Reviewer's report

Title: Assessing the Validity of Tuberculosis Surveillance Data in California

Version: 1 Date: 4 April 2006

Reviewer: David Durrheim

Reviewer's report:

General

The authors proposed to systematically assess the validity and completeness of TB surveillance data in California. Although I am unaware of a previous study assessing Californian TB surveillance data, the approach taken by the authors amounts to quality assurance of a legislated public health surveillance system.

The use of a stratified random sample of cases for considering system quality is appropriate. The time-frame is of considerable concern though, as the cases sampled were notified almost a decade ago and it takes a considerable stretch of the imagination to be convinced that this assessment is representative of current system performance. It could be argued that because nothing apparently has been done to improve the system during this period that it is unlikely that the system has improved, but if these findings are to be used to inform revision of the system then I would be sincerely concerned by the currency of the data.

The selection of certain of the test performance measures selected and their application to the data presented raises major concerns. Sensitivity, specificity and particularly positive predictive value and negative predictive value are ill-chosen attributes given the quality of the "gold standard" data. For the quality assurance exercise conducted it would have been more appropriate to report other attributes by the authors: concordance and misclassification, is highly appropriate and makes their case eloquently.

The nature and quality of data do not support all of the analytical techniques utilized. The construction of denominators for attribute determination appears flawed. This was a quality assurance exercise and not a research study.

The manuscript is professionally prepared, succinct and generally well presented.

The conclusions and recommendations are sound and rational, and generally supported by the data presented if surveillance system performance remains as it was nearly a decade ago. They represent what would be considered good practice in public health surveillance systems in developed countries. Fortunately, these recommendations for improving system quality would remain even after recommended exclusion of the dubious validity measures.

The title would probably be more accurate if it read, "Assessing the historical validity of tuberculosis surveillance data in California". The abstract is useful.

The paper is generally well written with important implications for improving the quality of tuberculosis surveillance systems in California.

• Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore)
Results – suggest removing “vs. lacking” in the last sentence.

Although completeness is indicated in the background this is not systematically presented.

Abstract Methods should indicate that stratification was by treatment outcome.

• Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

There is a problem with the proportions of treatment outcome provided in the final sentence of first paragraph of the Methods section: 46.7% death after treatment initiation.

It would have been commendable had a broader perspective been adopted with consideration of a broader range of attributes, including simplicity, flexibility, acceptability to users, representativeness and timeliness (Romaguera RA et al. Evaluating public health surveillance. In: Teutsch and Churchill (eds). Principles and practice of public health surveillance. 2nd edition. New York: Oxford University Press, 2000).

• Major Compulsory Revisions (which the author must respond to before a decision on publication can be reached)

Reasons, including cost, should be provided for why research staff accessed a secondary data source, local health department records, rather than primary data source, patient treatment file as a gold standard.

The selection of 80% concordance appears arbitrary and rather low for categorizing data “validity” as adequate. It would have been preferable to rather just provide actual concordance measurements and discuss the implications in terms of resource allocation, planning of public health programmes and programme evaluation.

Proportion of missing data by each RVCT variable should have been presented. This could have resulted from the quality of data entry but also poor agreement between the fields in the local health department TB record form and the RVCT. The two systems are clearly not independent with misclassification in the local health department record having a profound impact on the RVCT.

Generally when a gold standard is available, performance attributes considered for a surveillance system are sensitivity and positive predictive value (German RR. Sensitivity and predictive value positive measurements for public health surveillance systems. Epidemiology 2000;11:720–7). Determination of predictive values really demands presupposes that we have knowledge of the cases that actually have the health-related event under surveillance. As this is clearly not the case in the current study these should not be included. The current study does not attempt to determine actual sensitivity of the system in detecting community occurrence of tuberculosis nor was it designed with this purpose in mind. However, neither can it hope to report sensitivity given that the gold standard is not convincing (Centres for Disease Prevention and Control. Updated guidelines for evaluating public health surveillance systems. Recommendations from the Guidelines Working Group. MMWR Recommendation and Reports 2001; 50, RR-13). It is worth reiterating that analysis should rather be restricted to reporting on completeness, misclassification of unknowns (which is a very useful aspect of the current report), differences between key dates and concordance.

In the discussion it would be value to provide a relevant reference supporting the contention that the marginalized patient treatment groups mentioned require more complex case management. Similarly, it would be good to reference important statements, for example, “A prior episode of TB increases the risk of drug resistance and may be associated with treatment non-adherence.”

There is grammatical problem in the 6th paragraph in the Discussion, 2nd sentence. Should
probably read “Health departments generally do not implement these definitions strictly.

The inclusion of unknowns in the denominator for sensitivity and specificity calculations is not strictly correct (Last JM. A dictionary of epidemiology. 4th edition. Oxford: Oxford University Press, 2001). It would have been much more convincing to simply report the scale of misclassification, rather than shrouding this important finding in inappropriate analysis.

**What next?:** Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** No

**Declaration of competing interests:**

I declare that I have no competing interests