Response time as a sole performance indicator in EMS: Pitfalls and solutions

Sultan Zayed Khalifah Al-Shaqsi
Preventive and Social Medicine Department, Dunedin School of Medicine, University of Otago, Dunedin, Otago, New Zealand

Abstract: Response time (RT) has been liberally used as a process indicator to assess emergency medical services (EMS) performance around the world. It is objective, quantifiable, and easily understood by the public and policymakers alike. However, its correlation to better patient outcome is yet to be established. The evidence supporting the value of using RT is mostly lacking and the little existing evidence is conflicting. There is a concern that the RT notion is dominating the professional culture of ambulance providers. Quality of patient care is becoming a secondary target to RT. Solutions to shorten RT of ambulance services have prohibitive costs and risk the safety of patients, attending crew and the public. It is time to consider patient outcome as the main standard for performance of EMS in order to meet the public expectation of accountability and openness.

Keywords: performance, ambulance, EMS, response time, outcome

Introduction

In today’s health care environment the need for objective assessment and evaluation of the performance of health care agencies has created a demand for data-driven appraisal process. The performance evaluation of emergency medical services (EMS) is believed by O’Leary to be similar to mainstream medicine in being both a science and an art. The value of performance analysis and continuous quality assurance is beneficial internally to the organization and externally to the stakeholders of the system. Internally, it is critical for identifying areas of deficiencies and targets for improvements and continuous medical education. Externally, it is considered as a tool to demonstrate the value of EMS to stakeholders and implies accountability and transparency of the system to the public and funders. Performance evaluation aims to identify expectations of receivers of the services and ways to meet these expectations. In a recent study in Australia, performance evaluation has been rated as a top priority of EMS system activities among a panel of experts. This reflects the importance of EMS performance evaluation to the public, the policymakers, and the system itself in order to provide quality health care services.

The task of evaluating EMS and pre-hospital emergency care is difficult due to the multitude and complexity of variables in uncontrolled environments where EMS systems usually operate. Traditionally, EMS performance measurements are divided into three categories: input measures, process measures, and outcome measures. Input measures are metrics that gauge the capabilities of the system such as the number of ambulance vehicles per capita. Process measurements reflect the performance of procedures and protocols implemented by EMS system to reach the outcome. One well
known example of process measure of EMS performance is time delay from call to arrival widely known as response time (RT). Finally, outcome measures assess outcomes of the EMS operations for instance mortality rates of cardiac arrests attended by EMS services.

Rapid response of the EMS to a life-threatening event is a strong public expectation that EMS system strives to meet. Therefore, many EMS systems evaluate their performance according to how well they are able to meet this expectation in the form of their ability to respond promptly to EMS calls. RT is the time elapsed from the gathering of sufficient information about the event to the arrival of health care providers at the scene of the event. This elapsed time is used widely to proxy the performance of EMS. However, in recent years the validity and usefulness of RT as a performance indicator of EMS function has been questioned. This article discusses the major pitfalls and drawbacks of using RT in pre-hospital care performance assessment. First, there is a lack of evidence supporting the validity of RT as a performance indicator. Second, using RT gives an undeserved sense of high performance and dominates over the patient quality care. Third, aiming to meet RT standards induces financial and safety costs. Finally the article will points out some solutions to include outcome indicators in EMS performance evaluation.

**Conflicting evidence and immature generalization**

The ultimate goal of any EMS system is to improve the outcome of patients. However, patient outcomes such as death, disability, and discomfort are difficult to quantify and diverse in nature, making measurement of outcomes impractical in the short time frame of EMS operations. The concept of RT is based on the coverage concept which is the ability of the EMS system to cover and response to emergency calls. The notion of RT as a performance measure for EMS is attractive. It is easily quantifiable, objective and easily understood by the public and policy makers. The history of RT notion dated back to 1979, in a study from Seattle Ambulance that concluded the survival of patients suffering from nontraumatic cardiac arrest improved when basic life support (BLS) and advanced life support were administered within four and eight minutes from the event respectively. A couple of subsequent studies showed similar timeframes. After these initial studies, there was a rush from EMS systems around the world to adopt the attractive easily-understood notion of RT in all EMS missions. In few years, the eight-minute target became the rigid ultimate standard of EMS operations. In extreme cases, ambulance services have been penalized for not meeting the standard of eight minutes.

Since the early studies, there are many subsequent studies with conflicting results and many have challenged the validity of the eight-minute RT standard. Their argument is that the Seattle study and subsequent supportive studies might not be applicable to the current resuscitation protocols. These early studies (ie, Seattle Study and supportive studies) were conducted during a time when defibrillation could only have been performed by a trained paramedic and chest compressions received less emphasis during cardiopulmonary resuscitation (CPR). Current protocols allow lay people to use defibrillators. Chest compressions are now strongly emphasized during CPR. Thus, the validity of the eight-minute limit in cardiac arrest resuscitation is questionable, let alone its validity in other cases attended by EMS. Another argument is that the original Seattle study examined nontraumatic cardiac arrest, which represents only 1%–2% of ambulance missions in most EMS systems. Therefore, presenting the performance of 2% of EMS missions to the public as if it is the indicator for the entire system ignores 98% of EMS missions. It can still be argued that 2% of cardiac arrest cases are the most “save-able lives” and thus it is appropriate to use RT as a proxy for good EMS performance. That is true only in cardiac arrest cases because a good EMS performance in dealing with cardiac arrest does not guarantee good performance in other non-cardiac arrest cases such as trauma, burns, and strokes.

Recent studies show the lack of conclusive evidence supporting that shorter RT correlates to better outcomes in cardiac arrest, trauma calls, and general emergency life-threatening calls. Therefore, RT relationship to outcome even in cardiac arrest cases has been questioned. A Swiss study found that cardiac arrest patients defibrillated in hospital an average of 15.6 minutes after arrest were more likely to survive to hospital discharge, be alive one year following, and suffer less long-term neurological impairment than those defibrillated in the community an average of 5.7 minutes after arrest. Ergo, advocating reducing RT in order to improve outcome is still “speculative and unsupported.” Until evidence is robustly established, it is immature to generalize the value of RT of cardiac arrest as a performance indicator for all ambulance missions. An EMS RT as a sole performance indicator is with little value. It is time to incorporate other outcome-focused performance indicators along with RT in order to present a full picture of EMS performance to the public.
Which is more important: Vehicle’s speed or patient’s care?

The notion of RT is largely a public expectation-driven concept rather than evidence-based. The most common definition of the RT used by ambulance services is: “the time elapsed from the reception of sufficient information about the event to the arrival of first responding vehicle at the scene.” This definition does not reflect anything related to the patient’s outcome rather it reflects the physical features of the responding vehicle and the proximity of event site. The RT, in its current use, is just one of many time intervals in the chain of response to an event. Figure 1 shows the full spectrum of RT components.

Every time interval is equally essential for the overall performance of EMS. Thus, singling out one time interval as the indicator for the whole performance is not appropriate. Furthermore, the current use of RT as a sole performance indicator ignores the center of the paramedical profession; the patient. From a patient perspective, RT should be the “time elapsed from the call for help till the actual contact with the patient at the scene occurs”. The later definition is more inclusive of critical steps in EMS response such as call processing and access to patient, which can influence the overall outcome of the patient. This is particularly relevant when access to the patient is delayed as in urban high-rise areas and trapped patients. Table 1 illustrates a hypothetical example where RT does not represents patient’s outcome.

The strict implementation of rigid RT standards in most EMS systems led to serious concerns that RT notion is dominating the profession of pre-hospital care. A recent study from the United Kingdom showed that paramedics are more worried about the speed which they respond to patients rather than the quality of care they provide. This is a serious concern that will eventually lead to degradation in the pre-hospital care professionalism and morale.

The objectivity of RT is debatable as many paramedics believe that RT can be easily manipulated and fiddled to meet policymaker requirements. Another point is that, RT can easily mask serious suboptimal functions in EMS such as call processing and triage. For example, an EMS system that has poor call handling and triage will respond promptly to all cases regardless of the nature of the case to hide the poor triage. Hence, deep understanding of the value of RT as a sole indicator is required and the public has to be informed about the limited value of RT.

Safety and cost

Shorter RTs are not without financial and safety costs. Shorter RTs–especially if married with inappropriate triage, sirens, and lights in EMS–has well established ambulance crew and public safety risks. There is an increasing occurrence of ambulance crashes. Demanding shorter RTs is a strong contributing factor in many of ambulance crashes. Paramedics from the UK expressed a deep concern about the detrimental impact of
the solutions used to shorten the RTs of ambulance missions in their professional and social life. Demand shorter RTs increases the risk of ambulance crashes and exposes pedestrians and road users to increased risk of preventable crashes.

Shortening RTs is achieved by: 1) patients getting sick next to an ambulance station; 2) increasing the number of ambulances and paramedics; and 3) deploying ambulances to high demand areas in anticipation of an event happening. The first option is impractical as emergencies by definition are unpredictable. Increasing the number of ambulances and staff imposes financial costs. Reducing the RT by increasing the number of ambulance vehicles to meet the standard of eight minutes is estimated to cost US$400,000–500,000 per 24 hours of staffed ambulances. This is a prohibitive cost for most EMS providers around the world. The third option has attracted a great attention in recent years but current evidence regarding the cost-effectiveness of deploying ambulances in anticipation of events is lacking. In addition, standby deployment in anticipation for emergencies has safety risks especially if ambulance crew is deployed to high crime and violence sites. Nevertheless, there is some promising evidence that using computerized modeling to deploy ambulance vehicles to emergency calls maximizes survival rates. The concept of deployment modeling extends beyond the RT and takes a wide matrix of variables into account which increase its value and complexity in the same time.

What is the solution?
Many solutions have been proposed to create a more robust measure of EMS performance. The answer can not be a single measure rather EMS performance should be evaluated using an extensive set of metrics measuring different aspects of input, process and outcome of EMS operations. Fitch has proposed five fundamental strategies to improve the use of RT in EMS performance evaluation.

The use of ambulance deployment modeling
Using this approach will allow the matching of resources and demand. This will reduce RT and enhance its value. This is especially relevant in high demand busy EMS systems. It helps planning and outmost utilization of resources.

Measure response time components
Any successful implementation of RT should emphasize the measurement of the full spectrum of response mission time components (Figure 1). This allows a more comprehensive picture of the overall RT including critical processes such as triage and access to patient site.

Proper triage and call handling
Proper triage and using a dispatch system will distinguish cases that require prompt, light and siren response, and optimize clinical outcome of the patient, and avoid unnecessary safety risks.

Harness new technology and innovation
New technology such as geological information systems (GIS) and automated vehicle location (AVL) can provide valuable tools to safely shorten RT. Software programs, such

---

**Table 1: An illustration**

**Scenario A**
A car crash with a victim trapped inside the car. The call taker sent the closest ambulance to the scene that did not have an extrication tools. The ambulance was on scene six minutes post-event but they did not have extrication tool. The tools arrived after five minutes and the victim was extricated and released immediately. Unfortunately, he suffered severe lower extremity bleeding that lead to shock and circulatory collapse en-route to the trauma centre.

Performance:
1. Response time: within the standard as ambulance arrived within six minutes.
2. Outcome: patient died due to delayed extrication.

**Scenario B**
A car crash with a victim trapped inside the car. The call taker elected to send an ambulance with extrication tool even though it was not the closest vehicle to the scene. The ambulance was on scene nine minutes post-event. It took another minute to use the extrication tool and release the trapped patient. The patient suffered severe shock and was vigorously resuscitated on arrival to the trauma center.

Performance:
1. Response time: below the standard as the ambulance arrived nine minutes post-event.
2. Outcome: patient survived to hospital admission as the proper ambulance was sent rather than the closest.

Implication: proper telephone triage might improve care while appearing to degrade performance when measured solely by response time.
as Advanced Medical Priority Dispatch System (AMPDS) used by dispatch centers can enhance the effectiveness of triage and shorten time of call processing and provide attending paramedics with discreet and specific details about the event.

**Accountability and inclusion of all possible performance indicators**

Response time should be regarded as one of a group of performance indicators to evaluate the performance of EMS. The set of performance indicators should include indicators from input, process and outcome of EMS system. Accountability and transparency of EMS agencies should be a priority and performance evaluation should be a regular activity.

**Conclusion**

Response time has been used extensively in ambulance performance. However, its value as a sole indicator is questionable. The evidence for its validity as a performance measure is lacking and the available evidence is conflicting. There is a serious concern that the notion of RT as the target of ambulance mission is dominating the professional culture among paramedics. Patient care might be compromised if RT is regarded as the standard of ambulance performance. Finally, there are financial and safety costs associated with shortening RT. The British Health Commission for Improvement summarizes the nature of current RT in ambulance services as being “too simplistic and narrow”. Perhaps it is time to move on to more patient-centered performance indicators in EMS. The 2007 Consortium of US Metropolitan Municipalities’ EMS Medical Directors is a great example of the move to develop evidence-based performance measures. The EMS system is complex and multifaceted. Similarly, the evaluation of EMS performance has to be multifarious with a matrix of indicators. Table 2 lists some indicators that are patient-centered and can be included in EMS performance evaluation.

**Disclosure**

SA wrote and prepared the manuscript. Associate professor Robin Gauld (University of Otago) and Dr David McBride (University of Otago) reviewed the manuscript. The author reports no conflicts of interest in this work.

**References**


32. Fitch J. Response times: myths, measurement and management. JEMS. 2005;30(9):47.