In the current issue, Lalor and colleagues present the findings of their investigation into tuberculosis-related death reports in England and Wales using two large databases[1]. The UK Office for National Statistics (ONS) database collects information based on death certificates through vital registration; the Enhanced Tuberculosis Surveillance (ETS) system relies on clinical teams reporting TB cases. There were discrepancies between the TB deaths reported with only 57% of the deaths found in ETS present in ONS, and 53% of deaths in ONS present in ETS. Additionally, approximately one-third of the TB deaths recorded in ONS were not due to active TB on review of the data.

We are now in an era where data science and analysis of "big data" has become commonplace, and substantial progress has been made in harnessing the power of computing technology to explore datasets in ways that were previously not possible[2]. However, this paper highlights the unavoidable truth that our conclusions are dependent on the integrity of the data. For example, the authors concluded there were 5,961 deaths between 2005 and 2015 attributable to active TB, however ETS captured only 4,165 (70%) and ONS 4,606 (77%) of these deaths. A difference of almost 10% in the total deaths would translate into much larger numbers being missed if scaled up to reflect the current global TB burden.

The authors address the need for improved clinician training and the development of clear guidance for reporting. There is evidence that focussed TB teaching does improve knowledge among medical students[3]. The END-TB Strategy depends on an accurate assessment of the global state of the disease, as success is measured against the number of reported deaths from active TB. This study took place in a well-resourced setting where robust reporting methods should be achievable. A detailed analysis of the reasons behind the differences seen between the datasets and possible explanations for failure to notify some cases (when lack of infrastructure and resources were not limiting factors) may enable a "gold standard" framework to be agreed upon, which could be refined in high-income settings before being more widely implemented.

The WHO Global Tuberculosis Report[4] estimates the burden of disease using a combination of notification data, national surveys, and capture-recapture modelling but there are an estimated 3 million TB cases missed each year[5]. The majority of undetected or unreported cases and related deaths are to be found in areas with poor access to resources. Clinicians working in these areas are likely to have little time to think about matters beyond maintaining their TB service, and the collection of national level data could seem quite abstract and burdensome. Development of clear and evidence based criteria for the reporting of clinical TB data could help with this, and reaching agreement at the level of global policy could also provide some stimulus to direct funding towards this crucial activity. The global fight against TB continues, and to reach our goal of elimination we will need to make good decisions based on good data: as Napoleon Bonaparte famously said, "War is 90% information."

References

1. Lalor MK, Mohiyuddin T, Uddin T, et al. The challenge of estimating tuberculosis mortality accurately in England and Wales. Int J Tuberc Lung Dis 2018 ….

2. Donoho D. 50 Years of Data Science. J Comput Graph Stat 2017;26(4):745-766

3. Harrity S, Jackson M, Hoffman H, et al. The National Tuberculosis Curriculum Consortium: a model of multi-disciplinary educational collaboration. Int J Tuberc Lung Dis 2007;11:270–4

4. World Health Organisation. Global Tuberculosis Report. Geneva 2017

5. Herbert N, George A, Baroness Masham of I, et al. World TB Day 2014: finding the missing 3 million. Lancet 2014;383:1016-1018