Reviewer's report

Title: Evaluating the Risk of Patient Re-identification from Adverse Drug Event Reports

Version: 1 Date: 31 May 2013

Reviewer: Xiaoqian Jiang

Reviewer's report:

This paper aims at measuring re-identification risk of Canada’s post-marketing adverse drug event database (ADE) on different adversary scenarios. It is an important challenge for biomedical informatics. The paper did a good job in motivating the problem and thoroughly reviewed the literature. But the manuscript contains mistakes and there are inconsistency between lemmas and their proofs.

Major compulsory revisions

1. In Lemma 1, if n=1 and F_j=2, P_j^1 = 2p/F_j. But P_j^1 equals p/F_j according to the proof of the Appendix. Similarly, if n=1 and F_j =1, P_j^1 = 0 according to Lemma 1 whereas it equals 1 in the proof. It can be verified that such inconsistency exists in situations like n=2, F_j=3 and n=2, F_j=2. This needs to be double checked as Lemma 2 and theory 1 are both based on Lemma 1.

2. Lemma 2 is to quantify the risk of adversary getting a successful match in M_j attempts, which should be \( \sum_{i}^{M_j} P_j^i \) no matter what value F_j takes. But this is not the case in the proof, for example, P_j^{M_j+1} was counted when F_j=M_j and F_j=M_j+1. I do not understand why the probability of the successful match at the M_j+1 attempt should be considered when the goal is to calculate the probability of successful match in all M_j attempts.

Minor essential revisions

1. The symbol R_2^j used in the proofs was not mentioned in the main text and I believe it should be R_j instead.

2. I am pretty sure in the proof of Lemma 1 “the first was non-verifiable and the second was verified to be non-match” should be “the first was non-verifiable and the second was verified to be match”, which is misleading.

3. There is also a missing bib in page 4, “…suggests that this subset of individual is more likely to be re-identified [?]”.

Level of interest: An article of importance in its field

Quality of written English: Acceptable
**Statistical review**: No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests**: I have not competing interest.