

Lessons learned from initial efforts at multicenter integration of electronic health data from adult congenital heart centers: The CONGENERATE experience.

Background: Efforts to establish multicenter research rely on uniformity of data acquisition. Integration of existing health information from electronic health records (EHR) into a homogeneous format can facilitate such research. We sought to identify major obstacles and possible solutions for data acquisition in adults with congenital heart disease (ACHD).

Methods: 5 US institutions were engaged in an NIH-funded effort to integrate EHR data into an ACHD specific database (CONGENERATE). Clinical researchers developed a standard definition of ACHD patients and listed important target data variables to acquire, with proposed priorities and formats for each. Database analysts reviewed which variables were currently available within their institutions' EHR, and in which formats. Major issues were summarized.

Results: From a list of 462 potential target variables, only 31 were identified as being available in discrete formats at 4 institutions (10 of which are personal identifiers). Despite a standard definition of an ACHD patient, many subjects initially included did not fit this standard. Demographic variables commonly consisted of a standardized list at each institution and required basic mapping of variables, though each center used a different coding type. For example, the standard NIH categories for race/ethnicity were adopted, but alternative categories were in use across institutions. Mapping was required, which in some cases was not a one-to-one relationship. Uploading of numeric values with differing units for anthropometric measurements was accommodated. The multifaceted EHRs required multiple architects at each participating institution with varying degrees of familiarity with segments of their particular EHR. Over 13,000 diagnostic codes were uploaded, generally listed in ICD-9 format, yet with inter-institutional variability as to whether codes were encounter-specific or broadly acquired. For ACHD specific codes, 2 of the 5 centers coded anatomic diagnoses in IPCC terms that could then be mapped to ICD-9, whereas reverse mapping could not be performed. Documentation on procedural results has not yet been achieved given prohibitive variability in reporting formats. For pharmacologic data, over seven different medication schemes were available, with institutional uncertainty regarding which was in use, or whether existing schemes were appended with center-specific coding. For cross-mapping, purchase of mappings is required through private organizations. Despite the challenges, addressing these obstacles has allowed for acquisition of data on over 2,171 patients to date.

Conclusions: Large-scale multicenter automated data acquisition from existing ACHD databases is feasible. Challenges stem from variability in data formats, coding schemes, and adoption of non-standard lists. The success of large-scale multicenter research will require ongoing efforts to maximize homogeneous data collection at many different sites, and overcome the types of obstacles discovered.