7.2.1 Trend Analysis

Trend analysis is the practice of collecting information and attempting to spot a pattern, or trend, in the information. Trend analysis often refers to techniques for extracting an underlying pattern of behavior in a time series, which would otherwise be partly or nearly completely hidden by noise.

Detecting changes of either incidence or prevalence of a particular disease, treatment, procedure or intervention over time has major utility for population health monitoring, prediction of demand and effective resource allocation at enterprise and national levels. One challenge that is encountered when analyzing routinely collected patient data for trends, is distinguishing minor changes in coding style from real changes in disease incidence. Simply counting the use of individual concept identifiers may be highly misleading. For example, a fall in the use of the code 22298006 |myocardial infarction| might reflect a shift to using more specific codes (such as 314207007 |non-Q wave myocardial infarction| or 304914007 |acute Q wave myocardial infarction|), rather than a reduction in the incidence of myocardial infarctions. Use of subsumption testing on SNOMED CT encoded data (see section 6.2 Subsumption) can enable higher level trend analysis to be performed over more specific coded data.

SNOMED CT's polyhierarchy allows trends to be analyzed from multiple perspectives. However, deciding which level of aggregation to use for trend analysis can be arbitrary. Novel approaches to this task are emerging as the demand for trend analysis over SNOMED CT enabled data increases.

The UK Data Migration Workbench (case study 12.1.1 Data Migration Workbench (UK)), for example, includes a trend module which analyses the frequency with which individual SNOMED CT codes are used in the Electronic Patient Record (EPR) instance data, looking for those whose recording frequency has changed over the course of the data collection period. It also includes an Induce module, which performs a more sophisticated analysis of case mix and caseload trends within a clinical department. Instead of returning the most frequently used individual codes, the Induce module identifies the most frequently used types of codes. For example, an emergency department may use roughly 500 different SNOMED CT codes for a laceration in a particular anatomical location. While none of the site-specific codes may appear in a list of most common codes, the descendants of 312608009 |laceration| may collectively account for a significant part of the department's workload.

The algorithm used picks aggregation points at defined levels for analysis. The default setting finds roughly 100 sub-trees within the SNOMED CT hierarchy, where each sub-tree accounts for a more or less constant proportion of all coded episodes (around 1% of all coded events per sub-tree). The algorithm completes once the set of all codes within all identified sub-trees collectively accounts for the large majority of the dataset being analyzed. When applied to real emergency department attendance data, relatively low numbers of presentations (about 0.2%) were coded as occurring primarily as a result of endocrine disease. As a result, in order to get a big enough grouping of episodes, the algorithm chooses 362969004 |disorder of endocrine system| as the root of a single sub-tree covering these reasons for the patient's attendance. By contrast, a very high proportion (9.4%) of presentations relate to some subtype of 928000 |disorder of musculoskeletal system|. Therefore this part of the caseload is aggregated under multiple more granular sub-trees, including (separately) burns, abrasions, lacerations, blunt injury, crush injury and foreign body.

These code aggregations can then be tracked across time to reveal trends in demand, disease incidence or resource utilization.