International Journal of Population Data Science

Journal Website: www.ijpds.org

Validation & Vindication – Comparing Electronic Health Records with Hospital Notes

Hacker, A¹

¹University of Oxford

Introduction

No doubt your Electronic Health Records have been meticulously gathered, imported, validated and standardised. However, if you want to be certain that they are an accurate representation of reality, you can't beat physically going to hospitals and cross-checking their records against yours. Our biobank did exactly this.

Conclusion/Implications

This validation project was a huge and daunting undertaking, but repaid our investment with proof that our Electronic Health Records were generally very reliable, and also with much richer data about disease diagnosis and phenotyping. Other projects using Electronic Health Records may wish to adopt this approach.

Objectives and Approach

Our validation exercise encompassed all reported cases in our follow-up data of three key conditions: stroke, heart disease, and cancer. Key data about each hospitalisation was extracted and exported to tablet computers running custom software. Our staff then visited each hospital in this dataset seeking the corresponding medical notes, and collected additional data from those that they found including photographs of key documents. These results were then adjudicated by specialist physicians to determine the accuracy of the diagnosis, and identify disease phenotypes of interest. Finally, all these results were merged back into our follow-up data.

Results

Not only was gathering the data a huge logistical and technical challenge, integrating it back into the database presented its own difficulties. Our initial plan was to assign each sought event a status of 'validated', 'corrected' or 'unfound'. However, this proved inadequate for addressing the complexities of the data, as we will discuss, with examples.

Our solution was to initially treat the retrieved hospital notes as simply another source of follow-up data. We were thus able to use our existing systems for validating, standardising and aggregating events; and thus produce validated endpoints that were meaningfully comparable to our reported endpoints. We could then implement and test definitions of the required validation statuses at a participant level for each disease of interest.