SAFE: A Comprehensive Data Visualization System

Mehmet Basdere, Gabriel Caniglia, Charles Collar, Christian Rozolis
Northwestern University, Evanston, Illinois
{mehmetbasdere2016@u.northwestern.edu, gcan@u.northwestern.edu,
CharlesCollar2020@u.northwestern.edu, ChristianRozolis2017@u.northwestern.edu}

George Chiampas
Northwestern University, Bank of America Chicago Marathon, Chicago, Illinois
chiampas@cemevent.com

Michael Nishi
Bank of America Chicago Marathon, Chicago, Illinois
mike.nishi@cemevent.com

Karen Smilowitz
Northwestern University, Evanston, Illinois
ksmilowitz@northwestern.edu

This paper describes Situational Awareness for Events (SAFE), a comprehensive data visualization system for mass-participation endurance events. Working in partnership with the Bank of America Chicago Marathon and Chevron Houston Marathon, and their public safety partners, we developed SAFE to enhance logistics, medical preparedness, and response for mass-participation endurance events such as marathons. The system incorporates critical data into a user-friendly dashboard to serve as a centralized source of information during the events. SAFE uses historical and real-time data to provide pre-event and on-site analytics via descriptive, predictive, and prescriptive models. These models help race organizers and relevant stakeholders effectively manage and oversee all participants, monitor the dynamic location of race participants, and manage health and safety resources throughout the event. The system was deployed successfully at the 2014-2018 Chicago Marathon, 2014-2018 Shamrock Shuffle, and 2016-2018 Houston Marathon.

Key words: data analytics; simulation; situational awareness

History: This paper has been refereed.

Introduction

In 2013, a collective group of Northwestern University researchers and Bank of America Chicago Marathon (Chicago Marathon) organizers asked the following questions: 'Why is it important for event managers and related public agencies to visualize a marathon?' and 'What does it mean to visualize a marathon?' Initially, the answers to these questions focused on the need for situational awareness on race day, for example, knowing where the runners are on the course and understanding the demands of medical facilities. Over time, the value of visualizing a marathon was viewed more broadly, shifting from only a situational awareness tool to a comprehensive system that assists organizers in making decisions during the planning and implementation phases of the event. For a number of years, race organizers have employed real-time athlete tracking through radio-frequency chips and limited tracking of medical emergencies. However, existing race-management software that utilizes these data have been far from fully successful in visualizing a marathon.

This paper describes Situational Awareness for Events (SAFE), a comprehensive data visualization system for mass-participation endurance events. Working in partnership with the Chicago Marathon and the Chevron Houston Marathon, and their public safety partners, SAFE was devised to enhance marathon logistics and supply, medical preparedness, and responses by assisting organizers during the planning and implementation phases of an event. The system achieves this by (1) centralizing race-related information in one dashboard, and (2) providing pre-event and on-site descriptive, predictive, and prescriptive models. During the event, the system incorporates critical real-time data into a user-friendly dashboard with supporting features to provide a centralized source of information for situational awareness. The information presented is both an aggregation and summary of existing data streams in one environment and products of descriptive and predictive models developed by the SAFE team to gain a better understanding of the current situation. Centralized information reduces time-consuming interactions among participating agencies and provides a clear and uninterrupted flow of information. As we describe in this paper, the analytics tools developed for event-day situational awareness are also used in a prescriptive setting prior to the event to develop action plans for different scenarios and plan operations, such as start-line scheduling and medical-volunteer allocation.

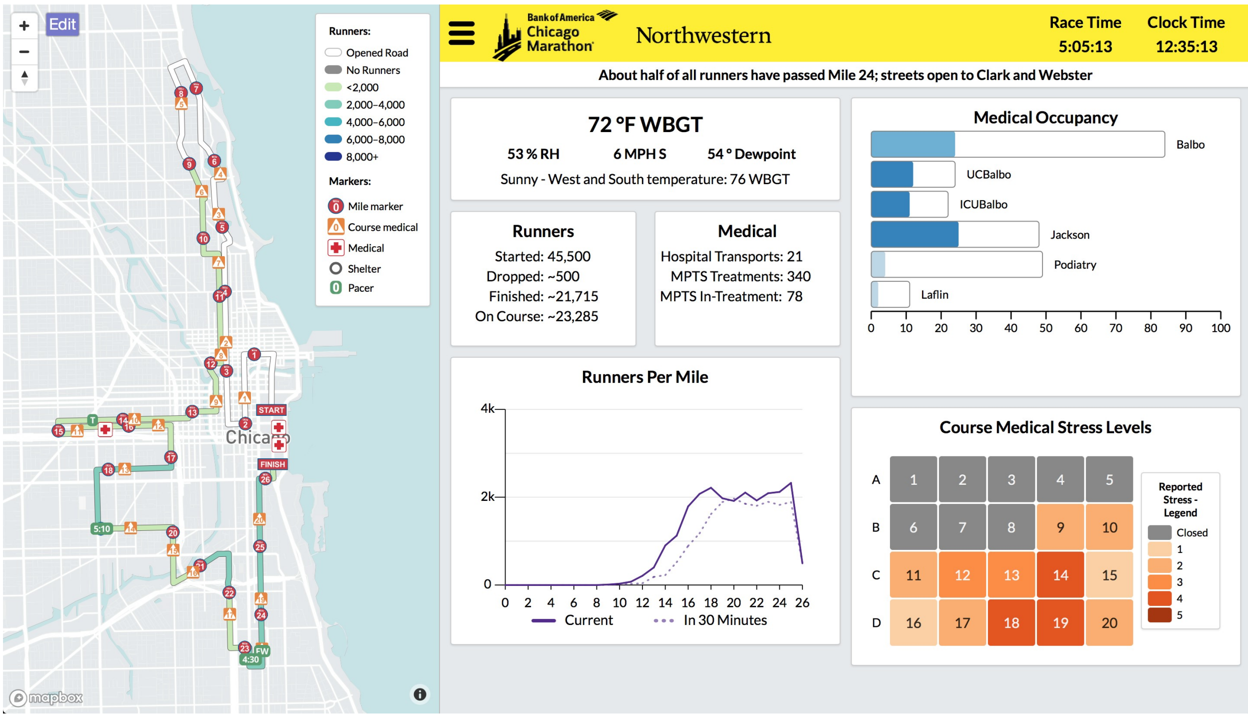
SAFE uses historical data to design descriptive and predictive algorithms, integrates these algorithms with real-time dynamic data from tracking devices, and presents the information on a variety of devices, including several large-screen displays in incident command facilities on a course and remotely in Chicago's Office of Emergency Management. The critical data streams incorporated into the system include weather, clock and race time, alert messages, medical-care statistics, runner tracking, and simulated runner density. Predicting runner location and density information is key because most race-related plans and decisions are made based on the location of the runners.

Figure 1 displays a screenshot of the desktop display for monitoring via personal device (e.g., laptop, tablet) during deployment at the 2017 Chicago Marathon. The system was deployed successfully at the 2014-2018 Chicago Marathon, the 2014-2018 Bank of America Shamrock Shuffle (8K), and the 2016-2018 Chevron Houston Marathon and Aramco Houston Half Marathon. The earliest deployments of the system at the Chicago Marathon and Shamrock Shuffle are described in detail in Hanken et al. (2016). The system has also been featured in Chicago Tribune, Runners' World, and Newsweek (Elahi 2015, Kuzma 2016, Gorman 2016).

The system has become integral to race operations. Wade Morehead, Executive Director of Houston Marathon Committee, states that 'The SAFE data visualization system has proven to be an invaluable resource for managing large endurance events. To receive data in real time which clearly depicts where all of our participants are on the course allows us to know exactly where our customers are should an incident occur at any time during the six-hour event. This is very important for our public safety planning and preparedness. But SAFE is also valuable for managing human and physical resources and assets. The data have allowed us to track the effectiveness of egress metering techniques implemented at the start line to reduce course crowdedness. This improves the participant experience and also assists us with volunteer staffing, water station preparation and runner post-race services, to name a few" (Morehead 2018).

Houston's medical director John Cianca echoes these comments from a medical perspective: "... SAFE data visualization system has been a functional part of my medical operations plan for three years. ... It gives me tracking information on my emergency medical services assets and bike medics and it keeps me updated on bed availability in the various tents we use to care for participants. ... Having all this information on one screen is invaluable to me as I manage the medical care and public safety for the event. Additionally, if there were ever a need to evacuate the course this system would drive the process of safely and efficiently moving people and assets to safe locales. In total this is a brilliantly devised and highly practical resource for event management" (Cianca 2018).

Figure 1. The Figure Displays a Screenshot of the Central SAFE Dashboard Taken at the 2017 Chicago Marathon



In this paper, we describe the key features of SAFE and how the system evolved from a simpler situational awareness tool into a more comprehensive data analytics tool for mass endurance events. We organized the remainder of the paper as follows: the *Background* section describes marathon and marathon planning operations in detail and presents the initial situational awareness tool. The *Development of Safe* section provides the details on enhancements that transformed the initial tool to a more comprehensive system. In the *Conclusions and Future Work* section, we conclude the paper and discuss potential improvements.

Background

Marathons and Marathon Planning

A marathon is a long-distance running event with an official distance of 26.22 miles. The annual running of a marathon is a test of preparedness for a city and involves a host of significant planning and implementation challenges, particularly in heavily populated areas. Given the scale of participation and the diversity of entities involved, planning for an event often begins immediately after completion of the prior year's event. Situational awareness has proven to be critical at a range of events from course rerouting because of a gas leak at the 2008 London Marathon to the tragic incident at the 2013 Boston Marathon. Table 1 provides a brief summary of incidents in recent marathons, highlighting the extreme challenges from a public safety perspective.

Marathon planning and implementation have been roughly categorized as a combination of (1) course design, (2) communication protocols and systems, and (3) resource management (Chiampas and Jaworski 2009). Specifically, course design focuses on the route to be followed during the race and the locations of runner services. Designing the course and service locations relies on accurate information regarding the flow of participants through the course on race day. Successful event management relies on effective communication channels among the participating agents, such as race organizers, medical team members, volunteers, and public agencies. Such channels must have the ability to rapidly and clearly disseminate race-related information. Finally, resource management focuses on estimating and securing necessary amounts of medical and other race-related supplies before the race, and efficiently redistributing resources depending on the progression of the event.

Table 1. The Table Lists Various Incidents in Recent Marathons

Event	Year	Incident
Quad Cities Marathon	2005	Runners were forced to wait for passing freight trains during the race (The Washington Post 2005).
Chicago Marathon	2007	Extreme heat forced cancellation of the event during the race (Chiampas and Jaworski 2009).
London Marathon	2008	The course was rerouted during the race because of a potential gas leak near mile mark 13 (BBC News UK 2008).
Pittsburgh Marathon	2010	The race was delayed briefly due to a bomb scare near the finish line (The Associated Press 2010).
New York City Marathon	2012	The marathon was cancelled because of Hurricane Sandy (Belson 2012).
Shamrock Shuffle	2013	The course was rerouted prior to the event because a man was threatening to jump off a bridge on the course (Basdere et al. 2014).
Boston Marathon	2013	Two homemade bombs were detonated near the finish line, causing three deaths and several hundred injuries (CBS / The Associated Press 2013).
Via Marathon	2016	Runners were forced to wait for passing freight trains during the race (Radzievich 2016).
Klang City International Marathon	2017	A marathon pacer was hit by a car during the race (Channel NewsAsia 2018).

The Chicago Model

In 2007, extreme heat forced the cancellation of the Chicago Marathon in the middle of the race, highlighting the challenges of event planning and management in the face of adversity. In response, the organizers of the Chicago Marathon developed a holistic approach to mass-participation event planning and management, referred to as the Chicago Model, which has become a standard for such events (Chiampas and Jaworski 2009, McCarthy et al. 2011). The Chicago Model brings together all major organizations (e.g., race organizers, fire and police departments, emergency management, the Red Cross) to coordinate preparation and response for the event and the surrounding areas impacted by the event, creating a collaborative environment in which all participating agencies share a common mental model of responsibilities and information requirements. This environment enables race organizers to coordinate and more efficiently operate the underlying events and allows them to facilitate faster and more flexible responses to rapid changes. The Chicago Model is based on three key components:

organizational structure, information systems, and communication (Chiampas and Jaworski 2009). The organizational structure of this model is based on the Incident Command System from the National Incident Management System, which is used to manage incidents and large-scale disasters (Federal Emergency Management Agency 2015). To further enhance this organizational structure, the Chicago Model ensures that all agencies are in the same physical environment, referred to as Forward Command, which serves as headquarters during the execution of the event. Information systems play a key role in the model's success because dynamic decision making is not possible without the clear and timely flow of information. Information, which comes from various data streams throughout the event, is communicated based on predetermined communication channels. A comprehensive medical tracking system to monitor medical coverage in real time complements the integrated organizational structure; see Basdere et al. (2014) for a detailed discussion of the Chicago Model and its use in responding to an acute event during the 2013 BASS. In a related study, Ross et al. (2015) examine the potential value of incorporating real-time patient tracking systems into the Chicago Model.

As noted by George Chiampas, Medical Director of the Chicago Marathon: "The Chicago Marathon has emerged in the last decade as the leader in participant safety, operations and logistics of large-scale mass events by establishing clear communication pathways, a structured command center and key engagement of private and public stakeholders, all while sharing best practices across the industry. The inclusion of innovation and technology to capture multiple streams of key decision-making information has without question transformed the space. Public safety officials as well as event organizers have been captivated by the utility as well as now find integral to the successful management of large scale and complex events" (Chiampas 2018).

Initial Situational Awareness Tool

Initial system development started in 2013 with a series of brainstorming sessions conducted with Northwestern University researchers and the Chicago Marathon organizers to explore various aspects of data visualization in a marathon setting and establish a common vision for data visualization. An essential feature was the ability to visualize runner density over time and space. In the event of race disruptions, runners can be held at the aid stations distributed along the course, which are equipped with supplies and medical volunteers, and connected to the Forward Command through a team of ham (amateur) radio operators. Therefore, it is critical for race organizers and public officials to know the progression of runners along the course at all times. The initial deployment of the system took place at the 2014 Shamrock Shuffle; deployment at the 2014 Chicago Marathon followed. Hanken et al. (2016) provide a detailed discussion of the initial system. These deployments provided opportunities to collect feedback on user experiences and identify key principles for a successful data visualization system.

Although the early deployments were successful in displaying essential data streams, the system had various limitations. The most pressing limitation was the way in which it represented runner density. A third-party provider operates timing mats at 5K intervals along the course. These timing mats record the cumulative count of runners as they pass each 5K mark. However, race organizers sought a more granular representation of runners to predict flow to the aid stations, which are roughly distributed every mile. This was done initially through simple interpolation, which uniformly distributed the runners within two timing mats. For example, if two consecutive timing mats recorded 20,000 and 17,000 runners, the difference of 3,000 runners was assumed to be uniformly distributed within the 5K-course segment between those timing mats. However, this crude interpolation failed to represent the runner locations with sufficient accuracy.

We made significant changes to the system back end to increase efficiency in processing and storing race data. Early versions of the system did not use a structured database and the data were stored and collected by using comma separated files (CSVs), which caused reliability issues and significantly slowed the overall system. Without a structured database, real-time race data were not archived effectively, preventing potential post-race analysis. Furthermore, the system relied on manual data entry and manual updates, resulting in interruptions and delays.

Finally, a single-page dashboard display was not sufficient to provide detailed information for all participating agencies. Some agencies, depending on their responsibilities, needed to visualize different data streams in detail throughout the race. Providing such information in a dashboard display was not achievable in the limited space of a single page, motivating the change to a modular structure with multiple pages.

Development of SAFE

In this section, we discuss the innovative enhancements that transformed the initial situational awareness tool into a comprehensive system, which provides descriptive, predictive, and prescriptive analytics. We divide the discussion into three categories: (1) runner tracking simulation, (2) back end improvements, and (3) interface updates.

Runner Tracking Simulation

The most important and innovative enhancement to the original system has been the development of a runner tracking simulation tool to provide a better representation of participant distribution over the course. The predictive and prescriptive capabilities of the enhanced simulation tool allow organizers to better plan and manage the event. These capabilities are employed at different stages of marathon planning and implementation to help with critical decisions, such as medical staffing at the aid stations, allocation of resources throughout the course, and establishment of evacuation routes in case of event cancellation. The simulation predicts the spatio-temporal density of runners along the course using past and current runner and event data. As the race evolves, the simulation responds dynamically as information inputs are updated.

During the implementation phase, the simulation accurately predicts the current and future locations of the runners. With current runner density information, each aid station is well-informed about the number of runners who have already passed the station and the number yet to arrive. This information enables staff at an aid station to more accurately evaluate its resource requirements. With future runner density information, race organizers are able to notify the aid stations and medical staff before they experience a surge of runners; therefore, they can preemptively allocate necessary resources. This information also allows city agencies to plan street reopenings to allow traffic to again flow through the streets that had been closed for the course.

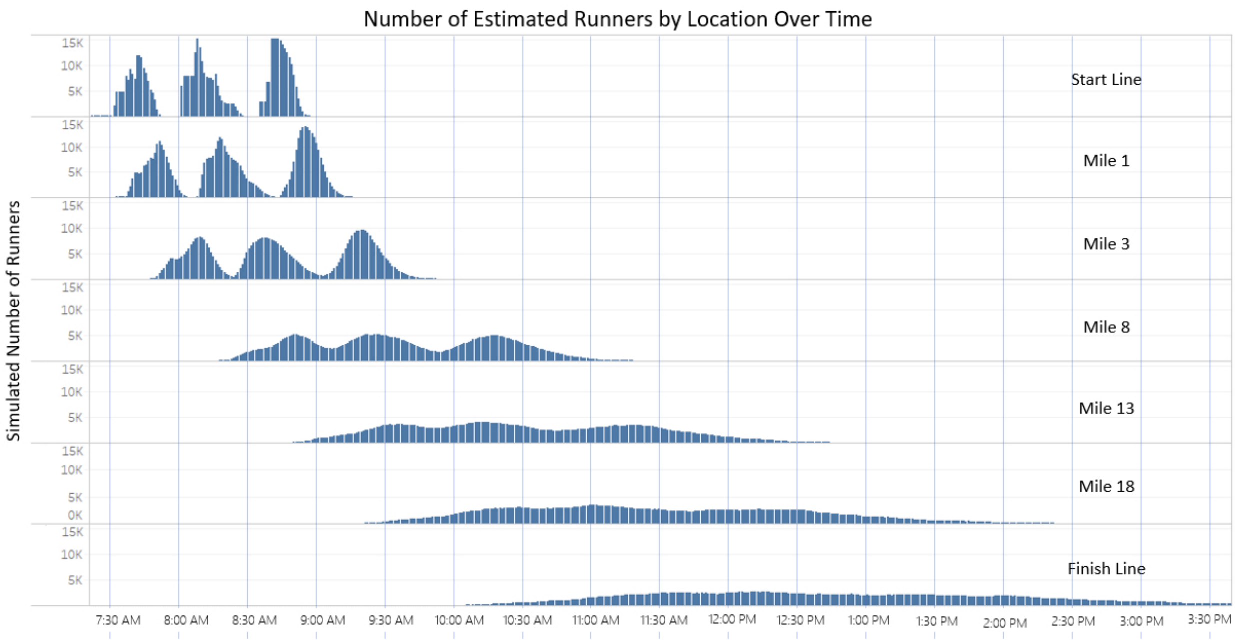
The simulation is also critical for planning prior to the race. For example, this interactive tool is capable of performing start-line analysis to determine the best start patterns and schedule for the corrals. Corrals are preassigned groups of runners based on previous (or estimated) individual runner times. Intelligently scheduling a race start plan is critical to maintaining safe and efficient levels of runner densities throughout the course. Large races use controlled starting waves, via staggered starts by corral, to manage flow through the course. From this perspective, the runner tracking simulation is a unique and effective planning tool, which is capable of simulating the entire race under various settings and assisting organizers in developing new action plans and policies. Figure 2 shows the distribution of runners over time, given the corral start times for the 2017 Chicago Marathon. In the graph, the x-axis represents the time from the 7:30 a.m. race start to 3:30 p.m. and the y-axis represents the simulated number of runners at different points in the race (e.g., the start line, miles 1, 3, 8, 13, and 18, and the finish line). Farina et al. (1989) perform a simplified version of this analysis to reduce crowding at the finish line at the annual Bolder Boulder 10K race. In their setting, the total race time of a runner is randomly generated from a race-time distribution estimated based on data from previous years.

The runner density simulation is driven by runner speed predictions. The predictions are based on multiple linear regression models with features pertaining to runner groupings, runner progression, and course conditions. The core output is a minute-by-minute distribution of runners across the 26-mile markers, which are displayed on the live SAFE dashboard. The speed regression is continuously updated and retrained on relevant race data for each event.

Simulation Inputs

The simulation is based on three key input streams: (1) corral groupings, (2) runner progression, and (3) course conditions.

Figure 2. The Graph Shows Simulation Results for Runner Distribution Based On Registered Participants for the 2017 Chicago Marathon at Seven Locations on the Course over a Period of Eight Hours

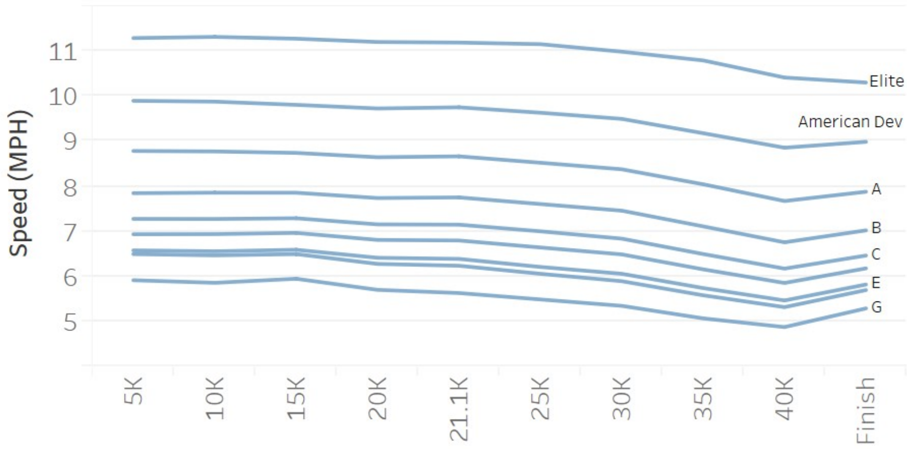


Corral groupings: In many long-distance events, runners are preassigned to corrals based on qualifying times and (or) anticipated marathon pace. Each corral is assigned a letter; earlier letters (in the alphabet) represent faster runner groups. Chicago has two additional corrals: *Elite* represents professional marathon runners and the *American Development Program* represents the fastest amateur runners with specific qualification times. Exploratory analysis of historical marathon data allowed us to model runner speeds and variances by corral. Intuitively, *Elite*, *American Development Program,* and *A* corrals run with the highest average speed throughout the race, and with relatively constant variances. The regression models reflect this intuition with these corral coefficients having the largest intercepts, and subsequent alphabetical assignments decreasing in coefficient size. Table 2 shows means and variances of runner speed for each corral.

Table 2. The Table Displays the Runner Speed Mean and Variance for Each Corral

	Corrals								
	Elites	**American Development**	**A**	**B**	**C**	**D**	**E**	**F**	**G**
Average speed (*mph*)	10.99	9.52	8.41	7.50	6.91	6.57	6.18	6.05	5.51
Variance (mph^2)	0.89	1.04	1.04	0.75	0.64	0.62	0.73	0.73	0.54

Figure 3. The Graph Shows the Average Speed Progression Calculated at 5K Timing Mats for the 2014-2016 Chicago Marathons by Corrals



Note. The x-axis represents the progression along the course and the y-axis represents the average speed.

Runner progression: Historical runner data indicate that a runner's progression (i.e., distance from the start) along the race impacts the runner's speed. Figure 3 plots the average speed by corral, averaged over the 2014-2016 Chicago Marathons, as a function of progression along the course. The average speed of an individual runner is calculated by dividing the distance between timing mats (5K) by the time spent during a 5K segment (e.g., the difference between the arrival time at the 10K mat and the arrival time at the 5K mat), and then averaged over three years. As expected, average speeds decrease over the course of the race—with the exception of the final 0.2 miles, which show a slight burst of speed by runners, the adrenaline stretch. Given this decrease in speed over distance, runner progression is included as a feature in predicting runner speed.

Course conditions: As noted by Chicago Marathon medical director George Chiampas, "One of the most important factors detrimental to running performance ... is heat and humidity" (Chiampas 2018). Previous simulation models did not account for temperature in speed predictions. This limitation was initially patched with a single race-day temperature; however, long-distance events often begin before sunrise and continue past midday, leading to large temperature and humidity differentials during the event. In event locations like Chicago, temperature swings of 20 to 30 degrees are not uncommon throughout the day. We match historical hourly weather observations to historic runner speeds, which were shown to be significant predictors of speed changes. We initialize the simulation with hourly temperature and humidity forecasts. Additionally, we change the simulation's weather variables uniformly across hourly observations to mimic real-world temperature changes and the true course conditions for runners. For example, the forecasted temperature for the 2016 Houston Marathon was 41 degrees at 9 a.m. and reached 46.4 degrees by 10 a.m. During the simulated hour from 9 a.m. to 10 a.m., the temperature was gradually increased in the simulation from 41 degrees to 46.4 degrees and the corresponding values were included into speed recalculations occurring per minute. The same analysis was applied to humidity values, using historical full-marathon and half-marathon data from 2014 to 2017 for the 2018 Houston Marathon. Using tenfold cross-validations, the individual full-marathon and half-marathon regressions attained average cross-validation training R^2 values of 0.82534 and 0.85044, respectively.

Object Representation and Efficiency

To maintain computational tractability, we do not simulate individual runners, but rather aggregate multiple runners into an object. The simulated object previously represented 50 runners of a race. In large races, this level of aggregation may be acceptable; in some events, however, 50 runners can equate to up to half of a corral. Estimating with a higher level of detail enables runner variance to be represented more accurately within corrals. We increased the granularity of the simulation by a factor of 10 (i.e., five runners to one object proportion) while decreasing the time needed to run the simulation by a factor of 20. The simulation can be run to completion in 45 seconds, enabling quick updates and adaptability during the event.

Speed Calculation

The runner objects are assigned to runner corrals according to the actual distribution of runners across corrals. Within each corral, an object has an assigned speed variance that corresponds to the estimated variance in speeds within a corral. Each object is then assigned a random speed variance, which is drawn from a Gaussian distribution with zero mean and a variance equal to the corral mean. Objects are assigned a start time for the simulated race given their assigned corrals (start times are uniformly distributed across a corral's start window). From these assigned start times, runner objects enter the course as the simulation progresses with the speed prediction function.

Runner Dropouts

During marathons and similar endurance events, all participants are not able to finish the race. In the early iterations of the simulation, runner dropout was captured with uniformly distributed reductions in runners, ignoring runner profiles. The current simulation has been updated to better address runner dropouts: Given historical data, it is possible to produce a probability of dropout matrix corresponding to runner corral and progression on the course. These probabilities are used to inform dropout rates by corral. Additionally, the simulation includes pre-event dropouts. For most events, a certain number of the registrants do not show up on the start line during the race day. We estimate pre-event dropout rates from historical pre-event data and use these numbers to initiate the simulation with a correct number of runners.

Race Day Simulation

The simulation is preloaded with runner registration numbers. At every simulated minute of the race, runner speed and location are recalculated using runner progression, corral, humidity, and temperature. Each runner object has its speed recalculated and its predicted location updated, given these inputs for every minute of the simulation.

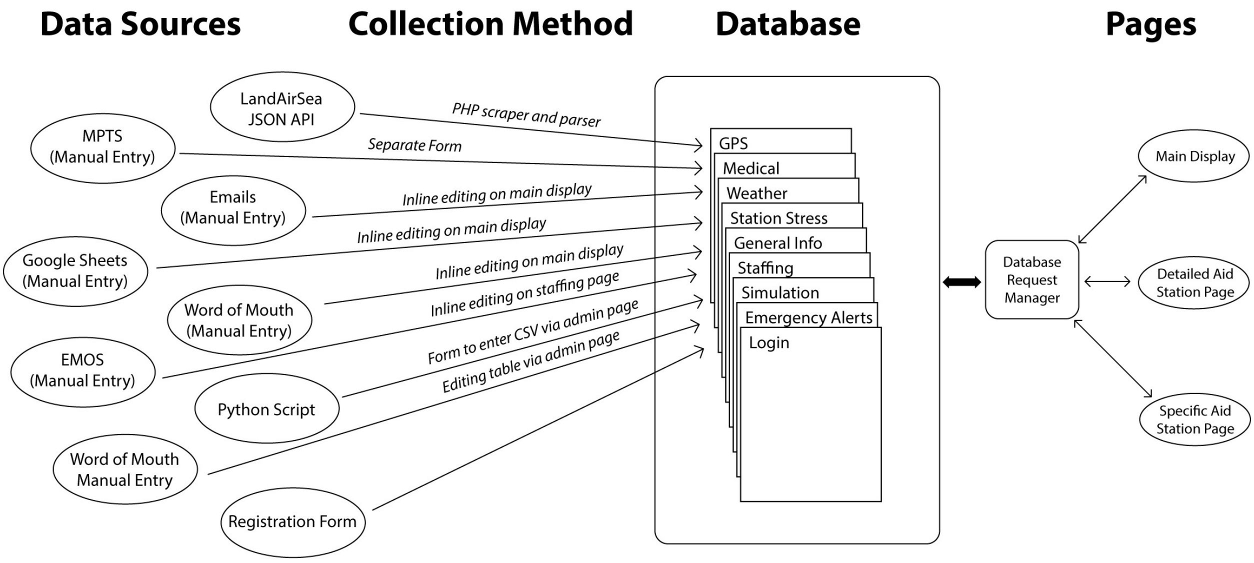
During the event, simulated runner densities are cross-checked against live numbers from timing mats at every 5K. Because the simulation takes less than a minute to run, we have the ability to resimulate and update predicted runner densities. During a race, many different scenarios may require resimulation. Some of these include unforeseen weather conditions, fewer than the expected number of runners beginning the race, race start-line delays, or changes to the number of runners per corral.

Back End Improvements

The back end of SAFE has been redesigned for a more reliable and uninterrupted experience. The updated system employs a structured database, providing faster, more reliable, asynchronous and organized data collection and processing. The new design minimizes data entry errors with automated and user-friendly data collection procedures. In addition, the system allows the SAFE team and race organizers to manage the access levels of users, protecting the display of critical data streams.

The system back end is coded in PHP with a MySQL database; Figure 4 provides an overview of the back end. To allow a user to see the display, the loaded page requests specific data from the server via Asynchronous JavaScript and XML (AJAX) with the database request manager. By using AJAX, the system is able to request data at regular intervals without refreshing the page, giving the user a continuous and uninterrupted experience. The database request manager verifies the user's privileges based on that user's unique and random ID in the request. It then draws from tables in the database and returns the data in JavaScript Object Notation (JSON), which the JavaScript in the page translates efficiently and displays.

Figure 4. The Figure Shows a Schematic of the SAFE Back-end System



The initial system had some data sources that we automated through PHP scripts: LandAirSea, the company that provides the GPS units used during the race, was able to provide JSON data in an XML file that we scraped, parsed, and entered into the database at a set time interval. Decreasing this interval allows for more robust updating, while increasing this interval reduces stress on the server. Scraping data by PHP from the server also reduces bandwidth usage in Forward Command by loading heavy network usage to another network on a remote server. Several data sources that have not been automated yet are entered via inline editing on the main display. With the new design, even the manual editing can be done more efficiently because a user can update data without leaving the main display.

In the initial versions of the system, data were stored and drawn from the server as CSV files with only login data being stored exclusively in the database. In the current system, we use a structured database mechanism and store data into a database specifically for each race. The new database implementation provides a number of essential advantages. First, by using a database, we are able to archive data throughout the race, creating a past record of the entire race, which allows us to perform post-race analysis by simply using the data with respect to their time stamps. Second, a database is far more secure than using text files on the server. The data stored in our database are accessible only through the PHP database request manager. By checking the submitted user's random ID with the database, the system guarantees that no other entity can access the database contents, even if that entity knows which data he (she) is seeking. Third, using a database and a single script to access the data enables us to provide reliability and ensures that our data are uncorrupted. Fourth, this implementation allows us to create levels for pages to request only what users need, minimizing stress on the database and the server. Finally, asynchronously drawing from tables in a database also maximizes speed, efficiency, and utilization.

Interface Updates

Users of SAFE vary widely in their roles related to the event. Thus, we have now tailored the system to meet a variety of needs. We have significantly expanded the system from a single-page dashboard display to an interactive multiple-feature system that can be modified according to user requests, providing a unique experience. In addition to the large dashboard display in Forward Command, we have extended the visualization experience to smartphones, tablets, desktop computers, and laptops. Below, we describe the changes in the main display and the newly added pages.

Main Display Changes

Next, we describe changes to the main display.

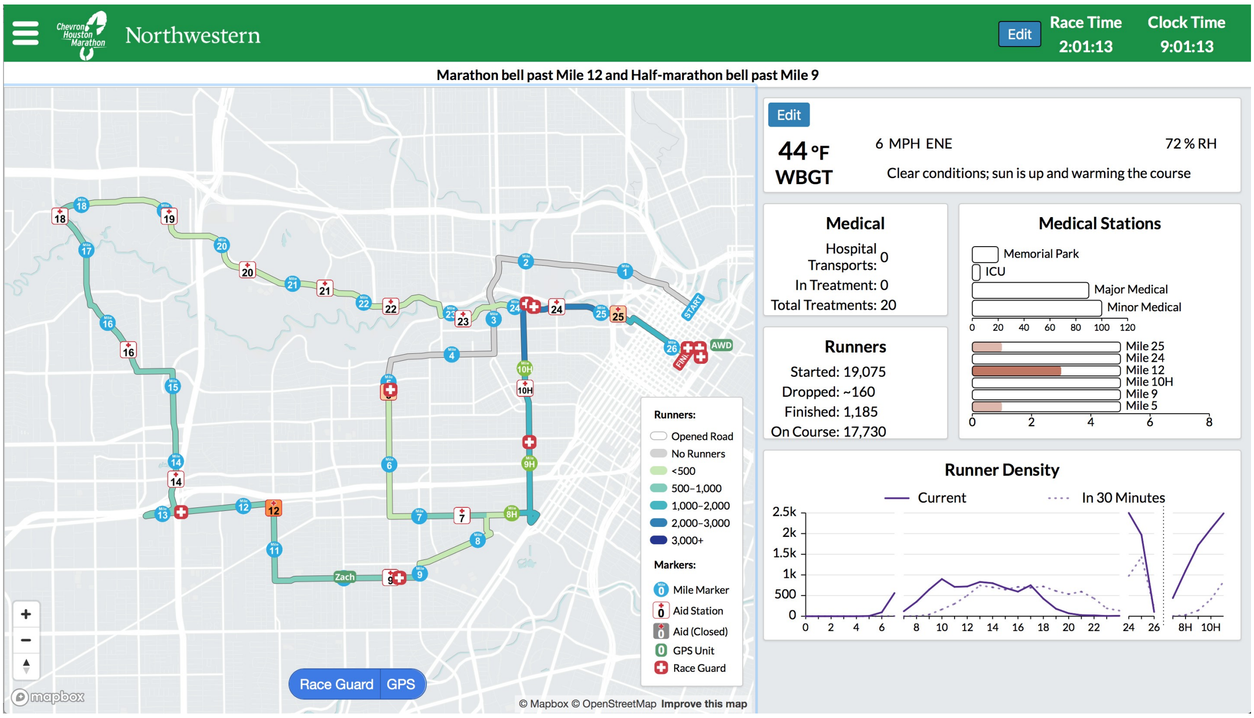
Event alert status and weather: The event status bar occupies the top section of the main dashboard. A color (e.g., white, green, yellow, red, or black) reflects the race's event alert status; see the event alert system table in Bank of America Chicago Marathon (2018). Course time and clock time are both displayed on the right side of the bar. The alert bar, which conveys event updates, such as race finishers, race status, and emergency messages via a text field, is at the bottom. Because of the universal importance of information contained in the event status bar, it is visible across all pages of SAFE.

The weather module displays the wet bulb globe temperature, the wind condition, dew point, relative humidity, and weather condition. The weather-condition text field can be used to identify weather risks, thus enhancing the usefulness of the system.

Course visualization: The importance of geodata for rapid data absorption has been recognized as one of the key principles for a successful data visualization system (Hanken et al. 2016); therefore, the course visualization represents one of the most important sections of the dashboard display. It displays the course route, which is marked by mile markers and aid station locations. Runner density is expressed by color-coded mile segments, showing the number of runners within each mile. The color of the segments also changes once the road is reopened to traffic. The colors used here, and throughout the system, were chosen to be color-blind friendly. Each module uses a different color scheme to distinguish the disparate data streams.

The course map also shows labeled GPS markers, such as the locations of the lead male and female runners. At the Houston Marathon, race guards (i.e., mobile paramedics) are also tracked with unique markers; see Figure 5. The display of GPS markers can be toggled using buttons at the bottom of the display. Aid station markers are updated dynamically to represent their patient occupancy, darkening in color to represent higher occupancy, and turning gray to represent a closed aid station. Aid stations and mile segments have tool tips to show more granular information. Mile segments can show a more precise runner density (to the nearest 20 runners), while aid stations can show exact patient, bed, and cumulative counts. The course map can also display emergency alerts at specific coordinates. A large red marker with an attached alert text field can bring attention to an emergency anywhere along the course.

Figure 5. The Figure Displays a Screenshot of the Central SAFE Dashboard Taken at the 2018 Houston Marathon



Runner tracking: Runner tracking is divided between two modules. The *Runners* module displays the number of runners who have started, dropped, or finished the race, and the number of runners who remain on the course. The *Runner Density* module displays simulated runner density along the course using a line graph. A solid line represents runner density per mile, while a dotted line represents the predicted runner density in 30 minutes. Tool tips display a more precise number for each mile segment, to the nearest 20 runners.

Medical tracking: Medical information is tracked across several modules. The *Medical* module displays key data points: patients who required a hospital transport, patients in treatment at medical facilities along the course, and cumulative patient treatments. The *Medical Occupancy* module displays labeled bar graph representations of bed occupancy at aid stations and medical tents along the course. Medical tents are divided into sections, such as podiatry and ICU, which are represented by separate bars. Tool tips display exact patient, bed, and cumulative counts. At the Chicago Marathon, aid stations are represented by stress levels instead of bed occupancy, because high patient load makes it difficult for medical volunteers to keep track of exact patient numbers. Therefore, the Chicago Marathon dashboard has a *Stress Levels* module, which displays the stress level from one to five for each aid station; stress levels are reported by ham radio operators located at each aid station. The accuracy of these stress levels has been evaluated relative to the medical records in Lin (2017).

Newly Introduced Pages

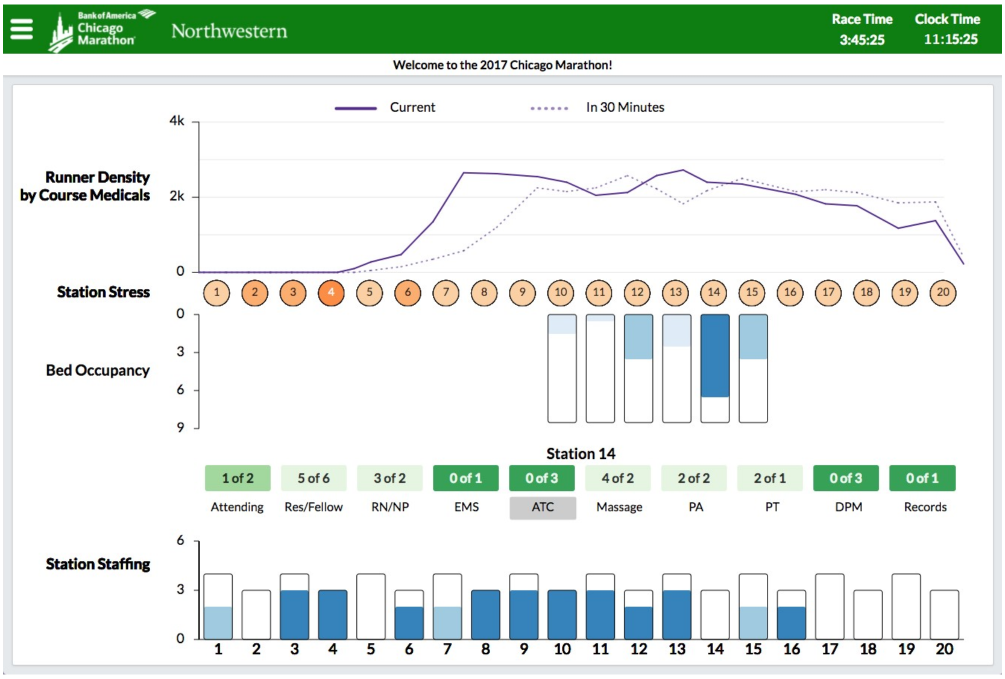
Working with race organizers, we identified additional information sources and added three new pages: (1) medical staffing, (2) detailed course medical, and (3) aid station. The goal of these pages is to help with staffing decisions on race day by monitoring staff arrivals prior to the race (medical staffing) and identifying needs during the race (detailed course medical and aid station).

Medical staffing: The medical staffing page consists of a large graph for all medical volunteer positions for each aid station, along with the staffing count for those positions. The staffing counts are updated in real time as volunteers check in, changing color as the staffing approaches the desired number of volunteers for that position. This page provides context on how many volunteers have showed up for their positions, and which aid stations are in need of staff before the race starts, giving organizers a view of surpluses and shortages to enable them to make real-time decisions. This page remains accessible on personal devices throughout the race.

Detailed course medical: The detailed course medical page provides a detailed summary of medical data and staffing during the race; Figure 6 shows a sample screenshot. The information on this page enables race organizers to identify problematic locations and respond dynamically to the needs of the medical team. The graph on top is similar to the runner density graph on the main display; however, this graph aligns the runners based on their positions relative to aid stations instead of on mile markers. Given the nature of the simulation and predictions for every minute of the race, we can deliver a predicted number of runners who have passed a specific marker at a given point in time. The horizontal axis of this graph has the labels for the aid stations and also assigns colors based on the stress level of the aid station. Stations under higher levels of stress are represented by darker colors. The graph immediately below the runner density graph shares the same horizontal axis and displays bed occupancy in the corresponding aid stations. Having this graph directly under the axis allows for a quick evaluation of the bed count relative to stress levels.

When hovering over the circles representing the stations, the user can see the reported staffing of the station for each position. Similar to stress levels, we use darker colors to indicate situations requiring attention: the darker the color, the larger the gap between the number of team members present and the number needed for that position. When the user selects one of the labels for a staffing position (e.g., physical therapy), the graph at the bottom shows the number of physical therapists throughout all the aid stations. If a station is short on physical therapists, this graph can quickly show race organizers which stations have a sufficient number of physical therapists and could potentially redeploy a member to help the station in need.

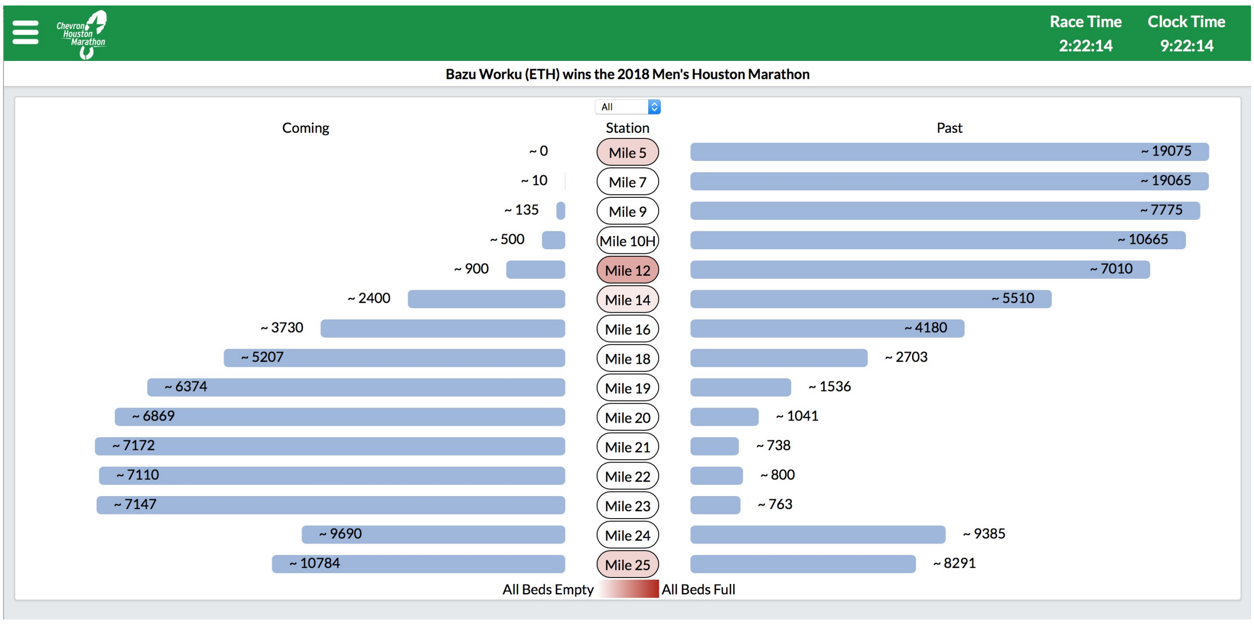
Figure 6. The Figure Displays a Screenshot of the Detailed Course Medical Page from the 2017 Chicago Marathon



The detailed course medical page quickly and efficiently draws data from multiple data sources to assist the organizers in making race-related decisions. For example, as Figure 6 shows, organizers notice that Station 14 already has a high bed count and is expecting a large number of runners. By hovering over the circle labeled 14, the user can then observe the staffing situation at that station. Upon identifying that no certified athletic trainers are present at Station 14, the user then clicks the athletic trainers field to view the athletic trainer staffing at other stations. The user then notices that Station 8 has little stress and most runners have passed this station. At this point, the user has identified the problem, found the corresponding solution, and can initiate the process of moving that additional staff member to help Station 14.

Aid station: The aid station page is designed to provide situational awareness at the aid stations and enables volunteers on-site at aid stations to see how many runners have passed and how many runners are expected to pass. These numbers help inform aid stations if they need to ask for help or increase stress levels if they are low on supplies. This feature of the predicted densities also enables city agencies in Forward Command to determine how many runners are left in a segment when considering road reopenings. Figure 7 shows a sample screenshot from a mobile device several hours into the 2018 Houston Marathon. The full page shows predictions of runner counts for each aid station. Note that Aid Stations 24 and 25 serve the finishers of both the half marathon and full marathon; therefore, a large number of runners have passed. The module allows the selection of a single station to reduce the visual noise of the other stations. The coloring of the station shows the occupancy ratio of the station; the darker the red, the higher the bed utilization, again keeping with the invariant that darker colors express critical situations.

Figure 7. The Figure Displays a Screenshot of the Aid Station Page from the 2018 Houston Chicago Marathon



Conclusions and Future Work

Data analytics tools have significant potential in the realm of public safety and planning for mass-participation events. Working closely with the organizers of the Chicago Marathon, the Houston Marathon, and their public safety partners, we developed SAFE, a comprehensive data visualization system, to complement their existing efforts to continuously improve planning and response for such events.

In this paper, we describe SAFE in detail. Our team is simultaneously continuing to explore new ways to improve the system. Because the simulation is core to SAFE, we are continuing to invest in the simulation and researching opportunities to merge live updates and parallel simulations to better capture the dynamics on the course. We are also exploring the use of additional data sources external to the event to include spectator dynamics. The Chicago Marathon draws an estimated 1.7 million spectators; the safety of these spectators is critical to organizers and public safety officials. Additionally, we are working with event organizers to expand our volunteer data visualizations; for example, we are looking to add general volunteer analysis to our work with medical volunteers. As we continue these efforts, we will look for ways to further leverage our descriptive and predictive analytics to provide decision-making tools for event organizers. Again, working with our industry and public agency partners, we believe that SAFE can become a standard for data visualization and analytics for mass-gathering events.

Acknowledgments

This research has been supported by Grants 1640736 PFI:AIR - TT: SAFE (Situational Awareness for Events): A Data Visualization System and 1405231 GOALI: Improving Medical Preparedness, Public Safety and Security at Mass Events from the National Science Foundation. The authors would like to thank Carey Pinkowski, Wade Morehead, John Cianca, and Brant Kotch and the entire staff and volunteers of the Bank of America Chicago Marathon and the Chevron Houston Marathon and their public safety partners. Additionally, the authors would like to recognize the work of earlier data visualization team members: Taylor Hanken, Sam Young, Rachel Lin, Bruno Peynetti, Ryan Rose, Matthew Zhang, Kate Larsen, Yasmeen Wood, and David Waskowski.

References

Associated Press (2010) Pittsburgh marathon disrupted by bomb scare. *NY Daily News Online* (May 2), http://www.nydailynews.com/news/national/pittsburgh-marathon-disrupted-bomb-scare-race-route-changed-device-found-microwave-article-1. 445850.

Bank of America Chicago Marathon (2018) Public safety. Accessed March 26, 2018, https://www.chicagomarathon.com/participant-information/public-safety/.

Basdere M, Ross C, Chan JL, Mehrotra S, Smilowitz K, Chiampas G (2014) Acute incident rapid response at a mass-gathering event through comprehensive planning systems: A case report from the 2013 Shamrock shuffle. *Prehospital Disaster Medicine* 29(3):320–325.

BBC News UK (2008) 'Gas leak' diverts marathon route. Accessed December 1, 2018, http://news.bbc.co.uk/2/ hi/uk_news/7345058.stm.

Belson K (2012) After days of pressure, marathon is off. *The New York Times Online* (November 2), https://www.nytimes. com/2012/11/03/sports/new-york-city-marathon-will-not-be-held-sunday.html.

Chiampas (2018) Email communication with Karen Smilowitz. March 6.

Cianca J (2018) Email communication with Karen Smilowitz. March 6.

CBS / The Associated Press (2013) Deadly bombs rock Boston Marathon. *CBS News Online* (April 15), https://www.cbsnews.com/news/deadly-bombs-rock-boston-marathon/.

Channel NewsAsia (2018) Malaysian dies 2 months after accident during marathon: Report. *Channel NewsAsia Online* (March 1), https://www.channelnewsasia.com/news/asiapacific/ malaysian-evelyn-ang-accident-klang-city-marathon-10001676.

Chiampas G, Jaworski CA (2009) Preparing for the surge: Perspectives on marathon medical preparedness. *Current Sports Medicine Rep.* 8(3):131–135.

Elahi A (2015) How a real-time simulation will keep the Chicago marathon on track. *Chicago Tribune Online* (October 9), http://www.chicagotribune.com/bluesky/originals/ ct-marathon-data-simulation-bsi-1009-story.html.

Farina R, Kochenberger GA, Obremski T (1989) The computer runs the bolder Boulder: A simulation of a major running race. *Interfaces* 19(2):48–55.

Federal Emergency Management Agency (2015) National Incident Management System (NIMS). Accessed March 26. 2018, //training.fema.gov/nims/.

Gorman M (2016) How the Chicago marathon will keep track of 1.7 million people. *Newsweek Online* (October 5), http://www.newsweek.com/2016/10/14/chicago-marathon-data-visualization-technology-506126. html.

Hanken T, Young S, Smilowitz K, Chiampas G, Waskowski D (2016) Developing a data visualization system for the Bank of America Chicago marathon (Chicago, Illinois USA) *Prehospital Disaster Medicine* 31(5):572–577.

Kuzma C (2016) Make your race fantastic. *Runners' World* 51(9):24.

Lin R (2017) Data analytics for medical preparedness and response: Applications to the Chicago marathon. Thesis, Northwestern University, Evanston, IL.

McCarthy DM, Chiampas GT, Malik S, Cole K, Lindeman P, Adams JG (2011) Enhancing community disaster resilience through mass sporting events. *Disaster Medicine Public Health Preparedness* 5(4):310–315.

Morehead W (2018) Email communication with Karen Smilowitz. March 7.

Radzievich N (2016) Train derails some runners at Via Marathon. *The Morning Call Online* (September 13), http://www. mcall.com/news/local/mc-theres-a-train-in-my-marathon-20160912-story.html.

Ross C, Basdere M, Chan JL, Mehrotra S, Smilowitz K, Chiampas G (2015) Data value in patient tracking systems at racing events. *Medicine Sci. Sports Exercise* 47(10):2014–2023.

Washington Post (2005) At marathon, ex-mayor stops the train dead in its tracks. *The Washington Post Online* (September 27), https://www.washingtonpost.com/archive/sports/2005/09/27/at-marathon-ex-mayor-stops-the-train-dead-in-its-tracks/1085cb44-cf5e-4858-88cc-dffb25c703b0/?utm_term=.7dbd3a5c1d60.