Practice of Epidemiology

Developing and Implementing Major League Baseball’s Health and Injury Tracking System

Keshia M. Pollack*, John D’Angelo, Gary Green, Stan Conte, Stephen Fealy, Chris Marinak, Edward McFarland, and Frank C. Curriero

* Correspondence to Dr. Keshia M. Pollack, Johns Hopkins Center for Injury Research and Policy, Johns Hopkins Bloomberg School of Public Health, 624 N. Broadway, Room 555, Baltimore, MD (e-mail: kpollac1@jhu.edu).

Initially submitted August 1, 2015; accepted for publication December 9, 2015.

In 2010, Major League Baseball and the Major League Baseball Players Association reached an agreement regarding the development and implementation of an electronic medical record system and a new league-wide injury surveillance system. The systems were developed to create a more efficient method to track medical histories of players longitudinally as they move across Major and Minor league affiliates, as well as to identify and monitor injury trends in the sport, identify areas of specific concern, and conduct epidemiologic research to better optimize player health and safety. The resulting injury surveillance system, the Health and Injury Tracking System (HITS), is a robust system that includes all players from the both the Major and Minor Leagues. HITS also allows for data linkage with other player- and game-level data to inform the development of injury prevention policies and programs. In the present article, we document the development and implementation of HITS; describe its utility for epidemiologic research; illustrate the potential analytic strength of the surveillance system and its ability to inform policy change; and note the potential for this new surveillance system to advance the field of sports injury epidemiology.

baseball; injury epidemiology; sports injury; surveillance

Abbreviations: AE, athlete exposures; AT, athletic trainer; EMR, electronic medical record; HITS, Health and Injury Tracking System; MLB, Major League Baseball; MTBI, mild traumatic brain injury.

Surveillance is “the ongoing systematic collection, analysis, and interpretation of health data, essential to the planning, implementation, and evaluation of health practice, closely integrated with the timely dissemination of these data to those who need to know” (1, p. 164). In the United States, surveillance of injuries has allowed the detection of important trends, identification of leading correlates of injuries, and development of prevention policies and programs (2). Injury surveillance is also critical for conducting descriptive epidemiologic studies, which help illuminate the patterns, affected populations, and distribution of disease and illness.

Sports and recreation are highly popular activities in the United States, and several sports injury surveillance systems exist to document trends and inform prevention efforts. For example, estimates of injuries resulting from sports and recreation for the general US population are captured by the National Electronic Injury Surveillance System, which includes injury data from a sample of emergency departments around the United States (3, 4). For college injury surveillance, data are collected using the National Collegiate Athletic Association’s Injury Surveillance Program (5–7). The National Collegiate Athletic Association’s Injury Surveillance Program collects injury and athlete exposure (AE) data from a representative sample of collegiate institutions and sports (7). Injuries among high school athletes are captured using the High School Reporting Information Online, which collects data on time-loss injuries among a national sample of specific US high school athletic teams (8–10). These surveillance systems have documented important injury trends in specific sports and helped to inform the development of prevention strategies among student athletes (11, 12). For example, analysis of data from the National Collegiate Athletic Association’s Injury Surveillance Program led to policy and rule changes in ice hockey to reduce the number of concussions,
The ability to identify leading injuries and monitor trends, especially to determine the impact of policy and rules changes, is an added benefit of these sports injury surveillance systems.

For professional baseball players, injury surveillance was primarily conducted using the Major League Baseball (MLB) Disabled List. The Disabled List has been analyzed to document injuries overall (13, 14), musculoskeletal injuries (15), and injuries resulting from collisions (16), among others. Limitations of the Disabled List have been noted, specifically that the Disabled List does not include all injuries and that decisions to place a player on the Disabled List might be based on nonmedical factors in order to maximize the chances of having the best available players on the active roster at any given time (13). The lack of a formal injury surveillance system for professional baseball has been noted as one reason little is known about the injury rates in MLB players (13).

In 2010, MLB and the Major League Baseball Players Association reached an agreement to develop and implement a new electronic medical record (EMR) and injury tracking system. The agreement included discussions about the types of data that would be collected, data storage, and research, among other legal and logistical topics. The primary goal of the new system was to create a more efficient method to track medical histories of players longitudinally as they move across Major and Minor league affiliates. A second goal was to identify and monitor injury trends in the sport, identify areas of specific concern, and conduct epidemiologic research to better understand injuries and optimize player health and safety through possible rules changes, equipment modifications, or medical education. The unique aspect of this surveillance system is that it includes all players from the Major and Minor Leagues, rather than just a sample of players from certain teams. Our purpose in the present article was to document the development of this new baseball injury surveillance system, describe the utility of a league-wide injury surveillance system for research, illustrate the potential analytic strength of the database and its ability to inform policy change, and note the potential for this surveillance system to advance the field of sports injury epidemiology.

**METHODS**

**Athlete population**

In the United States, professional baseball consists of the Major and Minor Leagues, which are administered by the MLB. The Major Leagues are divided into 30 clubs that play 162 games in 183 calendar days and have 25 active players on their rosters at any one time, for a total of 750 active players. Clubs can also have up to 15 additional players on their Major League rosters who are either not active or are optioned to the Minor Leagues; these players, combined with those on the 25-man active rosters, comprise the Major League 40-man rosters.

The Minor Leagues are a network of more than 200 clubs throughout the United States that are each affiliated with a Major League club. The Minor League clubs are organized into a number of different leagues based on both geography and level of play. Season length in the Minor Leagues differs depending on the league, with a range of 80 days for the low-level short seasons to approximately 150 days for the high-level AAA leagues. There are approximately 7,500 players assigned to the Minor Leagues each year, with nearly 6,500 actively playing at any particular time.

**MLB’s EMR system**

The MLB’s EMR system was built to provide a more standardized and streamlined way to enter, track, and transfer player medical records across professional baseball. All professional baseball players in both the Major and Minor Leagues provide their consent for their medical records to be included in the EMR system. The EMR system was released across baseball at the beginning of the 2010 season and is available on the MLB intranet. The system has a number of levels of security, with users needing to enter both a password and an RSA SecurID (EMC Corporation, Hopkinton, Massachusetts) token code in order to gain access.

The EMR system is linked to MLB’s electronic Baseball Information System. The electronic Baseball Information System is a web-based program that facilitates many functions of baseball operations for the 30 Major League clubs and the Office of the Commissioner. The system serves as the core method through which clubs carry out transactions, such as inputting information about contract terms. Linking the EMR system to the electronic Baseball Information System allows roster changes to be made automatically, which in turn facilitates the transfer of medical records as players move between and within organizations.

The core component of the EMR system is the entry of all injury, medical, treatment, and prevention data for each player on a club’s roster by the athletic trainer (AT). Once one of these occurrences (i.e., any injury, medical, treatment or prevention-related occurrence) has been entered into a player’s record, treatment notes, doctor’s notes, and diagnostics can also be entered into the system and linked to the applicable event. Files such as videos and magnetic resonance images can also be uploaded and attached to each event. The EMR system also allows the creation and input of standard physical examination forms, medical history forms, allergy and immunization information, exit questionnaire forms, and Health Insurance Portability and Accountability Act authorization forms for each player. Players complete a disclosure of health information form when signing their contracts that allows these approved system users access to the players’ medical records.

**Health and Injury Tracking System: overview and measures**

Injury data are included in Health and Injury Tracking System (HITS), a centralized database containing the de-identified medical data from the EMR system. All records were assigned a computer-generated unique identifier to preserve confidentiality. In addition, all individually identifiable information was stripped based on Health Insurance Portability and Accountability Act guidelines.
A research team from the Johns Hopkins University Bloomberg School of Public Health worked with the MLB to develop the specifics of the injury-tracking system and subsequently has provided epidemiologic and statistical support for the research. Development of the database involved reviewing existing sports injury surveillance systems and consulting with the ATs, as well as with experts who had experience with existing sports databases and team physicians. HITS was created in 2009, pilot tested during the 2010 season, and fully implemented for injury surveillance during the 2011 season.

HITS includes any injury or physical complaint sustained by a player that affects or limits participation in any aspect of baseball-related activity (e.g., playing in a game, practice, warm up, conditioning, weight training). For research studies, injuries are operationally defined as those that are work-related, did not occur in the off-season (i.e., occurred only spring training, the regular season, or the postseason), were a primary diagnosis, and resulted in at least 1 day out of play. Because of the multiple pieces of data collected for each injury, injuries can be investigated by nature, body part injured, mechanism, location, or activity, as well as by demographic fields such as player position or age. Data are inputted via dropdown menu choices for each variable. A summary of the variables included in HITS is presented in Table 1.

For sports injury research, the use of AEs is a well-established procedure for calculating rates (5, 7, 17). Injury research using HITS defines AE as the average number of players per team per game calculated based on analysis of regular season game participation via (publically available) box scores. This average number over a season multiplied by the number of team games at each professional level of baseball yields an estimate of AE.

HITS has a unique identifying number for each player; thus, data can be deterministically linked across various databases. Linking the data allows for investigation of injuries with regard to other key measures of exposure, such as pitch count and number of plate appearances, surgical outcomes data, personnel records to calculate cost data, and demographic data. Also, we can link injury occurrences to cost data and quantify the relative impact that different injuries have on player health to help us focus our efforts on research that has the greatest potential to address those areas. In addition, player data can be linked to previous injuries of the same or related type to better understand the influence of injury history on current or future injuries.

### RESULTS

#### Implementation of HITS

The primary research team from Johns Hopkins University Bloomberg School of Public Health oversees most of the...
epidemiologic research involving HITS, as well as other primary data collection efforts to supplement the HITS data (e.g., primary data collection using an online survey). The Bloomberg School of Public Health Institutional Review Board approves all study procedures, which include analysis of the de-identified HITS data. Centralizing the institutional review board approvals across the studies has ensured consistency with regard to the protection of confidential player data.

A system audit was initiated in 2010 to explore the initial HITS data. The audit focused on evaluating the consistency of data entry across teams and leagues, evaluating the consistency between variable definitions and variable completeness (i.e., missing data), and identifying anomalous entries in the database. All data from 2010 comprising approximately 30,000 records from both the Major and Minor Leagues were downloaded from HITS and analyzed using descriptive statistics. Summary tables describing distributional characteristics were generated for more than 30 variables. Results were also stratified by league and by team/club affiliate. In the audit, the type of data and level of detail required to support research initiatives were evaluated. For example, for injuries occurred while running a variable was added to distinguish whether the running activity was while on offense or defense. As of 2011, an annual audit of the HITS data was being performed to ensure continued consistency. Table 2 shows the total number of injuries per season, along with the average percent of missing values. Variables such as injury mechanism, which describes how the injury occurred (e.g., contact with ball, contact with ground, noncontact), injury activity, which describes what the player was doing when injured (e.g., batting, running, pitching), and injury location, which describes where on the field the injury occurred (e.g., home plate, specific base, outfield) are key to providing an overall injury summary. Because of increased efforts to reduce missing data, the percent of records with missing values for these important variables in at least 1 entry has substantially declined over the years.

Since the inception of HITS, several improvements have been made, including reducing the variable definitions (e.g., collapsing categories in the mechanism of injury variable to be focused around the type of contact) and enabling the system to automatically populate data entries of variables given data inputted to other variables, which dramatically improved completeness. For example, if a player sustains a sprain, a dropdown box will appear to allow the person entering the data to select the degree of the sprain as a measure of severity. We also learned from the ATs that a simplified user interface was needed; after this change was made, the percent of missing data declined.

Implementation of HITS also involved creating guidance for all data analysis. The initial data analysis approach for all research studies has been to complete a descriptive epidemiologic analysis. The benefits of this approach include: 1) conducting exploratory data analysis to describe the injury of interest as it relates to variables such as level of play, position, mechanism, activity, location, and time lost; 2) highlighting trends worthy of further investigation; and 3) providing a common starting point for all research studies to ensure a level of consistency across all ongoing research, as well as future planned studies. These descriptive epidemiologic analyses also included injuries reported as rates per 1,000 AEs, allowing our work to be based within the context of previous works on related injuries (baseball and otherwise) that used other databases and also reported injury rates (5–7, 17). Finally, data are aggregated across the entire league before dissemination of any analysis results.

Table 2. Frequency of Injury Eventsa by Level of Play for the 2010–2014 Seasons

<table>
<thead>
<tr>
<th>Level of Play</th>
<th>No. of Injuries by Year</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major League</td>
<td>2,076</td>
<td>1,641</td>
<td>1,347</td>
<td>1,270</td>
<td>1,249</td>
<td>7,583</td>
<td></td>
</tr>
<tr>
<td>Minor League</td>
<td>7,828</td>
<td>7,234</td>
<td>6,704</td>
<td>6,909</td>
<td>7,269</td>
<td>35,944</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9,904</td>
<td>8,875</td>
<td>8,051</td>
<td>8,179</td>
<td>8,518</td>
<td>43,527</td>
<td></td>
</tr>
<tr>
<td>Missing informationb</td>
<td>18.0</td>
<td>17.0</td>
<td>&lt;1.0</td>
<td>&lt;1.0</td>
<td>&lt;1.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Injury events were defined to include only events that were work related, did not occur in the offseason (e.g., events that occurred during spring training, the regular season, or the postseason), were a primary diagnosis, and resulted in at least 1 day out of play.

b Values are expressed as percentages. Values were based on the number of records with missing data for at least 1 of the entries for the variables injury mechanism, activity, or location.

Results from selected completed and current projects using HITS

After a review of the data to identify the most frequent factors correlated with injury and the body parts that were most frequently injured (Tables 3 and 4), a series of research projects was initiated to investigate injury patterns and identify data-informed solutions to prevent injury and disability to players. Decisions about which injury topics to explore were made jointly by the MLB and Major League Baseball Players Association, with input from the Bloomberg School of Public Health research team. Specific detail on methods and findings can be found in each of the articles referenced in the subsequent section.

Hamstring strains. The first publication using the HITS data was by Ahmad et al. (18); in that study, the authors explored hamstring strains during the 2011 season. They found that base running, specifically running to first base, was the activity that most frequently resulted in a hamstring strain. History of a previous hamstring strain in the prior year (2010) was found in 20% of the Major League players and 8% of the Minor League players. Based on this analysis, the authors concluded that hamstring strains are affected by prior history of hamstring strain, seasonal timing, and running to first base (18). Ahmad et al. reported a hamstring strain injury rate of 0.7 per 1,000 AEs for both Major and Minor League players, although there were more hamstring strains in players in the Minor League.

Mild traumatic brain injury. The second publication using HITS data was by Green et al. (19), who explored mild traumatic brain injury (MTBI). Although there has been extensive literature about MTBI in sports, their study was the first in which an entire population of professional baseball players was examined to determine the distribution and
correlates of MTBI in baseball. Green et al. concluded that MTBI is an important and understudied injury issue in professional baseball players. The overall game rate of MTBIs across both Major and Minor League clubs was 0.42 per 1,000 AEs, although the MTBI rate was 1.8 times higher among Minor League players (0.26 per 1,000 AEs) than among Major League players (0.26 per 1,000 AEs) (19). The study revealed that catchers were significantly overrepresented among players who sustained a MTBI relative to other positions and that concussions most often occurred from contact with the ball (i.e., foul tips) (19). This has resulted in follow-up studies to better understand the risk factors for concussions among catchers and whether these injuries can be prevented through changes to technique or equipment.

**Other ongoing research.** These studies by Ahmad et al. (18) and Green et al. (19) both built on previous studies in their respective areas by using data from an entire athlete population, unlike prior studies of smaller samples that relied on sampling. The HITS data also advanced thinking on the areas under study by providing in-depth investigation into injuries that have not been comprehensively examined. Results from studies in which injuries to the knee, shoulder, elbow, and hip/groin were examined are anticipated in the coming year. Findings from research exploring sliding injuries and a primary data collection effort are anticipated in the coming year. Findings from research portraying the ability to link data to ascertain individual-level variables; the ability to develop better measures of exposure; the creation of a record of all player illnesses and injuries in the EMR; and the creation of a record of injuries that result in lost time, which allows investigations of both medical and injury diagnoses. The system also provides an opportunity to conduct prospective research in a way that was impossible before. For example, for the current prospective study of elbow injuries, portrayed by the surveillance data. For example, an intervention study of hamstring injuries is currently being pilot tested. This intervention was implemented based on results from Ahmad et al. that identified factors correlated with hamstring strain (18). Future analyses will determine whether the trends in these injuries have been altered. Another example is the current research in which risk factors for elbow injury are being explored. A prospective epidemiologic study is underway to determine risk factors for an ulnar collateral ligament injury. In that study, investigators enrolled Minor League pitchers and collected baseline data, including biomechanical data, to determine risk factors for the development of an ulnar collateral ligament injury over time (Major League Baseball, unpublished data, 2015).

**DISCUSSION**

HITS is the first comprehensive injury surveillance system developed to explore injuries among professional baseball players. There are several analytic strengths of HITS, including the ability to link data to ascertain individual-level variables; the ability to develop better measures of exposure; the creation of a record of all player illnesses and injuries in the EMR; and the creation of a record of injuries that result in lost time, which allows investigations of both medical and injury diagnoses. The system also provides an opportunity to conduct prospective research in a way that was impossible before. For example, for the current prospective study of elbow injuries,
once the initial assessments were conducted, it was possible to continue to assess player’s injuries for the rest of their careers. This passive surveillance system reduces the burden on the trainers who do not have to actively search for cases.

One benefit of a robust data system is the ability to measure exposures. Identifying the best denominator for future analyses will propel the field beyond using crude measures of AEs for rates. Also, data on medical conditions will fill an important gap in knowledge and allow us to understand the occurrence of health conditions beyond injury. This is important because noninjury health issues and conditions can affect injury rates and risk, and vice versa. HITS will also allow us to advance research that validates sport injury surveillance systems. For instance, with access to the medical records and video recordings, investigators can conduct future studies that validate the reporting of injuries by analyzing video recordings from games.

Findings from these analyses have been used to create policies that inform play. For example, analysis of the HITS data revealed that players recovered from concussions in fewer than 15 days, which created a difficult situation for the player and the club because of the difference between recovery time and required Disabled List stint. This led to the implementation of a 7-day Disabled List for concussions that has promoted more players being placed on the Disabled List and having the appropriate time to recover from those injuries. HITS data have also been used to inform discussions about potential rule changes, including those related to collisions between base runners and fielders at home plate and other bases.

As previously noted, in addition to the HITS data, with the EMR there is an opportunity to identify leading illnesses and appropriate prevention strategies. There is a dearth of published research exploring medical issues in MLB. With the EMR data, studies in which leading medical conditions among players and treatment options are explored can advance knowledge about player health. Several research questions remain in this area, and MLB and the Major League Baseball Players Association in concert with the research team are prioritizing health conditions for future investigation, for example, the injury rate associated with sliding into bases, foul ball impacts on catchers, and the epidemiology of various throwing injuries. In 2011, Posner et al. noted that, “Surprisingly, few published articles are available that pertain to the epidemiology of MLB injuries” (13, p. 1679). With the development of HITS, this is changing; in recent publications, investigators have explored the epidemiology of hamstring strains and MTBI, and several other studies are anticipated in the near year.

It is important to also note that the creation of HITS has resulted in significant growth in the number of research initiatives jointly sponsored by MLB and the Major League Baseball Players Association. This growth has also resulted in a need for new standards and procedures for research and dissemination of data via publications and presentations.

Challenges in implementing HITS

The HITS database has in part revolutionized research on baseball injuries, but there have been some challenges. First, since the inception of the new system, data quality was a primary concern. Ensuring that the database is user-friendly for the ATs who are entering the data is critical for valid and reliable outputs. Part of ensuring quality has involved collaborating with the ATs to obtain their feedback on aspects that are not working and modifying the database accordingly. Because analysis of HITS data can only include data that are entered in the system, collaborative efforts have sought to minimize missing data and improve data reliability and validity.

Second, the AE metric previously defined is at best an approximate estimate, because in baseball, a player who only appears for 1 at bat (e.g., a pinch hitter) is counted as the equivalent of player who has played the whole game. This is a limitation, and it affects the measurement of risk (17). Moreover, there is currently no way to measure exposures during practice; the number of pitches thrown, number of at bats, etc., are not recorded during practice. Although our studies to date have shown that very few injuries occur in any given practice session, there could be a cumulative effect of exposures during practice, which could be associated with overuse injuries.

The next step currently under consideration is a refined exposure assessment to develop methods to establish more representative (sport- and injury-specific) exposure metrics for risk assessment and rate calculations for various injuries. For example, more appropriate denominators for use in studies of pitcher injuries could include number of games pitched, number of innings pitched, total pitches per inning or overall within a specific time period, pitch selection, and pitch velocity. HITS provides a level of information detail that supports these more refined exposure metrics. These results are also needed to support more advanced analytical methods based on statistical regression techniques to identify and quantify risk factors and risk factor interactions for injury research.

Finally, having the team from the Johns Hopkins Bloomberg School of Public Health lead the analysis of all research has helped to ensure consistency across research studies. Practically, this means that because of the desire to ensure consistency using the HITS data, there is a queue for research projects, which does not always align with the desire timeline of the investigators involved with a particular project.

Conclusions

HITS is the first comprehensive injury surveillance system developed to explore injuries among professional baseball players. The richness of the data is unprecedented and creates an opportunity to identify and monitor injury trends in baseball, identify areas of specific concern, and conduct epidemiologic research to better understand player risk and optimize player health and safety through possible rule changes, equipment modifications, or medical education.

The implementation of HITS has advanced sports injury research overall, and professional baseball research in particular. Future research involving HITS data will continue to optimize player health and safety for one of America’s greatest pastimes, as well as generate important overall lessons for injury surveillance.

ACKNOWLEDGMENTS

Author affiliations: Johns Hopkins Center for Injury Research and Policy, Bloomberg School of Public Health,
Johns Hopkins University, Baltimore, Maryland (Keshia M. Pollack); Office of the Commissioner, Major League Baseball, New York, New York (John D’Angelo, Gary Green, Chris Marinak); Division of Sports Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California (Gary Green); Conte Injury Analytics, Los Angeles, California (Stan Conte); Department of Sports Medicine & Shoulder Service, Hospital for Special Surgery, New York, New York (Stephen Fealy); Department of Orthopaedic Surgery, School of Medicine, Johns Hopkins University, Baltimore, Maryland (Edward McFarland); and Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland (Frank C. Curriero).

This research was supported by a contract from the Office of the Commissioner, Major League Baseball to Johns Hopkins Bloomberg School of Public Health for epidemiologic design, analysis, and statistical support.

We thank the Office of the Commissioner of Major League Baseball, New York, New York, especially Jon Coyles, and the Major League Baseball Players Association for their support and Dr. Bert Mandelbaum and Randy Dick, who provided input into developing and implementing Health and Injury Tracking System. Finally, we thank the athletic trainers for their dedication in collecting and entering data.

Conflict of interest: none declared.

REFERENCES