AMBULANCE DIVERSION AND EMERGENCY DEPARTMENT OFFLOAD DELAY: RESOURCE DOCUMENT FOR THE NATIONAL ASSOCIATION OF EMS PHYSICIANS POSITION STATEMENT


ABSTRACT

The emergency medical services (EMS) system is a component of a larger health care safety net and a key component of an integrated emergency health care system. EMS systems, and their patients, are significantly impacted by emergency department (ED) crowding. While protocols designed to limit ambulance diversion may be effective at limiting time on divert status, without correcting overall hospital throughput these protocols may have a negative effect on ED crowding and the EMS system. Ambulance offload delay, the time it takes to transfer a patient to an ED stretcher and for the ED staff to assume the responsibility of the care of the patient, may have more impact on ambulance turnaround time than ambulance diversion. EMS administrators and medical directors should work with hospital administrators, ED staff, and ED administrators to improve the overall efficiency of the system, focusing on the time it takes to get ambulances back into service, and therefore must monitor and address both ambulance diversions and ambulance offload delay. This paper is the resource document for the National Association of EMS Physicians position statement on ambulance diversion and ED offload time. Key words: ambulance; EMS; diversion; bypass; offload; delay

PREHOSPITAL EMERGENCY CARE 2011;15:555–561

INTRODUCTION

The emergency medical services (EMS) system is a component of a larger health care safety net and a key component of the emergency health care system. EMS providers render highly skilled care to patients who have acute illness and injury, with their efforts typically culminating in the transport of the patient to more definitive care at an acute health care facility, typically an emergency department (ED). The ability of EMS providers to respond to an acutely sick patient in a timely manner is dependent on many variables. One such variable that is often overlooked is the effect of ED volumes on the EMS system.

The fiscal burden of uncompensated care, along with other factors, has forced the unanticipated closures of some hospitals, especially in rural locations. EMS systems have adopted a practice of ambulance diversion, whereby ambulances are diverted to another ED in the community during periods of peak crowding. However, concerns that ambulance diversion may have untoward effects on patient care have led health care systems to adopt policies to limit periods of ambulance diversion.

While there may be benefits to limiting periods of ambulance diversion, without correcting the underlying problem of ED crowding, ED personnel may become overwhelmed when diversion is limited. This can prevent them from quickly assuming patient care from EMS personnel and can lead to increases in offload delay. Therefore, in order to maximize the ability of the EMS system to respond to the needs of the public in a timely manner, health care systems should focus on decreasing the ambulance offload delay, defined as the time it takes from arrival of the ambulance at the ED to transfer of the patient to a space in the ED and turnover of the care to the ED staff. Monitoring diversions and offload delay are important steps in assessing inefficiency in the system. Ultimately, EMS medical directors, ED directors, and hospital administrators should work together to identify factors that lead to decreased ED crowding and result in returning EMS crews back to service more efficiently.

Received April 15, 2011, from the Department of Emergency Medicine (DRC, HJW), SUNY Upstate Medical University, Syracuse, New York; the Department of Emergency Medicine (MGM), Johns Hopkins University School of Medicine, Baltimore, Maryland; the Department of Emergency Medicine (AC), Dalhousie University, Halifax, Nova Scotia, Canada; and the Department of Emergency Medicine (BJL, JVN), University of Maryland School of Medicine, Baltimore, Maryland. From the Standards and Clinical Practice Committee of the National Association of EMS Physicians, Olathe, Kansas. Revision received July 18, 2011; accepted for publication July 25, 2011.

The position statement based on this resource document appears in this issue on page 543.

The authors report no conflicts of interest.

Address correspondence and reprint requests to: Derek R. Cooney, MD, FF/NREMT-P, FACEP, Director, EMS & Disaster Medicine, 550 East Genesee Street/EMSTAT Center, Syracuse, NY 13202. e-mail: cooneyd@upstate.edu

555
This paper is the resource document for the National Association of EMS Physicians (NAEMSP) position statement on ambulance diversion and ED offload delay.3

**EMERGENCY DEPARTMENT CROWDING**

The problem of ED crowding has been described in the medical literature dating back to the mid- to late 1980s.4 In 1991, Andrulis et al. published a study on crowding in U.S. teaching hospitals, demonstrating that over a three-year period up to 75% of reporting hospitals showed increasing holding times for admitted patients, and increasing use of methods to decrease crowding.5 Articles published since that time have also shown increases in ED crowding.6–8

The study of ED crowding is quite complex and represents the potential for significant political ramifications. In a study of the causes of ambulance diversion, used as a surrogate marker for ED crowding, Schull et al. demonstrated that every admitted patient held in the ED added an additional 6 minutes of diversion time, a 3% increase over the mean.9 The findings of this study add credibility to the arguments made in an editorial by Derlet and Richards that lack of hospital beds, leading to ED boarding, is a significant cause of ED crowding.10 While there are many determinants of ED crowding, numerous authors have demonstrated that the root cause of ED crowding is overall hospital throughput and the inability to move admitted patients from the ED to hospital beds.11–16 Since the ability of an EMS unit to transfer a patient to an ED bed is determined by the availability of ED beds, which is determined by hospital throughput and availability of hospital beds, it follows that EMS unit availability is directly related to hospital throughput.

**AMBULANCE DIVERSION**

First described by Lagoe and Jastrzemski in 1990 as a novel approach to alleviating ED crowding in an urban environment, ED diversion is increasingly being utilized, in one form or another, by most busy emergency medical systems.17 While there are variations on its definition, essentially an ED that is placed on “diversion” is closed to incoming ambulances. This may mean diversion of all incoming traffic, or diversion of specific type(s) of patients (e.g., patients who will likely require cardiac monitoring, patients requiring psychiatric beds, or trauma patients) with concurrent acceptance of all other ambulance traffic.18 Exceptions to this rule can include patient demand and specialty services (trauma center, stroke center, percutaneous coronary intervention center, burn center). The original conceptual benefit of incorporating ambulance diversion into the system was a proposed decrease in mortality and morbidity as a result of a decrease in waiting times. Diversion was initially thought to be the logical solution to ED crowding, as this was thought to be a transient problem caused by periodic waves of patients who could overwhelm a single ED within a larger system. In fact, this may have been the case in a number of systems at the time Lagoe and Jastrzemski looked at this strategy as a system-based solution.

A later paper by Lagoe et al. in 2002 revealed diversion rates as high as 51% and a change from a seasonal pattern to a more fixed high rate of diversion during the observation period.19 This publication still cited a reduction in transports of 30–50% for an ED while on diversion. In 2003, Lagoe et al. published a paper acknowledging that diversion was a potential problem and demonstrated that a systemwide approach led to a 24.8–33.6% reduction in diversion hours during the study periods.20 The increase in ED utilization by the public has, in many ways, led to reexamination of the issue of ED crowding in general. As it is now clear that ED crowding is not an ED-specific problem, ambulance diversion is no longer considered the answer to what is now known to be a multifactorial problem.

A number of authors have raised concern over potential problems created by diversion. Some considerations include complex patients’ being transported to a different facility than where optimum care is typically delivered, increased turnaround times for EMS personal traveling outside normal service areas, and public perception of an institution’s turning away individuals requiring emergency care.21 While diversion is appealing in concept, actual implementation and experience with diversion have yielded conflicting research and brought the practice into question. Not only is the process being scrutinized by many different researchers and medical systems, but also the growing problem of ED crowding has clearly been shown to be much more complex than simply increasing numbers of patients presenting to the ED via ambulance.10,22

Research related to diversion has yielded seemingly conflicting results. A number of researchers have demonstrated some association between ambulance diversion and increased patient morbidity and mortality.23–25 However, a 2006 systematic review of the literature has suggested that no such association exists.21 In this review, Pham et al. found that diversion is associated with ED crowding, may be reducible through redesign or addition of resources, and in some cases may be associated with a small increase in patient transport and treatment times. Most significantly, the review showed that strategies involving ED management options, such as observation units and algorithms, yielded little reduction in diversion. Hospital-based throughput, laboratory, and staffing initiatives, on the other hand, were shown to decrease diversion as a consequence of reductions in ED crowding.21 Factors that may be predictive of ED ambulance diversion as outlined by Pham et al. are shown in Figure 1.
One diversion reduction study of particular note is a 2008 paper by Asamoah et al. that details reduction in diversion hours with a strict limitation policy. Probably more important than the primary finding of an 82% reduction in ambulance diversion is that “drop-off” time (i.e., offload delay) increased 32% as a side effect of the new strict diversion limitations on the system. The authors also cited other reports in the literature of long waits for EMS personnel attempting to transfer care of their patient during periods of ED crowding.

Some key publications related to ambulance diversion are noted in Table 1.

### AMBULANCE OFFLOAD DELAY

The term offload delay is best defined as corresponding to the “delivery interval” component of the EMS turnaround time as described by Spaite et al. Cone et al. further describe the hospital turnaround interval as a combination of six fixed events that follow the movement of the patient and EMS providers (Fig. 2), and a number of variable events that do not occur in the same sequence with every run but may still be required to get the ambulance back into service (notifications of unit status to dispatch, delivery of verbal and written reports to the ED staff, and completion of the ambulance call report). In all models, the delivery interval ends when the patient is placed on a stretcher or when verbal report ends, whichever is later and signifies complete transfer of care. The recovery interval ends either when the crew signals “available” or when the ambulance leaves the hospital; in some systems, the crews may be stationed at the hospital or may remain in the ED even after the unit is available for the next call.

<table>
<thead>
<tr>
<th>Study</th>
<th>Key Point(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lagoe and Jastremski, 1990&lt;sup&gt;17&lt;/sup&gt;</td>
<td>Described diversion as a novel approach to alleviating ED crowding.</td>
</tr>
<tr>
<td>Scheulen et al., 2001&lt;sup&gt;18&lt;/sup&gt;</td>
<td>Diversion had some limited effect on ED crowding in urban and suburban areas, and no effect in rural EDs; impact of alert policies may change over time.</td>
</tr>
<tr>
<td>Lagoe et al., 2002&lt;sup&gt;19&lt;/sup&gt;</td>
<td>Diversion was noted to have become common, and the citywide diversion rate was around 51%.</td>
</tr>
<tr>
<td>Lagoe et al., 2003&lt;sup&gt;20&lt;/sup&gt;</td>
<td>Showed that hospital administration response to diversion could lower diversion rate.</td>
</tr>
<tr>
<td>Pham et al., 2006&lt;sup&gt;21&lt;/sup&gt;</td>
<td>Diversion was associated with ED crowding and increase in patient transport and treatment times without significant increase in morbidity/mortality; the rate may be reducible through redesign or addition of resources.</td>
</tr>
<tr>
<td>Asamoah et al., 2008&lt;sup&gt;26&lt;/sup&gt;</td>
<td>A strict diversion policy led to a decrease (82%) in diversion but an increase (32%) in ambulance offload delay.</td>
</tr>
</tbody>
</table>

---

**TABLE 1. Key Publications Related to Ambulance Diversion**

ED = emergency department.
The results of papers on the topic of EMS turnaround intervals illustrate the importance of accurately studying this process. In the study by Cone et al. from 1998, the delivery time averaged 4 minutes 49 seconds. In 2006 study by Segal et al., the authors found intervals of 14 minutes before triage was complete, with 9 minutes spent waiting for triage. Carter et al. reported that in 2002 the 90th percentile for delivery interval was 24 minutes 37 seconds (median 12 minutes 45 seconds), and by 2007 it was 109 minutes 2 seconds (median 36 minutes 41 seconds). In New Haven in 2008, the delivery interval had a mean of 10 minutes 45 seconds and a 90th percentile of 19 minutes 21 seconds, while in Richmond the average was 17 minutes and 0 seconds and the 90th percentile was 35 minutes and 0 seconds. Breaking turnaround time down to specific processes, whereas the bulk of time used to be spent on charting, it is now spent waiting for an ED bed. In 2006, Segal et al. grouped together the times for all posttriage activities and found a mean time of 31 minutes 33 seconds (95% confidence interval [CI] 29 minutes 8 seconds–33 minutes 58 seconds). Eckstein and Chan showed a median of 27 minutes (interquartile range [IQR] 20–40) in their study from 2004; they described the longest single wait of 405 minutes (6.75 hours) in a study that was not isolated to large/urban/teaching hospitals.

The determinants of offload delay are much akin to the determinants of ED crowding. Schull et al. have shown the most significant of these to be the number of admitted patients being held in the ED, duration of ED holds, and duration of assessments. Derlet and Richards also outlined a number of key factors contributing to worsening ED crowding. Factors cited include increasing complexity, increasing volume overall, managed care (which paradoxically has increased ED volume in some areas by decreasing access/supply of primary care), lack of hospital beds for admitted patients, avoidance of admission by “intensive therapy” in the ED, delays in service by radiology, laboratory, and ancillary services, shortage of nursing staff, shortage of administrative/clerical support staff, shortage of on-call specialists, shortage of physical plant space in the ED, language and cultural barriers, increased medical record documentation requirements, and difficulty in arranging follow-up care. Because of these findings, Schull et al. state that the “search for solutions should be focused on increasing the institutional efficiency of the evaluation and disposition of complex patients, as...
opposed to encouraging ambulatory patients to seek care elsewhere."

Consequences that directly relate to offload delay can be broken down into two major headings: consequences to the patient and consequences to the EMS system. Patient-level consequences have not been well studied. Despite this fact, one might hypothesize that offload delay leads to delay to definitive care, poor pain control, delayed time to antibiotics, increased morbidity, and possibly even mortality. Ultimately, there is reasonable concern that ambulance offload delay will compromise patient safety.

Consequences to the EMS system are several. Offload delay relates directly to the system status plan and resource availability. This can affect response times and prolong time on task, resulting in decreased efficiency and the need for additional staffing.27 In larger service areas or in cases of mutual aid, this may result in units’ being drawn from outlying areas for coverage and/or moving a unit away from the crew’s home service area, potentially for the duration of their shift. This has public health implications and represents a potential decrease in surge capacity of the system. Editorial comments by Asplin on the topic state, “...part of optimizing emergency preparedness is providing assurance that the health care system has enough capacity to respond when called upon. Regrettably, it is dangerous to assume that capacity exists today.”23

Ambulance offload delay has been used as a proxy measure for crowding, much like diversion, and may also be a predictor of both. However, if offload delay is substituted for diversion as a marker for ED crowding, it is important to consider that decreasing offload delay directly, without improving throughput, also does not address ED crowding. Diversion may have an effect on ambulance resources by limiting offload delay in some circumstances, and the two are not independent variables.34 Because both may represent factors negatively impacting patient care and health care system efficiency, both should be monitored.

The Emergency Medical Treatment and Active Labor Act (EMTALA) gives some clarity as to who is responsible for the patient on the stretcher once on hospital grounds, and provides for penalties that may have political implications that could affect hospital response to offload delay. Under EMTALA, a patient must be granted a medical screening examination (MSE) upon presentation to a hospital with a request for an examination or treatment of an emergency medical condition.35 In response to reports of a lack of hospital attention to offload delay issues, the Centers for Medicare & Medicaid Services (CMS) issued the following statement: “The specific concern was that hospital ED staff deliberately delay the transfer of individuals from the EMS provider’s stretcher to an ED bed with the impression that the ED staff is relieved of their EMTALA obligation by doing so. This practice constitutes a potential violation of EMTALA.”36 The CMS have also stated, “A hospital’s refusal to accept responsibility for a patient who arrives via EMS...could be a violation of EMTALA.”37

However, despite the initial stance by CMS that strongly came out against offload delays, subsequent statements have introduced some confusion. In subsequent exchanges, CMS has stated, “...It was not the intent of the guidance in the Letters that there should be enforcement action against any hospital when the delay in the immediate provision of an appropriate MSE and/or stabilizing treatment is due to circumstances beyond the hospital’s control (e.g., the hospital does not have the capacity or capability at the time of presentation)...it could under those circumstances be reasonable for the hospital to ask the EMS provider to stay with the individual until such time as there were ED staff available to provide care to that individual.”36

However, EMTALA does require that even if the hospital cannot immediately provide an MSE, it must still triage the individual’s condition immediately upon arrival to ensure that an emergent intervention is not required and that the EMS provider staff can appropriately monitor the individual’s condition. This act clearly states that the patients are the hospital’s responsibility once the ambulance arrives on the grounds and further mandates a screening examination.

EMS and the Emergency Health Care System

Hospitals, their EDs, and EMS agencies all represent components of a larger emergency health care system. During response to emergencies, the efficiency and effectiveness of each component may drastically affect the others. While ambulance diversion may help one hospital deal with its ED crowding, the unintended effect may be to crowd another ED. Multiple medical communities have considered limitations on, or elimination of, ambulance diversion. One example of such a plan is detailed by the report of how Sacramento was able to drastically reduce ambulance diversion when a regional plan was implemented.38 Such programs, however, may result in longer offload times if other factors causing ED crowding are not addressed.25 Ambulance diversion, offload times, and factors leading to ED crowding must be addressed in a systemwide manner. EMS leadership must work with colleagues throughout the health care system to improve throughput on all levels, which should ultimately result in decreases in ambulance diversion and offload delay.

Future Research

It is clear that EMS systems do not exist in a vacuum isolated from the rest of the health care system. In
addition, it is also clear that the most efficient way to ensure that the EMS system is able to respond to the emergent needs of the public is to maximize hospital throughput strategies. However, there is still much that is not known about the relationships of hospital crowding, ED crowding, ambulance offload time, and EMS response times. Therefore, the NAEMSP endorses the need for further research in the relationships of hospitals, EDs, and EMS systems.

CONCLUSIONS

In the development of health care systems, it is important to recognize that no one element of the system exists in isolation of the rest of the system. All elements of a health care system must work together to ensure overall capacity to adequately respond to the emergency needs of the public. The NAEMSP position statement on ambulance diversion and ED offload delay reflects a number of principles that address the core elements of this complex problem. Restructuring hospital and ED processes and resource allocation toward improving hospital throughput and decreasing ED crowding will likely have the greatest impact on diversion and offload delays. In addition, surveillance of offload time and diversion is an important system status tool. Reliable data on prolonged delivery time, offload delay, and the impact on EMS systems can be used to leverage hospitals to improve ED throughput by reorganizing or committing additional resources. EMS and hospital system agreement on criteria for ED ambulance diversion is also important. Limiting offload delay may greatly impact the outcome for ill and traumatized patients and, therefore, benchmarks for offload delay should be adopted. Communication of system resource availability, as well as factors relating to offload delay, should also be integrated into every EMS system. Regular review and analysis of these benchmarks should be used to help guide the development of quality improvement programs and initiatives. Ultimately, the solution to ED diversion, ambulance offload delays, and overall hospital throughput is a matter of all entities in a given health care system working together in an integrated manner for the overall good of the public health.

References