

Professor Sir Bruce Keogh
National Medical Director
Skipton House
80 London Road
SE1 6LH

Jeremy Hunt
Secretary of State for Health
By email and hard copy

13 July 2017

Dear Jeremy,

Ambulance Response Programme

In recent weeks we have seen countless examples of the outstanding work done by the ambulance service in the most tragic of circumstances, from the response to terrorist attacks in London and Manchester to the devastating fire at Grenfell Tower. The extraordinary response to these terrible events came on top of the everyday heroics by paramedics that save countless lives day in, day out across the country.

We have also marked the 80th anniversary of the introduction of the 999 emergency telephone number. The ambulance service has changed beyond recognition during this time, from little more than vehicles transporting patients to hospitals, often staffed by volunteers, to the “mobile hospital” model we see today.

It is a timely reminder that the NHS is constantly evolving and, as leaders of the NHS, we must always ensure that we move with the times – supporting staff to provide the best possible service to our patients, rather than putting obstacles in their way.

Yet, in the case of the ambulance service, it has become increasingly obvious that we have failed to keep up. Since the mid-1970s most aspects of the service have changed beyond recognition: a large number of responses now focus on the frail elderly rather than traditional medical emergencies, half of all calls are now resolved by paramedics without the need to take patients to hospital, and for specialist care the focus of the ambulance service is increasingly on getting patients to the *right* hospital rather than simply the nearest. Over the last four decades, however, the service has remained organised around an eight minute response time target.

Amidst all of this change that standard has become an anachronism, with anxious callers placed into outdated categories that are no longer fit for purpose. Half of all calls are classed as urgent with an 8 minute response time target – but one that has to be met in only 75% of cases. The other half of calls are deemed non-urgent with no national response target at all. Response times for that second group of patients have, unsurprisingly, doubled in some trusts in the last two years alone.

For those covered by the 8 minute target the system is equally dysfunctional. Ambulance staff are given just sixty seconds to decide what resource each patient needs. While this may have worked many years ago, it is hopelessly unsuited to modern medicine. A stroke patient, for example, will gain little benefit from a paramedic on a motorbike when what they need is an ambulance that can rapidly convey them to a specialist treatment centre.

High quality care for all, now and for future generations

There is also the problem of “hidden waits” for those patients needing urgent hospital treatment. At present, the clock is “stopped” by the arrival of the first vehicle, not the arrival of the vehicle that the patient actually needs. A quarter of all patients who require hospital treatment have the clock stopped by a vehicle – often a motorbike – which is in fact incapable of taking them anywhere. There are few better examples of hitting the target and missing the point.

Most worryingly, the target can increase response times and cost lives. Multiple vehicles are often dispatched to the same patient in a race to “stop the clock”. When calls where a patient’s needs only become known after the one minute has elapsed are factored in, one in four ambulances dispatched are now stood down before they reach the scene. Every year hundreds of thousands of patients fail to get an immediate response because ambulances are dispatched in this wasteful and illogical manner. Correcting this anachronism would free up to fifteen thousand ambulance responses every week.

These criticisms are not new. They have been highlighted by the National Audit Office, by the Health Select Committee, and by countless paramedics and ambulance staff. So when I wrote to you in 2015, I said that we were determined to finally tackle this problem. I commissioned the Ambulance Response Programme (ARP) – an independently evaluated trial to test new ways of working for the service, led by Professor Jonathan Benger and Professor Keith Willett.

Over the last 18 months the ARP has covered over 14 million calls, testing a new operating model and new set of targets. Further details are annexed to this letter, but in summary this new system would:

- Change the dispatch model of the ambulance service, giving staff slightly more time to identify patients’ needs and allowing quicker identification of urgent conditions.
- Introduce new target response times which cover every single patient, not just those in immediate need. For the most urgent patients we will collect mean response time in addition to the 90th percentile, so every response is counted.
- Change the rules around what “stops the clock”, so targets can only be met by doing the right thing for the patient.

The results have been impressive. The trial has demonstrated that, should these changes be adopted nationally:

- Early recognition of life-threatening conditions, particularly cardiac arrest, will increase. Based on figures from London Ambulance Service, it is estimated that up to 250 additional lives could be saved in England every year.
- Up to 750,000 patients every year would receive an immediate ambulance response, rather than joining a queue.
- The differences in response time between patients living in rural areas and those in cities would be significantly reduced.

All of this has been achieved with no patient safety or adverse incidents attributed to the ARP in those 14 million calls.

Given this comprehensive and compelling evidence, I am writing to you formally to recommend the roll out of the Ambulance Response Programme to every ambulance service in England. Patients across the country deserve to benefit from the significant improvements seen in the trial areas, from ambulances reaching cardiac arrests in London 30 seconds faster to the one minute improvement on stroke responses in the West Midlands. These changes, together with ambitious new clinical standards for heart attack and stroke patients, will end the culture of “hitting the target but missing the point.” They will refocus the service on what actually counts: outcomes for patients.

These trials, the most extensive ever conducted, have provided us with an unrivalled evidence base for these changes. They also come with the strong endorsement of every expert organisation we have spoken to – whether the Royal College of Emergency Medicine, the Stroke Association, or the College of Paramedics.

If these recommendations are accepted then we intend to fully implement these new standards by the beginning of winter 2017, a little over six months before the NHS's 70th birthday. As we inevitably use this moment to reflect on both the achievements and challenges of the NHS, I am confident that the ambulance service would approach this landmark in a much stronger position to continue its remarkable work even more effectively.

Yours sincerely,

A handwritten signature in black ink, reading "Bruce Keogh". The signature is written in a cursive style with a long horizontal line extending from the end of the name.

Professor Sir Bruce Keogh KBE, MD, DSc, FRCS, FRCP
National Medical Director
NHS England

Annex 1 – Changes to the current national standards

Changes to triage questions

The “Nature of Call” system introduces three standardised pre-triage questions to increase the early recognition of cardiac arrest. Based on London Ambulance Service figures obtained by Sheffield University, it has been estimated that up to 250 additional lives will be saved in England every year.

Changes to clinical standards

To ensure the ARP changes drive improved clinical outcomes, we will be introducing a new set of clinical indicators.

For serious **heart attack** patients, who have specific ECG changes, we will measure the proportion of patients that receive definitive treatment (balloon inflation during angioplasty at a specialist heart attack centre) within 150 minutes of making a 999 call. We expect 90% of patients to meet this standard by 2022.

For **stroke patients**, we will measure the proportion of patients that complete their pathway of care (thrombolysis where appropriate, or first CT scan for those where it is not) within 180 minutes of making a 999 call – again with an expectation that 90% of patients will meet this standard by 2022, up from an estimated 75% of stroke patients currently completing their pathway of care within that timeframe.

Changes to dispatch practices, call categorisation and clock start/stop definitions

A comparison of the current operational standards and new operational standards is shown below.

Current standards

Category	Percentage of calls in this category	National Standard	How long does the ambulance service have to make a decision?	What stops the clock?
Red 1	3%	75% within 8 minutes	The clock starts at the point the call is connected to the ambulance service	The first ambulance service-dispatched emergency responder arriving at the scene of the incident
Red 2	47%	75% within 8 minutes	The earliest of: <ul style="list-style-type: none">•The problem being identified•An ambulance being dispatched•60 seconds from the call being connected	The first ambulance service-dispatched emergency responder arriving at the scene of the incident
Green	50%	No national standard	The earliest of: <ul style="list-style-type: none">•The problem being identified•An ambulance response being dispatched•60 seconds from the call being connected	The first ambulance service-dispatched emergency responder arriving at the scene of the incident

Proposed standards

Category	Percentage of calls in this category	National Standard	How long does the ambulance service have to make a decision?	What stops the clock?
Category 1	8%	7 minutes mean response time 15 minutes 90 th centile response time	The earliest of: •The problem being identified •An ambulance response being dispatched •30 seconds from the call being connected	The first ambulance service-dispatched emergency responder arriving 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 40 minutes 90 th centile response time	The earliest of: •The problem being identified •An ambulance response being dispatched •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 arriving at the scene of the incident stops the clock.
Category 3	34%	120 minutes 90 th centile response time	The earliest of: •The problem being identified •An ambulance response being dispatched •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 arriving at the scene of the incident stops the clock.
Category 4	10%	180 minutes 90 th centile response time	The earliest of: •The problem being identified •An ambulance response being dispatched •240 seconds from the call being connected	Category 4T: If a patient is transported by an emergency vehicle, only the arrival of the transporting vehicle stops the clock.