

Publication: [Term sets: A transparent and reproducible representation of clinical code sets](#)

**Publication details (Vancouver format)**

Williams R, Brown B, Kontopantelis E, van Staa T, Peek N. Term sets: A transparent and reproducible representation of clinical code sets. Olier I, editor. PLoS One [Internet]. 2019 Feb 14 [cited 2020 Apr 1];14(2):e0212291

**What are the most important findings/conclusions in this paper? Why are they important?**

Most patient health data used by researchers is 'coded' – that means, instead of a patient's record containing the term "type 2 diabetes", it would contain the clinical code "C10F". When we want to analyse patient data we must first make sets of these clinical codes, called 'code sets', to describe what we want to find out. We have developed a better way to create these code sets that is quicker, more reliable, and more accurate than existing methods. We call this a 'term set'.

**What did you do?**

First, we used maths to show that it would be possible to transform an existing set of clinical codes into a 'term set'. We then demonstrated that 'term sets' are better than standard code sets. We did this by finding sets of clinical codes that had been published by other researchers and then trying to replicate them using our 'term set' approach.

**Why did you conduct this research?**

Clinical code sets are the single most important part of research using patient health records. Mistakes in this early part of research are likely to change the results and conclusions of a study. Our method makes this step easier. It also makes it clearer to other researchers who are reviewing your work.

**What was known before your paper was published?**

Clinical code sets are important. However, they are not often made available to the public. Without this understanding, it is hard to trust the results of studies that have used clinical code sets.

**What is next? What is the potential impact of the work in this paper? What will change as a result of this paper (or the study it describes)?**

'Term sets' make it easier to produce reliable, accurate clinical code sets. All health research using electronic patient records could be done more quickly, and the quality could be improved, if 'term sets' were used.

**Does this paper link in to a particular study / project? If so, please summarise the study and explain how this paper has improved understanding, or will move the study forward.**

N/A