Supplementary information for

**Real-World Effectiveness, Its Predictors and Onset of Action of Cholinesterase Inhibitors and Memantine in Dementia: A Retrospective Health Record Study of 7400 Individuals using the UK CRIS Platform**

Nemanja Vaci1,2, Ivan Koychev1, Chi-Hun Kim1,3, Andrey Kormilitzin1, Qiang Liu1, Christopher Lucas1, Azad Dehghan4,5, Goran Nenadić5, Alejo Nevado-Holgado1,6,7

Author affiliations:

1. Department of Psychiatry, University of Oxford, Oxford, UK
2. Department of Psychology, University of Sheffield, Sheffield, UK
3. Department of Neurology, School of Medicine, Kyungpook National University Chilgok Hospital, Daegu, South Korea
4. DeepCognito ltd., Manchester, UK
5. School of Computer Science, University of Manchester, Manchester, UK
6. Akrivia Health, Oxford, UK
7. Big Data Institute, University of Oxford, UK

Corresponding author: Dr Nemanja Vaci, Department of Psychology, University of Sheffield, Cathedral Court, 1 Vicar Lane Sheffield, S1 1HD, Email: [n.vaci@sheffield.ac.uk](mailto:n.vaci@sheffield.ac.uk)

**This PDF includes:**

**Methods**

**Supplementary text**

**Supplementary results**

**Figs. S1 to S5**

**Tables S1 to S4**

**Methods**

The R code for the main analysis and the NLP model can be download from the OSF page (see <https://osf.io/4xev5/>).

**Rule based Natural Language Processing**

The main issue with the unstructured clinical records is that the data of interest, such as psychometric scores, medication and diagnosis information, is locked in the rich textual format. To unlock this information from the raw text format, in the first step of the analysis, we built a natural language processing (NLP) model for named entity extraction. The idea of this model is to statistically process the language used in the clinical notes and identify mentions of medication prescription, health scores and diagnosis information for each patient. In the case of this study, we developed rule-based NLP models in which we define common linguistical constructs how the medication, health score and diagnosis is coded in the clinical texts. The NLP models that we have developed consists of a number of rules that combine concepts of diagnosis outcomes, medication and health scores with lexical and orthography features. We used GATE software (<https://gate.ac.uk/>) as the analytical framework for text processing; including ManTIME13 for date normalisation.

**Annotation procedure and validation of the model**

In the first step of the analysis, two clinical experts (authors 3 and 6) developed a gold corpus that was used to build the natural language processing model. The corpus was consisted of 600 annotated Oxford Health notes, where for each note the experts annotated three crucial information: diagnosis, medication, and health scores. The annotation of diagnosis was divided in 5 different subcategories: Alzheimer’s disease, vascular dementia, mixed dementia, Lewy body dementia, and mild cognitive impairment. The medication was equally divided into five different categories: donepezil, rivastigmine, galantamine, acetylcholinesterase inhibitors (broad category), and memantine. Finally, the scale information was subcategorized into two constructs: MMSE and MOCA scores. For each identified concept the time information, that is, the date when concept occurred, was annotated. Besides annotating the occurrence of concepts in the text, the clinical advisors also indicated who is experiencing this event (e.g. patient, family, other), by whom it is prescribed (secondary care service, primary GP or other), how valid the information is in the text (factual, possible, probable, uncertain, or conditional), and whether it is affirmed or negated. Besides developing golden corpus of 600 documents, the clinical experts additionally annotated 37 documents that were used to calculate inter-annotator agreement. The NLP model was developed on the set of 360 notes that had 570 mentions of diagnosis, 216 mentions of medication, and 169 mentions of cognitive scores.

The performance of the model was tested on 240 Oxford Health clinical texts that collected 352 mentions of diagnosis, 153 of medication, and 87 of health score information. Additionally, we tested how well the NLP model generalized on Southern Health data on 20 separate notes with 26 mentions of diagnosis, 30 of medication, and 23 of health score information. To test the how well the model performs when extracting the information from the text we used combination of methodologies from “Message understanding conference” (MUC9) and “International workshop on semantic evaluation” (SemEval’1310). Based on the MUC categories the extractions from NLP model can be correct (COR), incorrect (INC), partial (PAR - not identical agreement between gold standard and extraction), missing extraction (MIS – model did not extract the concept), and spurious extraction (SPU – extracted concept that does not exist in the gold standard). Based on this we can calculate number of possible annotations (POS) in gold-standard corpus that contribute to the validation score by summarizing correct, incorrect, partial and missing outcomes (true positive + false negative). Equally, we can calculate total or actual (ACT) number of annotations that our NLP model produced by summarizing correct, incorrect, partial and spurious outcomes (true positive + false positives). Using these two measures we can calculate precision and recall of the system. The precision tells us how many extractions were correct out of the total number of extracted concepts and it is calculated as ratio between correct (COR) extractions and actual number of annotations (ACT). The recall indicates percentage of entities correctly identified in the corpus and is calculated as ratio between correct (COR) and all possible outcomes (POS). Finally, we can also calculate the overall performance of the model by using harmonic mean between these two values, where . The NLP model development in this study performed relatively well on all three main categories (see table 1). In the case of diagnosis, results show that model identified almost 95% of diagnosis mentions in the texts, while it correctly normalised 85% of them. Much better results are obtained in the case of mentions of medications where our system recognizes 98% of medication mentions and correctly normalised 97% of them. Finally, the scale information results in smaller percentage of recognized values, as approximately 73% of MMSE and MOCA mentions are recognized in the text. Besides calculating the performance of the model, we also estimated how well the two annotators agree when identifying information in the text. This was calculated on the separate 37 clinical texts that were annotated by both clinical experts. However, the final dataset used for development was adjudicated between the two annotators to ensure consistency. Results show almost perfect agreement between them, where not surprisingly they outperform the developed NLP model in the case of diagnosis and scale information. However, the devised NLP model identifies information about the name of the medication slightly better than clinical experts, due to the usage of extensive dictionary that lists all possible drug names.

Table S1: Validation of NLP model performance on the CRIS data

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Concept | Precision | Recall | F1-score | Human F1 | F1-Southern |
| Diagnosis | 89.6% | 96.3% | 92.8% | 95% | 84.8% |
| Medication | 98.03% | 98.03% | 98.03% | 96% | 78.3% |
| Scale | 92.85% | 74.7% | 82.80% | 100% | 82.6% |

Besides extraction of the general concepts, we also calculated how well the model performs when extracting the information on subtype of diagnosis, medication and health scores, agency (who is experiencing problems), relative time information and how valid is the information. Results show that factuality of the information and exact time relation represents slight problem when annotating and extracting the data, while all other categories reach over 80% of performance (Table 2).

Table S2. Validation of NLP performance on additional categories

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | NLP model | | | Human annotators | | |
| Concept | Diagnosis | Medication | Scale | Diagnosis | Medication | Scale |
| Modality | 71% | 78% |  | 73% | 94% |  |
| Experiencer | 90% | NA |  | 100% | 100% |  |
| Negation | 91% | 95% |  | 90% | 96% |  |
| Opinion by | 93% | 98% |  | 100% | 96% |  |
| Cause | 79% |  |  | 80% |  |  |
| Time relation | 91% | 83% | 62% | 93% | 90% | 95% |

The natural language processing model extracted 12,602 unique patients with mentions of diagnosis, 6,841 patients with mentions of medication and 9,396 patients with health scores. The extracted data was combined using the month and year information extracted from the raw text document or using the document creation date. Preliminary data screening was performed to ensure that potential data entry mistakes were eliminated. We restricted the tails of the age-related distribution, where we excluded MMSE and MOCA values collected 5 years before and 5 after medication was prescribed. The MMSE and MOCA information where extracted denominator was below 28 or above 30 was excluded from the analysis, while for the cases with the denominator of 28 or 29 one or two points were added to the numerator value, respectively (see Table S1). The Oxford dataset resulted in the total of 4,792 patients that contribute 14,838 observations with mentions of dementia, medication and health scores, while the Southern health data contributed 4,341 patients with 17,849 observations (See Figure S1 and online materials).

Table S3: Summary statistics for the variables used in the study (initial measurement for each patient)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Oxford Health | | | | | | | | | Southern Health | | | | | | | |
|  | MMSE  (3,330 patients with 9,583 observations) | | | | | MOCA  (2,398 patients with 5,255 observations) | | | | MMSE  (4,082 patients with 14,211 observations) | | | | MOCA  (1,789 patients with 3,618 observations) | | | |
|  | Min (1st Q) | Mean (SD) | | Median | Max (3rd Q) | Min (1st Q) | Mean (SD) | Median | Max (3rd Q) | Min (1st Q) | Mean (SD) | Median | Max (3rd Q) | Min (1st Q) | Mean (SD) | Median | Max (3rd Q) |
| Cognitive score | 1 (19) | 22 (5.3) | | 22 | 30 (26) | 1(13) | 17 (5.5) | 18 | 30 (21) | 1 (19) | 21 (5.5) | 23 | 30 (25) | 1 (15) | 18 (5.2) | 18 | 30 (22) |
| Age | 58 (77) | 82 (7.4) | | 88 | 102 (87) | 57(78) | 82 (6.5) | 82 | 101 (87) | 52(76) | 80 (7.4) | 81 | 99 (86) | 53 (76) | 80.2 (7.7) | 81 | 99 (85) |
| Type of medication | Without: | | 1768 | | | 320 | | | | 1,772 | | | | 852  522  195  19  852  82  13 | | | |
| Donepezil: | | 950 | | | 1,215 | | | | 1,374 | | | |
| AChEi: | | 355 | | | 453 | | | | 301 | | | |
| Galantamine: | | 61 | | | 45 | | | | 95 | | | |
| Memantine: | | 76 | | | 117 | | | | 326 | | | |
| Rivastigmine: | | 104 | | | 128 | | | | 200 | | | |
| Discontinued: | | 16 | | | 120 | | | | 17 | | | |
| Type of diagnosis | No diagnosis: | | 869 | | | 104 | | | | 1,129 | | | | 441 | | | |
| AD: | | 907 | | | 691 | | | | 1,001 | | | | 395 | | | |
| DLB: | | 53 | | | 32 | | | | 44 | | | | 21 | | | |
| FTD: | | 1 | | | 0 | | | | 10 | | | | 0 | | | |
| Mixed: | | 18 | | | 31 | | | | 116 | | | | 50 | | | |
| PPD: | | 2 | | | 0 | | | | 15 | | | | 0 | | | |
| Unspecified: | | 1,376 | | | 1,435 | | | | 1,630 | | | | 813 | | | |
| VaD: | | 104 | | | 94 | | | | 140 | | | | 56 | | | |
|  | Other | | 5 | | | 11 | | | | 0 | | | | 13 | | | |
| Gender | Female | | 2,064 | | | 1398 | | | | 2,469 | | | | 1,040 | | | |
| Male: | | 1,265 | | | 998 | | | | 1,616 | | | | 749 | | | |
| Ethnicity: | Asian | | 9 | | | 10 | | | | 10 | | | | 4 | | | |
| Black | | 10 | | | 7 | | | | 8 | | | | 2 | | | |
| White | | 2,510 | | | 1,275 | | | | 3,179 | | | | 1,333 | | | |
| Other | | 28 | | | 15 | | | | 22 | | | | 8 | | | |
| Missing information: | | 773 | | | 1,091 | | | | 866 | | | | 442 | | | |
| Number of observations per patient | 1 | | 972 | | | 880 | | | | 800 | | | | 816 | | | |
| 2 | | 801 | | | 779 | | | | 947 | | | | 527 | | | |
| 3 | | 575 | | | 407 | | | | 734 | | | | 243 | | | |
| 4 | | 422 | | | 175 | | | | 544 | | | | 102 | | | |
| ≥ 5 | | 560 | | | 157 | | | | 1060 | | | | 101 | | | |

Note: The 1st Q and 3rd Q indicates first and third quantile, while SD refers to the value of standard deviation.

**Effect on MOCA scores**

Identical model was estimated using MOCA scores, where results show identical effect as in the case of MMSE measures. Patients started to decline cognitively on average 1 year and 7 months before medication was prescribed (Figure S1). Similar to the MMSE-based models, the cognitive decline was interrupted at the time point of medication prescription. Finally, after being cognitively stable for approximately 7 months, the decline in MOCA scores continued at a similar rate to the one observed pre-medication.



Figure S1. Change in MOCA scores over time in relation to time point of medication prescription (year 0 on the X-axis). The red colour indicates periods with statistically significant change in MMSE scores relative to the trajectory projected by the analysis.

**Effect of Age and Time since diagnosis**

Besides estimating effect of medication prescription, the nonlinear model presented in the study included nonlinear predictors of age and time since diagnosis. Results show that both effects are significant, where older patients (see Figure S1) and the ones with the diagnosis for a long time (see Figure S2), score lower on MMSE questionnaire.



Figure S2: Nonlinear effect of age



Figure S3: Nonlinear effect of time since the diagnosis of dementia

**Effect of medication separated by NHS trust**

The main model reported in the study was also separately fitted to the Trust specific data, thus, we separated the analysis by the two sources of the data. Results show that we can identify preserving effect of medication in both Trusts (see Figure S3A and S3B). The preserving effect is stronger in the case of Southern Health NHS data in comparison to the Oxford Trust, showing 5 months of cognitive stabilization instead of 2 months.



Figure S4. Change in MMSE scale scores over time in relation to time point of medication prescription (year 0 on the X-axis) separated by NHS trust. The red colour indicates periods with statistically significant change in MMSE scores relative to the trajectory projected by the analysis.

Identical effect is observed using MOCA scale, where stabilization is replicated on both data sources. However, due to the fewer cases with MOCA observations, especially after medication is prescribed, the decline after initial medication treatment is not replicated in the case of Southern Health NHS trust (see Figure S4A and SB).



Figure S5. Change in MMSE scale scores over time in relation to time point of medication prescription (year 0 on the X-axis) separated by NHS trust. The red colour indicates periods with statistically significant change in MMSE scores relative to the trajectory projected by the analysis.

**Predicting the medication switch**

In the final step of the analysis, we used logistic regression to investigate which factors contribute to monotherapy regime versus medication switch patients. Results show that older patients and the ones without diagnosis at the first visit are more likely to be in the monotherapy group. Being on concurrent antidepressant therapy increases the changes of medication switch situation. In contrast to that, patients on diabetes related medication are more likely to be in the group of monotherapy medication. Finally, scoring lower on MMSE scale at the time of dementia-related medication prescription decreases the likelihood of being in medication switch category.

Table S4. Prediction of monotherapy versus medication switch group

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | z value | Pr(>|z|) |
| DiagnosisMixedADVaD | 0.07 | 0.35 | 0.84 | .39 |
| DiagnosisNoDiagnosis | -0.61 | 0.11 | -5.22 | <.001 |
| DiagnosisVaD | -0.12 | 0.16 | -0.77 | 0.44 |
| DiagnosisUnspecified | -0.01 | 0.06 | -0.15 | .87 |
| DiagnosisOther | 0.07 | 0.20 | 0.35 | .72 |
| AntipsychoticsYes | 0.18 | 0.09 | 1.89 | 0.058 |
| AntidepressantsYes | 0.29 | 0.06 | 4.21 | <.001 |
| DiabetesYes | -0.32 | 0.16 | -1.99 | <.05 |
| Age | -0.02 | 0.004 | -4.18 | <.001 |
| GenderMale | 0.04 | 0.06 | 0.69 | .49 |
| SeverityMild | -0.02 | 0.07 | -0.03 | 0.96 |
| SeverityModerate | -0.37 | 0.08 | -4.41 | <.001 |
| SeveritySevere | -0.77 | 0.13 | -5.75 | <.001 |