Using SNOMED CT to induce classifications for casemix analytics

Presenter: Dr Jeremy Rogers, NHS Health and Social Care Information Centre

Audience

Implementers and commissioners of healthcare analytics systems currently based on traditional clinical classification or dataset approaches.

Objectives

To demonstrate how casemix analytics can occur without the need to define a fixed classification or dataset coding frame *pre hoc.* SNOMED CT's rich taxonomy permits a system for aggregating real clinical data to be induced *post hoc* from the data itself.

Abstract

Standard healthcare reporting and statistical analytics approaches rely on the *pre hoc* development of fixed coding datasets[1-8] and/or monoaxial classifications of heathcare (ICD, CPT, OPCS etc). Limitations of this approach include: uncertainty which code to use when a case either spans more than one reporting category, or does not obviously correspond to any; a 'dual coding' or 'code translation' burden if clinical data is already coded using richer terminologies; inflexible reporting - the irreversible mapping of real clinical data onto a fixed classification prohibits further subanalysis of individual reporting categories.

A possible alternative *post hoc* approach is to leverage the richer taxonomy of SNOMED CT, the more expressive clinical coding it enables, and actual observed patterns of code use in order to induce a system of aggregating data, from the data itself: a set of abstract reporting categories is computed from the set of all possible SNOMED CT taxonomic ancestors of all SNOMED CT codes actually used by clinicians in the underlying data to be anlysed, weighted according to actual frequencies of individual code use.

Results will be presented of an experiment on real clinical data: an ED Department had experienced an unexplained 10% increase in caseload over 3 years. Each of 408,831 episodes of care occurring over the 3 years had been coded by attending clinicians using one of 12,022 different SNOMED CT codes. From this raw data, a data aggregation scheme comprising 113 abstract categories was induced. 14 of the induced categories exhibited a clear increase in casemix volume over the 3 years. However, because the induced categorization was both derived from and still linked to the raw and more detailed underlying SNOMED CT clinical data, cases assigned to one of just 36 individual diagnostic SNOMED CT codes were ultimately identifiable as collectively responsible for 60% of the observed rise in all-cause attendance.

References

1. NHS England Data Dictionary www.datadictionary.nhs.uk
2. NHS Wales Data Dictionary www.datadictionary.wales.nhs.uk
3. HDD Access/ 3M Healthcare Data Dictionary www.hddaccess.com
4. New Hampshire State Cancer Incidence Data Collection Manual
   http://www.dhhs.state.nh.us/dphs/hsdm/cancer/documents/cancerdictionary.pdf
5. National Data Dictionaries (Australia) meteor.aihw.gov.au
6. New South Wales Data Dictionaries www.cherel.org.au/data-dictionaries
7. New Zealand Ministry of Health Data Dictionaries
   www.health.govt.nz/nz-health-statistics/data-references/data-dictionaries
8. Palestinian Health Data Dictionary www.moh.ps/?lang=1&page=4&pid=100