Enhancement of Clinical Data Repository By Linkage to State Mortality Files

Robert D. Pates PhD, Kenneth W. Scully MS, Jonathan S. Einbinder MD, MPH, George J. Stukenborg PhD, Thomas A. Spraggs PhD, Bruce P. Dembling PhD
Department of Health Evaluation Sciences, University of Virginia Health System, Charlottesville, VA

Abstract
We report a method and preliminary results for merging state mortality data into a hospital Clinical Data Repository (CDR). A broadly inclusive set of potential matches for our CDR patient population was obtained from the state Death Registry at the Virginia Center for Health Statistics (VCHS). For each possible match, a total weighted score was calculated according to the level of agreement between seven identifiers. The scoring method was validated by consideration of that population of CDR patients known to have died in the hospital. We conclude that the method represents a practical way to enhance hospital data with important outcomes information.

Methods
Since the VCHS data are subject to certain confidentiality restrictions, we were unable to obtain a direct copy of the statewide data. We were authorized, however, to perform broad queries of the data (designed to produce many false positives) in an attempt to identify CDR patients within the statewide data. We performed three separate attempts to match the 509,162 patients of the CDR to individuals recorded in the VCHS Death Registry: a) Social security number; b) Last name / birth date; c) First name / last name. A total of 132,438 patients in the CDR were found to have a potential match in the VCHS data. Then, to assess the quality of each of these initial matches, a weighted scoring methodology was then applied to the following identifiers: SSN, last name, middle initial, first name, birth date, sex, and zip code.

Results
Once scored, the data were divided into one of three groups – "matches", "non-matches", and "uncertain". To validate the scoring process, we examined those scores obtained by patients reported in the CDR to have died in the hospital (n=5,220). Of these, 5,072 (97.2%) scored sufficiently high to be deemed a match. For the total sample (n=132,438), 28,008 (21.2%) patients were found to be matches. In addition to date of death, the VCHS data included in the CDR upon linkage comprise some 51 data elements including ICD codes for underlying cause of death, plus 3 accident codes, and up to 20 contributing causes of death (average approximately 3). The inclusion of mortality data from the state source is clearly advantageous for the study of all causes of death. However, the relative contribution of in-hospital death to the total mortality was found to vary considerably by clinical category of principal cause of death. For example, while the average was 18.6% (5,220 / 28,008), this value ranged from 12.7% for neoplasms to 40.0% for infections diseases.

Conclusions
Our internal control indicates that we can detect 97%+ of the total in-state deaths of CDR patients. Considering that 95% Virginia residents who die do so in-state, we estimate that we are able to include 92% of the total mortality among CDR patients. The combined clinical and cause of death data provide a substantial increase in epidemiological knowledge for the CDR patient population. Our preliminary results indicate that this may be particularly valuable for those clinical practices for treating patients with chronic conditions (e.g. cancer, heart disease). We conclude that the method yields a linked database that is acceptable for many research applications. Since all US states produce death records according to WHO and NCHS standards, this approach may be applicable in states other than Virginia.

References