

Nebraska Final Report for

State-based Cardiovascular Disease Surveillance Data Pilot Project

Principle Investigators:

Ming Qu, PhD

Public Health Support Unit Administrator

Nebraska Department of Health and Human Services

Ming.Qu@nebraska.gov

Jamie Hahn, MEd

Nebraska Cardiovascular Health Program Manager

Nebraska Department of Health and Human Service

Project Team Members:

Ying Zhang, PhD, Epidemiology Surveillance Coordinator, Public Health Support Unit

Jihyun Ma, MS, Health Surveillance Specialist, Public Health Support Unit

David DeVries, MA, Health Surveillance Specialist, Cardiovascular Health Program

Ge Lin, PhD, GIS coordinator, Public Health Support Unit

Project Summary

This CDC and CSTE funded pilot project linked 2005-2009 Nebraska state-wide EMS data, hospital discharge data and death certificate data. Integration and linkage of the three data source created new data sets with richer information that can be used by public health programs to inform decisions making and policy change.

In early April of 2010, the Public Health Support Unit and the Cardiovascular Health Program at Nebraska Department of Health and Human Services started to work on cardiovascular disease (CVD) related patients records from EMS, hospital discharge, and mortality data. This project has at least the following two public health benefits:

- 1) Develop and document methodology for data linkage, using public health accessible data;
- 2) Track CVD related morbidity and mortality in and out of healthcare settings.

The original proposed years of data used were from 2006 to 2008, but we were able to expand to 5 years of data from 2005 to 2009. In this way, we would be able to track more survivals and/or deaths over a prolonged period.

We essentially followed all planned schedule, and completed proposed activities in each stage. We used the probabilistic linkage method with most patient identifiers (e.g., first name, last name, date of birth, gender, and residence zip code). We linked CVD patients from EMS to the hospital discharge data and achieved a linkage rate of 89%. For hospital discharge data with known death, we had a linkage rate of 97.4%. We performed some initial linked data analysis and found that EMS responders tend to have highly consistent primary diagnosis (impression) with the

primary diagnosis obtained from the linked hospital discharge data. The linked dataset has already made some public health impact. Some initial results have been presented to the Nebraska Stroke Advisory Board in February 2011. Based on our model and findings, the epidemiology staff at the Division of Public Health started to analyze Electronic Health Records (EHR) data, and to develop a surveillance system which will allow tracking CVD related hospitalizations.

Linkage Methodology

The linkage process started with EMS data. The Nebraska EMS data collected from all EMS providers based on four reporting systems:

- 1) A statewide electronic reporting and uploading system (e-NARSIS), which allows 420 EMS agencies in Nebraska to collect and submit data to the state center system;
- 2) For those who opted not to use the system, we have NARSIS, a paper and key entry system;
- 3) Omaha EMS system. It is a legacy system for Omaha, the largest city in Nebraska;
- 4) Lincoln EMS system. It is a more recent EMS system for the city of Lincoln, the capital city of Nebraska.

The total number of records for the five year-period of 2005 to 2009 is 519,343.

To avoid potential disclosure, we performed linkage on site using Nebraska Hospital Association facilities. In this way, we do not have to perform the traditional way of data linkage without patient first and last names as the key linkage variable in our probabilistic linkage process. In addition, we can place more weights in first and last names with a high degree of accuracy. We brought both EMS data and death records to the central office of Nebraska Hospital Association, and the process that includes test runs, real linkage runs, and reassessment took about 3 weeks for 2 FTE. After the linkage process was completed, the linked dataset was released to the project team without personal identifiers.

EMS to hospital discharge data (HDD)

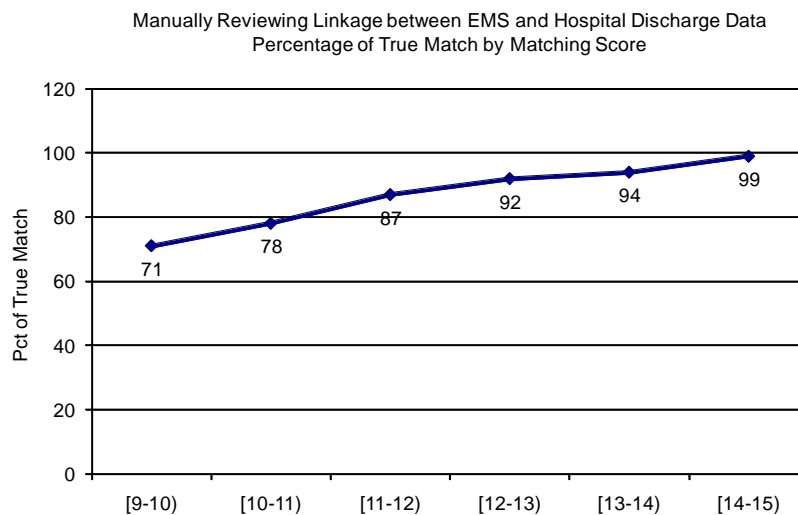
Before linking EMS data to hospital discharge data, we considered several factors.

First, we selected records of the EMS runs that sent patients to hospitals. In many cases, multiple ambulances responded to the same 911 call, but only one ambulance eventually sent the patient to a hospital. We used the variable “disposition” to check if a patient was sent to a hospital, died at scene or treated and released (those transported by private vehicles were not linked as they do not have a destination hospital). This process leads to 273,941 unique EMS runs that sent a patient to a hospital.

Second, we determined those 911 calls that related to CVD. One complication is that some systems use a text field, and some use ICD-9 codes. For those records with a text field, we used following text search *cardi, Chest pain, heart, cariac cardiac, cardiac, cest, chaest, cheast, chect ches, chet, chst, cjest, cvd* to derive CDV-like symptoms. For those using ICD-9 codes, we used following for CVD related symptom codes (ICD-9-CM 390-392, 393-398, 401, 402, 403,404,405,410-414, 415-417, 420-429, 430-438, 440-448, 451-459). Based on these text or code searches, we found a total of 49,796 CVD related records.

Finally, we used *first name, last name, date of birth, gender, residence ZIP Code, event date, and hospital code* as key variables for linkage. What is worth noting is that cleaning up hospital codes is critical, and eventually maintains a good list of legible hospital names by years. Hospital names often change, and sometime new hospitals are added; while EMS responders often write down hospital names in text rather than selecting from a list of codes.

We used LinkPlus to perform linkage. Once all parameters were set, the process ran about an hour, and yielded about half-million linked pairs. With an initial cut-off matching score set at 7, we randomly selected 100 pairs from each of the score range [9-10), [10-11), [11-12), [12-13), [13-14), [14-15) for manual review. Percentage of true matches based on manual review is shown in the following chart. Based on this chart and a group discussion, we decided to set the cut-off score at 12, since this was the point where percentage of true matches went over 90%. If multiple EMS records were linked to the same hospital records, only one link pair with the highest score will be retained. As a result, out of the 49,796 CVD related records, we matched 44,330 records to HDD, with a matching rate of 89%.



EMS to death data

From EMS data, we selected records that were indicated to be dead at scene, and obtained 2,851 records. Variables used for linkage were *first name, last name, date of birth, gender, residence zip code, event date, and event zip code*. The same manual review process as described earlier was applied. Finally we obtained 2,341 linked pairs, with a matching rate of 82%. Possible reason for a low linkage rate is that many deaths may not be reported to the Nebraska Vital Records, as they might have been transported to adjacent states for funerals or other arrangements.

Hospital discharge to death data

We followed the same procedure as described earlier. Two linkages between HDD and death data were performed. The first linkage was solely based on personal identifiers. We used *first name, last name, date of birth, gender, and residence zip code* as linkage variables. In this way, we were able to match those patients who did not die at hospitals, but later at home or other non-acute-care facilities. There was no match rate for this linkage, since the denominator was unknown.

The second linkage was based on person and event information. We added *date of death* in addition to the linkage variables used in the first linkage. Out of those with an indication of “expire” (24,368 records), we found 23,751 matches, with a matching rate of 97.4%.

Data Analysis and Dissemination Plan

We have linked Emergency Medical Service (EMS) data, hospital discharge data, and death data. Due to limited resources and time, we performed limited data analysis and the process is still ongoing on a funded basis. The most important barrier to analyzing the linked dataset is inconsistent coding among Nebraska EMS providers. When we merged EMS data from four sources to a single dataset, many variables were not consistently available, which greatly hindered our analysis capability. For instance, the providers use their own coding scheme to document patient conditions (e.g., chief complaint, primary impression) and service variables (e.g., procedures performed, and medication administered etc.), without further variable integration, we cannot use the linked data for analysis. Nevertheless, we piloted some analyses listed in the proposal.

First, the project intends to “increase knowledge of signs and symptoms for heart attack and stroke and the importance of calling 9-1-1”, and to “improve emergency response”, both of which are listed in the State Heart Disease and Stroke Prevention Programs priority areas. Hence, our initial analysis focused on the consistency of CVD diagnosis between EMS and HDD. Since we need to expand CVD from the EMS side to include potential CVD not captured by the EMS but identified by the HDD, the number may differ from the above linkage rate reporting. We compared the *primary impression* from EMS and the *primary diagnosis* from HDD. The results were listed in table 1. The table includes both ER and inpatient admissions. These two types of admissions were not exclusive; since there were about 63,192 (43.2%) inpatients went through ER admissions. Table 1 showed that the majority of both inpatient and ER admissions had a broadly consistent CVD diagnosis with the EMS primary impression, 79% for inpatients, and 91.7% for ER patients. However, there were some inconsistencies: 19% (11.8%+7.2%) for inpatients and 7.9% (5.7%+2.2%) for ER patients. The latter 7.9% is reassuring, as a low inconsistency rate between EMS and ER diagnoses was desirable. It may also help us to educate EMS responders to make efforts to reduce the inconsistency below 5% (note: we are planning to present our results to the Nebraska EMS annual meeting in April, 2011).

| | Inpatient admission | | Emergency room admission | |
|----------------------------|-----------------------------|--------------|-----------------------------|--------------|
| | EMS primary impression: CVD | | EMS primary impression: CVD | |
| HDD primary diagnosis: CVD | Yes | No | Yes | No |
| Yes | 818538 (79.0%) | 74163 (7.2) | 2110736 (91.7) | 131983 (5.7) |
| No | 121924 (11.8%) | 21442 (2.1%) | 50241(2.2) | 10110(0.4%) |

Note: percents in parentheses are based on the total matched records among four cells
The sample include all patients, not just Nebraska residents

Second, with these initial linkage rates, we can begin to answer some of CVD surveillance questions. For example, among 2007 ER patients with CVD as the primary diagnosis, 16.75% (10,110/60,351) were linked to EMS. In particular, among ER patients with “stroke” as the primary diagnosis, 30.22% (1,131/3,742) were linked to EMS, and 24.41% (2,046/8,382) of “coronary heart disease” were linked to EMS. We can therefore, examine EMS service time and treatment outcomes using the linked dataset. Please note that these results are preliminary, and are for demonstration purposes only.

Third, we also began some analysis from the HDD to death record linked data. One of Health People 2010 goals was to eliminate health disparities in terms of race, ethnicity, gender, geography, or socio-economic status. We had a corresponding goal in the CVD program. Since the race/ethnicity variables in the HDD do not have an acceptable quality, we used gender as an initial indicator to check. At this point, we are reluctant to release the results, as we found males were more likely to survive two years after their hospital admissions. For myocardial infarction (MI) patients, women were more likely to die in age groups 45-49, 50-54, 60-64, while results from all other age groups were not significant. The above results were based on the primary and secondary diagnoses. These results were basically retained for stroke patients too (Appendix I lists descriptive tables for both MI and Stroke patients). However, if we used death certificate data with “heart attack” as the primary and second causes of death, then those 60-64, and 70-74 men were more likely to die not having gone to the hospital. In addition, if we look at two-year survival among other CVD besides MI and stroke patients, those ages 30 and younger, and 50-94, in each age category, men were more likely to die than women after 2 years, so that is the reverse of what we had found for MI and stroke. We are stilling working on those seemingly inconsistent results, and hopefully present more coherent results in the Council of State and Territorial Epidemiologists (CSTE) conference in Pittsburg in June, 2011.

Based on our 2010 work, the division’s epidemiologists have begun to expand capability of syndromic surveillance to include chronic disease surveillance by using hospitalization data to identify and characterize cardiovascular disease (CVD). Nebraska has 92 hospitals that provide services for cardiovascular patients. The division’s epidemiologists have started to work with a number of hospitals by adding CVD indicators for inpatient hospital syndromic surveillance on cardiovascular disease. Part of funding is from CDC Public Health Infrastructure grant to Nebraska.

Status of All Project Objectives

We essentially followed our planned schedule, and completed proposed activities in each stage. We assembled the project team in April, 2010 and had monthly meetings among PIs and program coordinator. For more technical group, we had initially bi-weekly meetings for the first three months, and then met as needed.

Appendix I. Descriptive data for the initial analysis of MI and stroke patients in Nebraska

| Table A1 Descriptive statistics for MI Patients: 2005-09 | | | | |
|---|-------|-------|--------|-------|
| | Male | | Female | |
| N | 3557 | | 2151 | |
| Mean Age | 65.91 | | 76.21 | |
| Mean Stay Length | 3.89 | | 4.4 | |
| Mean Charge | 47706 | | 39747 | |
| Discharge Status from Hospital | N | % | N | % |
| Discharged home | 2127 | 59.8% | 802 | 37.3% |
| Discharged to other institution | 539 | 15.2% | 440 | 20.5% |
| Discharged to Nursing home | 203 | 5.7% | 309 | 14.4% |
| Dead at Discharge | 637 | 17.9% | 583 | 27.1% |
| Other Discharge | 51 | 1.4% | 17 | 0.8% |
| Type of Hospitalization | | | | |
| Inpatient | 2937 | 82.6% | 1867 | 86.8% |
| ER only | 450 | 12.7% | 213 | 9.9% |
| Other hospitalization | 170 | 4.8% | 71 | 3.3% |
| Primary Payer Code | | | | |
| Commercial Insurance | 1426 | 40.1% | 397 | 18.5% |
| Medicare | 1759 | 49.5% | 1637 | 76.1% |
| Medicaid | 74 | 2.1% | 45 | 2.1% |
| Self-pay | 177 | 5.0% | 52 | 2.4% |
| Other | 121 | 3.4% | 20 | 0.9% |
| Urban/Rural | | | | |
| Urban Metro | 1723 | 48.4% | 1016 | 47.2% |
| Micropolitan | 684 | 19.2% | 424 | 19.7% |
| Rural | 1150 | 32.3% | 711 | 33.1% |
| High blood pressure | | | | |
| Yes | 1698 | 47.7% | 1077 | 50.1% |
| No | 1859 | 52.3% | 1074 | 49.9% |
| Diabetes | | | | |
| Yes | 671 | 18.9% | 434 | 20.2% |
| No | 2886 | 81.1% | 1717 | 79.8% |

Table A2 Descriptive statistics for stroke Patients: 2005-09

| | Male | | Female | |
|---------------------------------------|----------|----------|----------|----------|
| N | 4287 | | 4658 | |
| Mean Age | 69.7 | | 74.67 | |
| Mean Stay Length | 4.8 | | 5.09 | |
| Mean Charge | 25796 | | 23133 | |
| Discharge Status from Hospital | N | % | N | % |
| Discharged home | 2235 | 52.1% | 1889 | 40.6% |
| Discharged to other institution | 827 | 19.3% | 1033 | 22.2% |
| Discharged to Nursing home | 492 | 11.5% | 891 | 19.1% |
| Dead at Discharge | 695 | 16.2% | 814 | 17.5% |
| Other Discharge | 38 | 0.9% | 31 | 0.7% |
| Type of Hospitalization | | | | |
| Inpatient | 3556 | 82.9% | 3841 | 82.5% |
| ER only | 443 | 10.3% | 486 | 10.4% |
| Other hospitalization | 288 | 6.7% | 331 | 7.1% |
| Primary Payer Code | | | | |
| Commercial Insurance | 1204 | 28.1% | 860 | 18.5% |
| Medicare | 2735 | 63.8% | 3514 | 75.4% |
| Medicaid | 151 | 3.5% | 136 | 2.9% |
| Self-pay | 104 | 2.4% | 73 | 1.6% |
| Other | 93 | 2.2% | 95 | 2.0% |
| Urban/Rural | | | | |
| Urban Metro | 2141 | 49.9% | 2283 | 49.0% |
| Micropolitan | 764 | 17.8% | 818 | 17.6% |
| Rural | 1382 | 32.2% | 1557 | 33.4% |
| High blood pressure | | | | |
| Yes | 2400 | 56.0% | 2707 | 58.1% |
| No | 1887 | 44.0% | 1951 | 41.9% |
| Diabetes | | | | |
| Yes | 882 | 20.6% | 895 | 19.2% |
| No | 3405 | 79.4% | 3763 | 80.8% |