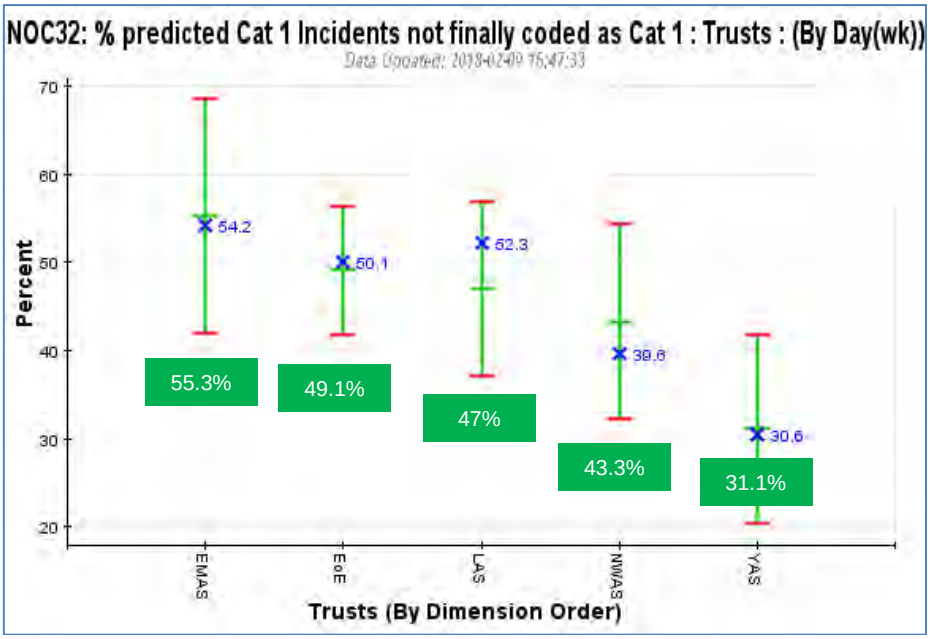
Predict accuracy

It is important to measure the accuracy of early predictions to Cat 1 patients to ensure that Trusts do not have a high degree of “false positives”. This is where a potential Cat 1 prediction is made and the call subsequently codes to a lower Category. High false positives can lead to an increase in response per incident in other categories.

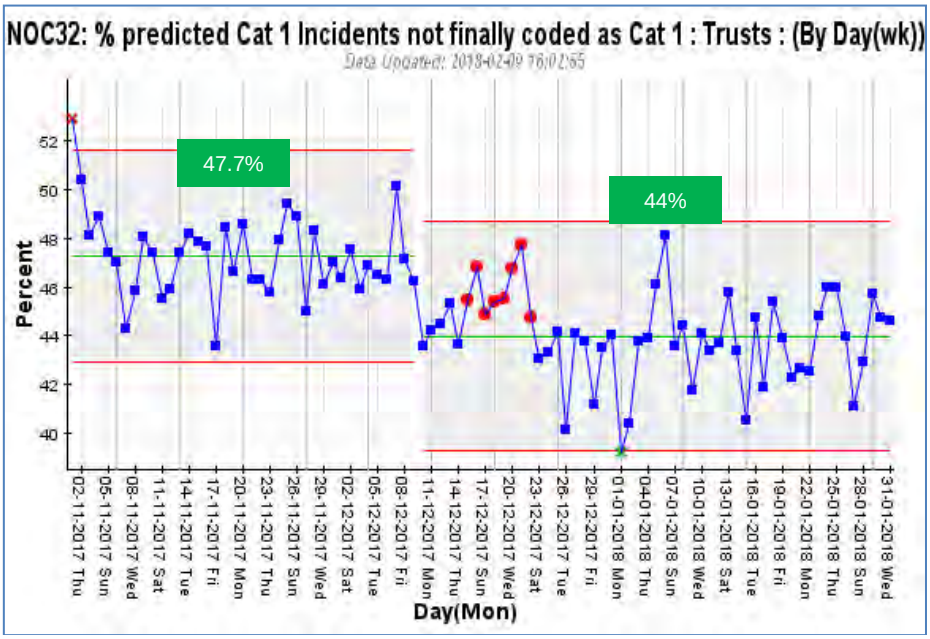
Percentage Cat 1 predicts, coding Cat 1 at MPDS determinant

There is large variation of Cat 1 predict accuracy between the five Trusts. As shown below the accuracy runs between 45.1% of Cat 1 predicts coding Cat 1 in EMAS and 70.3 in YAS. This provides further evidence that the YAS process of using the current mandated pre-triage sieve, NOC and keywords is the most effective at identifying Cat 1 patients early and with the most accuracy.

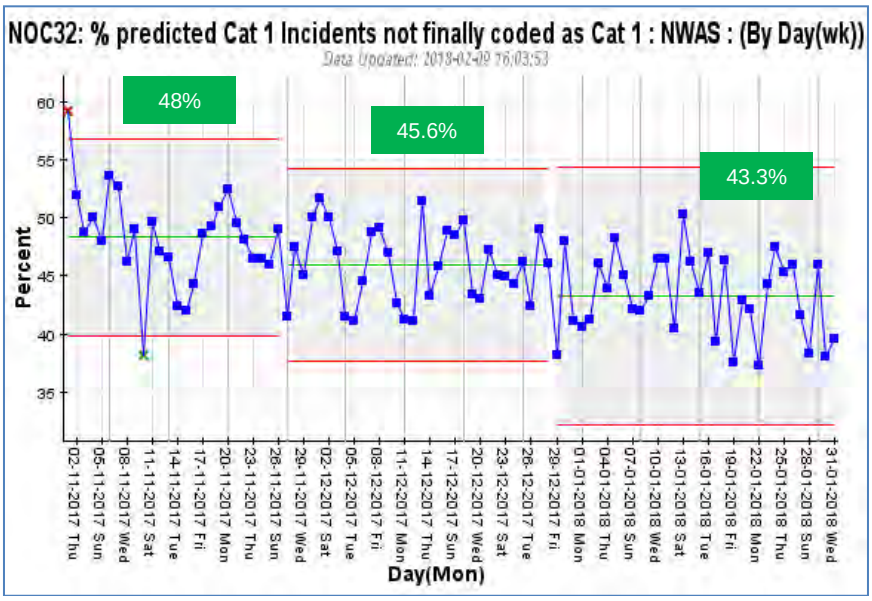




During the data period the percentage of Cat 1 predicts **not** finally coding as Cat 1 has reduced from an average (all 5 Trusts) of 47.7% to an average of 44%.

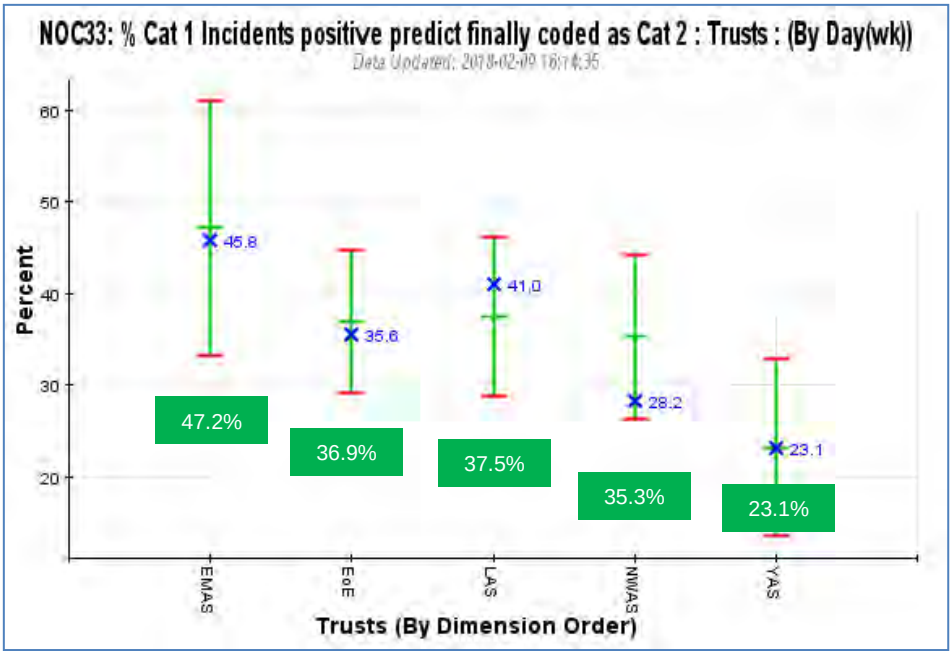


Of note is that as NWAS implement the various initiatives being worked through for Cat 1, they are seeing a step change in the volume of over predicts which have reduced from 48% to 43.3%.

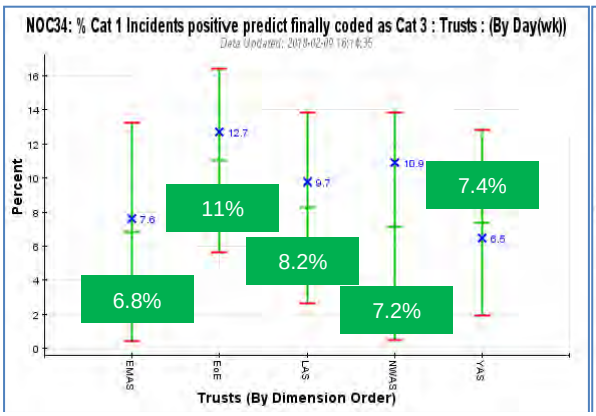


Where a Cat 1 predict does not subsequently code Cat 1 at MPDS determinant, the majority code Cat 2 as can be seen below. EMAS have the highest percentage of predicts that subsequently code Category 2 which is worthy of further investigation. It may be that the EMAS decision to early predict all patients declared unconscious in the pre-triage sieve is contributing to them having the highest conversion to Cat 2 from their early predictions.

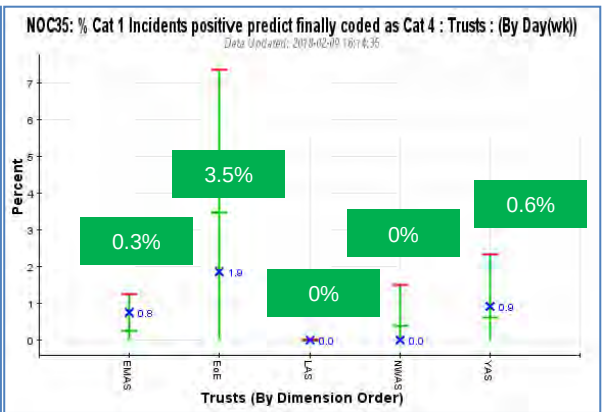
Cat 2



Cat 3



Cat 4



6. Conclusions and Recommendations

The table below outlines the conclusions made following the programme of work detailed within this report and the associated recommendations to NHS England. The NOC Task and Finish group members have jointly reviewed the recommendations and commend these to NHS England.

No:	Detail
1	NOC should be mandated by NHS England as per the recommended NOC types at appendix 2
2	Confirmation of the mandated PTS should be communicated by NHS England alongside the mandated NOC
3	Where CAD technology allows, the use of key words and phrases at the “problem” stage enhances the potential for early prediction of Cat 1 patients not picked up in PTS or NOC
4	Trusts should work together to share data and learning and spread best practice for both early prediction and overall Cat 1 patient response process improvement
5	A group should be established reporting to NASMeD in order to continually review the speed, accuracy and capture rates of early predict processes with recommendations made for future amendment to mandated PTS/NOC where appropriate. It is possible that this could utilise the same working groups and governance structure as outlined in the new code set review process.

Appendix 1) - Data measured during the trial period

NO	Data measure		
1	**Total Volume of Cat 1 Incidents**	nr	A8
1a	Volume of Cat 1 Incidents that are IFTs	nr	
	Volume of Cat 1 Incidents from 111		
2	Volume of Cat 1 incidents as % of overall emergency activity	%	A8/(A7-(A57-61))
3	% of Cat 1 incidents early identified	%	
4	% of Cat 1 incidents early identified from Pre Triage Sieve	%	Count incidents in the earliest measure that predicted Cat 1 only. *Include IFT Cat 1s*
5	% of Cat 1 incidents early identified from NOC	%	
6	% of Cat 1 incidents early identified from Key Words	%	
7	% of Cat 1 incidents not early identified	%	*Any upgrades by clinician to cat 1 count here*
	note 1) no. 3 = 4+5+6 note 2) 4+5+6+7 = 100%		
8	Cat 1 mean all Incidents	hh:mm:ss	A25
8a	Cat 1 90th centile all Incidents	hh:mm:ss	A26
9	Cat 1 mean predicted in Pre Triage Sieve	hh:mm:ss	Count incidents in the earliest measure that predicted Cat 1 only. *Include IFT Cat 1s*
9a	Cat 1 90th centile predicted in Pre Triage Sieve	hh:mm:ss	
10	Cat 1 mean predicted in NOC	hh:mm:ss	
10a	Cat 1 90th Centile predicted in NOC	hh:mm:ss	
11	Cat 1 mean predicted with key words	hh:mm:ss	
11a	Cat 1 90th centile predicted with key words	hh:mm:ss	
12	Cat 1 mean **no** early predict	hh:mm:ss	
12a	Cat 1 90th centile **no** early predict	hh:mm:ss	
13	Cat 1 mean Card 9	hh:mm:ss	
13a	Cat 1 90th centile Card 9	hh:mm:ss	
14	% of 999 calls answered within 5 seconds	%	
15	% of Cat 1 999 calls answered within 5 seconds	%	
16	Average time from Call pick up to early predict achieved (for all Cat 1 calls with an early predict)	hh:mm:ss	
17	Average time from Call pick up to early predict where achieved by Cat 1 Pre Triage Sieve	hh:mm:ss	Count incidents in the earliest measure that predicted Cat 1 only. *Include IFT Cat 1s*
18	Average time from Call pick up to early predict where achieved by Cat 1 NOC	hh:mm:ss	
19	Average time from Call pick up to early predict where achieved by Cat 1 key words	hh:mm:ss	
20	Average time from call pick up to Cat 1 AMPDS Determinant code - no early predict		

21	Average time from Call Pick up to Final AMPDS Determinant Code	hh:mm:ss	
22	Average T0 to first allocation average Cat 1	hh:mm:ss	
23	Average T0 to first allocation average Cat 1 predicted in Pre Triage Sieve	hh:mm:ss	Count incidents in the earliest measure that predicted Cat 1 only. *Include IFT Cat 1s*
24	Average T0 to first allocation average Cat 1 predicted in NOC	hh:mm:ss	
25	Average T0 to first allocation average Cat 1 predicted with key words	hh:mm:ss	
26	Average T0 to first allocation average Cat 1 **no** early predict	hh:mm:ss	
27	Average Call Pick Up to first allocation average Cat 1	hh:mm:ss	
28	Average Call Pick Up to first allocation average Cat 1 predicted in Pre Triage Sieve	hh:mm:ss	Count incidents in the earliest measure that predicted Cat 1 only. *Include IFT Cat 1s*
29	Average Call Pick Up to first allocation average Cat 1 predicted in NOC	hh:mm:ss	
30	Average Call Pick Up to first allocation average Cat 1 predicted with key words	hh:mm:ss	
31	Average Call Pick Up to first allocation average Cat 1 no early predict	hh:mm:ss	
32	% predicted Cat 1 Incidents not finally coded as Cat 1	%	
33	% Cat 1 Incidents positive predict finally coded as Cat 2	%	
34	% Cat 1 Incidents positive predict finally coded as Cat 3	%	
35	% Cat 1 Incidents positive predict finally coded as Cat 4	%	
36	% Cat 1 Incidents positive predict finally coded as Cat 1	%	
	note 3) no. 31 = 32+33+34 note 4) 32+33+34 +35= 100%		
37	Cat 1 Incidents - clock start trigger - T0 + 30 seconds	%	
38	Cat 1 Incidents - clock start trigger -First Assign	%	
39	Cat 1 Incidents - clock start trigger -T5	%	
	note 5) 36+37+38 = 100%		
40	% of Cat 1 incidents by chief complaint card	see by card tab	
41	% of Cat 1 incidents by chief complaint card not early predicted	see by card tab	

Appendix 2) - Recommended MPDS & NHSP mandatory Cat 1 NOC Types

The recommended NOC types below are based on the current MPDS V13 code set, as well as learning of the three Trusts using a NOC for MPDS as detailed within this report.

AACE has been involved in supporting two of these Trusts directly (YAS and LAS) in implementing the NOC process both in respect of Red 1 and Category 1. As part of that support, a workshop was undertaken looking at appropriate NOC types for Category 1 in MPDS, kindly supported by the WMAS Head of EOC Training and Development. This work took account both learning within WMAS in respect of NOC implementation within NHS Pathways, as well as a review of the Category 1 MPDS codes and what would be potentially required in an MPDS NOC.

NOC TYPE	Description / Prompts for choosing NOC types
Choking	- Any mention of choking or an airway obstruction
Drowning	- Any reference to a patient who is drowning or in the water
Fitting now	- Any patient who is fitting could be described as shaking, jerking, twitching etc..
Hanging	- Hanging if breathing status was unknown at pre-triage questions
Ineffective Breathing	- Barely breathing - Can't breathe - Fighting / gasping for air - Breathing just a little - Making funny noises - Turning blue or purple - Hardly / stopped breathing In the case of 1st party caller, select INBR if any of these can be heard when speaking with patient
Maternity Complications	Any maternity call where - Head out / visible - Breach presentation (i.e. hands feed or buttocks) - Cord presenting - Multiple births (e.g. twins triplets etc.) - 3rd trimester pregnancy with PV bleed
Unconscious Trauma	- Not conscious following fall / jump from extreme height - Not conscious following Assault - Gunshot or stabbing
Unconscious Medical	- Any overdose where patient not conscious - Diabetic unconscious - Not conscious after bleed e.g. PR / PV
Serious Haemorrhage	- For patients with serious, potentially life-threatening bleeding due to Trauma i.e. bleeding that fits the serious definition and is uncontrolled at the time of answering the NOC question i.e. where blood is spirting or pouring out from a wound
Serious RTC	Any RTC where - Patient / patients not conscious - ? Fatal - Abnormal or noisy breathing not picked up in pre-triage sieve

Severe Allergic Reaction	- History of severe allergic reaction - Difficulty breathing or swallowing - Medication administered e.g. EpiPen

At the request of NHS England, AACE also undertook a review of the Category 1 NoC codes in use by ambulance trusts using NHS Pathways, in order to review and revise the mandated NoC. A sub group including representation from Pathways Trusts and MPDS subject matter experts has agreed a minimum NoC list for trusts using NHS Pathways.

This NoC list was arrived at through a collective review of the NoC codes in use by trusts and a comparison of the proportion of calls that were early predicted as Category 1 that subsequently were finally coded as Category 1. The recommended minimum NoC list is as follows.

NHSP Cat 1 NOC Types
Arrest / Perri Arrest (Cat 1)
Choking
Fitting Now
Under 5 Severe Haemorrhage
Under 16 unconscious

NHS England Response to Nature of Call Report

NHS England, working with the ARP Development Group, has reviewed the findings, conclusions and recommendations of this report. NHS England recommends that all English ambulance services use a combination of pre-triage sieve questions (PTS) (set out below), a Nature of Call (NoC) and, where CAD systems allow, a keyword look up search to early predict the highest proportion of potential Category 1 emergencies.

In response NHS England will:

1. Confirm the three mandated PTS questions as:
    - Is the patient breathing?
    - Is the patient awake (conscious)?
    - Is their breathing noisy?

2. Mandate the use of a NoC in all English ambulance Trusts. NHS England will mandate the use of the Category 1 NoC types contained within the report recommendations at Appendix 2. The report outlines NoC codes both for AMPDS Trusts and NHS Pathways Trusts. These will be the minimum NoC types for Category 1 predict that Trusts are expected to use.

3. Where CAD technology allows, mandate the use of key words and phrases at the "tell me exactly what's happened (problem)" stage to act as an additional opportunity to support high early predict rates.

4. Support all Trusts to work together to share data and learning and spread best practice for both early prediction and overall Cat 1 patient response process improvement.

5. Promote the establishment of a sub- group reporting to NHS England through NASMeD in order to provide a continuous review of the speed, accuracy and capture rates of the Category 1 early predict processes with recommendations made for future amendment to mandated PTS/NoC where appropriate.

National Framework for Healthcare Professional ambulance admissions

This framework is intended for patients who require admission to hospital from a community setting following clinical assessment by a health care professional (HCP).

For the purposes of this framework a healthcare professional will be defined as HCPs working in General Practice, advanced practitioners, Paramedics, community matrons, community and district nursing teams, community midwifery teams, dentists and Approved Mental Health professional (mental health admissions only).

Patients who have immediate life-threatening injuries or illnesses must receive the same level of response in the community, irrespective of the source of the 999 call. HCPs in the community may require immediate clinical assistance from the ambulance service in addition to transportation in a timeframe appropriate to the patient's needs.

NHS England has asked the two NHS Pathways Beta sites (WMAS and NEAS) and two AMPDS sites (NWAS and LAS) to undertake an initial pilot of the following framework prior to national implementation. The pilot will allow individual systems to develop robust standard operating procedures and effective decision-making algorithms. The results of the pilot will be presented to NASMeD in the autumn.

A set of healthcare professional response levels will be described with a clear definition of the patient groups that would be allocated to each level. Those levels will be mapped to the current ARP categories and Ambulance Trusts would be expected to respond to these requests under the same response levels as other 999 calls.

There will be 4 levels of healthcare professional response.

HCP Level 1 (HCP 1) Category 1 (7 Minute mean response time)

This level of response should be reserved for those exceptional circumstances when an HCP requires immediate, additional clinical assistance from the ambulance service to treat a patient in need of immediate, life-saving intervention such as resuscitation. These requests should be processed through the Trust's 999 triage tool and only those that are deemed Category 1 under that assessment should receive a Category 1 response. Examples would include cardiac arrest, anaphylaxis, life threatening asthma, obstetric emergency, airway compromise and cardiovascular collapse (including septic shock). It would be expected that predominantly the HCP would be with the patient, however in exceptional circumstances may not be (for example relatives call surgery for patient in Cardiac Arrest).

HCP Level 2 (HCP 2) Category 2 (18 Minute mean response time)

This level of response is based on the clinical condition of the patient and their need for immediate additional clinical care in hospital in an emergency department or acute receiving unit (i.e. medical or surgical assessment unit, delivery suite).

Patients with a National Early Warning Score (NEWS2) 7 or greater may trigger a request for this level of response.

Patients with a NEWS2 score of 6 or less may be suitable for HCP Level 2 response by exception only and HCPs must be able to detail the clinical reason. Examples may be patients with sepsis, myocardial infarction, CVA, acute abdomen, acute ischaemic limb,

acute pancreatitis, major gastrointestinal haemorrhage and overdose requiring immediate treatment.

HCP Level 1 and Level 2 incidents are confirmed emergencies which require life-saving intervention and should be responded to as time critical emergencies and immediately allocated the nearest appropriate resources.

Healthcare professionals requesting a Level 1 or 2 response MUST remain with the patient until arrival of the emergency ambulance (where they are at the scene) to hand over the patient to the attending ambulance clinician.

If an ambulance is not immediately available for despatch to an **HCP Level 1 or HCP Level 2** this incident should be escalated within ambulance operations centres to ensure an appropriate response. Ambulance services should have in place appropriate clinical support and decision making for HCP requests requiring escalation. Clinical discretion should be applied in some cases where the patient's condition does not precisely meet the definition but additional considerations are involved such as with end of life care.

In the case of an HCP Level 1 incident, ambulance services should consider whether an HCP requires a solo responder in addition to an emergency ambulance (i.e. where the closest resource is a solo responder or the closest emergency ambulance crew are not a Paramedic crew)

There should be little or no variation in the proportions of the above categories across England.

HCP Level 3 (HCP 3) 2 hour locally commissioned response

This level may be commissioned for patients who require urgent admission to hospital. Examples may be patients who require urgent investigations to inform ongoing care such as CT, MRI, ultrasound or who need an urgent assessment by a specialist. Mental Health emergency admissions and patients with respiratory conditions, or suspected fractures (not due to major trauma) are examples that may be suitable for a Level 3 response.

Where this is commissioned a response timeframe of within 2 hours arrival at the patient (90^{th} centile) should be applied. This includes the option for both 1 and 2 hour responses (where commissioned).

HCP Level 4 (HCP 4) 4 hour locally commissioned response

This is for all other patients who do not fit the above definitions and require admission to hospital by ambulance for ongoing care but do **not** need to be managed as an emergency. Examples may be patients being admitted directly under specialty teams as well as those being admitted to emergency departments for further investigation who do **not** require emergency investigation or treatment immediately upon arrival.

Where this is commissioned a set timeframe of a within 4-hour response arrival at the patient (90^{th} centile)

General notes

Where immediate ambulance clinical support and/or transportation is requested by an HCP it is the responsibility of the attending clinician to make the request to the ambulance service. It is inappropriate for this to be delegated unless there are exceptional circumstances. Where delegation is unavoidable the individual making the request for support MUST be in a position to answer basic triage questions about the patient's condition. Only in exceptional circumstances should HCP Level 1, 2 requests be made by non- clinical staff.

In situations where systems operate a “bed bureau” arrangement and organise ambulance transport on behalf of HCPs these organisations must have in place the correct procedures so that they can provide all the clinical information including NEWS. HCP Level 1 and 2 incidents are not suitable to be requested via a bed bureau arrangement.

Healthcare professionals have a responsibility to be familiar with this framework and to utilise emergency ambulances responsibly. Ambulance services have a responsibility to ensure appropriate clinical support in control rooms and on scene for healthcare professionals dealing with patients with emergency conditions.

Alternatives to conveyance by ambulance (eg. local taxi service) are a matter for local determination and do not form part of this framework.

NEWS is intended for use in patients who are more than 15 years old. Patients under 16 years of age who require admission to hospital from a community setting, following clinical assessment by a healthcare professional, should be assigned to HCP Level 2-4 on the basis of the clinical judgement of the requesting clinician. The process for HCP Category 1 requests is identical for all ages (adults and children).

Question order for HCP bookings

Call Answer
Where the caller identifies themselves as an HCP (in line with above definition), the three mandatory pre-triage sieve questions and Nature of Call will be applied.

- Where an early identification of a potential Cat 1 occurs through the PTS or NOC, enter AMPDS/NHS Pathways
- Where the call codes as a Cat 1 in triage allocate resources as appropriate
- Where the call codes as a Cat 2 in triage allocate resources as appropriate
- Should the call code as a Cat 3 or 4, **upgrade** to an HCP Level 2 and allocate resources as appropriate
- If **no** Cat 1 early identification enter questions below

Only those patients who are categorised by the triage system as requiring a Category 1 response should receive one.

Q1
Is this a request for immediate lifesaving clinical support?

If YES – Process through Trust 999 triage procedure (AMPDS, NHS Pathways)

- Where the call codes as a Cat 1 in triage allocate resources as appropriate
- Where the call codes as a Cat 2 in triage allocate resources as appropriate
- Should the call code as a Cat 3 or 4, **upgrade** to an HCP Level 2 and allocate resources as appropriate

Only those patients who are categorised by the triage system as requiring a Category 1 response should receive one.

If NO – Go to Q2

Q2
Is there a need for immediate emergency admission?

For example: patient with chest pain and a likely diagnosis of acute coronary syndrome, unstable arrhythmia, Stroke, acute pancreatitis, acute ischaemia of a limb, suspected pulmonary embolus with respiratory distress.

If YES – GO to Q3
If NO – Where **HCP Level 3 (HCP 3)** is commissioned go to Q5

If NO – Where **HCP Level 3 (HCP 3)** is **not** commissioned **HCP Level 4 (HCP 4)** response

Q3
What is the patients NEWS score?

If 7 or greater respond as **HCP Level 2 Category 2** response - closest appropriate emergency ambulance to be despatched immediately.

If NO or UNKNOWN – Go to Q4

National Early Warning Score (NEWS2)

ALL Patients requiring an emergency ambulance MUST have their NEWS calculated, recorded and handed over to ambulance clinicians and upon arrival in the receiving unit.

National Early Warning Score (NEWS2)

Physiological parameter	Score 3	2	1	0	1	2	3
Respiration rate (per minute)	≤8		9–11	12–20		21–24	≥25
SpO_2 Scale 1 (%)	≤91	92–93	94–95	≥96			
SpO_2 Scale 2 (%)	≤83	84–85	86–87	88–92 ≥93 on air	93–94 on oxygen	95–96 on oxygen	≥97 on oxygen
Air or oxygen?		Oxygen		Air			
Systolic blood pressure (mmHg)	≤90	91–100	101–110	111–219			≥220
Pulse (per minute)	≤40		41–50	51–90	91–110	111–130	≥131
Consciousness				Alert			CVPU
Temperature (°C)	≤35.0		35.1–36.0	36.1–38.0	38.1–39.0	≥39.1	

Royal College of Physicians

Q4
Is there clinical reason as to why an emergency ambulance must be despatched immediately?

If Yes request HCP specifies the clinical reason and code **HCP Level 2 Category 2 r**esponse - closest emergency ambulance to be despatched immediately.

If NO – Where **HCP Level 3 (HCP 3)** is commissioned go to Q5

If NO – Where **HCP Level 3 (HCP 3)** is **not** commissioned **HCP Level 4 (HCP 4)** response

Q5
Does the patient require additional urgent clinical assessment upon arrival at hospital? (This may include emergency diagnostics or urgent specialist assessment)

If YES – Respond in line with locally commissioned service of **HCP Level 3 (HCP 3)**

If No – Respond in line with locally commissioned service of **HCP Level 4 (HCP 4)**

Reporting

Ambulance services should be able to report HCP Level 1 – 4 requests separately from core 999 emergency activity. Please refer to the Ambulance System Indicators guidance at: https://www.england.nhs.uk/statistics/statistical-work-areas/ambulance-quality-indicators/

NHS Commissioners and Ambulance Services should jointly audit compliance with this framework.

ASSOCIATION OF
AMBULANCE
CHIEF EXECUTIVES

National Framework for Inter-Facility Transfers

Final Version
July 2017

ASSOCIATION OF AMBULANCE CHIEF EXECUTIVES

National Framework for Inter-Facility Transfers

This framework is intended for patients who require transfer by ambulance between facilities due to an increase in either their medical or nursing care need. The definition of a facility which this framework applies to are healthcare facilities that provide inpatient services. In some locally determined situations an additional "facility" will be defined by the ambulance service as suitable to use the IFT process i.e urgent care centres with direct admitting rights to inpatient services.

Patients who have immediate life-threatening injuries or illnesses should be transferred, where necessary with an appropriate hospital escort, and within a set timeframe mapped to ARP categories defined below. Similarly, patients with serious or urgent healthcare needs should be transferred in an appropriately commissioned timeframe. Local systems should have commissioned arrangements in place for the return of personnel and equipment to facilities.

The following framework should be used so that individual systems can develop standard operating procedures and decision algorithms.

A set of inter-facility transfer levels will be described with a clear definition of the patient groups that would be allocated to each level. Those levels will be mapped to the current ARP categories and Ambulance Trusts would be expected to respond to these requests in the same way and in the same time frame as other 999 calls in that category.

There will be 4 levels of inter-facility response.

IFT Level 1 (IFT1) Category 1

This level of response should be reserved for those exceptional circumstance when a facility is unable to provide immediate life-saving clinical intervention such as resuscitation and requires the clinical assistance of the ambulance service in addition to a transporting resource. These requests should be processed through the Trusts 999 Triage tool and only those that are deemed category 1 under that assessment should receive a category 1 response. Examples would include Cardiac arrest, anaphylaxis, birth units requiring immediate assistance, or acute severe life-threatening asthma in an urgent care facility.

IFT Level 2 (IFT2) Category 2

This level of response is based on the need for further intervention and management rather than the patient's diagnosis. Immediately Life, Limb or Sight (Globe trauma) Threatening (ILT) situations which require immediate management in another facility should receive this level of response. For instance, patient going directly to theatre for immediate neurosurgery, immediate Primary Percutaneous Coronary Intervention, Stroke Thrombolysis, immediate limb or sight saving surgery or mental health patient being actively restrained.

These IFT level 2 patients would be mapped to category 2 response under ARP. A set of interventions fitting the definition of life, limb or sight saving should be strictly adhered to. The examples above are an indicative but not exhaustive list.

IFT Level 1 and Level 2 incidents are confirmed emergencies which require life-saving intervention and should be responded to as time critical emergencies and immediately allocated the nearest emergency ambulance.

There should be little or no variation in the proportions of the above categories across England.

IFT Level 3 (IFT3) Locally determined response

This level may be commissioned for patients who are not undergoing immediate life or limb saving interventions but require an increase in their level of clinical care as an emergency. Where this is commissioned a set timeframe for the level of response should be specified between 30 minutes and two hours.

This level of response may include mental health crisis transfers or those solely for the purpose of creating a critical care bed.

IFT Level 4 (IFT4) Locally determined response

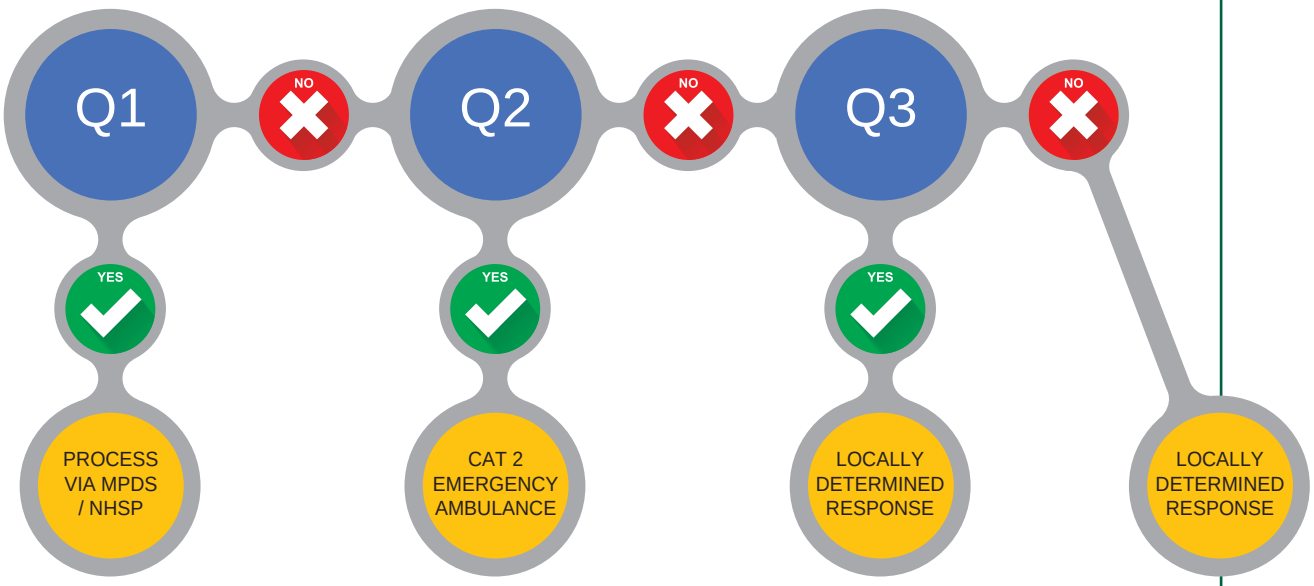
This is for all other patients who do not fit the above definitions and require urgent transport for ongoing care but do not need to be managed as an emergency transfer. Patients being transferred to inpatient wards for ongoing management or for elective and semi elective procedures or investigations. This category of patient will have a timeframe outside of the ARP standards and will be determined through their normal commissioning arrangements.

Patients who do not fit the definitions above are not appropriate for a Category 1, 2 or 3 response from the ambulance service. In some cases patients with immediately life or limb threatening conditions may not be ready for transfer within the Category 1 or 2 timeframe and require further management before being clinically suitable for transfer. In those cases a lower category will be allocated to reflect the time delay until the patient is ready for transfer.

Repatriations or step down transfers/discharges to non hospital facilities are not intended to be included in the IFT framework.

FLOW CHART

A Nature of call option "IFT" should be in operation in all ambulance Trusts to operate this framework.



The clinician in charge of patient determines that transfer to another facility is clinically necessary. It is the responsibility of requesting clinicians to ensure that this protocol is applied correctly. Inappropriately requesting IFT Level 1 or 2 transfers for patients that do not require immediate life or limb saving clinical intervention is a serious breach of protocol which should be reviewed by Ambulance Trusts in conjunction with the requesting Trust.

Where immediate ambulance clinical support and/or transportation is requested by a facility it is the responsibility of the attending clinician to make the request to the ambulance service. It is inappropriate for this to be delegated unless there are exceptional circumstances. Where delegation is unavoidable the clinician making the request for support must be in a position to answer basic triage questions about the patient's condition. IFT requests must not be made by non-clinical staff.

Q1 Is this a request for immediate life saving clinical support?

If YES – Process through Trust 999 Triage procedure (AMPDS, NHS Pathways)

Only those patients who are categorised by the triage system as requiring a category 1 response should receive one.

If NO – Go to Q2

Q2 Is there a need for an immediate intervention that cannot be carried out at the current facility and the patient is at immediate risk of death or life changing loss of a limb?

For example: transfer directly to theatre for immediate neurosurgery or PPCI, thrombectomy for a critically ischaemic limb.

If YES = Category 2 Response - closest emergency ambulance to be despatched immediately

If NO – Go to Q3

Q3 Does the patient require additional clinical management upon arrival at the new facility? This may include emergency diagnostics or emergency surgery Which is taking place within a set time period (to be defined by local commissioning) of less than 24 hours i.e 4 hours.

For example: Patient with sepsis going to a high dependency unit. Patient being transferred to a Burns centre. Referral for in-patient services not provided at current facility eg. Cardiology, Surgical speciality, ENT emergencies with no evidence of hypovolaemia, Step up in care to Specialist unit - Coronary Care, high dependency units or specialist nursing care. Patient in mental health crisis sectioned under mental health Act and requiring admission to specialist mental health facility.

If YES = Response in line with locally commissioned service of IFT Level 3

If NO = IFT Level 4

This category may also be used for issues with bed availability of critical or specialist care capability.

Patients being transferred to a normal ward environment do not fall into this category.

TIME CRITICAL TRANSFERS

Patients who require immediate life saving intervention as a result of assessment/diagnosis within a healthcare facility should have an ambulance immediately despatched to that facility. If an ambulance is not immediately available for despatch this incident should be escalated within ambulance operations centres to ensure an appropriate response.

Ambulance services should have in place appropriate clinical support and decision making for transfers requiring escalation or where there are additional factors that need to be considered in order for the patient to be matched to the correct clinical priority definitions.

Clinical discretion should be applied in some cases where the patient's condition does not precisely meet the definition but additional considerations are involved. Patients at end of life may well be included in this group.

REPORTING

Inter-facility transfers should be reported as part of 999 emergency activity. Additionally there will be a requirement for visibility of the 90th centile response time for each level of IFT. For further guidance on reporting of incidents under this framework see NHS England Ambulance Quality Indicators Technical guidance.

ASSOCIATION OF AMBULANCE CHIEF EXECUTIVES

AACE contact details

For more information please contact:

The Association of Ambulance Chief Executives
3rd Floor
32 Southwark Bridge Road
London
SE1 9EU

T: 020 7783 2043

E: info@aace.org.uk

W: www.aace.org.uk

Ambulance Emergency Call Categorisation and Prioritisation Governance Process

Introduction

NHS Ambulance Services are required to process 999 calls through an approved triage system. There are currently 2 systems approved in England for primary 999 assessments; NHS Pathways and Advanced Medical Priority Dispatch System (AMPDS). The outcome (disposition) reached at the conclusion of the initial assessment must be mapped to approved, contracted standards. There is a requirement to map these outcomes to the various categories set out within the NHS Constitution and Ambulance Service 999 contracts.

The production, maintenance, review and revision of this dataset is the responsibility of NHS England as the owner of the dataset. However the ambulance sector within England has a vital role in providing information, evidence and expert advice to NHS England regarding the dataset and the prioritisation of emergency calls. This requires a governance framework to ensure appropriate prioritisation, equity of access and uniformity of response across the English Ambulance Services. Where possible active participation in this process should reduce unwarranted variation. This governance process will ensure that the ambulance sector view of potential prioritisation changes is robust and with a sound evidence base.

The functionality of approved triage tools will vary but are approved on the basis of being able to determine as far as possible differing levels of acuity from immediately life threatening emergencies to patients with a urgent care need.

Triage Systems

There are currently two triage systems that are approved for use in English Ambulance Services:

- Advanced Medical Priority Dispatch System (AMPDS)

AMPDS determinant descriptors are the method by which the International Academy Emergency Medical Dispatch distinguishes the priority that should be attached to an incident. This priority is predominantly separated into time to respond and type of response required. Additionally, there are pre-arrival instructions that can assist until arrival of the ambulance service such as Basic Life Support, management of choking and haemorrhage.

MPDS determinant descriptors are assigned a code (e.g. 06E01 – Breathing Problems Ineffective Breathing). The combination of determinant descriptor and its associated code has a response priority allocated (currently Category 1 -4 under the Ambulance Response Programme) with an associated recommended timeframe for arrival of emergency ambulance assistance. This is known as the AMPDS dataset, which is used by AMDPS user Ambulance Trusts.

- NHS Pathways

NHS Pathways is a programme providing the Clinical Decision Support System (CDSS) used in NHS 111 and a number of ambulance 999 services. The system is hierarchical in nature, meaning that life-threatening problems assessed at the start of the call trigger

ambulance responses, progressing through to less urgent conditions which require a less urgent response (or disposition) in other settings.

The clinical content of NHS Pathways is under constant review and overseen by the National Clinical Governance Group (NCGG). The NCGG is an independent intercollegiate group chaired by the Royal College of General Practitioners (RCGP), made up of representatives from professional and lay stakeholders and experts in the delivery of urgent and emergency care. The group considers all aspects of the triage process, including the impact on services, as well as the evidence base for changes to the clinical content. Alongside this independent oversight of the NCGG, NHS Pathways ensures its clinical content and assessment algorithms are consistent with the latest advice from respected bodies that provide evidence and guidance for medical practice in the UK. In particular NHS Pathways is consistent with the latest guidelines from NICE (National Institute for Health and Clinical Excellence), UK Resuscitation Council and The UK Sepsis Trust.

Early identification of Category 1 emergencies

In addition to the triage systems there is a mandated process for all 999 calls within England which requires ambulances services to identify potential life threatening emergencies as early in the 999 call process possible. A set of pre-triage questions, a nature of call pre-determined list and key word search from the initial reason for the call are utilised by ambulance services to quickly identify potential Category 1 (life threatening) emergencies.

The national call categorisation dataset

Previously determinant codes (AMPDS) and dispositions (NHS Pathways) have been allocated a response priority/category based on a variety of levels of evidence and in the absence of clear evidence, expert opinion. This dataset has been endorsed by the national ambulance medical directors group (NASMED) and approved by the Emergency Call Prioritisation Advisory Group (ECPAG).

Review of this process is necessary in order to develop a more robust evidence base for the allocation of determinant codes and dispositions to response priorities. Additionally, a more detailed review and recommendation process with associated expert input where required should be utilised to inform discussion and endorsement of the dataset by ambulance service medical directors. The mechanism for this process will be the establishment of two sub groups; a dataset sub group and an early identification subgroup. The resultant draft code sets can then be recommended to ECPAG for approval and ratification.

It is an absolute requirement that ambulance services and other key stakeholders commit to sharing information, experience and evidence on all matters pertaining to clinical prioritisation so that a robust evidence base is developed and maintained. A feedback loop regarding incidents, concerns and legal proceedings within ambulance services to this national process must be established to ensure patient safety, equity and informed decision making. In particular where Trusts are considering responding to a particular code or disposition at variance from the national dataset they have an obligation to share the evidence and their decision making with the appropriate sub group and NASMED so that this change can be considered nationally.

A number of stages will be developed under the new process (outlined in the flow chart in Appendix 2). Any change in one system's prioritisation category should routinely be considered in the context of any other systems and any changes from them.

Expert User sub groups

Two expert user sub-groups of NASMED will be established to advise on changes to 'the dataset' and 'the early identification' process.

- Dataset sub group

Ambulance Trusts will be required to contribute personnel who have considerable clinical operational knowledge/experience of utilising their triage tool. Knowledge of AMPDS/NHS Pathways, control room procedures and ambulance operations will be required (maximum of 1 plus a nominated deputy per Trust). It is important that QGARD, NDOG, Lead Paramedics, and Heads of Control are adequately represented within this group, therefore the selection of the Trust representatives MUST ensure that national groups are represented within this pool.

Key external stakeholder groups such as College of Paramedics and the approved triage system agents must be appropriately represented. The role of this group will be:

- To propose a defined evidence gathering process to NASMED for their approval.
- To review evidence at regular intervals (6-12 monthly initially).
- To recommend any alterations/amendments to the current code set.

A member of NASMED will chair the group.

Additional experts/representatives would be invited to specific meetings if necessary to consider specific issues.

- Early identification sub group

A group of control room and clinical experts will review on an ad hoc basis any recommendations for changes to the mandated process of early identification. The methodology utilised in the ARP Review process of PTS, and NoC will be used for any future amendments.

Daniel Gore on behalf of AACE will chair this group.

It is imperative that all ambulance services submit information to these national groups for any changes they wish to make to either the dataset or the early identification process so that they can be considered on a national basis.

Evidence gathering

AMPDS user ambulance Trusts will be required to collect outcome data as set out by NASMED as part of six-monthly clinical audit cycle. Information that will be useful to the dataset sub group will include, but not be limited to, adverse incidents, complaints, concerns, coroners' cases, preventing future deaths reports, clinical interventions on scene, objective assessments of patient outcome, mortality and morbidity reviews and linked outcome data to hospital datasets (utilising NHS number).

NHS Pathways will continue to submit information through their national clinical governance group however any changes to the prioritisation of emergency calls must be submitted through the dataset sub-group for consideration by NASMED and subsequent recommendation (if appropriate), to ECPAG.

Routine change process

NASMED will receive recommendations from the sub groups a minimum of 1 calendar month before approval is required. Medical Directors will be responsible for raising issues from within their service or on behalf of stakeholder groups in a timely manner so that the sub group is able to review their recommendation as necessary. Evidence to support any changes will then be presented to NASMED for endorsement. NASMED will forward recommendations to ECPAG for routine change. It is important that NASMED include in their recommendation details of any pertinent discussion which ECPAG may need to consider, this will include but not be limited to whether a decision to recommend change was unanimous or a majority vote.

The dataset will be owned and published by NHS England on behalf of ECPAG. Terms of Reference for the sub groups will be approved by ECPAG.

The ECPAG Terms of Reference will detail approval of this governance process and the urgent change process.

Urgent change process

In exceptional circumstances, such as an HM Coroners' Prevention of Future Death report which may necessitate an urgent change to a determinant code or disposition, the response priority will be reviewed by the dataset sub group chair in discussion with NHS England (National Clinical Director for Urgent Care and Chair ECPAG) and Chair NASMED.

Collectively they will decide if an urgent ECPAG should be convened and will determine the immediate shortened process of evidence gathering, expert input, sub group recommendation and NASMED endorsement in order that a recommendation can be put to ECPAG.

Immediate changes which are necessitated to address immediate patient safety issues may be made before completion of this shortened process BUT must be endorsed by an emergency meeting of ECPAG at the earliest opportunity (within 10 working days). This decision would rest with the National Clinical Director for Urgent care, chair of NASMED and chair of the dataset sub group or their nominated deputies.

Where an ambulance Trust makes a decision to upgrade the priority of an incident they should as a matter of course escalate this change for consideration nationally. It is important evidence is made available so that consideration can be given to national change either using the urgent change process or through the routine audit cycle

Appendix 1

As part of the ARP Review process, a review of the current AMPDS dataset will be carried out in order to test the proposed process and also to enable additional evidence to be considered within a governance framework to inform any immediate changes to prioritisation.

This shortened process is necessary due to the high volume of Category 1 and 2 incidents within AMDPS Trusts in comparison to NHS Pathways Trusts

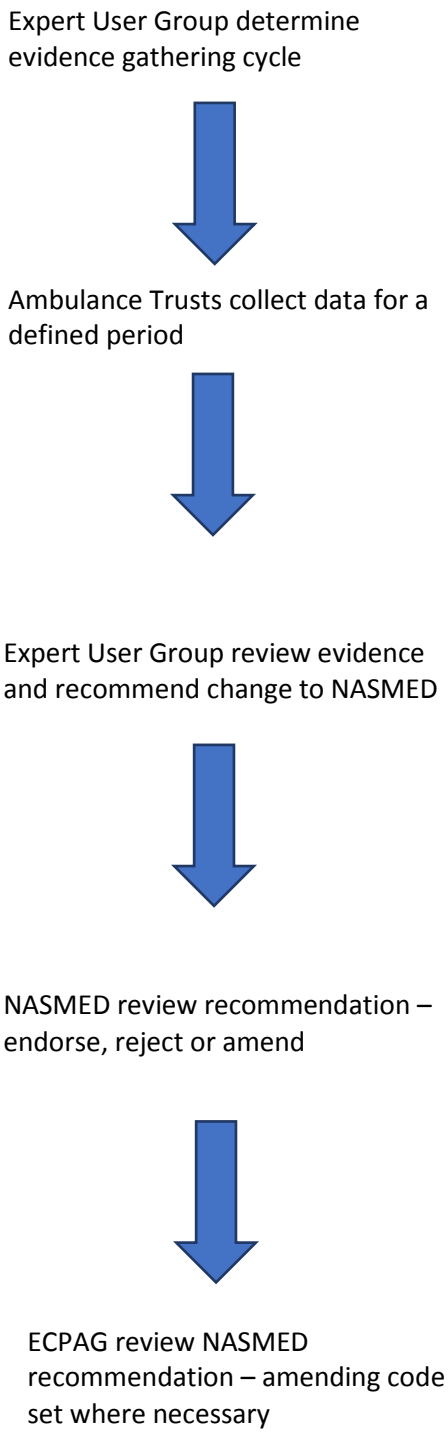
A small expert review group will be formed and will agree a limited number of determinant codes and the evidence that needs to be collected in order to consider any immediate changes.

Representatives from NASMED, NDOG, Lead Paramedic Group, NHS England and the AMDPS advisers will by the end of May 2018 have approved a list of codes that require consideration and the evidence provided by Trusts. Trusts will have 2 weeks to collate the required data evidence. The expert review group will convene (remotely if necessary) during the second week of June and will make a recommendation to the July meeting of NASMED for endorsement. An ECPAG meeting in September will consider these recommendations and, if required, produce a revised dataset for AMPDS Trusts.

Examples of evidence to be gathered:

- Cardiac arrest rate (survival to discharge)
- Non-conveyance rate
- STEMI and Stroke rate
- Complaint/Adverse/Serious incident information relevant to that code
- PFDs relevant to that code/pathway
- Paramedic pathfinder/ NEWS scoring
- Electronic patient record information of interventions
- Any objective information that Trusts may hold which would inform/ add to decision making

Appendix 2


Expert User Group determine evidence gathering cycle
Ambulance Trusts collect data for a defined period
Expert User Group review evidence and recommend change to NASMED
NASMED review recommendation – endorse, reject or amend
ECPAG review NASMED recommendation – amending code set where necessary

Emergency Call Prioritisation Advisory Group (ECPAG)
Terms of Reference

Introduction

NHS Ambulance Services are required to process 999 calls through an approved triage system. There are currently 2 systems approved in England for primary 999 assessments; NHS Pathways and Advanced Medical Priority Dispatch System (AMPDS). The outcome (disposition) reached at the conclusion of the initial assessment must be mapped to approved, contracted standards. There is a requirement to map these outcomes to the various categories set out within the NHS Constitution and Ambulance Service 999 contracts.

The production, maintenance, review and revision of this dataset is the responsibility of NHS England as the owner of the dataset. However the ambulance sector within England has a vital role in providing information, evidence and expert advice to NHS England regarding the dataset and the prioritisation of emergency calls. This requires a governance framework to ensure appropriate prioritisation, equity of access and uniformity of response across the English Ambulance Services. Active participation, where possible, in this process should reduce unwarranted variation. This governance process will ensure that the ambulance sector view of potential prioritisation changes is robust and with a sound evidence base.

The functionality of approved triage tools will vary but are approved on the basis of being able to determine as far as possible differing levels of acuity from immediately life threatening emergencies to patients with a urgent care need.

1.0 Purpose

The purpose of the Emergency Call Prioritisation Advisory Group (ECPAG) is to advise NHS England, NHS Improvement and Department of Health & Social Care (DHSC) on issues of ambulance call prioritisation. Its principal remit will be to recommend which codes from established ambulance triage systems should receive a Category 1-4 or other response based on robust clinical evidence. The group will also consider recommendations regarding changes to any early identification sets.

2.0 Objectives of the Group

2.1 To advise on issues of prioritisation of all ambulance 999 calls.

2.2 To escalate issues of call prioritisation to NHS England, NHS Improvement and the Department of Health and Social Care (DHSC) for further consideration as required, and particularly where they might have a significant clinical or operational impact.

2.3 To ensure any decisions made in relation to call categorisation are disseminated to ambulance services and system suppliers as required.

2.4 To provide a forum to inform and liaise with providers of ambulance triage systems as required/necessary.

2.5 To ensure national oversight of secondary clinical triage systems in use within ambulance control rooms.

3.0 Meetings

3.1 ECPAG is chaired by the National Clinical Director for Urgent Care, NHS England.

3.2 ECPAG will plan to meet twice a year, usually in March and September, with extraordinary meetings or teleconferences called as required.

3.3 Meetings will generally be held in London with administrative support provided by NHS England.

3.4 NHS England will arrange for circulation of the agenda and papers with direction from the Chair.

3.5 Draft minutes of meetings will be circulated within 14 days of the meeting following agreement of the Chair. The draft minutes of meetings will be formally adopted or amended at the next scheduled meeting of the Group.

3.6 In addition to ECPAG members, approved minutes will be circulated to:-

i) National Ambulance Service Medical Directors group
ii) National Directors of Operations group
ii) AACE Management Meeting (Chief Executives)
iii) NHS England Acute Care and Integrated Urgent Care teams
iv) NHS Improvement Urgent Care team
v) Department of Health & Social Care, Patient Flow and Access team

4.0 Membership

4.1 Membership consists of clinical experts and other subject matter experts from the English ambulance services and other relevant organisations as detailed below;

Organisation	Name	Job Role
NHS England	Prof Jonathan Benger (Chair)	NCD for Urgent Care
AACE	Dr Anthony Marsh	Chief Executive - WMAS
AACE	Martin Flaherty	Managing Director - AACE
Department of Health and Social Care	Adam McMordie	Patient Access and Flow Team (PAFT)
National Ambulance Service Medical Directors (NASMeD)	Dr Julian Mark	Medical Director - YAS
Ambulance Heads of Control Group	Jeremy Brown	General Manager – EOC WMAS
NHS England	Dr Gareth Stuttard	Interim National IUC Clinical Adviser

NHS Pathways	Darren Worwood	Deputy Clinical Director (Acting)
MPDS (Users)	Dr Fenella Wrigley Dr Dave Macklin	MPDS Advisors
College of Paramedics	TBC	TBC
National Ambulance Commissioners' Network (NACN)	Sue Sutton	Lead Commissioner SWASFT999
NHS England	Emma Hall	Deputy Director, Acute Care
NHS Improvement	Luke Edwards	Director of Sector Development

5.0 Responsibility and Accountability

5.1 Recommendations on call categorisations will be accepted during the course of the ECPAG meeting by the Chair, NHS England, NHS Improvement and DHSC attendees unless required to be referred to NHS England and/or DHSC for further scrutiny (see 5.2).

5.2 Recommendations will be referred to NHS England, NHS Improvement and/or DHSC for further scrutiny if there is:

- Impact on national policy
- Involve significant change/s to practice
- Involve significant potential cost or new burden to services
- Involve significant impact on performance against targets (for example, through major changes to the proportion of codes in Category 1)
- Impact upon other stakeholders interests (for example, major changes to heart or stroke-related codes)
- Likelihood of significant media attention
- Likelihood of significant political attention

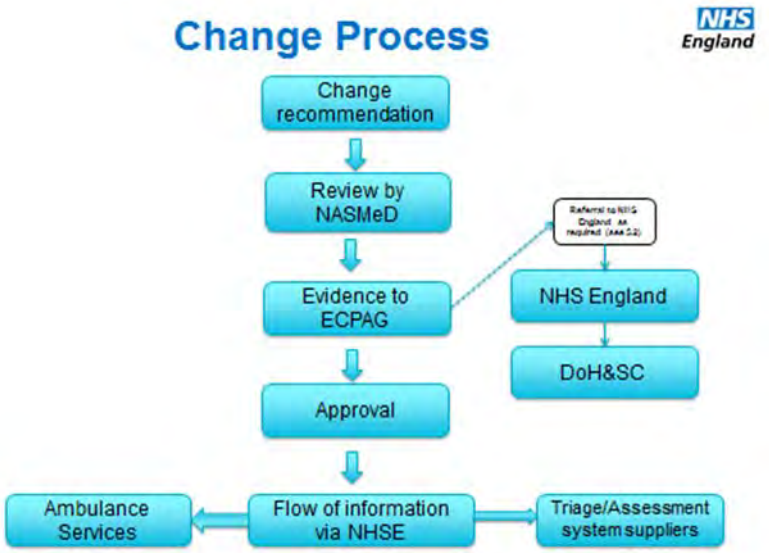
5.3 In the instances described in 5.2, above, further work may be undertaken by the NHS England National Clinical Director and the Acute Care Programme Lead in the first instance. Liaison may be required with NHS Improvement, DHSC and potentially the providers of triage software systems, to determine the impact of proposed changes.

NHS England and NHS Improvement will corporately consider clinical and operational issues and provide the assurances for these, in addition to any commissioning considerations. NHS England will consider whether its conclusions warrant referral to the DHSC and subsequently, the latter will determine whether ministerial briefing is required on the clinical and operational matters considered and recommended by ECPAG.

6.0 Change Process

6.1 ECPAG will consider changes to current coding based on receipt of a review of the evidence base with formal recommendations from the National Ambulance Service Medical Directors group.

6.2 Following approval change information will be disseminated to ambulance services in England and the triage/system suppliers by NHS England Acute Care team.



Operational Practices

Introduction:

As part of the ARP Review, AACE were commissioned by NHS England to undertake a number of workstreams including a review of certain operational practices namely:

- The use of blue lights and sirens when responding to incidents.
- Policies in respect of meal breaks and end of shift arrangements.
- The use of call handling scripts in control rooms.

The overarching intention was to establish a consistent position where this is practicable and adds benefit.

This paper summarises the position that has been reached in respect of this workstream.

The use of blue lights and sirens when responding to incidents

This action stemmed from questions raised by staff about whether it was appropriate to respond to Category 3 and 4 incidents on blue lights and sirens given that they were not described as emergencies under the original ARP descriptions.

NHS England requested that AACE sought to reach a consistent position with trusts to ensure common practice. On initial investigation it became clear that this is a question of law rather than a matter of policy. AACE commissioned comprehensive legal advice through CAPSTICKS. The advice is detailed and the issues are in places complex, but in very broad terms the key points include:

1. Determining the appropriate use of blue lights is ultimately the responsibility of the driver irrespective of the category of response.

2. On the face of it use of blue lights and sirens ought to be justifiable to most responses to calls fitting within Category 1 – 3.

3. Justification for the use of blue lights to most responses fitting within Category 4 becomes more difficult although it is possible that legal justification will exist in some instances.

4. The new categories have not changed the existing law in respect of when blue lights may be used.

5. Trusts will need to review their policies in light of this guidance.

6. National guidance may assist with this and a review of existing policies and assistance in drafting national guidance could be provided by CAPSTICKS if required.

This advice has been shared with Ambulance Trust CEOs and Operations Directors, ACEG and NDOG respectively.

The ARP Development Group was invited to discuss whether it was thought appropriate to develop national guidance. The group agreed this was required; the first draft of the guidance is expected after May 2018.

Policies in respect of meal breaks and end of shift arrangements

Following questions raised by staff representatives NHS England asked AACE to consider whether it would be possible to reach a nationally consistent position on these topics.

AACE have responded that such arrangements are the product of detailed and protracted local negotiations. They form a small part of a much broader range of operational and general working practices that are negotiated between employers and staff representatives. It would not be practicable to negotiate detailed national agreements for these working practices beyond that which is already set out in employment law.

AACE have previously offered assurance to the Secretary of State for Health and Social Care on matters relating to meal break policies and the response to higher acuity incidents.

999 call handler scripts

An unintended consequence of the introduction of ARP has been an increase in the number of duplicate calls received by some ambulance trusts. In the main this is due to members of the public calling back the ambulance service to establish an estimated time of arrival.

Some trusts have introduced call handling scripts that offer callers an ETA during the initial call in order to manage expectations with the intent of reducing duplicate calls. These scripts have been shared between trusts are we continue to share best practice through NDOG and the Heads of EOC Group.

In addition, NHS Pathways are exploring the potential to include the same script at the end of the 111 call handling process.

It is not felt that a NHS England mandated script is required as the design of the process is likely to evolve through local use and national sharing through existing structures. NHS England does support the use of key messaging in those ambulance services choosing to employ them. AACE will continue to support this work.

However, NHS England is working with NHS Pathways to develop the inclusion of a script for 111 providers transferring lower acuity calls to ambulance Trusts.

NAIG Peer Review Audits of Ambulance System Indicators.

Date: February 2018

Audits undertaken by : Chris Gresty, Tracy Rayment-Bishop

Introduction

Following the introduction of the new operating practices associated with the Ambulance Response Programme (ARP) a new set of Ambulance System Indicators were introduced in July 2017.

As part of the ARP process, NHS England is undertaking a review of the new Indicators in spring 2018, part of which is around the guidance and governance. In order to provide a level of assurance around data governance and reporting outputs, the National Ambulance Information Group (NAIG) were requested by Association Ambulance Chief Executives to support NHS England by undertaking a series of Peer Review audits of metric collation and supporting governance.

Aims

The aim of the audit process was to provide a confidence level in Ambulance System Indicator Reporting and governance, following the changes associated with ARP and level of assurance to AACE members in advance of the Spring Review planned by NHS England.

Methodology

A previous Peer review audit was undertaken in 2014 across all Trusts, comprising of 2 elements, Control Room Processes and Management Information Processes. NAIG undertook the Management Information processes review, with ambulance services working in pairs to undertake the reviews, looking at the system indicators.

In order to gain consistency in audit outputs, for these audits, it was requested that all 10 audits were undertaken by the same people.

Chris Gresty (Head of Informatics, NWAS) undertook the assurance side of the audits, focussing on the processes and governance around the production and publication of AQI figures.

Tracy Rayment-Bishop (Information and Performance Manager, WMAS) undertook the technical side of the audit, focussing on the underlying SQL code used to produce the AQIs and the correct application of the national guidance across the indicators.

Prior to the audit visit, Trusts were asked to complete a template providing their technical code for the AQIs, as well as answering a number of questions around both the technical and assurance sides.

Decembers AQI submissions were reviewed prior to the reviews to identify any areas where there were variations in returns or outliers.

Scope

All audits were undertaken on site at each Trust with a technical review focusing on reporting logic against the indicator guidance and supporting evidence provision. The audit also reviewed assurance processes from indicator production to submission and the reconciliations checks and sign offs employed by each Trust and supporting evidence provision.

Prior to undertaking the audits, AACE representatives, NHSE representatives, and NAIG members were asked if there were any specific things that wanted including in the audits. Feedback from this was incorporated where possible, and generated many of the consistency checks that were carried out.

Assurance Review

The Assurance audit covered the following areas, identifying evidence of these, and understanding the internal processes surrounding them:

Data Quality	Data Quality monitoring and assurance	assurance
Data Quality	Data Quality processes	assurance
Data Quality	Reconciliation of data to source systems	assurance

Governance	Retrospective time changes and monitoring	assurance
Governance	Data transfer	assurance
Governance	Critical System change management processes	assurance
Governance	AQI reporting process audits	assurance
Governance	Retrospective time changes and monitoring	assurance

System Assurance	System configuration processes	assurance
System Assurance	Operating system settings	assurance
System Assurance	Clock synchronisation	assurance

Reporting processes	Review processes to ensure compliance with national reporting standards	assurance
Reporting processes	Internal AQI reporting	assurance
Reporting processes	AQI production to submission sign off processes	assurance
Reporting processes	Procedural framework to reporting process	assurance

Control Practice	Mechanisms for review of practices in control environments	assurance
Control Practice	Processes for alignment of operational practices and reporting practices with control environments	assurance

Technical Review

The technical audit covered the following areas. Compliance checks were ensuring that Trusts were adhering to the national guidance, Consistency checks were areas which are not currently specified within the guidance but are areas of potential inconsistency.

Contacts	Out of area Exclusions	compliance

Calls	999 Line Identification	compliance
	Line Split	consistency
	111 Manual Transfers	consistency

Call Answer	Call Answer split	consistency
	Response Curve	consistency
	Connecting to tape	consistency
	Pt groups in 0 second call answer	consistency

Incidents	Out of area Exclusions	compliance
	Arrival before coding complete	compliance
	IFTs	compliance
	Running Calls	compliance
	HCP Emergency Calls	compliance
	Recategorisation - Subsequent call	compliance
	Recategorisation - Clinician Call back	compliance
	Call Linking	compliance
	Non Emergency Response	compliance
	Incident identification	compliance
	Code workarounds	consistency
	Exclusions from Incident Count	consistency
	Transfers from PTS	consistency
	Upgrades during call	consistency

NoC	NoC Codes list	consistency
	TimeStamp generation	compliance
	111 call inclusions	consistency
	PTQ/NoC/Keywords Times	compliance

H&T	National stop code list compliance	compliance
	H&T identification	compliance
	H&T Exclusions	compliance
	DX Code lists	compliance

CallBacks	Callbacks before resp on scene	compliance
	C4H with response on scene	consistency
	Obvious death/ DX91	consistency
	Taxis used for H&T	consistency

ClockStart	ClockStart timestamps	compliance
	DEFIB starting Clock	consistency
	CFR starting Clock (Cat 2 - 4)	consistency
	111 Clock Start	compliance

ClockStop	ClockStop rules	compliance
	C1 RRV DCA transport	compliance
	C1 RRV DCA non transport	compliance
	C2 RRV DCA transport	compliance
	C2 RRV DCA non transport	compliance
	C2 CFR RRV no transport	compliance
	DCA + DCA transport	compliance
	Double DCA logic	compliance
	Helicopters as DCAs	consistency
	HART team stopping clock	consistency
	CFRs as only response on scene	consistency

Performance	Performance calcuations	compliance
	Response curves	consistency

Responses	Vehicle Type exclusions	compliance
	Assignments for incidents only	consistency

Transports	Unknown Destinations	compliance
	MAU/SAU/EAU	compliance
	Stroke/PPCI	compliance
	S&C Identification	consistency

Summary of Findings

Exec summary:

The process was constructed in such a way to be a supportive exercise, that, while establishing an understanding of compliance and assurance from Trusts, would also identify and disseminate best practice in these areas to Informatics colleagues and Ambulance Trusts nationally.

The audits were split into reviewing compliance on the reporting metrics and also governance assurance from data collection information production to submission and the reconciliations checks and sign offs employed by Trusts.

The audits were very well supported by all Trusts in particular the Informatics functions. The process was felt to be very useful by the Trusts involved, both from undertaking the initial preparation and the experience and value gained from the audit day.

Overall, compliance to the guidance in all Trusts was extremely good. Whilst there are some areas of minor variation, and non-compliance these were not significant. There was some scope for improvements in governance processes, though areas of best practise were also identified.

Technical Audit

The technical audit reviewed each of the metrics in the reporting guidance considering reporting code, record level review and also outputs.

Detailed reports have been provided to each Trust informatics lead detailing the results of their audit, and a list of specific compliance actions, and additional consistency checks and actions. Many of these have already been actioned since the audits took place.

Many Trusts are still working with CAD suppliers to make new functionality available or compliant to the new indicators, and so the code and informatics process for generating some of the AQIs in some areas is more complex than necessary in some areas in order to be compliant.

A number of areas of guidance review were identified through this process, either new, additional guidance or review/clarification of existing guidance. These are detailed in the chart below.

Close relationship between informatics developers and analysts is essential for reporting to be understood and communicated wider and is an area which could be improved in some Trusts.

Findings:

Contacts (A0)

It was identified that there were a number of Trusts that were non-compliant with this metric. The reasons for non-compliance were due to incorrect inclusions/exclusions. Resolution of this issue is a simple process and is easy for Trusts to rectify. This metric is therefore deemed to have no cause for concern.

Calls (A1)

The majority of Trusts were compliant with the calls metrics. As with contacts any non-compliance issues were down to incorrect inclusions/exclusions. Resolution of this issue is a simple process and is easy for Trusts to rectify. This metric is therefore deemed to have no cause for concern.

Call Answer (A2-6)

It was identified that there is quite wide variation between Trusts around the speed of call answer within 0-2 seconds ranging from 0% to around 70%. The reasons for the variation are not known but are not reporting issues. The reporting of these metrics is compliant.

The call answer area and its variations will be addressed in more detail within the EOC phase of the peer audits.

Incidents(A7, A8-A12, A57-A61)

No issues with the incident metrics were identified. Incident counting is broadly consistent across Trusts with small variations within the consistency checks that would benefit from additional guidance.

The methodology for re-categorisations is not compliant with the current guidance wording in many Trusts, but the results are correct. This is mainly due to the way CAD systems allow changes to take place. This also affects the clock start for these calls, with most Trusts only being able to measure the time the call was upgraded, not when the call started.

The processes for recategorisation are varied across Trusts within control rooms, and will be addresses in more detail within the EOC phase of the peer audits.

NoC (A13-A16

There were no issues identified with the NoC metrics and all Trust demonstrated compliance based on the practices in place.

There is variation in the NoC codes lists used across all Trusts which is affecting the success rate of early identifying Cat 1s. ie the more you have the greater success of identification which was a concern raised by a number of Trusts.

Hear and Treat (A17,18,20,21,22)

There was no issues identified with the hear and treat metrics. The national stop code list has been implemented in 9 or the 10 Trusts audited, with the final Trust having mapped their existing codes to the national Stop code list. There were some very minor non-compliance identified at one Trust which is unlikely to affect any figures.

CallBacks (A20,A23)

Call backs were reviewed at all Trusts. It was identified that not all Trusts are reporting this due to lack of understanding of how to report this metric. Following the audits, reporting will be consistent although the guidance should be improved to increase understanding.

An issue was identified with C4H cases with a response on scene, which all Trusts have and isn't currently catered for within the guidance and so are being handled in different ways across Trusts. Guidance is required to be developed to cater for this.

Clock Start (A24+)

Extensive records checks were undertaken at all Trusts in regards to clock start logic.

All Trusts have correctly implemented the new Clock Start logic however additional guidance is deemed to be required around DEFIBs and CFRs starting the clock as there is none in place currently.

Clock Stop (A24+)

Extensive records checks were undertaken at all Trusts in regards to clock stop logic.

All Trusts were deemed to have correctly implemented the new logic correctly. One minor non-compliance issue was identified however this is being resolved.

A question was raised in many Trusts around urgent tier vehicles and how they should be counted, as this isn't currently set out in the guidance.

Performance (A24+)

There were no issues identified with performance calculations.

Responses (A39+)

Nearly all Trusts have implemented the inclusions and exclusions correctly associated with the responses metric. Minor non-compliance was identified at one Trust relating to specific vehicle types however this is being rectified. Therefore there is no cause for concern regarding the responses metric.

Transports (A55)

The majority of Trust have implemented the clarification around non A&E and specialist centres successfully.

Not all Trusts have the ability to separate out departments however this is being worked on to enable improved accuracy of reporting.

AQI Guidance Recommendations

The following potential guidance revisions or additions have been identified through the technical audit process. :

Metric+	Type	Detail	Current Position
A0	Wording clarification	111 calls - split into 2 paragraphs for clarity. Second paragraph to start ' For calls that are …'	
A0	Wording clarification	Specify abandoned calls as 'abandoned on telephony system'	
A1	new guidance - basic	Should 111 manual transfers be included?	6/10 don't include them 4/10 do. It depends on the methodology within EOCs as to how manual transfers are done
A7	current guidance discussion	Calls that are upgraded mid call. If transferred to clinician, this would restart the clock, rather than the elongated reporting time currently?	All Trusts currently report on original category
A7	current guidance discussion	Rewriting of the Recategorisation section due to the specification of running on second call and inability for some Trusts to work like this. Guidance to account for first or second cancellation and clock start capabiliities	Subsequent call by caller - 3/10 use original call, 7/10 use second call Clinician call back - 8/10 use original call, 2/10 use second call Dependant on CAD system for processes used and which call
A7	new guidance - detailed	Should cases with stop reasons of duplicate/ cancelled etc. be removed from the face to face counts even if they have a response on scene (may be DQ issue, may	5/10 include them 5/10 don't

		be run on wrong job, may be responded on 2 cases to 1 patient)	
A7	Wording clarification	Add in the section around cross border calls from section 2 to the top of section 5 as well for clarity	
A8-A12	Wording clarification	Include 'Face to Face' in metric titles eg C1 Face to Face Incidents'	
A56-A61	new guidance - detailed	Many Trusts now have urgent tier vehicles. These are not 'non-emergency' as per definition nor 'emergency'. How should these count for stopping the clock and counting as incidents. They will be the only vehicle attending scene.	
A56-A61	new guidance - basic	Should HCPs that are not transported be counted as guidance currently specifies 'transported by'	
A57	new guidance - basic	Should DX026 be excluded as delayed response? Some are 2 hour.	
A13	Wording clarification	Only says PTQ/NoC - Keywords should also be added to metric name.	All AMPDS Trusts using Keywords use the Keyword time if it's the earliest trigger
A13	current guidance discussion	Everyone is using different cat 1 NoC lists, ranging from 2 codes to around 25. As these will pick up different cases, this makes this metric non comparable across Trusts as it will affect success rates.	
A13/A8	current guidance discussion	111 Calls are included within A8 and A13 but they don't receive a NoC Code. Not an issue for A13-A16, but if A13/A8 done to see % of Cat 1s with Cat 1 NoC, then cant get 100%	
A20/A23	Wording clarification	The wording of this section is not clear and has led to some misunderstandings about how it could be counted. Perhaps context around aim of measure would help	
A23	current guidance discussion	Pathways Trusts only: Many of the C4H calls that go to a clinician for callback are those that have refused the initial disposition, or require clinician intervention in order to fully assess the patients. Therefore these are not true C4Hs (they haven't come out to any disposition) and counting them in this measure will falsely increase the volume of C4H with callback and then response.	
A24+ clock start	new guidance - basic	Should DEFIBs start the cat 1 clock if they are first assigned and earliest time stamp	6/10 do not allow defibs to start the clock, 1/10 does allow defibs to start the clock, 3/10 are still checking
A24+ clock start	new guidance - basic	Should CFRs start the clock on cat 2 - 4 if they are first assigned and earliest time stamp (and if so, which of their times should be used)	6/10 do allow CFRs to start the clock, 4/10 are still checking
A24+ clock stop	new guidance - detailed	If CFR (or other non clock stopping vehicle) is the only response on scene, should they stop the clock and count towards incident count and performance count, or should these cases be excluded from all counts?	6/10 do allow CFRs if only response on scene to stop the clock and count as incident. 4/10 do not currently.
A17/A8-A12	new guidance - basic	All Trusts have some C4H with responses at scene, for various reasons. Which category should these be under within the AQIs? (A8 - A12)	Reasons include DX91 responses on scene, arrival at same time as coding, no capacity on Clinician desk, control room processes,

A39+	Wording clarification	Should Officers be included within the allocations/response counts?	
A39+	new guidance - basic	Should Allocations only be to those incidents that had a response on scene, or should it be allocations to everything (including cases that get cancelled before anything arrives on scene and so case closed)	5/10 Trusts have filters to only include those in A7 incident count. 5/10 have no filter on.

Assurance Audit

The assurance audit covered 6 areas that support the effective introduction and implementation of new reporting standards into an organisation considering multi-disciplinary stakeholder review, decision making, data capture, system settings/configuration and reporting processes.

Data Quality

All Trusts demonstrated some level of data quality processes being in place. The level of processes adopted ranged from missing or incorrect data such as timestamps in the call cycle process to full reconciliation between core system data, AQI data and commissioning information.

Best practice is acknowledged as implementing both approaches with missing data reporting and full reconciliation reporting between data sources.

Governance

All Trusts have critical system change management processes in place however these varied greatly between multi discipline sign off, deemed to be best practice, and small sign off processes limited to key individual sign off. Changes such as those that have come through ARP need wider consideration which should incorporate Control, Emergency Service and Clinical functions.

Data management procedures varied greatly between Trusts with retrospective changes and monitoring of changes being supported by robust audited procedures at some Trusts and extremely limited at others.

Audit

Good audit practices were demonstrated at most Trusts with variation witnessed in the level of audits undertaken. Some Trusts focused on the compliance audits to reporting guidance and some gained assurance on procedures to support reporting production, some Trusts undertook both.

Utilising internal and/or external audit are deemed best practice.

System Assurance

The configuration of Command and Control systems is essential to achieving the correct reporting outputs. A number of Trusts have audits undertaken by system providers which review system settings to ensure they are in line with indicator guidance. IE clock start, clock stoppers, conveying resources. This approach was deemed to be best practice.

Clock synchronisation processes between time sources were in place and witnessed at most Trust, again this is again deemed to be best practice in assurance terms.

Reporting Processes

Procedures for report production and submission were reviewed and in most cases deemed to be robust. Checks and balances of outputs were witnessed in procedures in some Trust and deemed to be an important part of reporting process.

Value was achieved at a number of Trusts through regional or national benchmarking reporting of indicators.

Control Practice

The majority of Trusts referenced strong leadership and support structures in their control environments which oversaw operational practice and corrected poor behaviours. A number of Trusts also demonstrated the use of external review such as AACE Consultancy which was deemed to be best practice.

Reporting changes and control room practices are inextricably linked and need to be considered together to maximise efficiencies and benefits to the service.

Follow up

NAIG Audit leads will continue to follow up audit compliance actions and recommendations working with the Trust Informatics leads.

The results of these audits will for part of the wider audit reports prepared by AACE which will be circulated to Trust Chief Executives.

NAIG would recommended that the National Guidance is reviewed annually to ensure it continues to meet the needs of changing ambulance requirements with a direct link between NHSE and NAIG to ensure any new requirements can be incorporated into new versions of guidance.

Conclusion

These audits were well received and proved a valuable exercise for all Trusts. The compliance to the new guidance within Information departments has been extremely good, and only very minor non-compliance was identified. There was more variation within the consistency checks, which would benefit from additional guidance, but again, none of these will be impacting on accuracy of reporting to any concerning level.

There was variation in assurance practices, with many areas of best practice identified, many areas are still in development along with other areas that Trusts could improve on, but overall all Trusts have some level of assurance in place for the generation of their AQI reporting.

Ambulance Improvement Programme

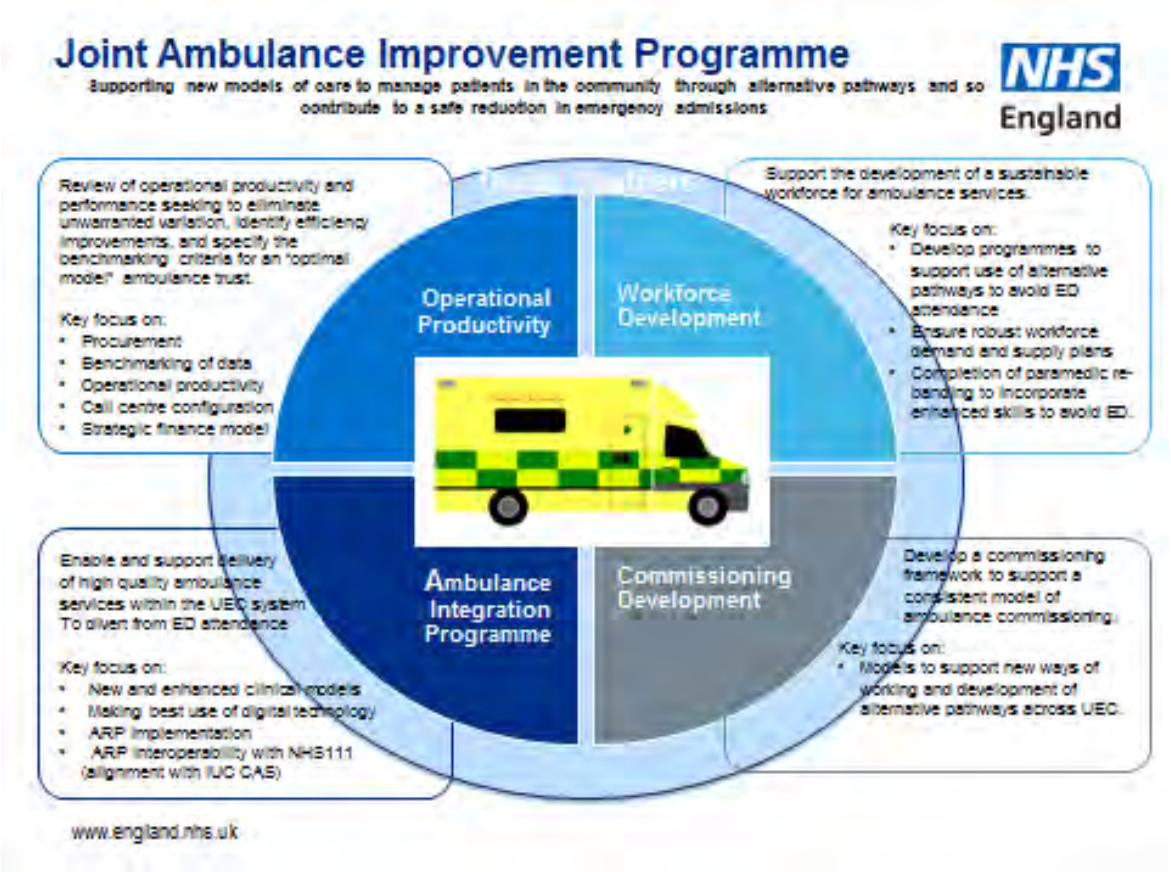
Since 2017 the NHS Improvement (NHSI) and NHS England (NHSE) Ambulance Improvement Programme has taken responsibility for the programme management of national ambulance development work on an integrated basis.

The Joint Ambulance Improvement Programme (JAIP) group is accountable to the executive teams of both NHSE and NHSI and provides quarterly updates to the two respective executive teams. The work of the group is undertaken in partnership with key stakeholders.

The workplan for the JAIP follows on from the foundations laid by the Ambulance Response Programme with the aims to ensure that by 2021 all ambulance services will:

- Meet all targets and deliver all patient outcomes;
- Have an efficient and effective financial balance;
- Have a satisfied, happy and productive workforce;
- Be integrated into the wider Urgent & Emergency Care System; and
- Be digitally enabled for the future.

The workplan is organised across five areas as illustrated below:



The progress of the JAIP will be monitored and communicated through:

- A quarterly Deep Dive by each workstream.
- Quarterly key data summary measures of:
  - Targets (ARP balanced scorecard);
  - Finance (cost improvement against plan);
  - Workforce (sickness, engagement); and
  - Digital Maturity assessment read out.

- A brief monthly update by each workstream is collated into a ‘Talking Points’ format and distributed to key stakeholders.

Transition to Business as Usual

As the Ambulance Response programme transitions into business as usual it will be important to ensure the principles, practices and interventions central to the programme are maintained at both local and national levels.

To support this NHS England has established the Ambulance Transformational Delivery Team to work closely with the regional teams, lead commissioners and CCGs to support effective and efficient delivery through partnerships by providing dedicated subject matter expertise to regions, Ambulance Trusts and lead commissioners.

The team is a new centrally managed support team that will provide dedicated subject matter expertise to regions, ambulance trusts and lead commissioners. This new team will give direct operational line of sight for the National UEC Director to both commissioners and providers, and resolve any lack of clarity about responsibilities. The team will provide support to all parties as they work to meet the required ambulance system and clinical quality standards, measures and trajectories by September 2018 and beyond. This will include a clear focus on the development of schemes and pathways to reduce ambulance conveyance to Emergency Departments as well as the development of the enablers required to deliver more hear and treat and see and treat interventions where clinically appropriate. The team will also deal with other ambulance specific trouble shooting issues e.g. BT call handling, delayed ambulance handover, working closely with colleagues in the year-round UEC operations function.

In practical terms NHS England will chair a quarterly meeting with representatives from each region supplemented by monthly meetings between the central team, each individual region and ambulance lead commissioners.

The NHS England operational support leads will be available to work closely with partners in the regions.

The central team will be reliant on regional teams developing the capacity for sufficient commissioning oversight to ensure that planning guidance asks are met, and that commissioners have ensured that local contracts reflect remedial actions required to ensure adequate performance by September 2018 and beyond.

The NHS England central team has developed a balanced scorecard to support timely identification of performance issues.

The scorecard is a combination of operational, safety and clinical measures that provide both an individual trust and benchmarking position. Operational and safety data is sourced from a weekly trust return, whilst clinical information is sourced from the published Ambulance Clinical Quality Indicators. The definitive source of operational data is still the Ambulance Quality Indicators (AQIs), published on the second Thursday of every month for the previous month.

The information in the dashboard includes:

- Mean Response Times (Category 1-3)
- 90th Centile Response Times Category 1-4 (including Category 1 Transport)
- Call answering times
- Call Closure Information
- Long Waits (Category 1 – 4)
- Call to Angiography for STEMI
- Call to Thrombolysis for Stroke
- Survival to discharge in the Utstein subset of cardiac arrest patients
- Untoward Incidents (from Autumn 2018)

The Scorecard will be used to give a more balanced and stable picture of ambulance services, for use by the central team, regional colleagues and lead commissioners in guiding longer term discussions and improvement over time (i.e. weeks to months). The monthly AQI data will continue to be the definitive source of performance data; this supports the reporting and monitoring of long-term performance and trends in ambulance services (i.e. months to years).

Ambulance Services
Balanced Scorecard

Trust WMAS

Weekly Indicators - Week Commencing: 30 April 2018

Monthly Indicators - Month Commencing: December 17

The overall operational RAG rating is calculated using the 10 indicators below for week commencing 30 April 18.
WMAS achieved 9 of these standards, with a further 0 RAG rated as amber

The overall clinical indicators RAG rating is calculated using the 2 available indicator(s) below for month commencing December 17

Please refer to the CQI Publication Schedule for further information

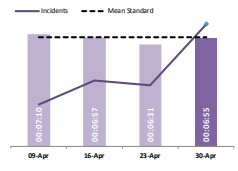
RAG Rating Guidance Notes
CQI Publication Shedule
Data Notes

Further information for each indicator is available by clicking inside the charts

Weekly Indicators
*** RAG Rating based on achievement of standards in the latest week's data**

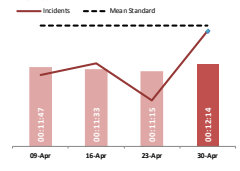
Mean Response

Category 1 RAG



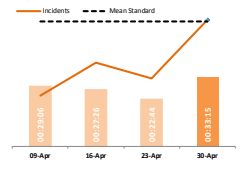
Data Source: Ambulance Trust Weekly Return

Category 2 RAG



Data Source: Ambulance Trust Weekly Return

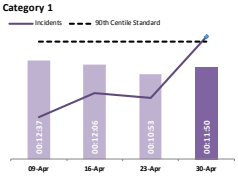
Category 3 RAG



Data Source: Ambulance Trust Weekly Return

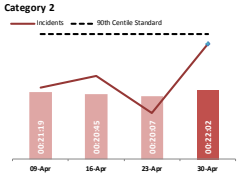
90th Centile Response

Category 1 RAG



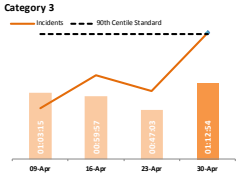
Data Source: Ambulance Trust Weekly Return

Category 2 RAG



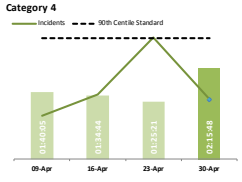
Data Source: Ambulance Trust Weekly Return

Category 3 RAG



Data Source: Ambulance Trust Weekly Return

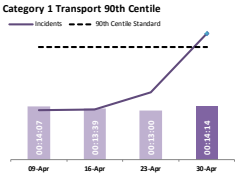
Category 4 RAG



Data Source: Ambulance Trust Weekly Return

90th Centile Response

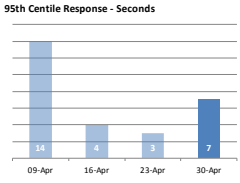
Category 1 Transport 90th Centile RAG



Data Source: Ambulance Trust Weekly Return

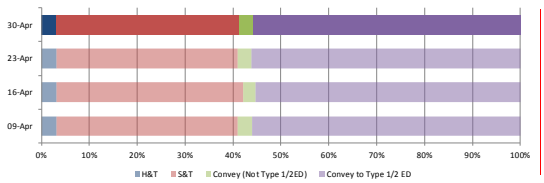
Call Answering

95th Centile Response - Seconds RAG



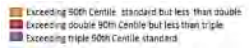
Data Source: Ambulance Trust Weekly Return

Call Closure

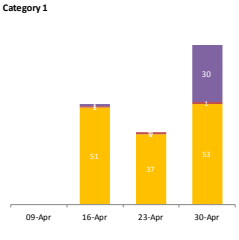


Data Source: Ambulance Trust Weekly Return

Long Waits
*** RAG Rating based on achievement of standards in the latest week's data - see notes for how RAG Ratings are calculated**

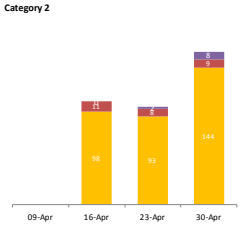


Category 1 RAG



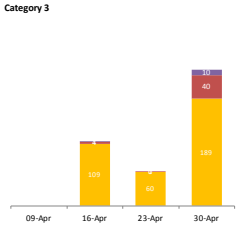
Data Source: Ambulance Trust Weekly Return

Category 2 RAG



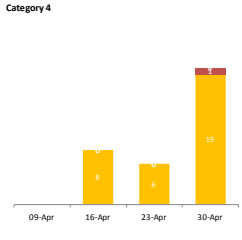
Data Source: Ambulance Trust Weekly Return

Category 3 RAG



Data Source: Ambulance Trust Weekly Return

Category 4



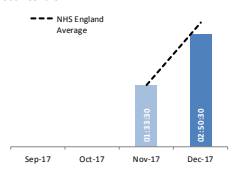
Data Source: Ambulance Trust Weekly Return

Monthly Data

Call to Angiography for STEMI

Data Source: Ambulance Quality Indicators Data
https://www.england.nhs.uk/statistics/statistical-work-areas/ambulance-quality-indicators/ambulance-quality-indicators-data-2017-18/

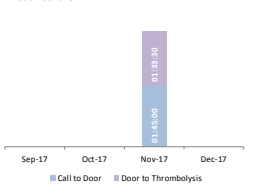
90th Centile RAG



Trust Rank* out of 11:

Call to Thrombolysis for Stroke

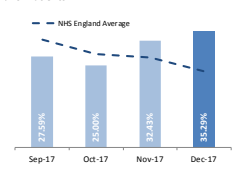
90th Centile RAG



Trust Rank* out of 11:

Survival to discharge in the Utstein subset of cardiac arrest patients

% of Patients RAG



Trust Rank* out of 11:

SUIs

Data not currently collected centrally - discussions ongoing with trusts

**Trust Ranking is defined as 1 - Worst Performing to 11 - Best Performing*