Paper: CB.30.03.2017/02

NHS ENGLAND – PRIVATE BOARD PAPER

Title: Ambulance Response Programme
Rationale for paper being discussed in the Private meeting: The Board is asked to discuss and approve the Ambulance Response Programme paper which seeks to change the operational delivery model and national metrics for ambulance services in England, ahead of publication in early April 2017.
Lead Directors: Professor Sir Bruce Keogh, National Medical Director Professor Keith Willett, Medical Director (Acute) Professor Jonathan Benger, National Clinical Director for Urgent and Emergency Care
Purpose of Paper: This paper presents the findings and recommendations arising from the Ambulance Response Programme's evaluation report, produced by the School of Health and Related Research (ScHARR) at Sheffield University. The paper recommends approval for implementation of these recommendations. This paper also sets out a recommendation for revision of ambulance response time standards arising from the Ambulance Response Programme's evidence-based evaluation. The Board is asked to consider the recommendations presented, and to support the proposed changes to ambulance quality indicators.
The Board is invited to: • Note the findings of the Ambulance Response programme (ARP) evaluation report, produced by the School of Health and Related Research (ScHARR) at Sheffield University; • Agree national adoption of the ARP interventions: Nature of Call, Dispatch on Disposition and the new ambulance prioritisation code set; • Consider the proposed changes to ambulance service standards, measures and indicators; and • Approve the standards for adoption into the national Ambulance Quality Indicator set.

Ambulance Response Programme

Background

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

2. Recognition of these issues, and a need to re-think how services are delivered, prompted NHS England to undertake a review of urgent and emergency care in 2013. Following publication of this review a number of work programmes were developed to support implementation of the review's key principles – the Ambulance Response Programme (ARP) is one of these initiatives.

3. The ARP aims to improve response times to critically ill patients. It will make sure that the best, high quality, most appropriate response is provided for each patient first time, and sets out to improve outcomes for all patients contacting the 999 ambulance service, with a generally reduced clinical risk throughout the whole patient group.

4. The key elements of the programme are:
    - The use of a new pre-triage set of questions to identify those patients in need of the fastest response at the earliest opportunity (Nature of Call);
    - Dispatch of the most clinically appropriate vehicle to each patient within a timeframe that meets their clinical need (Dispatch on Disposition);
    - A new evidence-based set of clinical prioritisation codes that better describe the patient's presenting condition and response/resource requirement; and
    - A full review of ambulance service measures and quality indicators.

5. Appendix 1 sets out the headline results of the programme and outlines the case for change.

6. In September 2015, NHS England engaged Sheffield University's School of Health and Related Research (ScHARR) to act as an independent academic partner to monitor, analyse and evaluate the ARP. Appendix 2 summarises this report and provides recommendations for change.

Evaluation

7. Dispatch on Disposition (DoD) gives call handlers more time to clinically assess all 999 calls that are not immediately life threatening (all calls except Red 1) before a resource is dispatched, ensuring a more appropriate response, based on clinical need. Following careful testing of various time intervals, in several pilot trusts, a maximum DoD window of 240 seconds has been in use in pilot form across England since October 2016 (except for the Isle of Wight, which commenced DoD in February 2017).

    We recommend the formal adoption of Dispatch on Disposition, with a maximum time of 240 seconds, by all ambulance services in England with immediate effect.

8. Dispatch on Disposition is supported by Nature of Call (NoC), which is a system that identifies the highest priority calls (particularly cardiac arrest) very early in a 999 call, so that a response is dispatched to life-threatening emergencies as quickly as possible.

   We recommend that all ambulance services adopt the three pre-triage questions. Those services using NHS Pathways should also adopt NoC. The early identification of potential cardiac arrest patients should be introduced as a new ambulance performance indicator.

9. The current call category system fails to take account of the increasingly diverse range of health problems, both in terms of type and urgency, that people request ambulance service help for.

10. The ARP tested, in two stages, a revised call category system that is designed to provide responses that are a better fit between urgency, clinical need and the most appropriate response. The revised call categories are presented at Appendix 2, Annex 2.

11. ScHARR found that the first stage coding trial made significant gains; the second stage refinements maintained these and enhanced efficiency in the ambulance service control room and dispatch, with greatly enhanced operational stability.

    We recommend national implementation of the new ambulance prioritisation code set in all ambulance services in England.

Ambulance standards

12. Red calls are currently divided into Red 1 and Red 2, and account for approximately 45% of all ambulance activity. All remaining calls are categorised as Green, and have no mandated response time. Green performance is reported locally, but is not visible and transparent to the public. There have been increasingly long waits for some of these categories of call, leading to adverse media headlines, and poor patient experiences, often for elderly and frail patients.

13. The current response standard, as described in the NHS Constitution Handbook, is: *“all ambulance trusts to respond to 75 per cent of Red 1 and Red 2 calls within eight minutes, and to respond to 95 per cent within 19 minutes of a request being made for a fully equipped ambulance vehicle (car or ambulance) able to transport the patient in a clinically safe manner.”*

14. When using the Sheffield evaluation, summarised in Appendix 2, to determine appropriate response time standards for the new operating model, we should consider the stated aims of the ARP:

    - To provide the fastest possible response to critically ill patients (principally cardiac arrest);
    - To provide the most appropriate response to a patient’s clinical condition; and
    - To reduce the long (and hidden) waits that some patients experience in the current system.

15. We should be careful to strike a balance between 'media friendly' short response times and those which are longer and come with ambulance service support, but which are ultimately difficult to sell to the public and to politicians. Our aim in this regard should be to devise a set of standards that:

    - Are sufficiently stringent to incentivise a change in operating model and fleet configuration within the ambulance service;
    - Encourage 'stretch' and improving performance over the first few years of operation;
    - Allow a more even spread of resource across all categories; improving patient outcomes for the majority, whilst accepting that some patients may wait longer than under the previous system;
    - Accept that 999 demand will continue to grow, and therefore the ambulance sector will remain under pressure to meet the new response time standards; and
    - Do not lead us back to where we started, with large groups of patients forgotten or 'hidden', and an unrealistic and clinically unjustified headline target applied to nearly half of all 999 calls, creating perverse incentives and operational inefficiencies.

16. With the above in mind, **we recommend that the Board agrees that all 999 call categories in the new system will have a nationally mandated response time, and that this should be set at the 90th centile, rather than the 75th centile as is currently the case.** In effect, the headline standard becomes; '90% (9 out of 10) patients will receive a response within x minutes'.

17. It is possible that a 90th centile response standard could be seen as unpalatable for the most critically ill patients (categories 1 and 2), due to the fact that it inevitably requires a longer time than the current 8 minute 75% target. For this reason, we believe there would be benefit to also having a mean standard for these categories of patients only. Using the mean will ensure that every patient counts towards the time target.

18. The suggested response time standards for each category of patient are as follows:

	90^{th} centile response time (minutes)
Category 1	15
Category 2	36
Category 3	90
Category 4T	180

	Mean response time (minutes)
Category 1	7 ½
Category 2	18
Category 3	
Category 4T	

Stakeholder management and public messaging

19. Stakeholder feedback to date has been broadly positive, and our stakeholder group has had active engagement from the unions and key charities including the Stroke Association. Unions have indicated their support and recognise that the changes allow for a more effective working environment. The College of Paramedics and Association of Ambulance Chief Executives have been involved in the design of the programme and are supportive.

20. Prioritising the sickest patients and reducing the longest waits for patients, aligned to improved rural response times, is expected to have the support of MPs.

21. In addition to the formal evaluation from the School of Health and Related Research at Sheffield University, a ‘lay’ summary of the report that explains what the ARP set out to achieve, the findings and our associated recommendations has been developed.

22. Public messaging will emphasise the need for standards that focus on a clinically appropriate response. This will deliver a more equitable and efficient service offer overall, freeing additional resources for allocation, reducing long waits, offering genuine transparency for performance in all categories and ensuing patients receive the response they need, rather than a response that merely ‘stops the clock’. The sickest patients will continue to be prioritised, and will receive the fastest possible recognition and response. This message can be aligned to a ‘reset’ of the public offer in the Delivery Plan in response to the funding settlement, whilst still providing reassurance that operational efficiencies will be delivered through implementation of these recommendations.

Legal/Regulatory

23. A change to the national time standards for ambulance services in England will require changes to secondary legislation.

Recommendations

24. The Board is asked to:

- Note the findings of the Ambulance Response Programme (ARP) evaluation report, produced by the School of Health and Related Research at Sheffield University;
- Agree national adoption of the ARP interventions: Nature of Call, Dispatch on Disposition and the new ambulance prioritisation code set;
- Consider the proposed changes to ambulance service standards, measures and indicators; and
- Approve the standards for adoption into the national Ambulance Quality Indicator set.

Author: **Professor Jonathan Benger**
Date: **March 2017**

APPENDIX 1

Ambulance Response Programme (ARP): Headline Results and the Case for Change

Background

1. The ARP is the largest study of an ambulance system ever completed anywhere in the world.

2. The ARP is safe. More than 10 million patients have been studied, and there have been no identified adverse incidents or associated patient safety concerns.

Dispatch on Disposition

3. Allowing ambulance call handlers a limited amount of extra time to decide what is wrong with a patient does not reduce the speed of response to the highest priority patients, however it significantly improves ambulance efficiency. When adopted nationally, there are 10,000 occasions when an additional ambulance vehicle is freed for response in England every week.

4. This yields the biggest improvements for patients who wait too long for an ambulance. The current target for Red 2 calls is that 75% should be responded to within 8 minutes, however for 1 in 20 Red 2 calls an ambulance may take as long as 20 minutes to arrive. The introduction of Dispatch on Disposition reduces this long wait by up to 3 minutes.

Nature of Call

5. The introduction of three standardised pre-triage questions increases the early recognition of cardiac arrest and will improve patient survival.

The New Call Prioritisation Code Set: Categories 1 to 4

6. The current prioritisation system divides all 999 calls into two groups of approximately equal size; red and green. The red calls have an 8 minute time target, even though we know that most red calls don't benefit from an 8 minute response. Yet trying to meet this target for millions of calls unnecessarily leads to huge inefficiencies in the ambulance service. Meanwhile the green calls can wait for hours because they have no national response time target.

7. The new prioritisation system has 4 categories of call, which matches the range and complexity of problems for which people dial 999. The most serious prioritisation group (Category 1) has more than doubled in size, to ensure that people with life threatening problems get the fastest possible response. For other calls the ambulance service will meet the needs of the patient, with the right response provided in the right timeframe.

8. The new system is more stable, and much better equipped to deal with unexpected events and peaks in demand.

9. The new system is more efficient: when adopted nationally, there are more than 6,000 occasions when an additional ambulance vehicle is freed for response in England every week.

10. Because the new system puts patients into different priority groups it is impossible to directly compare the old system with the new. For example, the old Red 2 category contained 1,842 patient codes, whereas the new Category 2 contains 1,092. However, mean and 95th centile response performance for the most serious emergencies in the ambulance services testing the new system (Category 1) is compared to performance for the most serious emergencies in those services still using the old system (Red 1) in Figure 1. Note that Category 1 contains more than twice as many patients as Red 1.

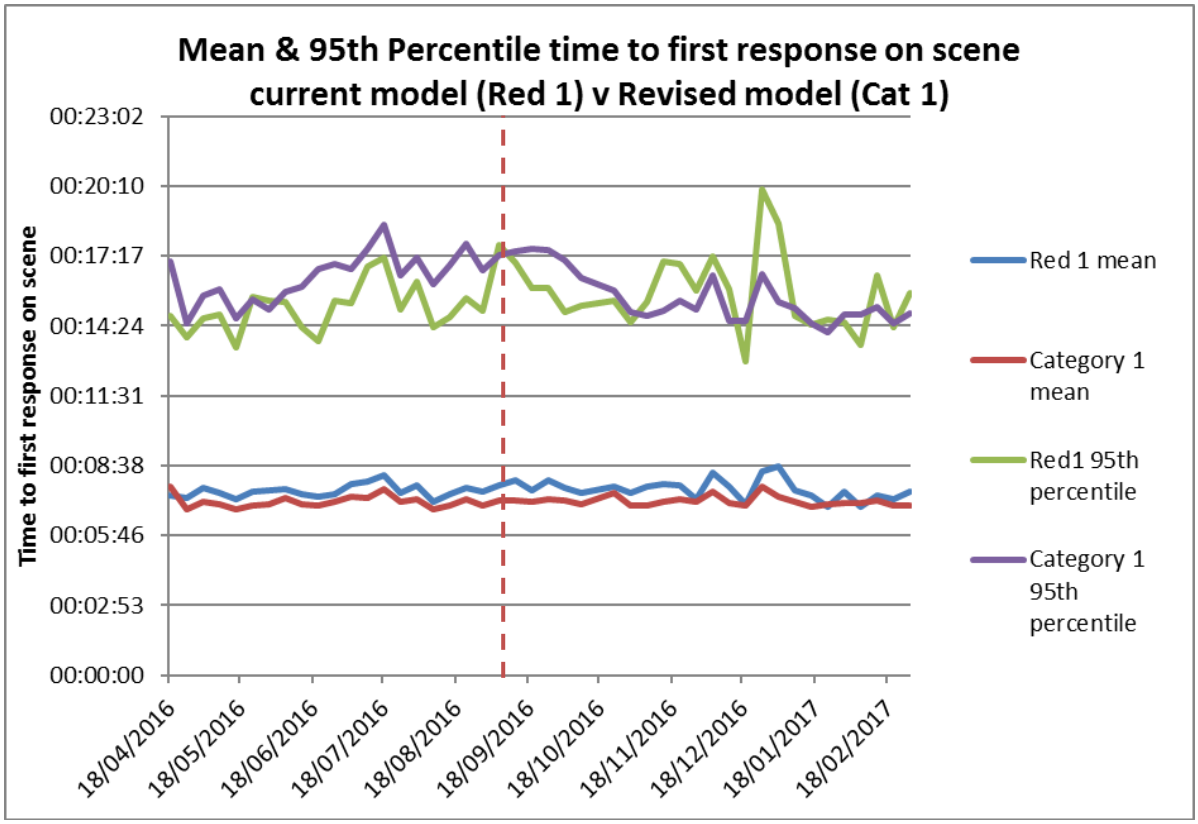

Figure 1; showing mean and 95th centile response times to the highest category calls under the new system (Category 1; red and purple lines) compared to the highest category calls in the old system (Red 1; blue and green lines). Note that lower values are preferable, and Category 1 has more than twice as many patients as Red 1. The new code set shows greater stability, with less week-to-week variability and, after the second phase trial was introduced in October 2016 (indicated by the dashed orange line), better 95th centile performance.

Rural Response

11. The impact on rural response times was measured in one trial site (West Midlands Ambulance Service). This showed that with the introduction of the new prioritisation system most of the existing disadvantage for rural areas was reversed, and for other measures variation in performance was substantially reduced. This translates to clear benefits for rural populations.

12. Under the old system 95% of patients in rural areas waited on average 5.5 minutes longer for an ambulance to arrive than those in urban areas, however under the new system 95% of patients in rural areas got a response on average 7.5 minutes *faster* than in urban areas. In other words, the patients living in a rural area will get an ambulance significantly sooner than they did before, and potentially faster than patients living in more urban areas.

Staff and Sector Support

13. Staff surveys have shown strong support for the principles that underpin the ARP, and the changes that have been tested during the programme.

14. The ARP is supported by ambulance services, trades unions, clinicians and patient groups, and is widely recognised as being “the right thing to do”.

New Standards for Ambulance Services

15. Current standards apply to no more than half of the patients who dial 999, and are set at 75%; this means that 1 out of every 4 patients can miss the time target but still meet the standard.

16. From now on we will set the bar at 90%, rather than 75%, so 9 in 10 patients have to hit the target in order to meet the standard. For the first time we will also measure mean, rather than median, response times, so every single patient counts towards the time target.

17. For the first time, a time standard will apply to every patient to whom a vehicle is sent, and when a patient needs to be transported only the arrival of an ambulance will “stop the clock”. This will reduce long waits for both a response and a transporting vehicle, which in the current system can sometimes be six hours or more.

18. We are also introducing a new set of clinical quality indicators, to measure the time between calling for an ambulance and receiving life-saving treatment for heart attack and stroke as well as survival following cardiac arrest. These will be accompanied by new measures for the care of patients with sepsis, and the response to people who have fallen and are still on the floor.

Summary

19. The Ambulance Response Programme (ARP) continues to ensure the fastest possible response for a larger number of people. It improves ambulance service efficiency and stability through a new system of call handling and prioritisation that provides time to establish the patient’s needs, before matching those needs to the right response in the right time frame. This reduces long waits and improves the speed of response in rural areas. A new set of time standards, that apply to all 999 calls combined with the measurement of clinical outcomes for serious conditions, ensure the service remains responsive, effective and accountable.

APPENDIX 2

Ambulance Response Programme: Summary of Phase 1 and Phase 2 Report, and Recommendations for Change

Introduction

1. This report summarises the Ambulance Response Programme (ARP) Phase 1 and Phase 2 Report, prepared by the School of Health and Related Research (ScHARR) at the University of Sheffield, and makes recommendations for changes to the operating model and performance management of ambulance services in England.

Background

2. It is widely acknowledged that whilst the vast majority of patients who receive a Red (8 minute) response do not derive any clinical benefit from this, the current ambulance standards have led to operationally inefficient behaviours including:

    i. Dispatch of multiple vehicles to a single 999 call.
    ii. Frequent tasking and re-tasking of ambulances and cars, with vehicles not actually arriving at a patient to whom they have been sent.
    iii. Use of response cars “to stop the clock”, followed by extended waits for the ambulance that actually conveys the patient to hospital; this additional wait is hidden in the present system, and can be very long.

3. There is also an increasing awareness that ambulance services in England are not assessed on those components of the service that reflect a patient-centred organisation, user experience and clinical outcomes.

4. In the face of steadily rising demand, the operational performance of ambulance services has declined over recent years. The current focus on response times in approximately 50% of the 999 cohort (red calls) has led to increasingly delayed responses and unacceptable waits for green category callers and those in more remote and rural areas.

The Ambulance Response Programme

5. The ARP encompasses three programmes of work:

    i. Dispatch on disposition (DoD). This allows the ambulance service additional time to determine what is wrong with a 999 patient, for all but the highest priority calls. Under the current system, the response time clock starts at 60 seconds, unless a disposition is reached (i.e. the caller’s precise problem is identified) or a vehicle is assigned to the call before then. For the highest category patients (actual, suspected or threatened cardiac arrest) the clock continues to start at “call connect” (the time the 999 call is connected to the ambulance switchboard); this is important because cardiac arrest is the only diagnosis that has ever been shown to have an improved clinical outcome if an ambulance arrives sooner. To further improve the ambulance response to cardiac arrest, DoD has been combined with “Nature of Call” (NoC); a set of actions in the ambulance control room to improve the early identification of actual or suspected cardiac arrest.

ii. A revised clinical code set. To reduce the size of the red group, and to ensure that the ambulance response reflects more closely the patient's actual needs, an evidence-based revision of the ambulance response code set has been completed and tested, through two iterations in three ambulance services.

iii. ScHARR has also led an evidence-based and expert-informed review to make recommendations for change to ambulance measures and indicators, and to be used as the basis of a future approach to a revised set of performance standards.

6. This is the largest prospective study of ambulance services ever completed anywhere in the world.

7. 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 contacting the ambulance service.

8. These objectives accord with the views of the Secretary of State for Health, who has indicated that any proposed changes must be beneficial for patients, operationally efficient, and supported by a clinical consensus within the ambulance sector.

9. Dispatch on Disposition (DoD). The independent ScHARR report demonstrates clear gains from DoD as follows:

i. No detriment to Red 1 performance.

ii. For Red 2 calls, the proportion of cases in which the patient's problem was identified before "clock start" increased by 18%.

iii. Consistent gains in efficiency with lower multiple vehicle allocation rates across all categories of call. This equates to 10,000 occasions when an additional ambulance vehicle is freed for response in England every week.

iv. Significant reductions in 95th centile response times, indicating that the additional freed resources are acting to reduce the longest waits.

v. No change in "hear and treat" rates overall, though the picture was mixed across England. The ARP is an enabler of future change, and we will follow this programme with further work to increase "hear and treat" rates where it is safe and appropriate to do so.

vi. The optimal additional time was found to be 240 seconds from call connect (i.e. the current 60 seconds plus a further 180 seconds). Only 5% of Red 2 calls achieve a disposition by 60 seconds, but this figure increases to 72% at 240 seconds. However, it is important to note that 240 seconds is the maximum clock start time, and the response clock starts if the patient's problem is

identified or a vehicle is allocated to the call before then. With a DoD time of 240 seconds the clock starts, on average, 60 seconds later for red calls, and this is balanced by the increased efficiency and availability of resource, so the average time between call connect and a resource arriving on scene is unchanged for all red calls.

Recommendation 1: That DoD is adopted as a permanent change for all calls except Red 1 (using the old code set) or Category 1 (using the new code set). This is consistent with current arrangements.

10. Nature of Call (NoC). The independent ScHARR report demonstrates benefit from NoC as follows:

    i. The introduction of three standardised pre-triage questions increases the early recognition of cardiac arrest.

    ii. Additional NoC processes, to identify calls that have a high probability of being a cardiac arrest, are effective in those ambulance services using the NHS Pathways triage system, but less effective in those ambulance services using the Advanced Medical Priority Dispatch System (AMPDS).

    iii. A continuing focus on the early identification of cardiac arrest improves recognition, and it is anticipated that this will translate into increased survival.

Recommendation 2: That all ambulance services adopt the three pre-triage questions as standard. Those services using NHS Pathways should also adopt NoC. The early identification of potential cardiac arrest patients, and the time taken to start cardiopulmonary resuscitation, should be introduced as new ambulance performance indicators to ensure an ongoing focus on the response to the sickest patients.

11. The revised clinical code set. It is difficult to draw clear conclusions from the independent ScHARR report due to the completely new nature of the revised code set, and the lack of a consistent comparator. However, the following benefits were observed:

    i. Much greater operational stability, suggesting that the revised call categories are very helpful in managing variation in demand and other factors that have an impact on performance, such as reduced resource availability caused by ambulances queuing at hospitals. It is notable that headline performance declined in non-code trial sites, whilst holding relatively steady in the code trial ambulance services.

    ii. The highest priority group (category 1; most like the current Red 1) increased in size considerably. For this group the 8 minute response time was retained and performance, though variable across the trial sites, remained largely steady. For all other calls, where an 8 minute standard did not apply, the average response time was largely maintained, but the 95th centile response time tended to reduce, suggesting fewer long waits.

    iii. Additional gains were achieved in operational efficiency and multiple vehicle allocation, though these were smaller than the benefits accompanying DoD. When scaled nationally, this equates to an additional 6120 resources freed for response in England every week. There was also some evidence of additional

benefit in the recognition of cardiac arrest.

iv. The impact on rural response times was measured in one trial site (West Midlands Ambulance Service). This showed that with the introduction of the code trial most of the existing disadvantage for rural areas was reversed, and for other measures variation in performance was substantially reduced, indicating clear benefits for rural populations.

v. Feedback from the trial sites has been positive, with examples of operational improvements and enhanced service efficiency. Further developments in the operating model and fleet mix are expected to yield additional improvements. There are opportunities to refine the code set and its implementation in the light of these findings, with an ongoing process of feedback and improvement based on the systematic collection and analysis of clinical outcome data; a process that will be embedded in the future code set.

Recommendation 3: That the new ambulance code set is adopted permanently in England.

12. Clinical safety. During the ARP, ambulance services closely monitored patient safety by scrutinising all adverse incidents and conducting regular reviews of all calls where there were long waits for an ambulance response. Throughout the evaluation, which encompasses more than 10 million 999 calls, there have been no identified adverse incidents or patient safety concerns associated with the ARP changes.

13. Staff feedback:

i. A staff survey completed following the introduction of Dispatch on Disposition and Nature of Call showed that these changes were generally viewed by both Operations Centre and Clinical staff as a positive development. The responses indicated that a substantial proportion of staff thought there had been improvements in triage, clinical assessment and operational efficiency (e.g. “the frequency of “stand downs”). 96% of Operations Centre staff felt that the pre-triage and nature of call questions identified immediately life threatening calls all of the time, most of the time or some of the time, and 78% felt that the additional time for call assessment made patient triage and allocation of the right response category more effective.

ii. Further staff surveys, completed in three trusts after the coding trial has been implemented, showed that the ability to dispatch resources more appropriately was maintained. Whilst the shift from an over-emphasis on 8 minute response time was seen as positive, with 87% of Operations Centre staff responding that this made the ambulance response more effective, there remained a perception of over-triage in the highest priority categories and suggestions that some calls in Category 2 could be offered a higher priority. There was also a view that there are sometimes long waits for lower priority calls.

14. Recommendations for future Ambulance Quality Indicators (AQIs)

I. Annex one contains initial suggestions for changes to ambulance standards and indicators in England. These standards are aligned to the revised clinical code set (Annex 2) recommended for national adoption. These proposals have been developed through the following process:

- Review of existing research literature;
- Comparison with similar systems in the devolved nations and overseas;
- Stakeholder engagement and consensus events;
- Early feedback from key stakeholders and the ARP Development Group.

II. The proposed measures, indicators and standards will be supported by a clear and unambiguous reporting tool and guidance in a revised AQI document. We have incorporated these into a step wise model that follows the patient journey in Annex 3.

III. The main changes proposed simplify the current set. In summary, we propose that:

- Call answering is retained as an indicator, using the 95th and 99th centiles;
- When measuring response times the 90th centile and mean response times should be reported, rather than using a single 75% cut-off. 90th centile (9 out of 10) is more readily understood by the public, and drives an improved response to more patients. Mean is representative of all ambulance responses, and provides an incentive to reduce long waits;
- These times should be reported for all categories (C1-C4), meaning the response to all incidents is visible. This avoids the current situation where response to Green calls (up to 50% of the 999 workload and often the elderly and frail) is unreported;
- For transported C1 calls an additional indicator ("C1T") has been added to measure the time that the vehicle that conveys the patient arrives at scene, as well as the initial ambulance service response.

Response Time Performance in the ARP, and Potential Future Standards

15. Code set performance in the three participating ambulance trusts is shown in Table 1. We have shown the 90th centile response times, reflecting our current recommendation. We have also included mean response times for Category 1 which we are proposing to adopt as a fifth national standard. The mean response times for other categories (shown in grey) would be reported as additional indicators.

Table 1: Average response times in the three code set trial sites from a clock start of Call Connect. If the DoD clock start were used instead all times would be reduced by approximately 2 minutes. Currently Call Connect is used for the Category 1 clock start, and DoD for all other categories. Our preference is to retain this approach, since it reduces the risk of response time "gaming".

Time from call connect to resource on scene: hh:mm:ss		WMAS	SWAST	YAS
Category 1	Mean	00:07:16	00:07:20	00:07:12
	90th centile	00:12:09	00:12:58	00:12:41
Category 2 R	Mean	00:10:48	00:20:02	00:16:54
	90th centile	00:18:58	00:42:03	00:35:24
Category 2 T	Mean	00:11:44	00:21:41	00:17:50
	90th centile	00:21:02	00:56:56	00:30:46
Category 3 R	Mean	00:21:41	00:48:12	00:33:22
	90th centile	00:52:43	01:49:34	01:18:32

Category 3 T	Mean	00:22:09	00:52:40	00:41:40
	90th centile	00:52:21	02:08:52	01:38:30
Category 4 T	Mean	00:36:11	01:28:12	01:05:31
	90th centile	01:52:32	03:35:10	02:30:12

16. Based on these data, national standards could be framed as shown in Table 2:

Table 2: Proposed Response Time Standards; 90th centile (All Call Categories)

	90th centile response time (minutes)
Category 1	15
Category 2	36
Category 3	90
Category 4T	180

17. We outline in Table 3 the recommended mean standards as a comparator.

Table 3: Proposed Time Standards; Mean (All Call Categories)

	Mean response time (minutes)
Category 1	7 ½
Category 2	18
Category 3	
Category 4T	

18. We recommend adopting a realistic set of standards that include all patients and which can be achieved and maintained by an effective ambulance service in the face of rising demand and relatively constrained resources. This approach avoids repeating the mistakes of the past, in which an excessive focus on a single highly demanding time-based standard drives system inefficiencies and disenfranchises 50% of patients contacting the ambulance service. Explicit time standards for each category of call allow a realistic dialogue with the public and will set clear expectations around ambulance response; as fast as possible for the very sickest patients, and tailored to urgency and clinical need for others.

19. Revised "clock stop" criteria will ensure that there are no more hidden waits for a transporting ambulance, and will have the additional benefit of strongly incentivising "see and treat" and new models of care; if the patient is transported the clock is stopped by the conveying resource, however if the patient is not conveyed the clock stops when the first ambulance response arrives.

Recommendation 4: That the current ambulance performance indicators are replaced with those proposed in Annex 1, and that the 90th centile response time standards shown in Table 2 (plus mean response for Category 1) are adopted in England.

Clinical Quality Indicators (CQIs)

20. The final ARP report from ScHARR includes broader recommendations for a better and more comprehensive rolling clinical audit programme that is separate to, but runs alongside and complements, these proposed system indicators. However, in the short term, we recognise the need to maintain clinical quality indicators (CQIs) within the national reporting framework. We therefore propose to retain CQIs for STEMI (heart attack), cardiac arrest and stroke, whilst adding new CQIs relating to sepsis and patients who have fallen and are still on the floor. We also recommend that CQI reporting moves to quarterly.

Recommendation 5: That a new, updated and expanded Programme of Clinical Quality Indicators is adopted in England.

Realising the ARP's Objectives

21. We believe the ARP's three key objectives are met by the proposals set out in this paper as follows:

    I. The sickest patients (those in Category 1) will be recognised more swiftly than ever before by Nature of Call. New indicators of "Time to identify a C1 call" and "Time to CPR in cardiac arrest" have been added, with Category 1 response standards that apply to 90% of patients, instead of 75%.

    II. Dispatch on Disposition and the new clinical code set increase ambulance service efficiency by reducing multiple allocation and dispatch. Under the new code set, more patients will get the right response, first time, and in a clinically appropriate time frame. This will put an end to perverse incentives and the consequent inefficiencies created by targets, and will be measured and monitored.

    III. The new ambulance code set, and a requirement to report response times for all calls, including the 90th centile and mean, will put an end to unacceptably long delays in both urban and rural areas.

Implementation

22. If approved, these changes could be implemented in approximately 3-4 months using a staggered transition (since technical suppliers cannot change all services at once). This means that if a decision to proceed were made in March 2017 implementation and "bedding-in" could occur over the summer, so the new system has fully settled into place before next winter.

23. As the system adapts to the new indicators and standards, and the revised definitions that support these, along with the introduction of further initiatives that are enabled by the ARP (such as additional work to support "hear and treat" and "see and treat") we expect to see and will work to support:

    I. Progressive fleet reconfiguration over 1-2 years, with fewer rapid response cars and a higher proportion of double-crewed ambulances.

II. Gradual changes in workforce with an enhanced skill mix providing clinical care at scene.

III. Increased rates of “hear and treat” and “see and treat”.

IV. Improved performance against the proposed indicators and standards, though this will be attenuated significantly if demand continues to rise in line with current trends.

V. Improved rural response times.

VI. The further development and introduction of an effective set of clinical quality indicators, supporting enhanced clinical outcomes for ambulance services in England.

Annex 1: Proposed Ambulance System Indicators

Revised System Indicators
Activity (Call volumes) All calls All incidents For all categories (C1; C2R, C2T; C3R, C3T; C4T, C4H) **Of all incidents:** Calls not receiving a face to face response Calls receiving a face to face response Patients transported Patients not transported
Call Answering time (seconds; median, 95th, 99th percentile) **Time to identify C1 call** 90th centile; mean **Time to CPR in cardiac arrest** 90th centile; mean **Hear and Treat – all no vehicle response** % closed with advice % referred to alternative service % returned for ambulance response (index call) (Divided by C4H and all other categories)
Response performance Response time for all categories (C1, C1T; C2; C3; C4T); 90th centile and mean (Clock stop on arrival of the vehicle that conveys the patient; if the patient is not conveyed the clock stops on the arrival of first ambulance dispatched response) **Allocation rate:** Number of ambulance-dispatched responses allocated per incident and arriving on scene for all categories (C1, C1T; C2; C3; C4T); mean.
Calls receiving an ambulance dispatched response % Not transported % Transported to type 1 or 2 ED % Transported to other facility

Annex 2: ARP Call Coding Descriptors

Call Type Definition	Recommended Response and Resource
Category 1 -Life-threatening Time critical life-threatening event needing immediate intervention and/or resuscitation e.g. cardiac or respiratory arrest; airway obstruction; ineffective breathing; unconscious with abnormal or noisy breathing; hanging. Mortality rates high; a difference of one minute in response time is likely to affect outcome and there is evidence to support the fastest response.	**C1** Ambulance clinician who can assess and deliver advanced life support Standard 15 mins (90th centile); 7 ½ mins (mean) C1T Transporting vehicle where transport required. Operational response plan to deliver fastest suitable resource
Category 2 - Emergency Potentially serious conditions (ABCD problem) that may require rapid assessment, urgent on-scene intervention and/or urgent transport. Mortality rates are lower; there is evidence to support early dispatch. Standard 36 mins (90th centile), 18 mins (mean) **(For calls that need conveying clock stop is by the vehicle that actually conveys)**	**C2T Assess; Treat; Transport** e.g. Probable MI, serious injury, stroke, sepsis, major burns Suitably qualified clinician who can assess and treat **and** vehicle that does transport
	C2R Assess; Treat e.g. Fits; unconscious with normal breathing Nearest available resource (any type) with suitably qualified clinician who can assess and treat
Category 3 – Urgent Urgent problem (not immediately life-threatening) that needs treatment to relieve suffering (e.g. pain control) and transport or assessment and management at scene with referral where needed within a clinically appropriate timeframe. Mortality rates are very low or zero; there is evidence to support alternative pathways of care. Standard 90 mins (90th centile); **(For calls that need conveying clock stop is by the vehicle that actually conveys)**	**C3T Assess; Treat; Transport** e.g. serious injury modalities without systemic compromise; burns (not major); non-emergency late pregnancy/childbirth problems.
	C3R Assess; Treat Calls within scope of advanced clinical practice and suitable for treat and leave. e.g. uncomplicated diabetic hyper/hypoglycaemia; not immediately at risk drug overdoses; non-emergency injuries; abdominal pain.
Category 4 – non-urgent Problems that are not urgent but need assessment (face to face or telephone) and possibly transport within a clinically appropriate timeframe.	**C4T Assess; Treat; Transport** 999 or 111 calls that may require a face to face ambulance clinician assessment, or Requests for transport by health care professionals Standard 180 mins (90th centile)
	C4H Non-ambulance response Calls that do not require an ambulance response but do require onward referral or attendance of non-ambulance provider in line with locally agreed plans or dispositions, or can be closed with advice (Hear & Treat)
Type S – Specialist response Incidents requiring specialist response i.e. hazardous materials; specialist rescue; mass casualty.	**Locally agreed plans apply**

Annex 3: Stepwise Model of the Proposed Measures and Indicators, Reflecting the Patient Pathway