Accident and Emergency Informatics
T.M. Deserno et al. (Eds.)

doi:10.3233/SHTI220004

Improving Emergency Medical Services Information Exchange: Methods for Automating Entity Resolution

Robert W. TURER[a,b,1], Graham C. SMITH[c], Faroukh MEHKRI DO[a], Andrew CHOU [a], Ray FOWLER[a], Ahamed H. IDRIS[a], Christoph U. LEHMANN[b,d], and Samuel A. MCDONALD[a,b]

[a] *University of Texas Southwestern Medical Center, Department of Emergency Medicine. Dallas, TX, USA.*
[b] *University of Texas Southwestern Medical Center, Clinical Informatics Center. Dallas, TX, USA.*
[c] *University of Michigan Medical School, Department of Emergency Medicine. Ann Arbor, MI, USA.*
[d] *University of Texas Southwestern Medical Center, Department of Pediatrics. Dallas, TX, USA.*

Abstract. The 21st century has seen an enormous growth in emergency medical services (EMS) information technology systems, with corresponding accumulation of large volumes of data. Despite this growth, integration efforts between EMS-based systems and electronic health records, and public-sector databases have been limited due to inconsistent data structure, data missingness, and policy and regulatory obstacles. Efforts to integrate EMS systems have benefited from the evolving science of entity resolution and record linkage. In this chapter, we present the history and fundamentals of record linkage techniques, an overview of past uses of this technology in EMS, and a look into the future of record linkage techniques for integrating EMS data systems including the use of machine learning-based techniques.

Keywords. emergency medical services, emergency department, entity resolution, health information exchange, record linkage

1. Introduction

High-quality prehospital care relies on the ability to quickly and accurately retrieve and document data into a patient's medical record. Emergency medical services (EMS) often interface with multiple hospital systems and individual EMS agencies. The challenge of obtaining and synthesizing multiple sources of medical information for patients is heightened by the unpredictable, emergent nature of EMS care. These challenges are exacerbated further by the need to utilize accurate, standardized, and complete data in

[1] Corresponding Author; Robert W. Turer, Emergency Medicine, UT Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, Texas 75390. Email: Robert.Turer@UTSouthwestern.edu

broader systems-of-care. Overcoming these obstacles requires rigorous entity resolution methods.

During pre-hospital care, patients and their data first encounter EMS clinical information systems, which record and document the patient's history, condition, vital signs, presumptive diagnoses, treatments received, and their disposition. Subsequently, downstream clinical systems involved in real-time clinical care, quality assurance, public health, disaster planning and response, and/or research perform better when EMS clinical systems communicate with electronic health records (EHRs), clinical registries, and public databases. Bidirectional data exchange allows downstream systems to be primed with important patient information, improving accuracy and reducing duplicate therapies and documentation, and can provide valuable outcome data to EMS information systems. However, to date, information exchange of health information technology systems with EMS systems remains woefully underutilized.

Patient handoffs between EMS and emergency department (ED) teams stand to benefit from better interoperability between EMS and hospital-based information systems [1]. In an ideal environment, EMS information systems would be prospectively interoperable with one another and with hospital systems. This would include end-to-end integration of emergency-call management, computer-aided dispatch systems, EMS patient care records, and hospital charting to facilitate improved real-time handoff of information as proposed by Schooley and Hikmet in Santa Clara County, California [2] or as proposed by Park *et al.* after a formal needs assessment [3]. The ability to uniquely identify a patient is a requisite first step in this process which utilizes record linkage methods, defined by Shah *et al.* as "the ability to compare and match data records from multiple sources in order to determine which sets of records belong to the same person, object or event" [4]. However, given the chaotic nature of pre-hospital care and the challenges of achieving agreement across systems and organizations, data in EMS systems and other relevant data repositories suffer from missingness, inaccuracy, and inconsistency in representation, thus creating challenges in record linkage.

An exponentially increasing number of publications on record linkage since the early 2000s demonstrates the importance of the field [5]. In this chapter, we introduce fundamental concepts in record linkage, provide examples of record linkage involving EMS systems, and summarize the current and future state of the field. We do not explore EMS integration with health information exchanges. The current state of EMS health information exchange was recently reviewed by Martin *et al* [6] and current barriers to implementation were articulated by the National EMS Management Association [7]. We similarly do not discuss hardware, network, or EMS vehicle related architectural concerns. These challenges were well summarized by Landman and colleagues [8].

2. Fundamental Concepts in Record Linkage

The literature describing probabilistic algorithm development for linking records with imperfect identifiers or keys dates back to the 1950s [4]. Shah and colleagues provided an excellent review of the basics of record linkage techniques [4]. There are three overarching classes of techniques used in record linkage systems: deterministic (direct/exact), probabilistic record linkage, and machine learning (ML)-based linkage.

Because of its accuracy, deterministic or rules-based link linkage is the preferred method when uniquely identifying information is uniformly available, which unfortunately is rarely the case in real-world EMS systems. Deterministic record linkage traditionally requires an exact match for each linked data element. Various methods attempt to improve matching performance by normalizing data elements based on spelling variances or common data entry errors, but these methods ultimately return a dichotomous outcome (matched or not matched) [9]. Most commercial and open-source database applications can perform deterministic linkages between data sets. Semi-deterministic approaches include the use of truth-tables, which create unique identifiers from several variables (name, date of birth, address, etc.) when a truly unique key is unavailable. A "master patient identifier" has been proposed both in the US and the European Union, though has not been universally implemented. Denmark's efforts to unify its EMS systems with other medical systems using a civil registration number provide a representative example of what is possible using deterministic linkage when political and administrative will and privacy concerns align to make such a process feasible [10].

Probabilistic record linkage was introduced by Newcombe *et al.* in 1959 [11]. In their work, which linked birth and marriage records to explore connections between family fertility and the presence or absence of genetically inherited diseases, the authors used an array of error handling methods to account for spelling mistakes. To prepare for linkage, they included phonetic encoding of surnames and the inclusion of maternal and paternal surnames, provinces or countries of birth, and first initials, which were available on all births and most marriage records. Probabilities of appropriate record matching were then calculated using probabilities of matching in a target population either via previously published statistics or based on the characteristics of the population itself. The probabilities by different predictors were then logged and combined to generate the odds of a match. Fellegi and Sunter built on this work in 1969 to establish the mathematical theory behind modern probabilistic record linkage, which remains foundational in the field [12]. Their paper first established that an empiric approach using likelihood ratios was in accordance with classical inferential hypothesis testing. Later work showed that this technique could be derived using Bayes theorem [13]. For these techniques and those that followed, probabilistic linkage techniques required manual calibration of parameters, match thresholds, and linkage weights. These techniques typically require human adjudication when the probabilistic linkage system fails to classify with adequate confidence.

Several studies have compared deterministic and probabilistic approaches. Gomatam *et al.* explicitly compared deterministic and probabilistic approaches in linking records between a Florida intensive care registry with educational records from the Florida Department of Education database, using social security numbers (present in both databases) as the gold standard [14]. Specifically, they compared a stepwise deterministic method with the AUTOMATCH algorithm previously described by Jaro [15] and determined that the probabilistic method had a far greater match rate and sensitivity at the expense of a slightly lower positive predictive value, since deterministic methods are less susceptible to false positives given their more exact nature. Jamieson and colleagues compared a deterministic strategy, a probabilistic strategy, and a mixed strategy for linking Public Health Nursing Services with income assistance and found the mixed approach most effective with specificity of 0.98 and sensitivity of 0.94 [16]. Roos *et al.* linked 96.6% of records between Manitoba hospital encounters and physician claims using a staged approach with a deterministic stage followed by multiple probabilistic stages on several identifiers [17]. Many open source and commercially available software packages implement these record linkage techniques for commonly used statistical and analytics packages [4,18].

The third category of record linkage systems uses machine learning. The parallels between the Fellegri-Sunder probabilistic approach and ML-based classifiers such as naïve Bayes alongside the increasing availability of adequate computing power for handling large data sets have contributed to a rapid expansion of these alternate techniques for record linkage [5]. Supervised classification approaches, including decision trees, support vector machines, ensemble methods (*e.g.* random forests), or conditional random fields; and unsupervised methods, including *k*-means clustering, have been applied [4,5]. The literature is still emerging on the optimal approach for healthcare-based record linkage systems using machine learning [19].

Machine learning based approaches, which historically were only available to expert statisticians and data scientists, are becoming more approachable with the use of freely available packages such as *RecordLinkage,* an open source R package, which allows users to choose between traditional Fellegri-Sunder typed probabilistic, supervised ML-based, and unsupervised ML-based approaches [20]. Future techniques seeking to handle language translation, sound or video-based information, visual representations, and multiple-dimensional surfaces might also prove useful someday to projects integrating audio or video footage from EMS providers or surveillance cameras.

In practical use, many projects combine manual, deterministic, and probabilistic approaches, especially in stratified probabilistic approaches, where a deterministic match facilitates stratification or blocking to split a large population into smaller subgroups for probabilistic assessment [13]. Often, non-technical practical challenges represent the greatest barriers to successful record linkage, including the need for extensive data cleaning and standardization, imputation for data missingness, and ethical/privacy concerns [5].

2.1. Examples of Record Linkage Projects Involving EMS Information Systems

One of the earliest examples linking EMS data to hospital outcomes using modern probabilistic methods was presented by Dean *et al.*, who used probabilistic linking methods in Utah between EMS and hospital discharge records [21]. Since, there have been a wide array of specific uses of record linkage involving EMS data that we will discuss here.

2.1.1. Cardiac Arrest and Acute Coronary Syndrome

An early applied example of probabilistic record linkage techniques for EMS usage was conducted by investigators in Toronto, who were able to compare rates and survival outcomes of cardiac arrest survivors from different area hospitals by linking three databases: the Metro Toronto Ambulance database, the Canadian Institute of Health database, and the provincial Vital Statistics Information System database [22]. Their manuscript provides an excellent account of how to handle many of the data cleaning tasks that arise during record linkage. Another team working with the California EMS Information Systems database attempted to link EMS data to California's Office of Statewide Health Planning and Development Inpatient and Emergency Outcomes database using probabilistic linkage. This attempt demonstrated the challenges with very high degrees of missingness often found in EMS data sets [23]. A United Kingdom-based study used deterministic record linkages between EMS, ED, intensive care, and administrative data sets to evaluate the feasibility of registry-based follow-up on cardiac arrest patients, who had been cluster randomized to receive mechanical assisted chest compressions vs. usual therapy [24].

In North Carolina, deterministic matching was used to link a Pre-Hospital Medical Information System with the Acute Coronary Treatment and Intervention Outcomes Network Registry. The goal was to assess if hospitals participating in the Regional Approach to Cardiovascular Emergencies (RACE) quality improvement program achieved an adequate performance in guideline suggested timeliness for percutaneous coronary intervention [25].

A mixed probabilistic and deterministic record linkage study in New Jersey facilitated the joining of patient-level EMS data, hospital utilization, and death information to evaluate outcomes between centers adopting therapeutic hypothermia and those who did not using a wide array of process and clinical outcomes [26].

2.1.2. Trauma and Traffic Safety

Cook et al. in 2000 probabilistically linked Utah hospital discharge databases with Utah Department of Transportation data on reportable motor vehicle crashes to facilitate a logistic regression analysis comparing outcomes by age adjusted for alcohol or drug involvement and features related to the crash itself, yielding valuable insight into accident patterns and predictors of morbidity and mortality [27]. Subsequently, the same team, working with the US National Highway Traffic Safety Administration (NHTSA) between 1993 and 2013, created the Crash Outcomes Data Evaluation System (CODES), which linked crash, vehicle, and behavior characteristics to medical and financial outcomes by collaborating with States to develop data linkage programs, with systems becoming autonomous in 2013 [28]. This system used an augmented probabilistic approach that included multiple imputation steps for missing data with a Markov Chain Monte Carlo step to ensure sample independence [29]. Ultimately, they used this system with data from 11 States to evaluate medical outcomes based on motor vehicle collision circumstances, including the impact of age on clinical outcomes, the impact of safety restraints on children, the impact of helmets on motorcycle head injuries, and the impact of differences in driver's education program rating on teen injuries. This project also yielded new techniques for generating test data for assessing record linkage algorithms and for Bayesian record linkage approaches for multiple imputation of missing links [30,31].

A Massachusetts team used rule-based deterministic linking to match the Massachusetts Crash Data System (CDS) and Massachusetts Ambulance Trip Record Information System (MATRIS), facilitating evaluation of the relationships between crash characteristics and injury patterns [32].

Bianchi Santiago *et al.* compared deterministic to probabilistic matching algorithms when working to link the Puerto Rican CRASH database to data from a state-run personal injury insurance company [33]. They found a 20% improvement using probabilistic matching and used the linked records to develop a crash-cost estimation model for traffic-related personal injuries.

A study in Alabama, USA, demonstrated that deterministic linkage can facilitate the comparison of Glasgow coma scale (GCS) scores between EMS and trauma center clinicians [34]. Differences between EMS and hospital staff assigned GCS scores were seen in the patient population with moderate to severe head injury, suggesting some improvement in the patient's condition during transport.

Newgard combined deterministic matching of a unique identifier with probabilistic matching of secondary data to link a community EMS service database with a state trauma registry with high accuracy [35]. He concluded that matching ambulance records to a trauma registry without the use of patient identifiers is possible using probabilistic linkage. However, sensitivity of identifying true matches depends on the number and type of common variables (e.g., age, gender, race, county, hospital, date, rural setting, call and arrival times, mechanism, penetrating injury, vital signs, intubation, and intoxication) included in the analysis. Subsequently his team expanded these efforts to include multiple imputation, ultimately generalizing the process across 11 trauma registries and 94 EMS agency databases [36].

2.1.3. Mental Health and Substance Abuse

A team in Melbourne, Australia used deterministic linking techniques to match records from the Melbourne Metropolitan Ambulance Service with National Death Index data, yielding actionable knowledge about the association between non-fatal opioid overdoses attended by EMS teams and subsequent fatal overdose, providing an opportunity for outreach and prevention [37]. More recently, an iterative deterministic approach with substantial clerical review facilitated improved linkage of EMS runs with emergency department data in patients receiving naloxone for presumed opioid overdose over prior efforts, facilitating outcome studies [38].

In Scotland, researchers used the National Health Service Information Systems Division Unscheduled Care Data Mart, which is created via a mixed probabilistic and deterministic approach to integrate Scottish EMS data, EDs, inpatient and psychiatric admission data, and death records – providing insight into the frequency of patient medical contact prior to a self-harm event or a suicide attempt [39].

2.1.4. Stroke Care

A team striving to build a better registry for evaluating stroke care quality compared deterministic and probabilistic matching when linking Michigan's EMS Information System with the Michigan Coverdell Acute Stroke Registry – matching 46% of records using deterministic matches and 68% using probabilistic matches [40]. Another effort in North Carolina using deterministic linkage to match the North Carolina Emergency Medical Services Data System to the North Carolina Stroke Care Collaborative database yielded a 63% potential match rate with a positive predictive value of 89% [41].

2.1.5. EMS Protocol Safety and Quality Improvement

A Western Australian team used probabilistic matching techniques with multiple passes followed by clerical review of doubtful matches to match EMS data to hospital morbidity, emergency department, and death register data to facilitate a safety evaluation of patients receiving methoxyflurane in the pre-hospital setting [42,43].

The University of Rochester linked EMS dispatch data to hospital records using probabilistic matching, which revealed a small subset of dispatch codes that were highly associated with admission or mortality, facilitating better allocation of resources using dispatch codes [44].

2.1.6. Disease Severity Prediction

In King County, Washington, probabilistic linkage techniques were used to match EMS records to the Washington State Comprehensive Hospital Abstract Reporting System (CHARS), facilitating the derivation of a generalized non-traumatic, non-cardiac arrest disease severity prediction score for deployment by EMS services [45]. A similar illness severity score for the elderly was derived for seven counties in Oregon, Washington, Colorado, California, and Utah using EMS records probabilistically linked to hospital records [46].

2.1.7. Integrated Systems

An Oregon team successfully linked all EMS agencies in the state to state hospitalization and motor vehicle collision registry data using probabilistic matching, which required an immense effort to normalize and link data from 14 distinct electronic patient care record (ePCR) systems [47]. The same team worked in Oregon and Washington States to link nine databases, including EMS ePCR systems, state trauma registries, discharge registries, vital statistics registries, a Medicare claims database, and the Oregon Physician Order for Life-Sustaining Treatment registry [48]. They used probabilistic linkage and multiple imputation to match these data and handle missingness, which allowed for one year follow-up of this large cohort.

2.1.8. Machine Learning for Record Linkage in EMS

Several studies have demonstrated the feasibility of various supervised and unsupervised algorithms for record linking with EMS systems. A team from Beth Israel Deaconess in Boston used logistic regression with 5-fold cross-validation and L2 regularization to link Boston's EMS ePCR with their hospital's electronic health record (EHR) [49]. The coming years will likely reveal expansion in the use of these techniques in novel ways.

2.2. Future Steps

The prior sections demonstrate the myriad record linkage techniques investigators have harnessed to solve a multitude of problems and develop novel knowledge only possible by "marrying" various data sets with EMS data. The biggest challenges with current techniques are the extensive data cleaning and validation efforts required to facilitate entity resolution. As a result, real-time record linkage to date has required pre-defined unique fields that can be deterministically linked (as are used in most health information exchanges). Advances in matching techniques, especially unsupervised classifiers that can be iteratively updated with new data, may reduce the need for data cleaning and supervised turning. However, there remain privacy, information standard, and cultural barriers that make this process challenging. The United States National EMS Information System (NEMSIS) – one of the many projects of the NHTSA (National Highway Traffic Safety Administration) – was designed as "a standardized system for electronic documentation and sharing of EMS data [that] allows local agencies to measure performance and support more effective quality improvement programs. Standardized national data have also facilitated a growth in EMS research, which is essential to ensuring EMS continues to provide the best care to patients across the country" [50]. The development of NEMSIS has provided a road map for the design of ePCR software that specifically accounts for the challenges in data collection during pre-hospital care while integrating nationally and internationally scalable EMS data standards. Many hoped that NEMSIS would facilitate tight integration between EMS and hospital systems. While integration efforts continue, most EMS ePCR systems remain uncoordinated with hospital systems, resulting in data silos that will require joining via the use of record linkage techniques for the foreseeable future.

Ample opportunity awaits to optimize the care and outcomes of patients managed in the pre-hospital arena with improved entity resolution. Incentives must arise to compel system designers to link data, in the interest of assuring quality of care while providing valuable sources of data to public health professions. We want to leave you here with a scenario that demonstrates the potential for the future – a patient with heart failure is managed by a small community health care system and occasionally receives specialty care at an academic medical center. The community paramedics are dispatched to a post-discharge follow-up visit entailing history taking, physical examination, 12-lead electrocardiography, vital sign and weight checks, and a medication reconciliation. While on scene, the paramedics seamlessly connect with both the local community health system and the academic medical center to review the recent clinical course. They document their findings so that the patient's on-campus providers can review the visit. The patient's medications are adjusted in real-time, keeping the patient safely at home.

References

[1] Meisel ZF, Shea JA, Peacock NJ, et al. Optimizing the patient handoff between emergency medical services and the emergency department. *Annals of emergency medicine.* 2015;65(3):310-317. e311.

[2] Schooley B, Hikmet N. Design of an enterprise architecture for electronic patient care record (ePCR) information exchange in EMS. 2013.

[3] Park E, Kim JH, Nam HS, Chang H-J. Requirement analysis and implementation of smart emergency medical services. *IEEE Access.* 2018;6:42022-42029.

[4] Shah GH, Lertwachara K, Ayanso A. Record linkage in healthcare: Applications, opportunities, and challenges for public health. *International Journal of Healthcare Delivery Reform Initiatives (IJHDRI).* 2010;2(3):29-47.

[5] Asher J, Resnick D, Brite J, Brackbill R, Cone J. An introduction to probabilistic record linkage with a focus on linkage processing for WTC registries. *International journal of environmental research and public health.* 2020;17(18):6937.

[6] Martin TJ, Ranney ML, Dorroh J, Asselin N, Sarkar IN. Health information exchange in emergency medical services. *Applied clinical informatics.* 2018;9(04):884-891.

[7] Gunderson MR, Florin A, Price M, Reed J. NEMSMA Position Statement and White Paper: Process and Outcomes Data Sharing between EMS and Receiving Hospitals. *Prehospital Emergency Care.* 2020;25(2):307-313.

[8] Landman AB, Rokos IC, Burns K, et al. An open, interoperable, and scalable prehospital information technology network architecture. *Prehospital Emergency Care.* 2011;15(2):149-157.

[9] Christen P. Data Matching Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection. In: 1st ed. Berlin, Heidelberg: Springer Berlin Heidelberg : Imprint: Springer; 2012.

[10] Lindskou TA, Mikkelsen S, Christensen EF, et al. The Danish prehospital emergency healthcare system and research possibilities. *Scandinavian journal of trauma, resuscitation and emergency medicine.* 2019;27(1):1-7.

[11] Newcombe HB, Kennedy JM, Axford SJ, James AP. Automatic linkage of vital records. *Science.* 1959;130(3381):954-959.

[12] Fellegi IP, Sunter AB. A Theory for Record Linkage. *Journal of the American Statistical Association.* 1969;64(328):1183-1210.

[13] Clark DE. Practical introduction to record linkage for injury research. *Injury Prevention.* 2004;10(3):186-191.

[14] Gomatam S, Carter R, Ariet M, Mitchell G. An empirical comparison of record linkage procedures. *Statistics in medicine.* 2002;21(10):1485-1496.

[15] Jaro MA. Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida. *Journal of the American Statistical Association.* 1989;84(406):414-420.

[16] Jamieson E, Roberts J, Browne G. The feasibility and accuracy of anonymized record linkage to estimate shared clientele among three health and social service agencies. *Methods of information in medicine.* 1995;34(04):371-377.

[17] Roos LL, Walld R, Wajda A, Bond R, Hartford K. Record linkage strategies, outpatient procedures, and administrative data. *Medical care.* 1996:570-582.

[18] Karr AF, Taylor MT, West SL, et al. Comparing record linkage software programs and algorithms using real-world data. *PloS one.* 2019;14(9):e0221459.

[19] Ramezani M, Ilangovan G, Kum H-C. Evaluation of Machine Learning Algorithms in a Human-Computer Hybrid Record Linkage System. Paper presented at: AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering2021.

[20] Sariyar M, Borg A. The RecordLinkage Package: Detecting Errors in Data. *R J.* 2010;2(2):61.

[21] Dean JM, Vernon DD, Cook L, Nechodom P, Reading J, Suruda A. Probabilistic linkage of computerized ambulance and inpatient hospital discharge records: a potential tool for evaluation of emergency medical services. *Annals of emergency medicine.* 2001;37(6):616-626.

[22] Waien SA. Linking large administrative databases: a method for conducting emergency medical services cohort studies using existing data. *Academic Emergency Medicine.* 1997;4(11):1087-1095.

[23] Mumma BE, Diercks DB, Danielsen B, Holmes JF. Probabilistic linkage of prehospital and outcomes data in out-of-hospital cardiac arrest. *Prehospital Emergency Care.* 2015;19(3):358-364.

[24] Ji C, Quinn T, Gavalova L, et al. Feasibility of data linkage in the PARAMEDIC trial: a cluster randomised trial of mechanical chest compression in out-of-hospital cardiac arrest. *BMJ open.* 2018;8(7):e021519.

[25] Fosbøl EL, Granger CB, Peterson ED, et al. Prehospital system delay in ST-segment elevation myocardial infarction care: A novel linkage of emergency medicine services and inhospital registry data. *American heart journal.* 2013;165(3):363-370.

[26] DeLia D, Wang HE, Kutzin J, et al. Prehospital transportation to therapeutic hypothermia centers and survival from out-of-hospital cardiac arrest. *BMC health services research.* 2015;15(1):1-9.

[27] Cook LJ, Knight S, Olson LM, Nechodom PJ, Dean JM. Motor vehicle crash characteristics and medical outcomes among older drivers in Utah, 1992-1995. *Annals of Emergency Medicine.* 2000;35(6):585-591.

[28] Crash Outcome Data Evaluation System (CODES). U.S. National Highway Traffic Safety Administration. http://www.nhtsa.gov/crash-data-systems/crash-outcome-data-evaluation-system-codes. Accessed 09, 2021.

[29] Cook LJ, Thomas A, Olson C, Funai T, Simmons T. *Crash Outcome Data Evaluation System (CODES): An examination of methodologies and multi-state traffic safety applications.* U.S. Department of Transportation, National Highway Traffic Safety Administration;2015.

[30] McGlincy MH. Using test databases to evaluate record linkage models and train linkage practitioners. *Proceedings of the 29th American Statistical Association, Survey Research Method Section, Seattle, WA.* 2006:3404-3410.

[31] McGlincy MH. A Bayesian record linkage methodology for multiple imputation of missing links. Paper presented at: ASA Proceedings of the Joint Statistical Meetings2004.

[32] Tainter F, Fitzpatrick C, Gazillo J, Riessman R, Knodler Jr M. Using a novel data linkage approach to investigate potential reductions in motor vehicle crash severity–An evaluation of strategic highway safety plan emphasis areas. *Journal of Safety Research.* 2020;74:9-15.

[33] Bianchi Santiago JD, Colón Jordán H, Valdés D. Record linkage of crashes with injuries and medical cost in Puerto Rico. *Transportation research record.* 2020;2674(10):739-748.

[34] Kerby JD, MacLennan PA, Burton JN, McGwin Jr G, Rue III LW. Agreement between prehospital and emergency department glasgow coma scores. *Journal of Trauma and Acute Care Surgery.* 2007;63(5):1026-1031.

[35] Newgard CD. Validation of probabilistic linkage to match de-identified ambulance records to a state trauma registry. *Academic Emergency Medicine.* 2006;13(1):69-75.

[36] Newgard C, Malveau S, Staudenmayer K, et al. Evaluating the use of existing data sources, probabilistic linkage, and multiple imputation to build population-based injury databases across phases of trauma care. *Academic Emergency Medicine.* 2012;19(4):469-480.

[37] Stoové MA, Dietze PM, Jolley D. Overdose deaths following previous non-fatal heroin overdose: record linkage of ambulance attendance and death registry data. *Drug and alcohol review.* 2009;28(4):347-352.

[38] Fix J, Falls D, Proescholdbell S, Ising A, Fernandez T, Waller AE. Optimization of Linkage between North Carolina EMS and ED Data: EMS Naloxone Cases. *Online Journal of Public Health Informatics.* 2019;11(1).

[39] Duncan EA, Best C, Dougall N, et al. Epidemiology of emergency ambulance service calls related to mental health problems and self harm: a national record linkage study. *Scandinavian journal of trauma, resuscitation and emergency medicine.* 2019;27(1):1-8.

[40] Oostema JA, Nickles A, Reeves MJ. A Comparison of Probabilistic and Deterministic Match Strategies for Linking Prehospital and in-Hospital Stroke Registry Data. *Journal of Stroke and Cerebrovascular Diseases.* 2020;29(10):105151.

[41] Mears GD, Rosamond WD, Lohmeier C, et al. A link to improve stroke patient care: a successful linkage between a statewide emergency medical services data system and a stroke registry. *Academic Emergency Medicine.* 2010;17(12):1398-1404.

[42] Sprivulis P, Silva JAD, Jacobs I, Jelinek G, Swift R. ECHO: the Western Australian emergency care hospitalisation and outcome linked data project. *Australian and New Zealand journal of public health.* 2006;30(2):123-127.

[43] Jacobs IG. Health effects of patients given methoxyflurane in the pre-hospital setting: a data linkage study. *Open Emerg Med J.* 2010;3:7-13.

[44] Hettinger AZ, Cushman JT, Shah MN, Noyes K. Emergency medical dispatch codes association with emergency department outcomes. *Prehospital Emergency Care.* 2013;17(1):29-37.

[45] Seymour CW, Kahn JM, Cooke CR, Watkins TR, Heckbert SR, Rea TD. Prediction of critical illness during out-of-hospital emergency care. *Jama.* 2010;304(7):747-754.

[46] Newgard CD, Holmes JF, Haukoos JS, et al. Improving early identification of the high-risk elderly trauma patient by emergency medical services. *Injury.* 2016;47(1):19-25.

[47] Newgard CD, Zive D, Malveau S, Leopold R, Worrall W, Sahni R. Developing a statewide emergency medical services database linked to hospital outcomes: a feasibility study. *Prehospital Emergency Care.* 2011;15(3):303-319.

[48] Newgard CD, Malveau S, Zive D, Lupton J, Lin A. Building a longitudinal cohort from 9-1-1 to 1-year using existing data sources, probabilistic linkage, and multiple imputation: a validation study. *Academic Emergency Medicine.* 2018;25(11):1268-1283.

[49] Redfield C, Tlimat A, Halpern Y, et al. Derivation and validation of a machine learning record linkage algorithm between emergency medical services and the emergency department. *Journal of the American Medical Informatics Association.* 2020;27(1):147-153.

[50] National EMS Information System (NEMSIS). National Highway Traffic Safety Administration (NHTSA)'s Office of EMS. https://www.ems.gov/projects/nemsis.html. Accessed 11, 2021.