**Improving ambulance services: Changes to the way performance and quality are measured**

**XX Clinical Commissioning Group Governing Body meeting**

**1 December 2017**

## 1. Purpose

The purpose of this paper is to update the Clinical Commissioning Group (CCG) Governing Body on the recent national work undertaken to improve ambulance service performance and develop a consistent framework for measurement of performance.

## 2. Executive summary

Following an extensive review of the way ambulance services in England assess 999 calls and allocate a response, in July 2017 the Secretary of State for Health approved a number of changes to the way ambulance services prioritise their work and their performance is measured. The changes made are designed to ensure:

- The sickest patients receive the fastest response
- Patients get the response they need first time, and in a timeframe that is appropriate for their condition
- Resources are spread more equally amongst all patients contacting the ambulance service and waits are reduced
- That people living in rural areas receive a more equitable response

These changes started being measured by the Yorkshire Ambulance Service (YAS) from 1 September 2017.

## 3. Background

Since the 1974, ambulance services across the UK have been assessed on their ability to achieve time based response times and meet various clinical quality indicators. This was borne out of a belief that the quicker someone was seen and/or taken to hospital the better the outcome. Over time a number of factors have changed that thinking:

- The training, clinical education and development of the ambulance workforce has improved meaning greater capability to treat on scene rather than convey to hospital.
- Patients would prefer to be treated on scene where this is appropriate and clinically safe.

- Clinical effectiveness has improved. Whilst some people dialling 999 will still need immediate treatment, for example to stem blood loss and others with serious conditions will need treatment in an acute care setting e.g. stroke patients others some can be managed safely and to a high standard of care without conveyance to hospital.
- The nature of the reasons people dial 999 for has changed over time due to the aging UK population and the availability of alternative services e.g. NHS 111. Many 999 calls now relate to urgent care for people with long term conditions rather than life threatening emergencies.

In addition, emergency ambulance providers have themselves found the historic service standards have distorted behaviour. This can include:

- Dispatching resources to a 999 call, on blue lights and sirens, before it has been determined what the problem is, and whether an ambulance is actually required.
- Dispatching multiple ambulance vehicles to the same patient, on blue lights and sirens, and then standing down the vehicles least likely to arrive first.
- Diverting ambulance vehicles from one call to another repeatedly, so that ambulance clinicians are chasing time standards rather than focussing on patients care needs.
- Using a 'fast response car' to 'stop the clock', when this unit may provide little clinical value to the patient (e.g. stroke patient), who then has to wait a long time for a conveying ambulance to arrive.
- Patients receiving very long waits for lower priority (green) calls that nevertheless need assessment and conveyance to hospital, and some of which concern time sensitive problems

## 4. Development process

As part of the review of urgent and emergency care led by Sir Bruce Keogh, http://www.nhs.uk/NHSEngland/keogh-review/Documents/UECR.Ph1Report.FV.pdf a service improvement programme called the Ambulance Response Programme (ARP) was designed around three overarching improvement objectives:

- Prioritising the sickest patients to ensure they receive the fastest response.
- Driving clinically and operationally efficient behaviours, so the patient receives the response they need first time and in a clinically appropriate timeframe.
- Reducing long waits by ensuring resources are distributed more equitably amongst patients

The ARP established a number of working groups that looked at:

### 4.1 Despatch on Disposition (DoD)

Ambulance trusts felt that the existing national target of 60 seconds to triage and dispatch a resource wasn't enough time to do this effectively. Pilots were established within a small number of ambulance trusts to allow them longer to determine the nature of the call before deciding to despatch a response. The changes were piloted and evaluated and gradually expanded to a larger number of ambulance services. The objectives were:

- Better responses to Red1 incidents: faster dispatch to the most critical calls.
- Ensure resources more available (through less multiple dispatches to single incidents) to respond to life threatening immediate calls.
- Dispatch of the most clinically appropriate resource (single-crew, double-crew, local responder) to patients by taking a little more time to triage the call.
- Increase 'Hear & Treat' during the call – no ambulance is dispatched if inappropriate.
- Increase 'See & Treat' on scene with no conveyance from scene of incident.
- Improve speed of conveyance for patients for whom time of conveyance is critical e.g. cardiac and stroke patients.

### 4.2 Coding changes

ARP also looked at how 999 calls were coded once received. The coding determines the priority of response. Traditionally calls were allocated a response code: red, amber or green and an associated response timeframe.

The working group found the coding model used was no longer for purpose with many calls being coded 'higher' than was necessary. Work was undertaken to review the data and reach consensus as to which codes (presenting clinical conditions) could be moved safely to a different response category. These codes were tested with an external reference group, including patient representatives and other stakeholders.

### 5.0 Evaluation of pilots

The ARP development work was overseen by a group of leaders (including commissioners) working in the field and was monitored for safety and evaluated on a regular basis.

Early on NHS England identified Sheffield university as an academic partner to conduct a thorough evaluation. The University of Sheffield published their final report in July 2017 and it can be found at:
https://www.england.nhs.uk/wp-content/uploads/2017/07/ARPReport_Final.pdf

In summary, the main findings of the ARP are that:

- Giving call handlers more time to assess a call is beneficial
- The most urgent emergency calls do not receive a slower response when call handlers have this extra time available
- The ambulance service becomes more efficient, and there are less long waits for an ambulance to arrive
- Patients in rural areas get a response more like urban areas
- In over 14 million 999 calls analysed no patient came to harm as a result of the ARP
- Ambulance staff acknowledge that the changes are beneficial to patients and to staff. They said:

    - Additional time for call assessment is a positive step;
    - Workload management is clearer for the people who allocate and send ambulance vehicles. This helps them to prioritise the sickest patients;
    - Ambulance crews in rapid response cars feel they are sent to more appropriate patients;
    - There are fewer “stand downs”, where vehicles are cancelled before arrival;
    - Fewer stand downs are better for public safety as there are fewer ambulance vehicles travelling at high speeds on the roads.
    - ambulance services are better able to manage their workload, and are less vulnerable to peaks in demand.
    - Staff recognise that the changes are designed to be more focused on patients than simply chasing a time target. This leads to staff being more satisfied that they are doing the right thing for patients.

**6.0 New ambulance performance and clinical quality measures**

On 13th July 2017 the Secretary of State for Health accepted five recommendations:

- That DoD is adopted as a permanent change in England.
- That all ambulance services adopt the three pre-triage questions. The early identification of potential cardiac arrest patients should be introduced as a new ambulance performance indicator.
- That the new clinical prioritisation set is adopted permanently in England.
- That the current ambulance performance indicators are replaced, and that a 90th centile response time standard is adopted in England.
- That a new updated and expanded programme of Clinical Quality Indicators (CQIs) be introduced in England.

999 calls will be prioritised into one of four new categories:

–Category 1 (Purple) – Life threatening
–Category 2 (Amber) – Emergency
–Category 3 (Yellow) – Urgent
–Category 4 (Green) – Less Urgent

ARP sets the bar at 90%, rather than 75%, so 9 in 10 patients have to hit the target in order to meet the standard. Measurement will be around the mean, rather than median response times, so every single patient counts towards the time target.

A time standard will apply to every patient to whom a vehicle is sent. Apart from the most urgent cases (Cat 1), when a patient needs to be transported only the arrival of the conveying ambulance will 'stop the clock'. This aims to reduce long waits for both a response and a transporting vehicle.

NHS England will also be introducing a new set of Ambulance Quality Indicators (AQIs) to measure the time between the 999 call and receiving life-saving treatment for heart attack and stroke, as well as cardiac arrest survival. These will be followed by new measures for patients with sepsis, and people who have fallen and are still on the floor.

New AQIs will include reporting of data across the patient pathway as ambulance Trusts begin to utilise national outcome databases. Reporting will move to a quarterly schedule to better monitor trends ahead of full publication in April 2018 due to the preparatory work required for a new stroke indicator. The STEMI and Cardiac Arrest outcome data, which are unchanged, will be available ahead of this.

The new quality indicators include revised and standardised definitions for "Hear and Treat" and "See and Treat" which will impact on reporting of these two interventions.

The revised quality indicators are being adopted by ambulance trusts as they move to the ARP system during the period July to November 2017. During this period there will be two sets of national standards in place and ambulance services will report on those associated with their operating system; ARP or non-ARP. These two sets are not comparable.

**7.0 Next steps**

a. Nationally, the ARP work will continue under the auspices of the Ambulance Integration Programme (AIP) which has commissioner representation. New Ambulance System Indicators have been published https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/04/20170811-Ambulance-System-Indicators.docx and a full suite of ambulance reporting can be found at https://www.england.nhs.uk/statistics/statistical-work-areas/ambulance-quality-indicators/

b. NHS England has updated the NHS Constitution to accommodate the changes to the way ambulance performance is measured https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/642390/NHS_constitution_addendum_September_2017.pdf and confirmed that ambulance trusts must not have financial sanctions directly applied in 2017-19 for non-delivery of the new performance standards.

c. YAS have indicated to commissioners that they are in the process of undertaking an impact assessment of the new changes. YAS commissioners have similarly indicated that they wish to receive a trajectory for performance for 2018-19 against all the new categories.

d. Commissioners and YAS remain in dialogue regarding agreement of a contract variation embedding the new standards into the YAS 999 contract.

e. Nationally, ambulance commissioners will continue to encourage NHS England to review the four national ambulance currencies to bring them into line with the new standards.

f. At a Y&H level performance against the new measures will be monitored, as now, through the 999 Contract Management Board which has representation from the coordinating commissioner and Sustainability and Transformation Partnership (STP)/ Accountable Care System (ACS) sub regions. Y&H CCGs are sent performance updates on a weekly basis with a more detailed report being made available monthly.

g. Over the next few months further engagement with Y&H CCGs/STPs/ACS will be undertaken regarding the ARP changes. This will happen at two levels:

h. YAS and a 999 coordinating commissioner representative will attend will attend STP/ACS meetings to discuss the changes.

i. A workshop for CCG Business Intelligence leads/ CCG Directors of Performance will be held in each of the STP/ACS sub regions in January/February 2018 to discuss in detail the ARP changes and to inform the narrative within CCG Governing Body Board reports. These workshops will be supported by Y&H CCGs, YAS, Embed Health Consortium and North of England Commissioning Support.

**8.0 Action for Governing Body/Recommendations**

The x Governing Body is asked to note the positive outcomes of the ambulance response programme and the changes to reporting from 1 September 2017.

## Appendix 1 – Ambulance Performance Indicators (Abridged)

| Category | Percentage of calls in this category | National Standard | When does the clock start? | What stops the clock? |
|---|---|---|---|---|
| Category 1 | 8% | 7 minutes mean response time<br>15 minutes 90th centile response time | The earliest of:<br>•The problem is identified<br>•An ambulance response is dispatched<br>•30 seconds from the call being connected | The first ambulance service-dispatched emergency responder arrives at the scene of the incident (There is an additional Category 1 transport standard to ensure that these patients also receive early ambulance transportation) |
| Category 2 | 48% | 18 minutes mean response time<br>40 minutes 90th centile response time | The earliest of:<br>•The problem is identified<br>•An ambulance response is dispatched<br>•240 seconds from the call being connected | If a patient is transported by an emergency vehicle, only the arrival of the transporting vehicle stops the clock. If the patient does not need transport the first ambulance service-dispatched emergency responder arrives at the scene of the incident |
| Category 3 | 34% | 120 minutes 90th centile response time | The earliest of:<br>•The problem is identified<br>•An ambulance response is dispatched<br>•240 seconds from the call being connected | If a patient is transported by an emergency vehicle, only the arrival of the transporting vehicle stops the clock. If the patient does not need transport the first ambulance service-dispatched emergency responder arrives at the scene of the incident |
| Category 4 | 10% | 180 minutes 90th centile response time | The earliest of:<br>•The problem is identified<br>•An ambulance response is dispatched<br>•240 seconds from the call being connected | Category 4T:<br>If a patient is transported by an emergency vehicle, only the arrival of the transporting vehicle stops the clock |

**Coding categories explained:**

| Identifier | Clock start | Response measure | Definition | Comment |
|---|---|---|---|---|
| Category 1<br>Purple | Problem identified<br>Allocation<br>30 seconds | Mean (7 mins)<br>90th Percentile (15 mins) | Time critical life-threatening event needing immediate intervention and/or resuscitation.<br>Example – Cardiac or respiratory arrest; airway obstruction; ineffective breathing; unconscious with abnormal or noisy breathing; hanging.<br>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. | Sub-category of C1T (transport) will be monitored with a view to a future standard regarding transport. |

| Identifier | Clock start | Response measure | Definition | Comment |
|---|---|---|---|---|
| Category 2<br>Amber | Allocation<br>MPDS Code<br>240 seconds | Mean (18 mins)<br>90th Percentile (40 mins) | Potentially serious conditions (ABCD problem) that may require rapid assessment, urgent on-scene intervention and/or urgent transport.<br>Example – Probable MI, serious injury, stroke, sepsis, major burns, fits, unconscious with normal breathing.<br>Mortality rates are lower; there is evidence to support early dispatch. | |

| Identifier | Clock start | Response measure | Definition | Comment |
|---|---|---|---|---|
| Category 3 yellow | Allocation MPDS Code 240 seconds | 90th Percentile (120 mins) | 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. Example – serious injury modalities without systemic compromise; burns (not major); non-emergency late pregnancy/childbirth problems; uncomplicated diabetic hyper/hypoglycaemia; not immediately at risk drug overdoses; non-emergency injuries; abdominal pain. Mortality rates are very low or zero; there is evidence to support alternative pathways of care. | |

| Identifier | Clock start | Response measure | Definition | Comment |
|---|---|---|---|---|
| Category 4T (Transport) GreenT<br><br>Category 4H (Hear & Treat) GreenT | Allocation MPDS Code 240 seconds | 90th Percentile (180 mins) | Problems that are not urgent but need assessment (face to face or telephone) and possibly transport within a clinically appropriate timeframe.<br><br>**C4T** – 999 or 111 calls that may require a face to face ambulance clinician assessment, or requests for transport by health care professionals.<br><br>**C4H** – 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 and treat). | |