OXFORD

Data and text mining

On the discovery of hospital admission patterns—a clarification

Ognjen Arandjelović

School of Computer Science, University of St Andrews, St Andrews, UK

Associate Editor: Jonathen Wren

Received and revised on 19 January 2016; accepted on 20 January 2016

Abstract

Contact: ognjen.arandjelvoic@gmail.com

This article concerns my publication 'Discovering hospital admission patterns using models learnt from electronic hospital records', which has recently appeared in *Bioinformatics* (Arandjelović, 2015a) and the comments regarding its contributions made thereafter by Lopez-Campos *et al.* (2016).

Lopez-Campos *et al.* (2016) begin by observing a mismatch between the descriptions of certain ICD codes as stated in my article (Arandjelović, 2015a) and those published by the World Health Organization. This observation is correct and, needless to say, while I greatly regret the error, I am genuinely grateful for having this brought to my attention so that a timely erratum can be published.

It is important to stress that the mismatch in question in no way affects the main content of the article, its contribution, or any of the results: the method described in the original publication does not use any semantic information about the codes and their meanings are stated merely for the sake of illustration and context. The unfortunate error in the description of the codes is likely to have taken place in the editing stages of the original article when the presentation and the content of the relevant table and the related information in the text underwent changes several times.

More pertinent to the actual methodological contribution of my work are the opinions of Lopez-Campos *et al.* (2016) on the manner in which the aforesaid codes are used and in particular what Lopez-Campos *et al.* (2016) refer to as the intended use and the spirit of the codes. There are several arguments which speak against this stance. First, on the fundamental philosophical level, the intention behind the design *per se* is inherently inconsequential. Rather the relevant question is whether the codes, such as they are, contain the structure and information which permit prediction. The results I present which show that the proposed method achieves far better prediction than more primitive models—adapted to the present problem from the work of others as described in Arandjelović (2015a,b)—demonstrate this conclusively (Arandjelović, 2015a; Vasiljeva and Arandjelović, 2016). This is also far from an isolated

piece of evidence. For example in 'A Special Survey of Deaths' (RGI-CGHR Collaborators, 2009), a major large-scale and long-term health care initiative in India, the same manner of coarse coding as described in my work is used too. In short, the concern expressed by Lopez-Campos *et al.* (2016) is neither warranted by theoretical principles nor empirical evidence.

Last, I certainly appreciate the viewpoint of Lopez-Campos *et al.* (2016) in desiring more detailed information on the data set used. Equally, I trust that they will appreciate the alternative argument which comes from the need for a compromise imposed by the rather restrictive page limit of *Bioinformatics* of only seven pages. Considering that the contribution of my work does not pertain to any specific medical findings as such but rather to an algorithm (which when deployed may and indeed should lead to findings of the aforementioned nature) I considered it reasonable to place a greater emphasis on the description of the methodological aspects of the work, wherein the key novelty lies.

Conflict of Interest: none declared.

References

Arandjelović,O. (2015a) Discovering hospital admission patterns using models learnt from electronic hospital records. *Bioinformatics*, **31**, 3970–3976.

Arandjelović,O. (2015b). Prediction of health outcomes using big (health) data. *In: Proc. International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2543–2546.

Lopez-Campos,G.H. *et al.* (2016) Comment on 'Discovering hospital admission patterns using models learnt from electronic hospital records'. The importance of using the right codes. *Bioinformatics*.

RGI-CGHR Collaborators. (2009). Report on the causes of death in India: 2001–2003. *Office of the Registrar General of India*.

Vasiljeva,I. and Arandjelović,O. (2016). Prediction of future hospital admissions—what is the tradeoff between specificity and accuracy? *In: Proceedings of the International Conference on Bioinformatics and Computational Biology*.