Assessing the information difference between coded conditions and their related clinical notes

**Methods**

**Data:** Integrated Primary Care Information (IPCI) Dutch general practitioner EHR database.

**Concept extraction:** MedSpacy with Dutch resources.

**Setup:** We applied the concept extraction framework to clinical notes related to commonly occurring ICPC-1 coded conditions.

**Semantic similarity:** SNOMED CT concept (text and ontology-based) embeddings were used to calculate the similarity between coded conditions and extracted concepts.

**Results**

Figure 2A and B illustrate the similarity scores between extracted concepts and the coded condition using text-based and ontology-based embeddings. In 73.5% of condition occurrences, a concept either semantically similar or related to the coded condition was found in the surrounding text, as shown in Figure 2C and D.

**Conclusion:** Our study successfully adapted an open-source English concept extraction framework to Dutch clinical text, accurately matching structured data in most condition occurrences, yet many conditions are not described in the surrounding text.

Tom Seinen, Erik van Mulligen, Jan Kors, Peter Rijnbeek
Department of Medical Informatics, Erasmus MC, The Netherlands
t.seinen@erasmusmc.nl