E33 Standardizing a Large-Scale, Whole Body Computed Tomography (CT) Image Database

Heather J.H. Edgar, PhD*, Anthropology, MSC01 1040, 1 University of New Mexico, Albuquerque, NM 87131; and Shamsi Berry, PhD, University of Mississippi Medical Center, 2500 N State Street, Jackson, MS 39216

After attending this presentation, attendees will understand the benefits of standardizing data and limiting or eliminating free text answers in databases used in medicolegal investigation. Attendees will also be informed of the process by which standards are chosen and created.

This presentation will impact the forensic science community by exploring the benefits of standardized data formats, especially in medicolegal investigations. While data standardization meets several challenges in medicolegal studies not encountered in many other fields, standardization improves potential applications of large-scale data in epidemiology.

The Office of the Medical Investigator (OMI) is a state-wide, centralized medical examiner's office for New Mexico. In 2010, 5,249 deaths were routed to the OMI; 51% of those underwent autopsy, accounting for 35% of the deaths within the state.[1,2]

In 2010, the National Institute of Justice (NIJ) awarded the Center for Forensic Imaging at the OMI a grant to evaluate whether CT scans can supplant or supplement the traditional autopsy. To this end, every decedent that underwent an autopsy received a high-resolution, head-to-toe CT scan. Scanning has continued as standard practice at the OMI, even after the end of this project. So far, more than 11,000 whole body CT images have been created; however, as the original scanning was performed to address specific research projects, and not to create a general research resource, there are no indexes or tags associated with the images that might allow additional research.

In 2016, an NIJ grant was awarded to create a free-access Decedent CT Database, which will make these 11,000 whole body CTs available to the research public. Work is currently underway to contact next of kin and query the OMI's database to populate the new database. This query presents numerous challenges in that the majority of fields within the database are free text fields without any limitations. As a result, even sex can be recorded in multiple ways (e.g., Male, male, M, m), limiting the ability of future researchers to query efficiently. This is especially true as the complexity of the metadata increases. In order to combat this issue, the CT database being developed will utilize data standards, terminologies, and classification systems. Using a modified Delphi technique, experts from varying fields determined the 59 metadata variables to associate with each image.[3] This data is being captured from the OMI investigations database as well as next-of-kin interviews.

A step-by-step process is currently underway to determine the best standards to implement. First, Unified Medical Language System (UMLS) is searched to identify all of the standards for a particular concept (e.g., race) that exist. Each standard is then identified and compared for usefulness in this particular database. A standard can be implemented as is, modified, or all can be rejected. If all current standards are rejected for use in the database (or none are found), a new standard will be proposed and implemented.

A wide range of terminologies, standards, and classification systems may be implemented. This includes LOINC, SNOMED CT, CDISC, and multiple nursing standards. By comparing the standards and determining the best fit, we will eliminate duplication of standards, utilize current health care, and research terminologies will be eliminated.

For example, one of the variables included in the Decedent CT Database is medical diagnoses. This field is often recorded as free text in medicolegal investigation; however, this limits the value of this data in later research because the same condition could have hundreds of synonyms and abbreviations. Standards exist for recording actual data (Logical Observation Identifiers Names and Codes (LOINC), Clinical Data Interchange Standards Consortium (CDISC)), but the responses are of importance here. Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT) and the International Classification of Diseases (ICD) both record diagnoses; however, one standard is used to identify medical problems (SNOMED CT) within electronic health records, while the other is used for billing (ICD-10). Both standards require coding the response from next of kin or the information from the OMI database as a computer-readable number. Due to the vast number of SNOMED CT codes and the lack of everyday use of them by physicians and researchers, ICD-10 was chosen to represent the medical diagnoses of the decedents for this project.

The free-access Decedent Database is currently under development and is slated to be available by the end of 2018.

Reference(s):

1. OMI. Office of the Medical Investigator 2010 Annual Report. University of New Mexico: 2010.
2. US Census Bureau. *2010 US Census.* 2010 [cited 2011 January 19, 2011]; Available from: http://2010.census.gov/2010census/pdf/2010_Questionnaire_Info.pdf.
3. Berry, S. *Metadata Determination for a Cadaveric Collection. Master.* (Thesis).Albuquerque: University of New Mexico; 2014. Available from: UNM Digital Repository.

Data Standardization, Computed Tomography, Research Database

*Presenting Author