Appendix 2: Statistics of the domain effect

Figure 1. Statistics of the domain effect. (a) shows the counts that in-domain or out-domain models performed better in 3 corpora. In-domain models indicate EhrBERT for the MADE corpus, BioBERT for the NCBI and CDR corpora. (b) shows the counts that biomedical-clinical-domain models (BioBERT or EhrBERT) and general-domain models (BERT) performed better in 3 corpora.

(a) In-domain vs Out-domain
- in-domain models performed better
- in-domain models performed significantly better
- out-domain models performed better
- out-domain models performed significantly better

(b) BioBERT or EhrBERT vs BERT
- BioBERT or EhrBERT performed better
- BioBERT or EhrBERT performed significantly better
- BERT performed better
- BERT performed significantly better