**Supplementary materials**

**Supplemental Table 1. ICD-9 and ICD-10 codes used to identify cases of delirium**

|  |  |
| --- | --- |
| **ICD-9 Codes** | **ICD-10 Codes** |
| **Code** | **Description** | **Code** | **Description** |
| 290.11 | Presenile dementia with delirium | F03.91 | Unspecified dementia with behavioral disturbance |
| 290.3 | Senile dementia with delirium | F05 | Delirium due to known physiological condition |
| 290.41 | Vascular dementia with delirium | F01.54 | Vascular dementia with behavioral disturbance |
| 291.0 | Alcohol withdrawal delirium | F10.121, F10.221, F10.231, F10.921 | Alcohol dependence with withdrawal delirium |
| 292.81 | Drug-induced delirium | F11.121, F11.221,F11.921F12.121, F12.221, F12.921F13.121, F13.221, F13.231, F13.921, F13.931F14.121, F14.221, F14.921F15.121, F15.221, F15.921F16.121, F16.221, F16.921F18.121, F18.221, F18.921F19.121, F19.221, F19.231, F19.921, F19.931 | Other psychoactive substance use, unspecified with intoxication with delirium |
| 293.0 | Delirium due to conditions classified elsewhere | F05 | Delirium due to known physiological condition |
| 293.1 | Subacute delirium | F05 | Delirium due to known physiological condition |

**Supplemental Figure 1. Overlap between methods for identifying true cases.** Results indicate overlap between methods for identification of cases compared to manual chart review, ICD-codes, and NLP algorithm.



**Supplemental Appendix**

**NLP Sources.** The NLP algorithms for extracting delirium status have been previously described1. Briefly, the infrastructure of the NLP system was an open-source NLP pipeline MedTaggerIE2, which was developed using an open-source unstructured information management architecture–based information extraction framework3. The pipeline separates task-specific NLP knowledge engineering (i.e., individual CAM concepts) from the generic routine NLP, which enables words and phrases containing clinical information to be directly defined by subject matter experts (e.g., geriatricians and psychologists). To enhance research reproducibility and transparency, we released the NLP infrastructure, definition, and algorithm under the Open Health Natural Language Processing Consortium (**Table** **A**).

**Table A**. Access to Delirium Models and Resources

|  |  |  |
| --- | --- | --- |
| **Category** | **Description** | **Link** |
| Infrastructure | Open-source NLP pipeline MedTaggerIE | <https://github.com/OHNLP/MedTagger> |
| Algorithm Definition | Definition of the NLP-CAM and NLP-mCAM | <https://github.com/OHNLP/AgingNLP/tree/master/delirium> |
| Definition of the CAM-related Clinical Concepts | <https://github.com/OHNLP/AgingNLP/tree/master/delirium> |
| Annotation guideline | <https://github.com/OHNLP/AgingNLP/tree/master/delirium/annotation_guideline>  |
| Algorithm | Lexicons and regular expression | <https://github.com/OHNLP/AgingNLP/tree/master/delirium/DELIRIUM/regexp> |
| Context algorithm | <https://github.com/OHNLP/AgingNLP/tree/master/delirium/DELIRIUM/context> |
| Rulesets | <https://github.com/OHNLP/AgingNLP/tree/master/delirium/DELIRIUM/rules> |
| Other support | Installation Guide | <https://vimeo.com/392331446> |

**Gold Standard.** The gold standard corpus for the evaluation of the NLP system was established through a standard corpus annotation process4. One geriatrician and one psychologist manually reviewed clinical notes from the Mayo Clinic Rochester (MCR) EHR for 300 patients from the Mayo Clinic Biobank. Half of the records were from patients who had received a delirium ICD-9 code and half who did not. The annotation was conducted using a previously validated annotation guideline and scheme. An iterative process was executed involving annotation training, organizing consensus meetings, measuring agreements, updating guidelines, and conducting adjudication. Following the same process, a trained nurse abstractor manually reviewed 400 patients (3287 notes) randomly stratified by the presence of delirium ICD-9 code from the Olmsted Medical Center (OMC).

**NLP Development and Evaluation.** We used 50% of the MCR data to develop the initial algorithm and used the other 50% of MCR data to evaluate the final performance. The validated NLP system was then deployed to the OMC data. Similarly, we used the first 50% of the OMC data to refine the NLP algorithm and report the final performance on the second half of the data. The final NLP performance is shown in Tables B and C.

**Table B.** Performance Evaluation of Delirium Identification on Mayo Clinic Rochester (MCR) clinical notes

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Gold standard |  |
|  |  | Delirium | No Delirium | *Total* |
| NLP | Delirium | 57 | 0 | *57* |
| No Delirium | 5 | 88 | *93* |
|  | *Total* | *62* | *88* | *150* |

**Table C.** Performance Evaluation of Delirium Identification on Olmsted Medical Center (OMC) clinical notes

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Gold standard |  |
|  |  | Delirium | No Delirium | *Total* |
| NLP | Delirium | 22 | 0 | *22* |
| No Delirium | 0 | 172 | *172* |
|  | *Total* | *22* | *172* | *194* |

**References**

1. Fu S, Lopes GS, Pagali SR, Thorsteinsdottir B, LeBrasseur NK, Wen A, et al. Ascertainment of delirium status using natural language processing from electronic health records. The Journals of Gerontology: Series A. 2020.

2. Liu H, Bielinski SJ, Sohn S, Murphy S, Wagholikar KB, Jonnalagadda SR, et al. An information extraction framework for cohort identification using electronic health records. AMIA Summits Transl Sci Proc. 2013;2013:149.

3. Ferrucci D, Lally A. UIMA: an architectural approach to unstructured information processing in the corporate research environment. Nat Lang Eng. 2004:1-26.

4. Fu S, Leung LY, Raulli A-O, Kallmes DF, Kinsman KA, Nelson KB, et al. Assessment of the impact of EHR heterogeneity for clinical research through a case study of silent brain infarction. BMC Med Informatics Decis Mak. 2020;20:1-12.