

Appendix A

In this section we provide scatter plots showing the relationship between the frequency of an ICD-9 code in a dataset (i.e., number of patients or number of citations with the code) and the number of associations to which the code belonged in a dataset. Scatter plots were made for each of the six datasets used in our analysis in the main manuscript. For each dataset we provide two scatter plots. The first (left side) shows the frequency of all associations for each ICD-9 code. The second (right side) shows only those associations that had a p-value $< 1.0 \times 10^{-10}$ (Chi square statistic ≥ 41.8). These plots show that more frequently occurring ICD-9 codes tend to belong to a larger number of associations.

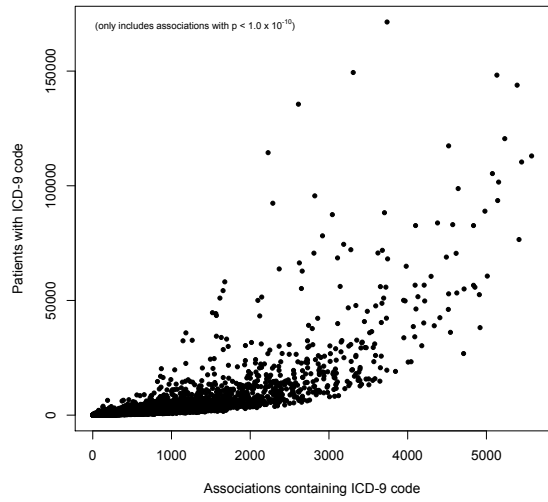
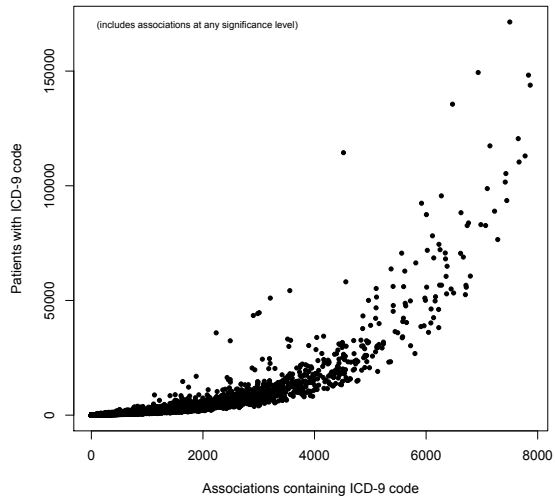


Figure A1. Clinical (original) dataset

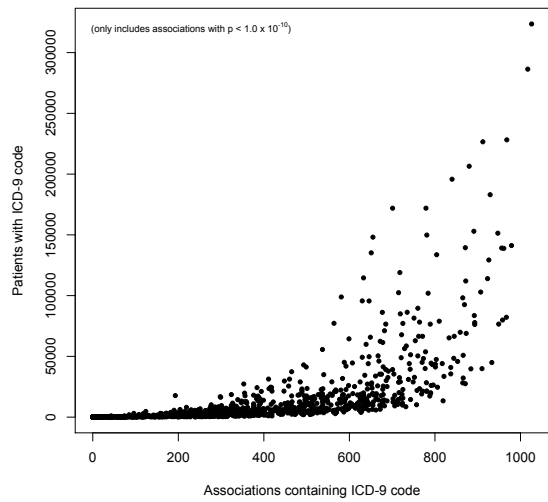
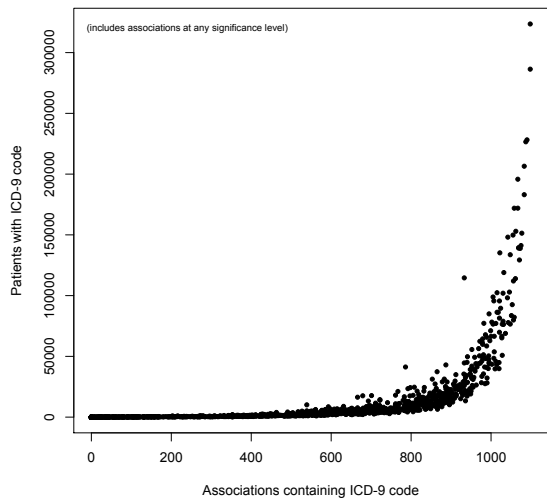


Figure A2. Clinical (simplified) dataset

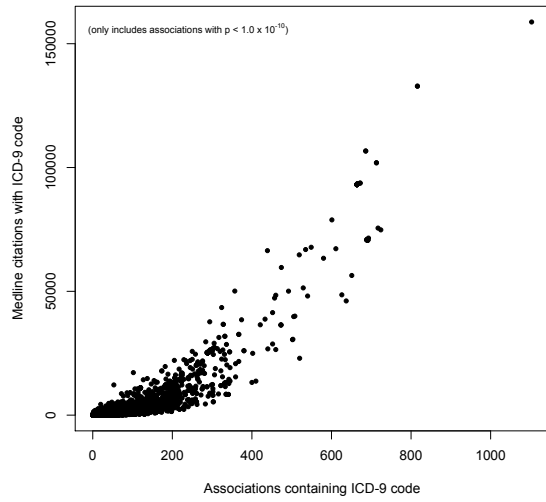
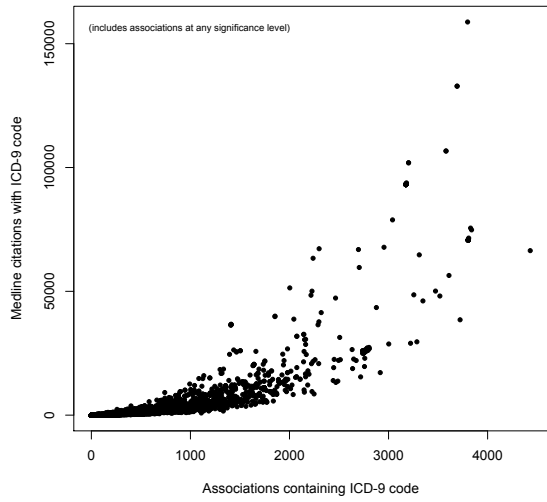


Figure A3. Medline 1000 (original) dataset

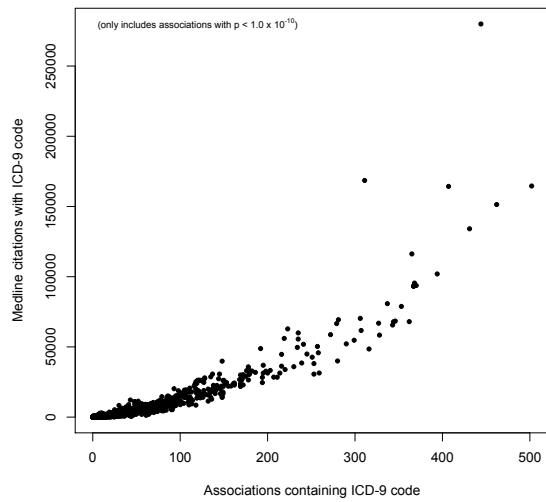
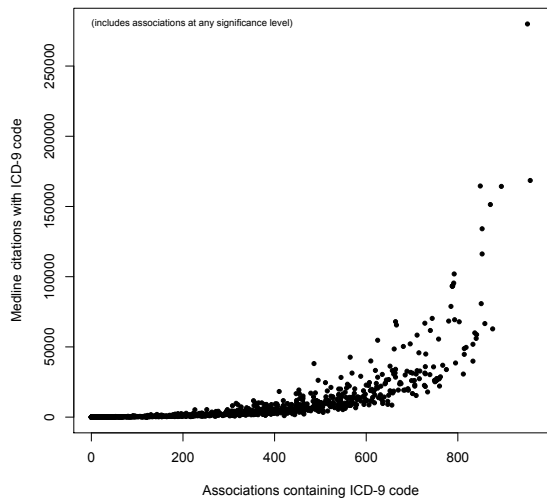


Figure A4. Medline 1000 (simplified) dataset

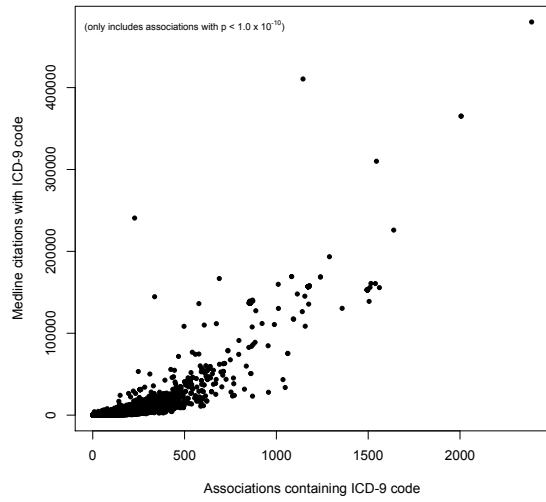
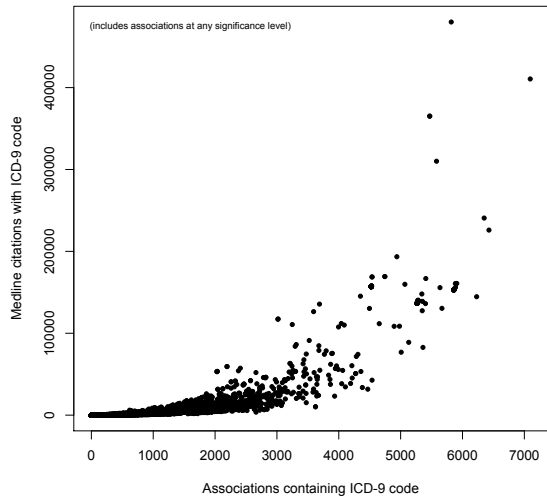


Figure A5. Medline 600 (original) dataset

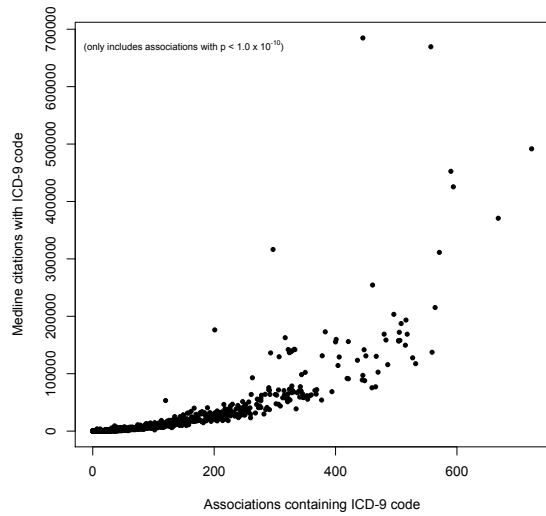
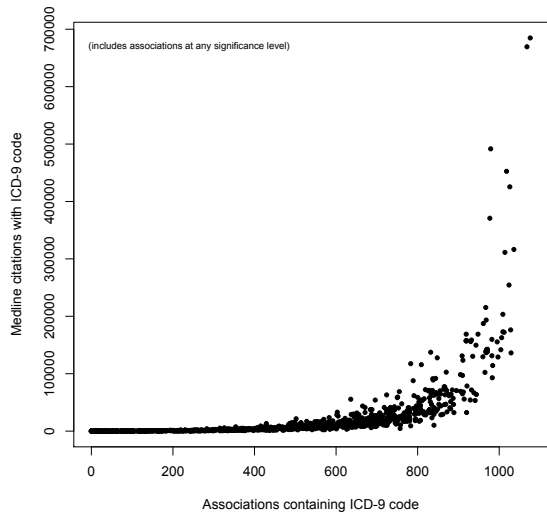


Figure A6. Medline 600 (simplified) dataset