International Journal of Population Data Science

Journal Website: www.ijpds.org

Panning for gold: finding medical treatment data in insurance records

Avery, Daniel[1]

[1]University of Oxford

Objective

In our Chinese biobank of half a million people, we use data gathered from health insurance agencies to supplement our follow-up. We have 217,000 participants with insurance records including a breakdown of what the insurance paid for, totalling 1.6 million insurance records and 60 million chargeable items. The objective was to find ways of using this information to enhance our Electronic Health Records (EHRs) by adding usable and reliable treatment data, as a basis for future research.

Approach

Machine translations of every charge description were produced so that early investigation could be done by analysts who were not Chinese speakers. Key phrases were produced by specialist clinicians in an iterative process. We began by focussing on haemodialysis treated ESRD, heart failure, and coronary revascularisation. With our refined techniques, key phrase searches were developed which could be tied into ongoing validation procedures elsewhere in the study (e.g. cancer) or which could be validated using existing data from other sources (e.g. death reporting).

Results

Machine translation provided both problems and unexpected solutions. While it could be inaccurate ('Divine Comedy', 'semen', 'corpse cuisine'), more often than not it provided unexpected advantages, converting regional, archaic, or otherwise uncommon Chinese terms into the most common English equivalent.

The majority of chargeable elements in our insurance records are not treatment data per se, but instead hospital fees, generic care, and records of tests without result data. This makes identification of relevant treatment data challenging. Targeted key phrase searches proved successful, demonstrating that it was possible to use this data to answer research questions, even teasing out details which would otherwise not be available to us (e.g. ESRD, location and type of revascularisation).

Validation of these findings is ongoing. For example, we found that 395 of our participants have been charged for 'corpse cuisine' (more accurately 'corpse preparation'). Comparing these figures to our death records (an independently gathered source) we confirmed that 326 are known to be dead, and we added the remaining 69 to our list for active follow-up. Similarly, will we be seeking hospital records for the 528 patients who are receiving cancer treatment with no record of cancer.

Conclusion

Our methods for dealing with treatment data are still being refined, but early results are looking promising. We are investigating standardisation to ICD-10-PCS codes, developing more treatment-based diagnoses, and feeding our findings back into our ongoing validation program.

*Corresponding Author:
Email Address: daniel.avery@ndph.ox.ac.uk (D. Avery)

http://dx.doi.org/10.23889/ijpds.v1i1.198