Rutgers

Center for State Health Policy

A Unit of the Institute for Health, Health Care Policy and Aging Research

Technical Report on the Creation of a Linked Prehospital, Hospital, and Mortality Database in New Jersey

Derek DeLia, Ph.D.
Jose Nova, M.S.
Kristen Lloyd, M.P.H.
Daisuke Goto, B.S., B.A.

September 2013

Project Team

<u>Rutgers Center for State Health Policy</u>

Derek DeLia, Ph.D. (Principal Investigator)
Associate Research Professor

Joel C. Cantor, Sc.D.
Distinguished Professor and Director

Jose Nova, M.S.
Research Analyst

Kristen Lloyd, M.P.H.
Senior Research Analyst

Dorothy Gaboda, Ph.D.
Associate Director for Data Analysis

Daisuke Goto
Ph.D. Candidate – Department of Economics

<u>NJ Department of Health – Center for Health Statistics</u>

Ping Shi Zhang, M.S.
Health Data Specialist
New Jersey Department of Health

<u>Clinical Partners</u>

Henry E. Wang, M.D., M.S.
Associate Professor
Department of Emergency Medicine
University of Alabama at Birmingham

Jared Kutzin, R.N., M.P.H., D.N.P., E.M.T.
Director of Simulation, Winthrop University Hospital
Staff Nurse, Englewood Hospital and Medical Center (NJ)

Mark Merlin, D.O., E.M.T.-P., F.A.C.E.P.
Adjunct Associate Professor
University of Medicine and Dentistry of New Jersey Piscataway
Attending, Emergency Medicine, Newark Beth Israel Medical Center

This project was supported by grant number R01HS020097 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality.

Technical Report on the Creation of a Linked Prehospital, Hospital, and Mortality Database in New Jersey

Derek DeLia, Ph.D., Jose Nova, M.S., Kristen Lloyd, M.P.H., and Daisuke Goto, B.S., B.A.

Hospital billing records serve as an important resource for conducting a wide variety of health service research and health policy studies. These records are typically maintained by state health departments and are often accessed through the Agency for Healthcare Research and Quality (AHRQ) Healthcare Cost and Utilization Project (HCUP) databases. Despite their common use, however, the depth and scope of research that can be conducted with these data are limited by the lack of clinical detail that is common in administrative records. In response, AHRQ funded several projects to enhance the clinical content of hospital billing data with particular emphasis on building databases to support comparative effectiveness research (CER) projects that are needed to guide more effective healthcare delivery in the United States.

As part of this effort, AHRQ funded a collaborative of academic researchers, state agencies, and clinical partners to create a linked database of prehospital, hospital, and mortality records for the state of New Jersey. The collaborative was led by the Rutgers Center for State Health Policy (CSHP) in partnership with the NJ Center for Health Statistics (CHS), NJ Office of EMS (OEMS), and clinical partners from University of Alabama – Birmingham, Newark Beth Israel Medical Center, and Englewood Hospital and Medical Center.

The linked database is designed to serve as a resource for CER studies on patient- and system-level interventions where prehospital and hospital services are tightly connected. These interventions include treatment of out-of-hospital cardiac arrest (OHCA), trauma, and critical illness. More broadly, this project responds to repeated calls from the Institute of Medicine (IOM) and the National Highway Traffic Safety Administration (NHTSA), and others to improve the evidence base for emergency care delivery, which remains underdeveloped in many ways (IOM 2007a, 2007b, 2007c; NHTSA 2001).

Because the use of emergency care is often a barometer for the performance of other parts of the health system (DeLia and Cantor 2009), the database is also designed to inform the planning and evaluation of broader healthcare delivery reforms (e.g., enhanced care coordination for frequent 9-1-1 callers). Moreover, the data linkage provides opportunities to enhance risk adjustment in existing hospital quality measures by using critical information from prehospital data (e.g., prehospital vital signs, response times) as well as information about post-discharge survival outcomes.

For a variety of reasons, NJ provides an ideal setting for the creation of a linked database of prehospital, hospital, and mortality records. For many years (predating HCUP), the state has maintained all-payer uniform billing (UB) records for inpatient discharges and same-day surgeries statewide. In 2004, NJ added emergency department (ED) visits that did not result in inpatient admission to its UB data collection. In 2005, the New Jersey Hospital Association (NJHA) began a joint initiative with the New Jersey Department of Health (DOH) to improve the accuracy of data elements recording patient race, ethnicity, and primary language spoken (NJHA 2008). The initiative included a set of guidelines to assist hospitals with uniform data collection and reporting procedures as well as training of hospital intake workers.

NJ is currently reforming its prehospital EMS delivery system statewide. In 2007, the state completed the legislatively mandated *EMS Systemwide Review*, which made 55 recommendations for EMS reform (TriData 2007). One of the major recommendations was for the Office of EMS (OEMS) (within the DOH) to create a statewide EMS data warehouse with the eventual goal of universal EMS agency participation. After completing a small pilot program in 2006, the OEMS began the process of statewide data collection in 2007.

NJ also maintains a vital records system that includes mortality data for all state residents. Through agreements with other states, NJ captures death records for state residents who died in other parts of the U.S.

This report documents the methodologies used to create NJ's linked prehospital, hospital, and mortality database for 2009–2010. (When it becomes available, data for 2011 will also be linked into the database.) It also provides insights and lessons learned to guide future use of the data and to inform similar efforts that may be undertaken in other states. The report has three specific aims:

1. Describe the component databases
2. Outline the data linkage and assembly processes
3. Illustrate additional data processing needed to conduct specific research projects using the example of emergency response to out-of-hospital cardiac arrest

Component Databases

The linked database is built from three separately maintained databases: 1) New Jersey Discharge Data Collection System (NJDDCS), 2) NJ EMS Data Warehouse (EMSDW), and 3) mortality records for NJ residents (Table 1). NJDDCS contains patient-level information for all inpatient, same-day surgery, and emergency department (ED) encounters that occur in NJ hospitals. These data are derived from hospital billing procedures and reflect the universe of services delivered. They also serve as the source of information contributed to AHRQ's HCUP-SID (State Inpatient Databases) and HCUP-SEDD (State Emergency Department Databases).

Table 1: Contents of Component Databases

Component Database	**Main Contents**
New Jersey Discharge Data Collection System (NJDDCS)	▪ Patient identifiers: Name, date of birth, Social Security Number ▪ Other patient information: Sex, race/ethnicity, residential zip code ▪ Dates of admission and discharge ▪ Discharge status (e.g., home, other hospital, nursing home, hospital, expired) ▪ Expected payer and detailed charges (e.g., ICU) ▪ Diagnoses ▪ Procedures
EMS Data Warehouse (EMSDW)	▪ Patient identifiers: Name, date of birth, Social Security Number ▪ Other patient information: Sex, race/ethnicity, residential zip code ▪ Symptoms, vital status, medical conditions, complaints ▪ Procedures attempted ▪ Cardiac arrest indicators & procedures ▪ Trauma indicators & procedures ▪ Medications administered ▪ EMS personnel credentials ▪ Type of response unit (e.g., basic and advanced life support) ▪ Dates & times of dispatch, scene arrival & departure, destination arrival ▪ Scene location and description ▪ Destination type (e.g., hospital) and associated identifiers
Mortality Records	▪ Patient identifiers: Name, date of birth, Social Security Number ▪ Other patient information: Sex, race/ethnicity, residential zip code ▪ Date of death ▪ Includes NJ residents who died in other states

The EMSDW contains patient-level information for EMS encounters within NJ. This database includes detailed information about EMS agencies and their personnel, and encounter-level data on patient vital signs, run times, diagnoses, and procedures performed. Participation in EMSDW is universal among Advanced Life Support (ALS) agencies (i.e., staffed by paramedics) and rapidly growing among Basic Life Support (BLS) agencies (i.e., staffed by basic emergency medical technicians (EMT-B)). BLS participation reached approximately 50% by 2009 (Halupke and Gruber 2009). Although NJ BLS agencies include a mix of professional and volunteer units, BLS participants in the EMSDW are predominantly professional units.

Data from the NJDDCS and EMSDW indicate whether a patient died in the hospital or prehospital setting, respectively, but do not provide information about survival after hospital discharge. This information is available from NJ's vital records system, which includes information for all deaths of NJ residents anywhere in the US.

Data Linkage and Processing

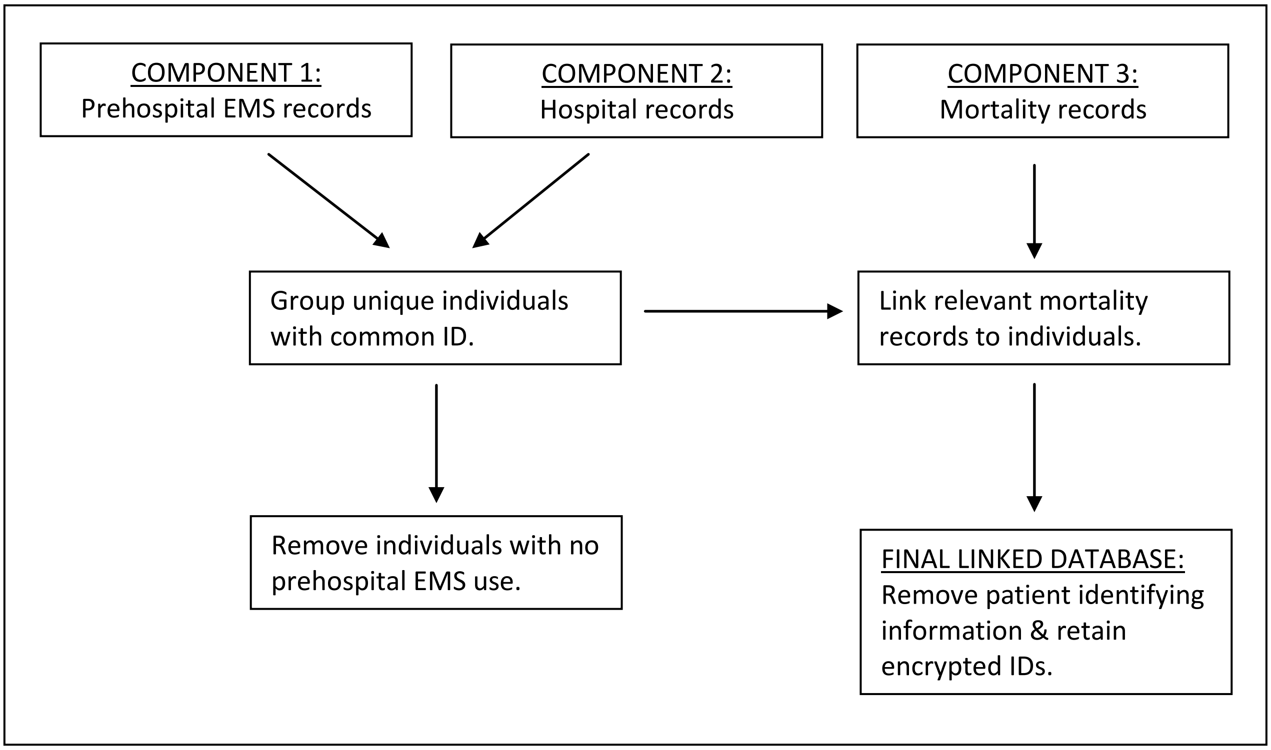
Overview of Linkage Strategy

Linkage of the component databases is enabled by the existence of fields recording name, date of birth (DOB), and Social Security Number (SSN) in each of the component databases (Table 1). Although these personal identifiers are not available on public use files, they are routinely available to the NJ CHS. Under a special agreement for this project, the CHS implemented the required linkages using patient identifiers under state auspices, and delivered a de-identified, linked database to Rutgers CSHP for further preparation and analysis. This agreement creates a clear "firewall" between confidential information in possession of the state CHS and the data available to Rutgers researchers and has been approved by Institutional Review Boards (IRBs) at DHSS and Rutgers University.

Despite the existence of personal identifiers across the three component databases, these identifiers are often incomplete and subject to a variety of recording errors (e.g., misspelled names, transposition of SSN digits). To overcome these challenges, the data were linked using LinkKing© software, which was developed under a grant from the Substance Abuse and Mental Health Services Administration (SAMHSA) (Campbell et al., n.d.). LinkKing uses a combination of deterministic and probabilistic methods to find potential record matches and classify the strength of matches into different confidence levels for building the final linked database.

The linkage process, outlined in Figure 1, proceeded in four steps. In step 1, unique individuals were found across the EMS and hospital records by grouping records that are "highly likely" to be for the same individual and assigning to them a common person-level ID number. In step 2, individuals who have hospital records but no EMS records were removed from consideration (since the data are being linked to study care episodes that begin with a prehospital EMS call). In step 3, a similar process was used to group mortality records with linked EMS-hospital records that are "highly likely" to be for the same individual and the prior person-level ID number was assigned to the relevant mortality record. In step 4, all patient identifiers were removed to produce the final linked database.

Figure 1: Linkage of Component Databases



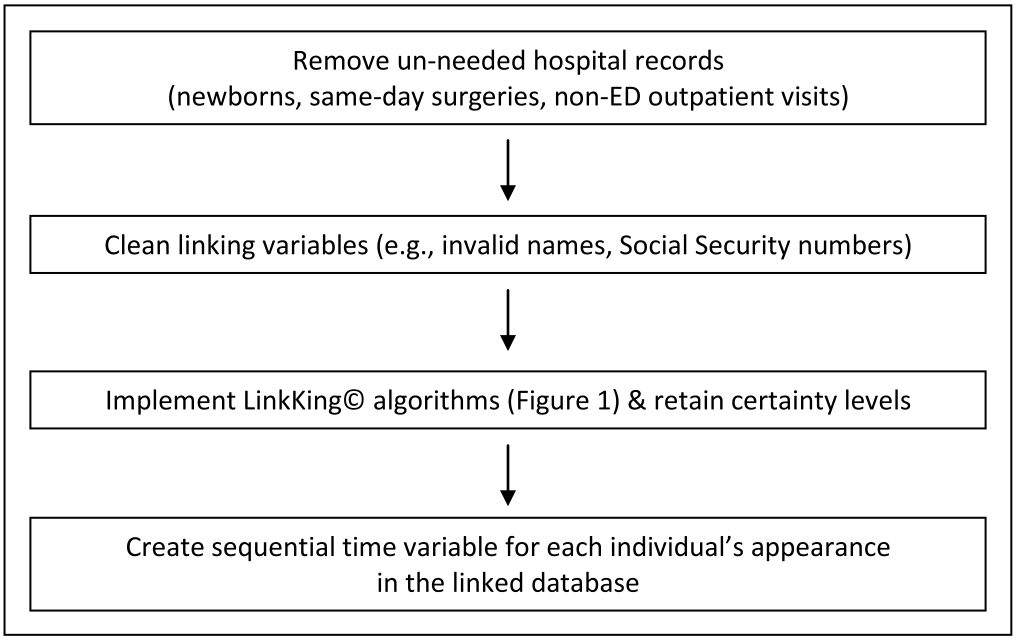
Linking Variables and Match Scores

In the execution of steps 1 and 3 above, LinkKing creates match scores that determine the strength of the match between a given record pair based on the available combinations of name, DOB, and SSN. Although these identifiers are the primary determinants of whether two records are highly likely to be a true match, LinkKing allows additional linking variables to be added at the user's discretion. For this linkage, we added three additional linking variables – patient sex, residential zip code, and race/ethnicity. Weighting scores given for agreement and disagreement were assigned based on the default options in LinkKing where name, DOB, and SSN exert the greatest influence on the final matching score for each record pair and the secondary variables exert smaller influence.

Additional Data Processing

Figure 2 summarizes a broader set of data processing procedures that were implemented to improve the efficiency and accuracy of the data linkage. First, hospital records that typically would not involve prehospital EMS transport transport (e.g., newborns, same-day surgeries, non-ED outpatient visits) were removed from consideration to reduce computer processing time needed to create match scores. This step reduced the total number of hospital records from 8,935,036 to 8,166,644. Second, clearly invalid values for linking variables were changed to missing (e.g., name="John Doe", SSN=999-99-9999).

Figure 2: Data Processing Procedures



Although LinkKing can accommodate missing values for some of the linking variables, no link can be established if name, DOB, and SSN are all missing from a particular record. As a result, some records were excluded from consideration for linkage. Although this problem did not occur in any hospital or mortality records, it reduced the number of EMS records that could be considered for linkage from 1,253,208 to 899,318 (a reduction of 28.2%). This reduction is not surprising, since EMS personnel often transport patients who are unconscious or in critical conditions where no time can be allowed for administrative data collection. But as explained below, some EMS incidents involve multiple EMS units. Therefore, if the record from at least one of these units includes the necessary patient information, the elimination of records from the other units would not prevent the incident from being included in the linkage.

Quality of Record Matches

After taking the steps above, we implemented LinkKing's record matching process as follows. The first time an individual appeared in the EMS dataset, their record was labeled as the "reference record". Then match scores were derived between the reference record and other records that might belong to the same individual. These match scores were converted to one of six certainty levels that are defined by LinkKing to assess the quality of linkage between pairs of records. Levels 1-3 are considered extremely rigorous with minimal invalid links. Linked records at Level 4 are also highly likely to be valid but LinkKing recommends manual review of these records where possible to ensure that some clearly invalid links are not retained. Matched records at higher-numbered certainty levels are often valid but the chances of invalid links are sufficiently high that these levels should not be used without manual review to screen out clearly inappropriate matches. Given the large size of the databases involved and the study team's lack of direct access to personal identifiers, it was not possible to conduct manual reviews. Thus, matched records at certainty levels exceeding 4 were considered not matched.

Record pairs can also be matched through a process called "cross linkage." For example, suppose there is no direct match between record A and record B but there is a direct match between each of these two records and a third record, called record C. In this situation, LinkKing would consider records A and B to be cross-linked and include them in the final set of linked records. Table 2 shows the linkage classifications for both stages of the linkage process described above.

Table 2: Final Linkage Classifications

	Initial linkage of prehospital EMS and hospital records		Subsequent linkage of mortality records to linked prehospital EMS and hospital records	
Linkage classification	**Number of records**	**Percentage of records**	**Number of records**	**Percentage of records**
Reference record	514,656	57.2%	N/A	N/A
Level 1	295,746	32.9%	51,494	91.38
Level 2	63,573	7.1%	3,019	5.36
Level 3	7,689	0.9%	743	1.32
Level 4	7,321	0.8%	1,088	1.93
Cross linked	10,329	1.2%	10	0.02
Total	899,314	100%	56,354	100%

Accounting for Patients with Repeated Episodes of Care

Under the data use agreements obtained to implement the data linkage, the Rutgers CSHP was not given access to exact dates of service, which are considered confidential (although year and month are provided). To understand the sequence of repeated appearances by the same patients over time, the Rutgers CSHP worked with the CHS to create a variable that identifies the first appearance of the patient in the database and another variable to indicate the number of days since the first appearance for all subsequent records. These variables in combination with the unique patient ID were used to associate each EMS transport with the appropriate hospital stay and to create time-based utilization or outcome measures (e.g., 30-day hospital readmission, 90-day mortality). A similar strategy was used to identify separate EMS incidents for the same individual.

A potential drawback to this approach is that a patient with several distinct EMS episodes in a single day would be indistinguishable from a patient who had one episode that attracted multiple EMS units. This drawback is fairly minor, however, since multiple calls on a single day are very rare.

Final Results of Data Linkage

The number of observations in the final linked database for different units of analysis are shown in Table 3. Although most EMS incidents involve one ambulance unit, 2-unit responses are also fairly common (Table 4). Incidents with 3 or more units are possible but rare. Nevertheless, some incidents, particularly those of greater severity (e.g., cardiac arrest, major motor vehicle accidents), are more likely to have a multiple-unit response than others. In these situations, information about a specific patient encounter might be spread over multiple records. Multiple-unit responses also add complexity to the verification of appropriate linkages between EMS incidents and hospital records (i.e., where at least one applicable EMS record indicates that the patient was treated by EMS and transported to the hospital).

Table 3: Number of Observations for Different Units of Analysis in the Final Linked Database, 2009–2010

Unit of Analysis	Number of Observations
Total EMS records	899,314
Records with EMS treatment and transport	706,584
Subset of above records linked to a hospital record	535,771
Unique EMS incidents	780,297
Unique EMS incidents linked to a hospital record	490,068
Unique individuals	512,560
Unique individuals with linked mortality record	56,354

The resolution of issues raised by multiple-unit responses is often study specific. In the next section, we provide examples of specific strategies that were used to extract needed information from multiple units attending to the same patient.

Table 4: EMS Units per Incident, 2009–2010

EMS Units	Number of Incidents	Percentage of Incidents
1	664,657	85.2%
2	112,601	14.4%
3	2,731	0.35%
4 or more	308	0.04%

Additional Data Processing for Cardiac Arrest Analysis

After creating the linked database, further data processing was conducted to create an analytically useful dataset for a study of out-of-hospital cardiac arrest (OHCA). Although the data processing decisions were made specifically for the OHCA analysis, they are included in this report to illustrate the kinds of decisions that would be needed in other similar applications of linked prehospital, hospital, and mortality data.

Identification of Cardiac Arrest Cases

Several variables from the prehospital EMS data were used to identify OHCA cases. Specifically, at least one of the following conditions had to be met to designate a case as OHCA:

1. Cardiac arrest was identified in a designated field
2. At least one procedure code indicated that CPR or a defibrillator was used
3. The first monitored cardiac rhythm was recorded as ventricular fibrillation, ventricular tachycardia, pulseless electrical activity, or asystole.

With this procedure, 8,901 OHCA incidents applying to 8,886 unique individuals were identified on the basis of 12,057 separate records.

Information from Multi-unit Responses

EMS responses to OHCA are more likely than EMS responses overall to involve multiple ambulance units (Table 5). In some cases involving multiple-unit response, we searched all relevant records to obtain required information such as implementation of a procedure or administration of medication. In other cases, such as those involving final patient diagnosis or disposition, it was important to extract one unique value from the most complete and reliable record.

Table 5: EMS Units per OHCA Incident, 2009–2010

EMS Units	Number of Incidents	Percentage of Incidents
1	5,838	65.6%
2	2,977	33.5%
3	79	0.9%
4	7	0.1%

In cases where one unique value was needed, we defined certain records as primary records using a 2-part hierarchy based on the type of EMS response and the service level of the responding ambulance unit. In part 1, information from EMS field responses took priority over information found in EMS records for transports between healthcare facilities. In part 2, we extracted information from the highest ranking EMS unit according to the Center for Medicare and Medicaid Services (CMS) service level, which indicates the clinical level of care and determines Medicare payment for ambulance services. If the required information was found at the highest service level, the search would end. If the information was unavailable at the first level, then we searched the next level, and so on. (If two or more records were tied for the highest CMS service level, we chose one at random to decide which should be searched for information first.)

Table 6: Hierarchy for Defining Primary OHCA Records in Multiple-Unit Responses

		OHCA incidents with multiple-unit response		**OHCA incidents with single-unit response**	
Hierarchy level	**Record type***	**Number of incidents**	**Percentage of incidents**	**Number of incidents**	**Percentage of incidents**
1	ALS2	1,547	50.5%	3,566	61.1%
2	ALS1 Emergency	441	14.4%	1,507	25.8%
3	ALS1	1	0.03%	0	0.0%
4	Specialty Care Transport	0	0%	0	0.0%
5	BLS Emergency	1,043	34.1%	689	11.8%
6	BLS	6	0.2%	2	0.0%
7	Rotary Wing Helicopter	8	0.3%	5	0.1%
8	Inter-facility transfer	1	0.03%	3	0.1%
9	Type unknown	16	0.5%	66	1.1%
Total		**3,063**	**100.0%**	**5,838**	**100.0%**

*Defined by Medicare and Medicaid Services (CMS) reimbursement categories.

Table 6 shows the specific hierarchical categories and their frequency of occurrence in the final linked database. For comparative purposes, the Table shows the distribution of record types for single-unit responses versus primary record types for multiple-unit responses. In both cases, almost all records are classified as ALS2, ALS1-emergency, or BLS-emergency although BLS-emergency plays a larger role in multiple-unit responses.

EMS Time Intervals

As mentioned above, the data use agreement with the NJ DOH did not allow exact dates and times to be included in the final linked database. Therefore, the CHS calculated important time intervals for every EMS record in the database – specifically, response time, on-scene time, and patient transport time. For single-unit responses, these intervals can be used in a straightforward way. For multiple-unit responses, it was impossible to determine which unit arrived first and whether on-scene times overlapped. In these circumstances, we used time intervals from the primary record defined as the highest ranking record in the hierarchy in Table 6. As shown in Table 7, these times do not differ substantially from the corresponding times for single-unit responses.

Table 7: EMS Time Intervals in Single-Unit versus Multiple-Unit Records for OHCA Incidents

			Percentile				
Time Interval in Minutes	**Number**	**Mean**	**10th**	**25th**	**50th**	**75th**	**90th**
Response							
Single-unit	5,826	9.3	4.0	6.0	8.0	12.0	15.0
Multiple-unit	3,061	9.5	3.0	5.0	7.0	10.0	13.5
On-Scene*							
Single-unit	4,277	19.7	9.0	13.0	18.0	24.0	30.8
Multiple-unit	2,346	19.6	8.0	13.0	18.0	23.7	30.0
Transport*							
Single-unit	3,799	9.4	3.0	5.0	8.0	12.2	17.0
Multiple-unit	2,186	9.4	3.0	4.0	7.0	10.0	15.0

*Fewer observations are available for on-scene and transport times due primarily to the removal of patients who died at the scene.

Conclusion

The linked prehospital, hospital, and mortality database for NJ is designed to serve as an ongoing resource for comparative effectiveness research and other health services research and policy analyses. At the time of this writing, data from 2011 is being prepared for inclusion in the linked database. Other states contemplating prehospital and mortality linkages with their hospital billing records may use similar methods and benefit from lessons learned from the NJ data linkage experience.

References

Campbell K, D Deck, A Cox, and C Broderick. n.d. *The LinkKing User Manual: Version 5.2*. http://the-link-king.com/user_manual.zip.

DeLia D, and J Cantor. 2009. *Emergency Department Utilization and Capacity*. Research Synthesis Report, no. 17. Princeton, NJ: Robert Wood Johnson Foundation.

Halupke K, and D Gruber. 2009. "New Jersey: Using Research to Improve Practice and Policy." Presentation at Bridging the Gap between EMS and Health Services Research: A Conference for Researchers and Practitioners, New Brunswick, NJ, July.

IOM (Institute of Medicine). 2007a. *Emergency Care for Children: Growing Pains*. Washington, DC: National Academies Press.

———. 2007b. *Emergency Medical Services: At the Crossroads*. Washington, DC: National Academies Press.

———. 2007c. *Hospital-Based Emergency Care: At the Breaking Point*. Washington, DC: National Academies Press.

NHTSA (National Highway Traffic Safety Administration). 2001. *National EMS Research Agenda.* Washington, DC: U. S. Department of Transportation.

NJHA (New Jersey Hospital Association). 2008. "Health Research & Educational Trust of New Jersey (HRET): Research: Ongoing Projects; Patient Race, Ethnicity and Primary Data Collection; Background & Previous Studies." NJHA. http://www.njha.com/research/PREPLBackground.aspx.

TriData. 2007. *The State of New Jersey EMS System Review, DHSS, OEMS*. Arlington, VA: TriData.

Rutgers
Center for State Health Policy

Center for State Health Policy
Rutgers, The State University of New Jersey
112 Paterson Street, 5th Floor
New Brunswick, NJ 08901

p. 848-932-3105 f. 732-932-0069
cshp_info@ifh.rutgers.edu
www.cshp.rutgers.edu