|  |
| --- |
| **Outcome Adjudication**  Work Package 9 |
| **Objective(s)**: |
| To develop methods for dementia case identification at scale using the data related to the UK Biobank (UKB) cohort and produce an online system for remote classification.  The methods were to be validated using UKB and other DPUK cohorts and more specific objectives were to:   1. Provide up-to-date assessments of the accuracy of routine data sources for dementia, Parkinson’s Disease (PD) and motor neurone disease (MND) outcomes 2. Develop algorithms based on routine data sources for ascertaining, confirming and sub-classifying these outcomes 3. Develop protocols for remote confirmation and sub-classification using an online system 4. Apply these protocols for adjudication (i.e. confirmation according to standard criteria) the dementia, PD and MND outcomes occurring in UKB and other cohorts |
| **Overview Summary:** |
| This work investigated the accuracy of using routinely-collected primary care, hospital admissions and mortality datasets to identify dementia, motor neurone disease (MND) or Parkinson’s disease (PD) cases. Systematic reviews on the accuracy of these datasets for dementia, MND and PD were published. These reviews summarised the findings of existing validation studies, and identified gaps in current knowledge, particularly with respect to a UK setting. Following on from this, a validation study of UK primary care, hospital admissions and mortality datasets was conducted using data from [UK Biobank](https://www.ukbiobank.ac.uk/) In this study, coded diagnoses from routine datasets were compared to the full text medical record, leading to the identification of algorithms that can be employed by UK cohorts to identify participants who develop dementia during follow up. UK Biobank has used these findings to provide derived dementia outcome variables for researchers, thereby facilitating dementia research using the resource. These algorithms were made freely available to the research community (as the publications are Open Access) enabling other UK cohorts to adopt this approach to dementia case identification. Similar approaches have been used for Parkinson’s Disease ascertainment.  A separate project from the same group, with collaborators at Swansea University, has used routinely-collected healthcare data to create an electronic cohort (e-cohort) for dementia research. Utilising the SAIL Databank (https://saildatabank.com/), this e-cohort contains data for 1.2 million people, of whom 130,000 developed dementia during follow-up. Access to the e-cohort is facilitated through DPUK |
| **Executive Summary:** |
| The aim of this work was to develop methods for accurately identifying neurodegenerative conditions including dementia, motor neurone disease (MND) and Parkinson’s disease (PD) at scale within UK cohorts, using UK Biobank as an exemplar. The main objectives were:  Objective 1. Provide up-to-date assessments of the accuracy of routine data sources for dementia, PD and MND outcomes  Objective 2. Develop algorithms based on routine data sources for ascertaining, confirming and sub-classifying these outcomes  Objective 3. Develop protocols for remote confirmation and sub-classification using an online system  Objective 4. Apply these protocols for adjudication (i.e. confirmation according to standard criteria) of the dementia, PD and MND outcomes occurring in UKB and other cohorts.  The work produced by the group was particularly important in showing that the coding of dementia and Parkinson’s Disease by General Practice is generally accurate and can be used to determine cases of these diseases. |
| **Summary of Outputs**: (as per Researchfish categories) |
| **Publications:** |
| * Horrocks S., et al. (2017) **Accuracy of routinely collected healthcare data for identifying motor neurone disease cases: A systematic review.** PLoS ONE 12(2): e0172639.<https://doi.org/10.1371/journal.pone.0172639>   *In this novel systematic review, the team summarised evidence from validation studies that used routine datasets to identify motor neurone disease cases, allowing researchers around the world who are using these datasets to understand their accuracy. This work also identified gaps in knowledge, indicating where further validation research is required.*   * Wilkinson T., et al. **Identifying dementia cases with routinely collected health data: a systematic review.** Alzheimer’s Dement. 2018 Aug;14(8):1038–51.   *The team systematically reviewed the international literature on the accuracy of using routine datasets to identify dementia cases - a previously unaddressed area of evidence synthesis. It identified high accuracy datasets suitable for dementia research and highlighted areas requiring further research, advising cohorts to conduct validation studies where such gaps exist.*   * Harding Z., et al. **Accuracy of routinely collected health data for identifying Parkinson’s disease and parkinsonism cases: a systematic review.** PLOS ONE. 2019 Jan 31;14(1):e0198736.   *The team systematically reviewed studies validating the accuracy of using routinely collected healthcare datasets to identify Parkinson’s disease cases. It found relatively wide-ranging accuracy estimates and identified gaps in current knowledge. It highlighted areas where further, context-specific validation studies are warranted plus identified additional datasets that may assist future algorithm development.*   * Wilkinson T., et al. **Identifying dementia outcomes in UK Biobank: a validation study of primary care, hospital admissions and mortality data.** Eur J Epidemiol. 2019 Feb 26. doi: 10.1007/s10654-019-00499-1.   *Using UK Biobank (UKB) data, the team conducted a validation study of using routinely collected healthcare datasets (primary care, hospital admissions and mortality data) to identify dementia cases. It identified three algorithms that are likely to be highly predictive of all-cause dementia. UKB will use these to generate derived dementia outcome variables for researchers.*   * Wilkinson T., et al. **Cohort Data resource profile: The Secure Anonymised Information**   **Linkage Databank Dementia e-cohort (SAIL-DeC).** International Journal of Population Health Sciences, v5 (i1),2020. <https://doi.org/10.23889/ijpds.v5i1.1121>  *The team aimed to apply reproducible methods to create the SAIL dementia e-cohort (SAIL-DeC). It created SAIL-DeC with a view to maximising its utility for a broad range of research questions whilst minimising duplication of effort for researchers. From 4.4 million unique participants in SAIL, 1.2 million met the cohort inclusion criteria, resulting in 18.8 million person-years of follow-up. Among the dementia cases, the median duration of observation time was 14 years.*  ***Further publications***   * Wu H, Hodgson K, Dyson S, Morley KI, Ibrahim ZM, Iqbal E, Stewart R, Dobson RJ, Sudlow C. Efficient reuse of natural language processing models for phenotype-mention identification in free-text electronic medical records: a phenotype embedding approach. JMIR Med Inform 2019;7:e14782. doi: 10.2196/14782. * Denaxas S, (13 authors), Sudlow CLM, Hemingway H. UK phenomics platform for developing and validating EHR phenotypes: CALIBER. Journal of the American Medical Informatics Association 2019;26:1545-1559. doi: 10.1093/jamia/ocz105. * Alex B, Grover C, Tobin R, Sudlow CL, Mair G, Whiteley W. Text Mining Brain Imaging Reports. Journal of Biomedical Semantics 2019;10(Suppl 1):23. doi: 10.1186/s13326-019-0211-7. * Davis K AS, Bashford O, Jewell A, Shetty H, Stewart RJ, Sudlow CLM, Hotopf M. Using data linkage to electronic patient records to assess the validity of selected mental health diagnoses in English Hospital Episode Statistics (HES) PLoS One 2018; 13:e0195002. * Denaxas S, Gonzalez-Izquierdo A, Direk K, Fitzpatrick NK, Fatemifar G, Banerjee A, Dobson RJB, Howe LJ, Kuan V, Lumbers RT, Pasea L, Patel RS, Shah AD, Hingorani AD, Sudlow C, Hemingway H. UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. J Am Med Inform Assoc. 2019 Dec 1;26(12):1545-1559. doi: 10.1093/jamia/ocz105. * Calvin CM, Wilkinson T, Starr JM, Sudlow C, Hagenaars SP, Harris SE, Schnier C, Davies G, Fawns-Ritchie C, Gale CR, Gallacher J, Deary IJ.. Predicting incident dementia 3-8 years after brief cognitive tests in the UK Biobank prospective study of 500,000 people. Alzheimer's & Dementia 2019 Oct 13. pii: S1552-5260(19)35137-4. doi: 10.1016/j.jalz.2019.07.014. [Epub ahead of print] * Denaxas S, Gonzalez-Izquierdo A, Direk K, Fitzpatrick NK, Fatemifar G, Banerjee A, Dobson RJB, Howe LJ, Kuan V, Lumbers RT, Pasea L, Patel RS, Shah AD, Hingorani AD, Sudlow C, Hemingway H. UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. J Am Med Inform Assoc. 2019 Dec 1;26(12):1545-1559. doi: 10.1093/jamia/ocz105. * Calvin CM, Wilkinson T, Starr JM, Sudlow C, Hagenaars SP, Harris SE, Schnier C, Davies G, Fawns-Ritchie C, Gale CR, Gallacher J, Deary IJ.. Predicting incident dementia 3-8 years after brief cognitive tests in the UK Biobank prospective study of 500,000 people. Alzheimer's & Dementia 2019 Oct 13. pii: S1552-5260(19)35137-4. doi: 10.1016/j.jalz.2019.07.014. [Epub ahead of print] * Giebel, C., Hollinghurst, J., Akbari, A., Schnier, C., Wilkinson, T., North, L., Gabbay, M. & Rodgers, S., , Socio‐economic predictors of time to care home admission in people living with dementia in Wales: A routine data linkage study: International Journal of Geriatric Psychiatry12 Oct 2020. doi: 10.1002/gps.5446. * Schnier C, Duncan S, Wilkinson T, Mbizvo GK, Chin RFM. A nationwide, retrospective, data-linkage, cohort study of epilepsy and incident dementia. Neurology. 2020 Sep 22;95(12):e1686-e1693. doi: 10.1212/WNL.0000000000010358. Epub 2020 Jul 17. PMID: 32680951   **In preparation / submission:**   * Bush K., et al. **Using routinely collected healthcare data to identify Parkinson’s disease in UK Biobank.** * Ly A., et al. **Dementia case identification using linkage to electronic health data in population-based cohort studies: the Whitehall II cohort study.** |
| **Collaborations & Partnerships** |
| The work has only been possible due on the team’s close collaboration with UK Biobank and its permission to access linked healthcare records. |
| **Further Funding** |
| N/A |
| **Next Destinations** |
| Tim Wilkinson is now a Clinical Lecturer in the Centre for Clinical Brain Sciences at the University of Edinburgh. |
| **Engagement Activities** |
| * UK Biobank annual conference. T Wilkinson. Identifying dementia cases in UK Biobank using routinely collected health data. June 2016 * DPUK/EMIF joint conference. T Wilkinson. Identifying dementia in prospective population-based studies. March 2017. * Informatics for Health conference. A Ly. Dementia case ascertainment in population-based cohort studies –lessons from the Whitehall II study. April 2017 * DPUK annual conference. A Ly. Dementia case ascertainment in population-based cohort studies –lessons from the Whitehall II study. May 2017 * DPUK annual conference. C Sudlow. Ascertaining dementia outcomes. May 2017 |
| **Influence of policy, practice, patients & the public** |
| Researchers can now apply the developed algorithms for a whole range of use cases and could in future use them to create e-cohorts for PD and MND. Further advances will be possible by recreating the Dementia e-cohort in other UK nations as the work of HDRUK and others progresses to link up General Practice and hospital records across the four UK nations. |
| **Research Tools & Methods** |
| **ALGORITHMS**   * Dementia algorithm – for details see Wilkinson T., et al. Identifying dementia outcomes in UK Biobank: a validation study of primary care, hospital admissions and mortality data. Eur J Epidemiol. 2019 Feb 26. doi: 10.1007/s10654-019-00499-1. * Parkinson’s Disease algorithm - <https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/alg_outcome_pdp.pdf> * Motor Neurone Disease algorithm - <https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/alg_outcome_mnd.pdf> |
| **Research Databases & Models** |
| See above re algorithms |
| **Intellectual property & licencing** |
| N/A |
| **Medical products, interventions & clinical trials** |
| N/A |
| **Artistic & creative products** |
| N/A |
| **Software & technical products** |
| N/A |
| **Spin outs** |
| N/A |
| **Awards & recognition** |
| Cathie Sudlow was awarded an OBE in the Queen’s Birthday Honours 2020. |
| **Other outputs & knowledge/future steps** |
| The e-cohort of dementia patients for the whole Welsh population associated with this work has already been used as a resource to investigate other aspects of dementia including its interrelationship with epilepsy and the socio-economic data associated with dementia patients in care homes. |
| **Use of facilities & resources** |
| The work was reliant on access to linked healthcare records. |
| **Most successful outcome and what it means for future dementia research**: |
| The ability to access GP data and select the codes that indicated a diagnoses of dementia, Parkinson’s disease and Motor neurone disease led to validation studies. This work has shown that the GP codes are reliable diagnoses for these diseases.  The creation of an e-Cohort of dementia patients in the SAIL databank. This work will be an ongoing resource for researchers all over the world. For details of this work see the separate final report on DPUK Discovery Award: Building an e-cohort for dementia research. |
| **Lessons learned**: |
| * Working across multiple cohorts with different information governance and consent procedures has been challenging. The only cohorts that collaborated with this work package were UK Biobank, Whitehall II and Generation Scotland although many other DPUK cohorts were contacted and asked if they would work with the team. * Working to access NHS data has been challenging as well. Using consented UK Biobank participants the team was able to manually access electronic medical records. However, the NHS can be resistant to allowing new natural language processing programmes into their IT systems. This reluctance has meant that the team was unable to extend the work looking at free text using scalable NLP techniques. |
| **Other:** |
| Report completed 3 December 2020 |