International Journal of Population Data Science

Journal Website: www.ijpds.org

Validating a novel deterministic privacy-preserving record linkage between administrative & clinical data: applications in stroke research

Alisia Southwell[1], Susan Bronskill[2,3,4], Tom Gee[5], Brendan Behan[6], Susan Gilbert Evans[5], Tom Mikkelsen[6], Elizabeth Theriault[6], Kirk Nylen[6,7], Shannon Lefaivre[6], Nelson Chong[2], Mahmoud Azimaee[2,3], Natasa Tusevljak[2], Doug Lee[2,3,9], and Richard H. Swartz[1,2,4,8,10]

Submission History

Submitted:	27/04/2022
Accepted:	16/08/2022
Published:	22/11/2022

[1]Department of Medicine (Neurology), Sunnybrook Health Sciences Centre, TorontoToronto, M4N 3M5
[2]ICES, Toronto, G1 06, 2075 Bayview AvenueToronto, ON M4N 3M5
[3]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, M5T 3M6
[4]Hurvitz Brain Sciences Research Program, Sunnybrook Research Institute, Toronto, M4N 3M5
[5]Indoc Research, Toronto, M5H 3W4
[6]Ontario Brain Institute, Toronto, M5H 3W4
[7]Department of Pharmacology & Toxicology, University of Toronto, Toronto, M5S 1A8
[8]Department of Medicine, University of Toronto, Toronto, Ontario, M5S 1A8
[9]Division of Cardiology, Peter Munk Cardiac Centre, UHN, Toronto, M5G 2N2
[10]Heart and Stroke Foundation Canadian Partnership for Stroke Recovery, K1G 5Z3

Abstract

Introduction
Research data combined with administrative data provides a robust resource capable of answering unique research questions. However, in cases where personal health data are encrypted, due to ethics requirements or institutional restrictions, traditional methods of deterministic and probabilistic record linkages are not feasible. Instead, privacy-preserving record linkages must be used to protect patients' personal data during data linkage.

Objectives
To determine the feasibility and validity of a deterministic privacy preserving data linkage protocol using homomorphically encrypted data.

Methods
Feasibility was measured by the number of records that successfully matched via direct identifiers. Validity was measured by the number of records that matched with multiple indirect identifiers. The threshold for feasibility and validity were both set at 95%. The datasets shared a single, direct identifier (health card number) and multiple indirect identifiers (sex and date of birth). Direct identifiers were encrypted in both datasets and then transferred to a third-party server capable of linking the encrypted identifiers without decrypting individual records. Once linked, the study team used indirect identifiers to verify the accuracy of the linkage in the final dataset.

Results
With a combination of manual and automated data transfer in a sample of 8,128 individuals, the privacy-preserving data linkage took 36 days to match to a population sample of over 3.2 million records. 99.9% of the records were successfully matched with direct identifiers, and 99.8% successfully matched with multiple indirect identifiers. We deemed the linkage both feasible and valid.

Conclusions
As combining administrative and research data becomes increasingly common, it is imperative to understand options for linking data when direct linkage is not feasible. The current linkage process ensured the privacy and security of patient data and improved data quality. While the initial implementations required significant computational and human resources, increased automation keeps the requirements within feasible bounds.

Keywords
data linkage; stroke; feasibility; privacy; personal health information

*Corresponding Author:
Email Address: rick.swartz@sunnybrook.ca (Richard H. Swartz)

https://doi.org/10.23889/ijpds.v7i4.1755

Introduction

Health administrative data contain routine clinical information collected from patients' hospital, clinic, and laboratory visits [1]. These data can be unique and valuable due to their ability to cover geographic regions spanning large populations, and to track individual patients across different health care visits and over time [2]. These data are particularly advantageous in epidemiological research, as they often offer a large selection of variables and provide a robust and representative picture of real-world populations [1, 3]. Conversely, health administrative data can be limited by the specificity of diagnostic codes, can lack clinically important medical history, and can be limited to admission rather than discharge details [1, 4]. As such there are specific health research questions that administrative data answer well, and questions where primary data collection is required. This can, in turn, limit the types of analyses that can be conducted [1, 4–6].

Integrating administrative data with prospectively collected data from research studies can provide reliable, large samples and inexpensive long-term outcome data (e.g., mortality, hospital readmissions), as well as detailed and robust patient-level data [2]. The value of linking 'broad' health administrative data and 'deep' clinical data from research studies is that it enriches the context necessary to better understand the multi-factorial nature of chronic disease and can help address both health and social policy needs [7]. However, merging large, clinical administrative databases with research-driven databases is not always feasible. Data collected specifically for research projects may have limitations regarding which data can be stored due to ethics approvals or may be conducted in institutions that are not permitted, or willing to, store personal health information due to the potential liabilities associated with privacy breaches. In a society with an ever-increasing concern about privacy protection, there may also be an increase in projects circumventing the collection and storage of certain private health information. These differences in which data can be stored and how is a key feature of data governance.

In Ontario, the Personal Health Information Protection Act (PHIPA) states that only Prescribed Entities are authorized to collect, store and use personal health information, such as health card numbers, without explicit consent (whether for research purposes or not) [8]. This means that project-specific cohort data collected by non-Prescribed Entities may not be allowed to store the personal health information inherent in clinical administrative datasets without encryption or de-identification of personal health information. Additionally, entities allowed to use health administrative data for research purposes must often follow strict privacy protocols related to how data can be linked and shared. Conflicting data governance in these cases can pose problems when merging datasets [9].

Data can be matched using exact or probable matches and via direct or indirect identifiers. A deterministic linkage relies on matching exact strings of data (i.e., a record with the value "12345" will only match to another record whose value is "12345") whereas a probabilistic linkage can match approximate strings of data (i.e., "12345" could match with "12344" or "12346"). Direct identifiers are data points that are unique to an individual, such as personal health card numbers, where one number can only be associated with one individual [10, 11]. Conversely, indirect identifiers are data points that may be applicable to many individuals, such as name, sex, and year of birth. If project-specific data governance does not allow for the storage of certain personal health information, it may not be feasible to conduct linkages with direct identifiers. When dealing with health care data, in particular, the lack of direct identifiers often means that a privacy-preserving record linkage (PPRL) is required to link the databases [12, 13]; this method ensures that no personal data are revealed in the process of combining the datasets. Due to the potential errors and variation in indirect identifiers (e.g., a patient's name which could match as "Elizabeth", "Elisabeth", or "Liz"), probabilistic privacy-preserving linkages, often using Bloom filter encoding [14, 15], have shown great success in health care datasets [13, 16–19]. Deterministic PPRLs, or combinations between probabilistic and deterministic algorithms, have also become more common and have had demonstrated success using healthcare data [20–22]. Homomorphic encryption is one method of encrypting sensitive data that allows users to perform computations while data remain encrypted [23]. Given that homomorphic encryption uses complex algorithms to transform raw data into its encrypted form, the only way to ensure accurate linking of the encrypted data is via exact matching or a deterministic linkage.

In the present study, we aimed to combine a project-specific Ontario-based cohort of stroke patients with a provincial health administrative database. Project-specific data were collected and stored with the Ontario Brain Institute, a non-Prescribed Entity, where homomorphic encryption was performed on direct identifiers at the time of data entry [12]. As a result, our project necessitated a deterministic, privacy-preserving record linkage (PPRL) and thus our primary goal was to determine if we could complete a governance agnostic data linkage using a single direct identifier. Reliability of this PPRL protocol has previously been established via several billion test comparisons [12]. In this case, considering that we did not have access to unencrypted direct identifiers to complete a gold standard evaluation of the PPRL, we instead opted to perform a silver standard evaluation of the PPRL by comparing multiple indirect identifiers (e.g., data points that characterize individuals but are not unique to them, such as sex and date of birth combined) [24].

Our objectives are twofold: (1) to evaluate the feasibility of a PPRL using encrypted direct identifiers; (2) to examine the validity of the linkage using multiple indirect identifiers. We hypothesize that our PPRL will be both feasible and accurate.

Methods

A good record linkage can be said to have a sensitivity over 95% [25]. Thus, we established that feasibility will be fulfilled if at least 95% of the records successfully match during the linkage, and validity will be fulfilled if at least 95% of the records' indirect identifiers accurately match.

There are several steps required to perform the PPRL: (1) the cohort project team must clean and prepare the cohort dataset for linkage, (2) the administrative project team needs to prepare the administrative dataset for linkage, (3) a

deterministic linkage software will then create a comparison file of the de-identified datasets, and (4) a third-party will identify matching encrypted records. To establish feasibility, both the cohort and administrative datasets must share at least one direct identifier; to validate the linkage both datasets also need to share (at least) two indirect identifiers. Shared indirect identifiers will allow the study team to ascertain the accuracy of the dataset once all direct identifiers have been removed. In the present study, we used encrypted government-issued health card numbers as the shared direct identifier and date of birth and sex as shared indirect identifiers. A detailed outline of the algorithms and methodology for this specific project can be found in Supplementary Appendix 1. This project has been approved by the Sunnybrook Research Ethics Board.

Cohort dataset

The depression, obstructive sleep apnea, and cognitive impairment (DOC) screen (www.docscreen.ca) was developed as a clinical tool to facilitate routine screening for these common conditions after stroke [26, 27]. The DOC Utility project is a multi-centered observational study to determine whether symptoms of depression, sleep apnea, and cognitive impairment can add to known risk factors for recurrent stroke and mortality. The DOC screen (Supplementary Appendix 2) is administered at patients' initial stroke prevention clinic visit following a referral. The screen is administered on all incoming patients; however, the population of interest in the DOC study is patients with a confirmed diagnosis of ischemic stroke or probable transient ischemic attack (TIA). In addition to screening scores of depression, apnea, and cognitive impairment, trained abstractors collect clinically relevant demographic and medical history details for the trial (Supplementary Appendix 3). Participants' health card numbers went through double-entry and validity checksum [28] into an electronic data capture tool where they were homomorphically encrypted at source.

Prior to the linkage, the DOC study team performed manual and automated data cleaning checks for missingness, dates, numeric variables, and free text fields. Manual flags consisted of checking for study-specific inconsistencies (e.g., date of referral occurring after the initial clinic visit day) and automated flags consisted of univariate outlier analyses for the numeric variables over/under two standard deviations. All flags resulted in 3,311 queries across all sites participating in the study, all of which were resolved before locking the site in preparation for data linkage. There were 8,128 participant records to match.

Administrative dataset

Administrative data were required to ascertain the long-term outcomes (e.g., hospitalizations, mortality post-stroke) of the DOC cohort. In the present study, our administrative project team had access to health system interactions for over 14 million Ontarians. These datasets were linked using unique encoded identifiers and analyzed at ICES. ICES is an independent, non-profit research institute whose legal status under Ontario's health information privacy law allows it to collect and analyze health care and demographic data, without consent, for health system evaluation and improvement. Matching a 14 million participant dataset to an eight thousand participant dataset would expend enormous computational resources; accordingly, the administrative team worked to reduce the health administrative dataset, without eliminating any potential matches by filtering based on age range, diagnosis, date of diagnosis, place of residence and additional project-specific details to reduce the potential match database to 3.2 million records.

Deterministic linkage software

A PPRL requires software capable of performing a deterministic linkage with large datasets of encrypted identifiers. In this project we used Brain-CODE Link, software developed by the Indoc Consortium, in partnership with the Ontario Brain Institute [12, 29, 30]. Brain-CODE Link uses exponential ElGamal public-key cryptography for homomorphic encryption. Using the Ontario Brain Institute public key, this encryption was applied at source to the cohort data identifier. The same encryption algorithm and public key was used to encrypt the administrative data identifier during the linkage process - this allowed for comparisons between datasets without requiring (or even enabling) decryption to the original values. The software can calculate the exponential difference between the encrypted source values and the encrypted administrative values. These differences are calculated for each cartesian product of the two imported databases and are subsequently provided to the third party in a comparison file. The third party, who also holds the private decryption key, uses that key to decrypt the differences between the source and administrative identifiers, allowing exact matches to be determined.

Even with a reduced database of 3.2 million records, the cartesian match product of both datasets still surpasses 26 billion records for the comparison file. As such, it is important that the third-party computer has adequate storage, computer memory and processing capabilities to accept and perform data transfers.

Role of third-party

A neutral third-party is needed to maintain both the matched and unmatched datasets in the same space during decryption. The third-party received the full comparison file and a decryption key for the encrypted, calculated differences. In the present study, this consisted of data linkage personnel who did not have the deterministic linkage information. Once decrypted, the differences in the compare file will equal zero if the records are in fact a match; any *non*-zero number would represent a *non*-match. The third-party was thus responsible for identifying all records whose 'difference' was zero, and then transferring the randomized IDs of those matched records back to the study team at the administrative data centre for final cohort creation. With this method, the third party will also remain blind to the personal health information from the cohort and administrative datasets.

Final study dataset creation

At this stage, the cohort and administrative teams worked jointly to create the final study dataset. The third-party

transferred two sets of randomized identifiers (IDs) for each matched record, the IDs provided by the cohort team, and the IDs provided by the administrative team. The cohort team populated clinical and project-specific data based on their randomized IDs and provided that information to the administrative team. The administrative team populated the clinical variables requested by the research team and merged the datasets in a specified location that the research study team could access for later analyses.

Outcomes

The primary aim of this methodology is to ensure that the deterministic privacy-preserving linkage is in fact feasible. This means that the third-party can find matched records based on the provided comparison file and decryption key. Secondly, it is important that these matches are accurate. Thus, validity will be met if the matched records returned from the third-party can be verified as accurate based on a combination of indirect identifiers. We quantified feasibility and validity as follows:

1. Feasibility: at least 95% of the records provided to the third party are returned as matched records (i.e., matched health card numbers).
2. Validity: at least 95% of the indirect identifier pairs accurately match in the final dataset (i.e., date of birth *and* sex).

Results

The final study dataset resulted in 8,117 individuals having linked cohort and administrative data which allowed for further analyses with the final study dataset.

Feasibility

The linkage process began 3$^{\mathrm{rd}}$ June 2019 and, with a combination of manual and automated transfer, ended on 9$^{\mathrm{th}}$ July 2019 (36 days). The data were transferred in 55 packets of approximately 500GB each, having between 144 and 190 records per packet. In total, 8,117 of the 8,128 (99.9%) participant records shared with the third-party were successfully linked with the 3.2 million records provided by the administrative team; 11 health card numbers did not match.

Validity

Validation occurred after linkage and cohort creation and involved investigating whether the indirect identifiers provided by the cohort and health administrative datasets matched. During the first iteration of validation, 7,832 (96.5%) of the 8,117 records were exact matches for both indirect identifiers (sex and date of birth). Although this met our 95% threshold for validity, a granular review of some of the mismatched records revealed many near matches, where most but not all indirect data points matched (e.g., sex, year and month of birth matched however the date did not). As a result, the cohort team decided to further investigate the 285 records identified as mismatches. The study team audited the original dataset (i.e., raw data obtained prior to the data linkage) which allowed them to verify and amend 267 of the 285 mismatched records while maintaining the personal health information of the administrative dataset. Most remaining mismatched cases could not be verified with source data. The values of the few that could be verified with source data still returned as mismatched. These 18 cases were ultimately excluded from the final dataset. A second iteration of validation resulted in 8,103 (99.8%) records with exact matches of both indirect identifiers.

Incidental findings

The data linkage also uncovered 72 duplicates; this meant that there were 72 records in the health administrative dataset that each matched to two records in the cohort dataset. To ensure data quality, the cohort team investigated these duplicates with the goal of reducing the final dataset to wholly unique participants (once again this required returning to source data). The study team confirmed that 49 pairs of records included patients who had been screened more than once for the same, qualifying event (i.e., stroke or TIA); here we kept the earlier of both records. Fewer than twenty pairs included one individual who had experienced multiple qualifying events; in these cases, we also kept the earliest record of both events. The final pairs ($<$6) included individuals who had different referral events, one being a true qualifying event and the other a non-qualifying event (e.g., abnormal CT); in these cases, we removed records with a non-qualifying event.

A breakdown of these results can be found in Figure 1.

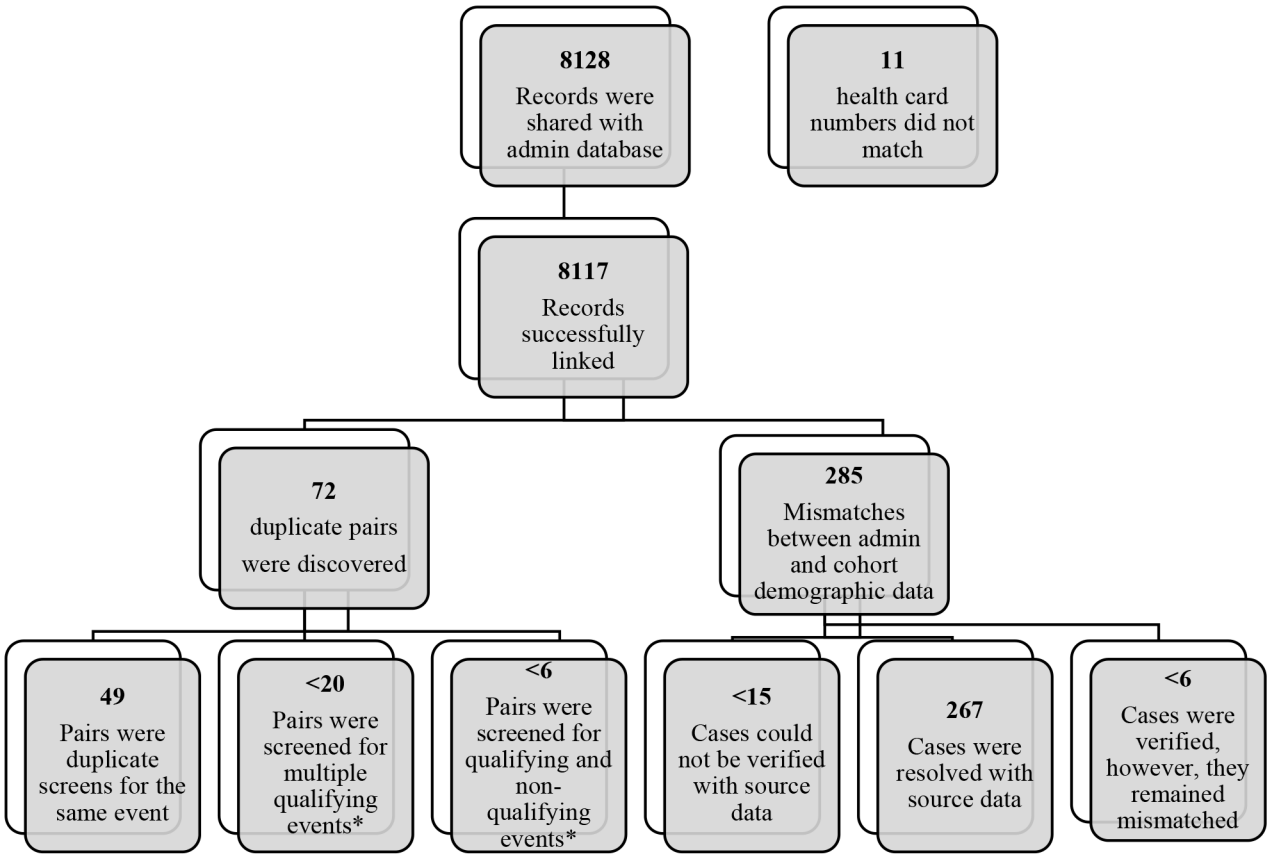
Discussion

The main advance described here is the ability to perform a direct match without storing or decrypting personal health information at any point. With 99.9% of records successfully matching, we demonstrated that privacy-preserving linkage between homomorphically encrypted identifiers is feasible on a large scale. We have also been able to demonstrate the accuracy of this linkage, with 99.8% of the records matching to two indirect identifiers (sex and date of birth). Thus, we can also have confidence that, while maintaining data privacy and security with encrypted identifiers, this PPRL meets our criteria of a silver standard validation of data linkage.

There were 11 records that did not match, and a few suspected reasons for this. First, despite processes meant to mitigate errors (e.g., double entry of OHIP numbers), it is possible that our direct identifier was entered incorrectly at the source of the project-specific data. In that case, though the same participant might exist in both databases, the direct identifier would not have matched during data linkage. An alternative explanation is that, in reducing the health administrative dataset (from 14 million to 3.2 million) to reduce computational resources, we may have filtered out potential matches. It is not possible to distinguish between these two possibilities due to the deterministic nature of the PPRL (i.e., we cannot ascertain whether any unmatched health cards were approximate matches or not).

Moreover, the linkage process facilitated data cleaning and improved the quality of the DOC dataset by identifying multiple types of source data errors. The linkage uncovered

Figure 1: Results of the privacy-preserving record linkage



*Qualifying events in this study were considered a diagnosis of 'stroke' or 'transient ischemic attack'.

72 duplicate records, all of which could be resolved. These cases were an artifact of the multicenter, multi-year design of the DOC Utility study. Seeing as health card numbers were encrypted at source, these repeat entries could not be screened out. Yet, in a high-risk stroke/TIA population, recurrent events or second opinions across institutions, or even patients moving from site to site, are not uncommon. Overall, our rate of 0.89% of duplicates is quite low.

The linkage also revealed 285 demographic mismatches between the health administrative and cohort datasets. Of those, 267 cases of data entry errors for sex and date of birth were confirmed, despite extensive data monitoring and cleaning prior to matching and resolving over 3,000 identified data queries. Given that there are four data fields (sex, month, year and date of birth), the rate of error (267/32,468 fields = 0.8%) is also relatively low. Comparably, Goldberg et al. investigated two clinical research databases and found the error rates to range between 2.3% and 5.2% for demographic data, with double-entry auditing catching only a portion of these types of errors [31]. The process of automated and manual screening checks identified many of these, but some date errors (e.g., 25$^{\text{th}}$ March vs. 3$^{\text{rd}}$ day, 25$^{\text{th}}$ month) are easy to identify at entry or in cleaning, while others (e.g., 4$^{\text{th}}$ March vs. 3$^{\text{rd}}$ April) would be missed by these approaches. Most cases that could not be verified with source data were removed from the final dataset. Additionally, there were a few cases that were verified with source data; however, their demographic data points were still different than the health administrative dataset and we suspect that in these cases our direct identifiers were incorrectly entered at the source of the project-specific data. As such, those records may have matched to different patients in the health administrative database. Those records were also excluded (considered failed matches) from the final dataset.

To complete a full investigation into the 11 unmatched and fewer than six mismatched records, we would need to have access to the identifiable information inputted in the cohort dataset at source (rather than the encrypted values) to verify that there were no typographical errors in the data entry of HCNs. Alternatively, we would need to re-enter the health card numbers (our direct identifiers) of each patient into the cohort dataset and then repeat the entire data linkage. Dedicating significant resources to validate what amounts to 0.2% of our data was not feasible. However, increased automation, which may be combined with future efforts in continuous linkage processing and the use of 'crosswalk' linkage tables, could make this kind of data checking more achievable.

As data were homomorphically encrypted at source, it was not feasible to employ more traditional methods of encryption, such as Bloom filter [15] or multiple match-key encoding [32]. Although we may have gathered additional information about our unmatched health cards using one of these probabilistic methods, we may also have encountered a larger number of false matches, which are more likely in probabilistic rather than deterministic linkages [33]. In some cases, the margin of

error associated with probabilistic linkage may be above the acceptable threshold for smaller research databases (e.g., with just a few hundred records), particularly when matching with a single identifier and not having access to the raw data needed to investigate approximate matches. Moreover, in cases where linkage relied solely on indirect identifiers, it would indeed be more prudent to employ probabilistic methods; as we noted in our dataset, there were many mismatched records due to errors in entry or variations in data formatting that may not have been caught had our validation been performed at the same time as data linkage.

Another key consideration is the security and privacy offered by different linkage methods. While Bloom filter and multiple match key encoding are more efficient methods of linking, there may be some trade-offs between data linkage quality and vulnerability to privacy attacks [32, 34–36]. Homomorphic encryption algorithms offer exceptional security, and the proposed methodology delivers a highly accurate data linkage, with the trade-off instead being computational resources. Though presently strenuous, homomorphic encryption has progressed substantially in the last few decades, and is increasingly being used in the big data and healthcare spaces with ever-improving efficiency [37–39].

Limitations and future direction

This PPRL method does have drawbacks; though the proposed methodology may be feasible, it may not always be practical in cases where data governance allows institutions to share direct identifiers. In those cases, more traditional methods, such as Bloom Filter encoding, [15] would require significantly fewer resources. The resources required to complete this PPRL were extensive, in terms of staff (person hours) and computational requirements (e.g., high storage, RAM, CPU). Therefore, this type of linkage may not be feasible for smaller research teams with fewer resources. Further, while many record linkage protocols have access to both encrypted and unencrypted data and can then compare the results of each linkage with this "gold standard" approach, resulting in metrics of sensitivity, specificity, this was not possible in the current protocol.

The current study is the largest cohort used to perform a data linkage with the Brain-CODE Link software. Consequently, one of the most time-consuming processes was transferring data from the third-party's server to our administrative team. Initially, this was a completely manual process; however, our project teams were able to improve the process by facilitating a semi-automated transfer of data, which markedly sped up this pipeline. The automation allowed the system to run 24 hours a day and more than doubled the number of matches that could be performed each day. The automated transfer also permitted the system to operate over the weekend without operator oversight. When fully enabled, the automated transfer mechanism combined with the medium-sized trusted third-party server would permit up to 4 billion comparisons a day to be evaluated. The linkage described in this paper could be performed optimally in as little as 7 days, although this initial project required 36 calendar days because the process was entirely manual at first, and there were a number of initial software and hardware restrictions to be addressed. Moving forward, this increased automation could reduce some of the staff hours required and make this linkage more feasible on an even larger scale. Of note, the number of comparisons required for this project is due to the sizeable administrative dataset – the proposed methodology may prove even more feasible when linking two or more moderate-sized databases (e.g., with hundreds of thousands of records rather than millions). Further, seeing as we were only linking using a single direct identifier, we were not able to utilize blocking and filtering techniques that may have reduced the cartesian match product in the comparison file.

This PPRL protocol is also working towards real-time comparisons, where matching patient records would occur concurrently with data entry. This would allow researchers to identify data entry errors/duplicates at a much earlier stage, although this technological capability will need to be established within the bounds of an appropriate governance model.

Conclusion

The growing focus on pragmatic and/or registry embedded trials, [40, 41] privacy protection, and the increasing use of clinical administrative datasets in health and epidemiological studies is likely to result in greater demand for secure, privacy-protected linkages between traditional research datasets and administrative health records. Though it may be computationally intensive for some projects, deterministic linkage using encrypted personal health information can provide more robust datasets and data quality validation when data governance procedures do not align.

Acknowledgments

The DOC project is supported by the Canadian Institutes of Health Research (Funding Research Number: 137038), the Heart and Stroke Foundation and the Alternative Funding Plan from the Academic Health Sciences Centres of Ontario. This study was supported by ICES, which is funded by an annual grant from the Ontario Ministry of Health and the Ministry of Long-Term Care (MOHLTC). Parts of this material are based on data and/or information compiled and provided by CIHI. However, the analyses, conclusions, opinions and statements expressed in the material are those of the author(s), and not necessarily those of CIHI. The Ontario Brain Institute funding is provided in part by the Government of Ontario. Richard Swartz receives salary support for protected time for research from the Heart and Stroke Foundation (Clinician-Scientist Phase II Award), the Ontario Brain Institute through the Ontario Neurodegenerative Disease Research Initiative (ONDRI), and the Department of Medicine, University of Toronto and Sunnybrook HSC.

Conflicts of interest

There are no conflicts of interest to report.

Ethics statement

This research study was approved under the Sunnybrook Research Institute's Ethics Department, File number 141-2014, with a waiver of consent to participate.

References

1. Hashimoto R, Brodt E, Skelly A, Dettori J. Administrative database studies: Goldmine or goose chase? Evid Based Spine Care J. 2014;05(02):074–6. https://doi.org/10.1055/s-0034-1390027

2. Harbaugh CM, Cooper JN. Administrative databases. Semin Pediatr Surg. 2018;27(6):353–60. https://doi.org/10.1053/j.sempedsurg.2018.10.001

3. Gavrielov-Yusim N, Friger M. Use of administrative medical databases in population-based research. J Epidemiol Community Health. 2014;68(3):283–7. https://doi.org/10.1136/jech-2013-202744

4. Rabinstein AA. Administrative medical databases for clinical research: The good, the bad, and the ugly. Neurocrit Care. 2018;29(3):323–5. https://doi.org/10.1007/s12028-018-0625-6

5. Van Walraven C, Austin P. Administrative database research has unique characteristics that can risk biased results. J Clin Epidemiol. 2012;65(2):126–31. https://doi.org/10.1016/j.jclinepi.2011.08.002

6. Tirschwell DL, Longstreth WT. Validating administrative data in stroke research. Stroke. 2002;33(10):2465–70. https://doi.org/10.1161/01.STR.0000032240.28636.BD

7. OECD. Dementia Research and Care [Internet]. Anderson G, Oderkirk J, editors. Paris: OECD; 2015. 112 p. https://doi.org/10.1787/9789264228429-en https://www.oecd-ilibrary.org/social-issues-migration-health/dementia-research-and-care_9789264228429-en

8. Government of Ontario. Personal Health Information Protection Act, 2004, S.O. 2004, c. 3, Sched. A. 2019.

9. Harron K, Dibben C, Boyd J, Hjern A, Azimaee M, Barreto ML, et al. Challenges in administrative data linkage for research. Big Data Soc. 2017;4(2):1–12. https://doi.org/10.1177/2053951717745678

10. U.S. Department of Health & Human Services. Methods for de-identification of PHI [Internet]. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. 2015 [cited 2021 Dec 21]. https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html

11. Information and Privacy Commissioner of Ontario. De-identification guidelines for structured data. Toronto, Ontario; 2016.

12. Gee T, Behan B, Lefaivre S, Azimaee M, Dharsee M, El Emam K, et al. Designing and implementing a privacy preserving record linkage protocol. Int J Popul Data Sci. 2018 Sep;3(4):23889. https://doi.org/10.23889/ijpds.v3i4.831

13. Schmidlin K, Clough-Gorr KM, Spoerri A. Privacy Preserving Probabilistic Record Linkage (P3RL): A novel method for linking existing health-related data and maintaining participant confidentiality. BMC Med Res Methodol. 2015;15(1):1–10. https://doi.org/10.1186/s12874-015-0038-6

14. Schnell R, Bachteler T, Reiher J. Privacy-preserving record linkage using Bloom filters. BMC Med Inform Decis Mak [Internet]. 2009 Dec 25;9(1):41. https://doi.org/10.1186/1472-6947-9-41 https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-9-41

15. Christen P, Ranbaduge T, Schnell R. Bloom Filter based Encoding Methods. In: Linking Sensitive Data [Internet]. Cham: Springer International Publishing; 2020. p. 193–219. https://doi.org/10.1007/978-3-030-59706-1_8 http://link.springer.com/10.1007/978-3-030-59706-1_8

16. Randall S, Wichmann H, Brown A, Boyd J, Eitelhuber T, Merchant A, et al. A blinded evaluation of privacy preserving record linkage with Bloom filters. BMC Med Res Methodol [Internet]. 2022 Dec 16;22(1):22. https://doi.org/10.1186/s12874-022-01510-2 https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-022-01510-2

17. Nguyen L, Stoové M, Boyle D, Callander D, McManus H, Asselin J, et al. Privacy-Preserving Record Linkage of Deidentified Records Within a Public Health Surveillance System: Evaluation Study. J Med Internet Res [Internet]. 2020 Jun 24;22(6):e16757. https://doi.org/10.2196/16757 https://www.jmir.org/2020/6/e16757

18. Heidt CM, Hund H, Fegeler C. A Federated Record Linkage Algorithm for Secure Medical Data Sharing. In 2021. https://doi.org/10.3233/SHTI210062 https://ebooks.iospress.nl/doi/10.3233/SHTI210062

19. Brown AP, Randall SM. Secure Record Linkage of Large Health Data Sets: Evaluation of a Hybrid Cloud Model. JMIR Med Informatics [Internet]. 2020 Sep 23;8(9):e18920. https://doi.org/10.2196/18920 http://medinform.jmir.org/2020/9/e18920/

20. Ark TK, Kesselring S, Hills B, McGrail KM. Population Data BC: Supporting population data science in British Columbia. Int J Popul data Sci [Internet]. 2020 Mar 26;4(2):1133. https://doi.org/10.23889/ijpds.v5i1.1133 http://www.ncbi.nlm.nih.gov/pubmed/32935036

21. Kho AN, Cashy JP, Jackson KL, Pah AR, Goel S, Boehnke J, et al. Design and implementation of a privacy preserving electronic health record linkage tool in

Chicago. J Am Med Informatics Assoc [Internet]. 2015 Sep 1;22(5):1072–80. https://doi.org/10.1093/jamia/ocv038 https://academic.oup.com/jamia/article/22/5/1072/930113

22. Bian J, Loiacono A, Sura A, Mendoza Viramontes T, Lipori G, Guo Y, et al. Implementing a hash-based privacy-preserving record linkage tool in the OneFlorida clinical research network. JAMIA Open [Internet]. 2019 Dec 1;2(4):562–9. https://doi.org/10.1093/jamiaopen/ooz050 https://academic.oup.com/jamiaopen/article/2/4/562/5585396

23. Zanussi Z. Privacy Preserving Technologies Part Two: Introduction to Homomorphic Encryption [Internet]. Statistics Canada. 2022 [cited 2022 Jul 10]. https://www.statcan.gc.ca/en/data-science/network/homomorphic-encryption

24. Setoguchi S, Zhu Y, Jalbert JJ, Williams LA, Chen CY. Validity of deterministic record linkage using multiple indirect personal identifiers linking a large registry to claims data. Circ Cardiovasc Qual Outcomes. 2014;7(3):475–80. https://doi.org/10.1161/CIRCOUTCOMES.113.000294

25. Dusetzina S, Tyree S, Meyer A, Meyer A, Green L, Carpenter W. An overview of record linkage methods. In: Linking Data for Health Services Research: A Framework and Instructional Guide. Rockville: Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services; 2014.

26. Swartz RH, Cayley ML, Lanctôt KL, Murray BJ, Cohen A, Thorpe KE, et al. The "DOC" screen: Feasible and valid screening for depression, Obstructive Sleep Apnea (OSA) and cognitive impairment in stroke prevention clinics. Romigi A, editor. PLoS One. 2017 Apr;12(4):e0174451. https://doi.org/10.1371/journal.pone.0174451

27. Eskes GA, Lanctôt KL, Herrmann N, Lindsay P, Bayley M, Bouvier L, et al. Canadian Stroke Best Practice Recommendations: Mood, Cognition and Fatigue following Stroke Practice Guidelines, Update 2015. Int J Stroke. 2015 Oct;10(7):1130–40. https://doi.org/10.1111/ijs.12557

28. Ministry of Health and Long-Term Care. Technical Specifications Interface to Health Care Systems [Internet]. Ministry of Health and Long-Term Care. 2017. https://www.health.gov.on.ca/en/pro/publications/ohip/docs/techspec_interface_hcsm.pdf

29. Vaccarino AL, Dharsee M, Strother S, Aldridge D, Arnott SR, Behan B, et al. Brain-CODE: A secure neuroinformatics platform for management, federation, sharing and analysis of multi-dimensional neuroscience data. Front Neuroinform. 2018 May;12. https://doi.org/10.3389/fninf.2018.00028

30. Lefaivre S, Behan B, Vaccarino A, Evans K, Dharsee M, Gee T, et al. Big data needs big governance: Best practices from Brain-CODE, the Ontario Brain Institute's neuroinformatics platform. Front Genet. 2019 Mar;10. https://doi.org/10.3389/fgene.2019.00191

31. Goldberg SI, Niemierko A, Turchin A. Analysis of data errors in clinical research databases. AMIA Annu Symp Proc. 2008;242–6.

32. Randall S, Brown AP, Ferrante AM, Boyd JH. Privacy preserving linkage using multiple dynamic match keys. Int J Popul Data Sci [Internet]. 2019 May 23;4(1). https://doi.org/10.23889/ijpds.v4i1.1094 https://ijpds.org/article/view/1094

33. Harron K, Dibben C, Boyd J, Hjern A, Azimaee M, Barreto ML, et al. Challenges in administrative data linkage for research. Big Data Soc [Internet]. 2017 Dec 5;4(2):205395171774567. https://doi.org/10.1177/2053951717745678 http://journals.sagepub.com/doi/10.1177/2053951717745678

34. Christen P, Ranbaduge T, Schnell R. Attacking and Hardening Bloom Filter Encoding. In: Linking Sensitive Data [Internet]. Cham: Springer International Publishing; 2020. p. 221–51. https://doi.org/10.1007/978-3-030-59706-1_9 http://link.springer.com/10.1007/978-3-030-59706-1_9

35. Vidanage A, Ranbaduge T, Christen P, Randall S. Privacy Attack on Multiple Dynamic Match-key based Privacy-Preserving Record Linkage. Int J Popul Data Sci [Internet]. 2020 Aug 11;5(1). https://doi.org/10.23889/ijpds.v5i1.1345 https://ijpds.org/article/view/1345

36. Brown AP, Borgs C, Randall SM, Schnell R. Evaluating privacy-preserving record linkage using cryptographic long-term keys and multibit trees on large medical datasets. BMC Med Inform Decis Mak [Internet]. 2017 Dec 8;17(1):83. https://doi.org/10.1186/s12911-017-0478-5 http://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-017-0478-5

37. Hamza R, Hassan A, Ali A, Bashir MB, Alqhtani SM, Tawfeeg TM, et al. Towards Secure Big Data Analysis via Fully Homomorphic Encryption Algorithms. Entropy [Internet]. 2022 Apr 6;24(4):519. https://doi.org/10.3390/e24040519 https://www.mdpi.com/1099-4300/24/4/519

38. Munjal K, Bhatia R. A systematic review of homomorphic encryption and its contributions in healthcare industry. Complex Intell Syst [Internet]. 2022 May 3; https://doi.org/10.1007/s40747-022-00756-z https://link.springer.com/10.1007/s40747-022-00756-z

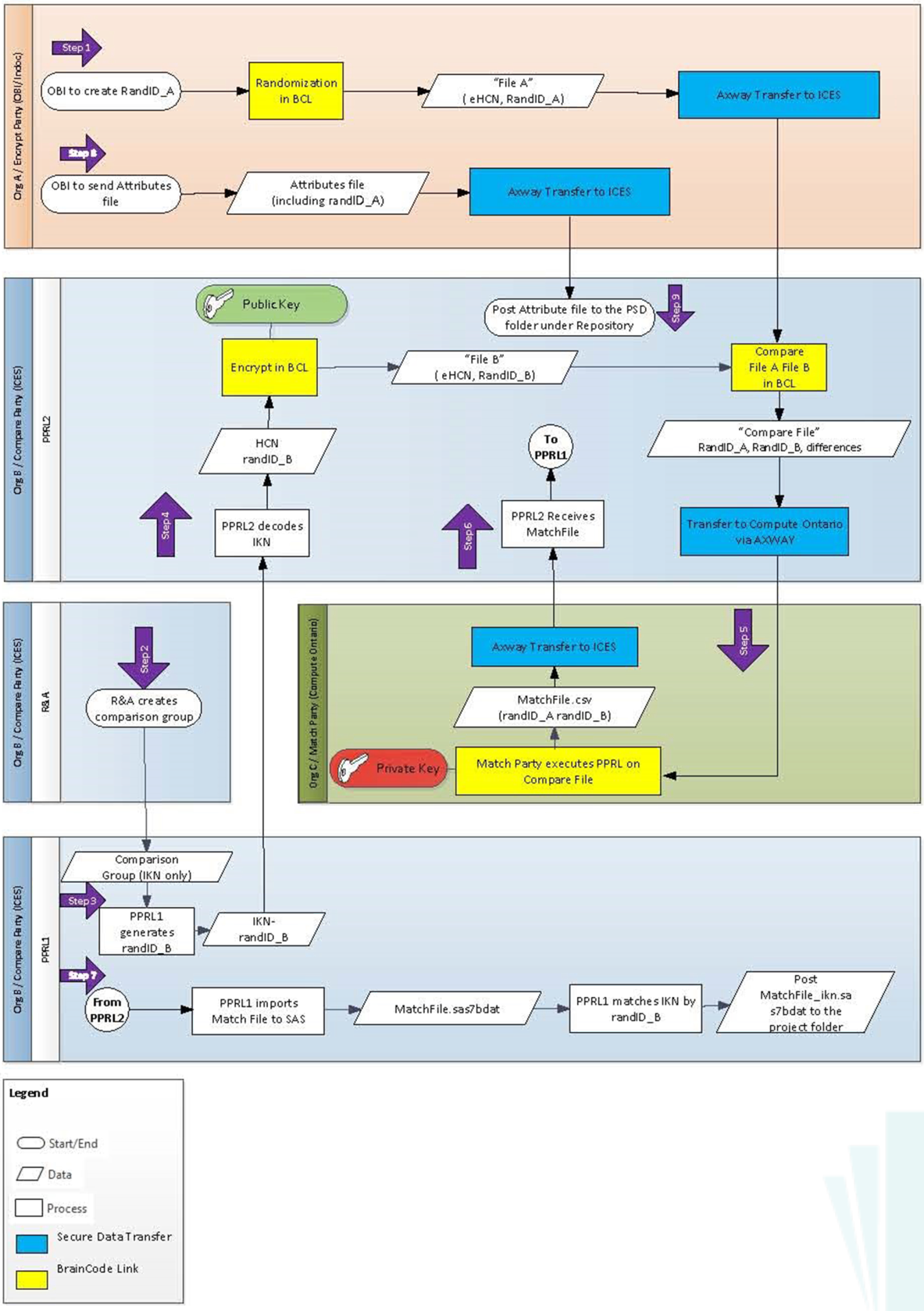
39. Alloghani M, M. Alani M, Al-Jumeily D, Baker T, Mustafina J, Hussain A, et al. A systematic review on the status and progress of homomorphic encryption technologies. J Inf Secur Appl [Internet]. 2019 Oct;48:102362. https://doi.org/10.1016/j.jisa.2019.102362 https://linkinghub.elsevier.com/retrieve/pii/S2214212618306057

40. Mc Cord KA, Al-Shahi Salman R, Treweek S, Gardner H, Strech D, Whiteley W, et al. Routinely collected data for randomized trials: Promises, barriers, and implications. Trials. 2018;19(1):1–9. https://doi.org/10.1186/s13063-017-2394-5

41. Mentz RJ, Hernandez AF, Berdan LG, Rorick T, O'Brien EC, Ibarra JC, et al. Good Clinical Practice guidance and pragmatic clinical trials. Circulation. 2016 Mar;133(9):872–80. https://doi.org/10.1161/CIRCULATIONAHA.115.019902

Supplementary Appendix 1

Supplementary Figure 1: Data Flow and Process for Brain-CODE Link (BCL) at ICES



The detailed algorithm (Supplementary Figure 1):

1. OBI: Ontario Health Insurance Plan (OHIP) numbers were entered into the Subject Registry at Brain-CODE over several years. These values were encrypted using a public key in the browser via JavaScript, and only the encrypted text left the user's computer to be stored in Brain-CODE. The encryption technique is such that the encrypted text is highly-variable, even if the same OHIP number is entered repeatedly. For a specific linkage study, a set of test subjects within Brain-CODE was identified for linking. Their encrypted OHIP numbers were extracted and paired with randomized subject IDs (RandID_A) via the 'link-randomize' utility of the BCL package, while the original participant IDs were retained on the central system. This extracted data was transmitted to ICES via the Axway secure transfer system.

2. ICES - Research and Analysis department (R&A): A cohort of potential individuals was identified and their encoded OHIP numbers (called ICES Key Numbers or IKNs) extracted from ICES data repository. This list was passed to the group 'PPRL1', which represents the group of individuals within the ICES Data Quality (DQ) team that are only permitted to work with encoded information.

3. ICES – Data Quality (PPRL1): PPRL1 assigns a random ID (RandID_B) to each IKN and passes the two variables to the group 'PPRL2'. The PPRL2 is the group of individuals within Information Management (IM) team that are authorized to view and process direct personal identifiers (DPI), such as the OHIP number. To respect the privacy-preserving aspect of the process the creation of random ID must not be shared with PPRL2.

4. ICES – IM (PPRL2): PPRL2 group decodes IKNs to raw, original OHIP numbers. The raw OHIP numbers are then encrypted in BCL using the public key (provided by the 'Match Party'). The resulting file (File B) is ready for the Compare Process. The PPRL2 group will import File A and File B into the BCL application to create the Compare file. The Compare file which will be a Cartesian match product of file A and file B will contain RandID_A, RandID_B, and the differences (generated by the software). When the size of the compare file is greater than the size limit of an Axway transfer, an adequate number of splits must occur and steps 4 and 5 will need to be repeated iteratively.

5. Trusted third-party (CS): The difference file was transferred securely via SFTP to an autonomous server run by a trusted third party. Once the file transfer was complete, the file integrity was verified by the server, which proceeded to run the Link utility 'link-match' against the encrypted differences, decrypting these values using the Brain-CODE secret key. When decrypted, the values were either 0, indicating a match, or a random value, indicating a non-match. An output file consisting of the lists of matched ID pairs is created by 'link-match,' which also creates a flag file when the run has finished.

6. ICES – IM (PPRL2): The PPRL2 group receives the encrypted MatchFile via Axway and passes it to the PRL1 group.

7. ICES – DQ (PPRL1): The PPRL1 group retrieves the study subject IKN by merging with the RandID_B and posts the file directly to the project folder.

8. OBI: OBI sends the clinical variables (attributes file) for the matched RandID_A (received from CS) to ICES via the Axway secure transfer system.

9. ICES – IM (PPRL2): The PPRL2 group receives the attributes file(s) (clinical variables) and makes them available to the project team through the ICES Data Repository. This attributes data will be linkable to ICES repository by the research team through the crosswalk posted under project folder in step 6.

Supplementary Appendix 2 – DOC Screen

DOC SCREEN PART 1

Date: _____

Patient of: _____

Completed by:
- ☐ RN
- ☐ MD
- ☐ RA
- ☐ Medical Student
- ☐ Other_____

Language:
- ☐ English 1st language
- ☐ ESL, fluent
- ☐ ESL, not fluent

Sex:
- ☐ M
- ☐ F

Unable to complete due to:
- ☐ Language
  - ☐ Aphasic/Dysphasic
  - ☐ Unable to translate
- ☐ Motor
- ☐ Vision
- ☐ Too ill
- ☐ Other: _____

Resides at:
- ☐ Home
- ☐ Nursing Home/ LTC/CCC
- ☐ Inpatient rehab facility
- ☐ Retirement Home
- ☐ Other Residential Facility
- ☐ No Fixed Address
- ☐ UTD

Education: highest grade (1-13) _____ # of undergraduate years _____ # of graduate years _____

Systolic BP: _____

Diastolic BP: _____

Height: _____

Weight: _____

Waist/Hip Circumference: _____ /_____

Modified Rankin Scale: _____

0 – No symptoms
1 – No significant disability despite symptoms; able to carry out all usual duties and activities
2 – Slight disability; unable to carry out all previous activities, but able to look after own affairs without assistance
3 – Moderate disability; requiring some help, but able to walk without assistance (**with or without cane or walker)**
4 – Moderately severe disability; unable to walk without assistance and unable to attend to own bodily needs without assistance
5 – Severe disability; bedridden, incontinent and requiring constant nursing care and attention

Clinical Frailty Scale: _____

1 – Very fit - Robust, active, energetic, well-motivated and fit
2 – Well - Without active disease, but less fit than those in 1
3 – Well, with treated comorbid disease - Disease symptoms well controlled
4 – Apparently vulnerable - Not dependent, but common complaints about being 'slowed up' or having disease symptoms
5 – Mildly frail - Limited dependence for IADLs
6 – Moderately frail - Help required for both IADLs/ADLs
7 – Severely frail - Completely dependent on others for ADLs, or terminally ill

Please see reverse side!
Version Date Mar 13 2014

Please see reverse side!
Version Date Mar 13 2014

DOC SCREEN PART 2

Time to complete: _____ min _____ s

Memory (Registration)						
Read list of words, subject must repeat them. Do 2 trials. Do a recall after 5 minutes.		FACE	VELVET	CHURCH	DAISY	RED
	1st trial					
	2nd trial					

DOC Mood[1] "Over the last 2 weeks, how often have you been bothered by any of the following problems?"				
	Not at all (0)	Several days (1)	More than half the days (2)	Nearly every day (3)
Little interest or pleasure doing things				
Feeling down, depressed or hopeless				

Score (D): /6

DOC Apnea[2]	**Yes**	**No**
Do you snore loudly (louder than talking, heard through a door, or bother other people)?		
Do you often feel tired, fatigued or sleepy during the daytime?		
Has anyone observed you stop breathing during your sleep?		
Do you have, or are you being treated for high blood pressure?		

Score (O): /4

Draw a Clock (Ten Past Eleven)

Contour [ ]
Numbers [ ]
Hands [ ]

Score (C): /3

Abstraction "What is the similarity between: (e.g. banana – orange = fruit)"
A train and a bicycle? [ ] A watch and a ruler? [ ]

Score (C): /2

Memory (Delayed Recall)						
What were those 5 words?	WITHOUT CUE	FACE [ ]	VELVET [ ]	CHURCH [ ]	DAISY [ ]	RED [ ]
	Category cue					
	Multiple choice cue					

Score (C): /5

D (DOC-Mood) : /6	**O (DOC-Apnea) : /4**	**C (DOC-Cog[3]) : /10**
TOTAL DOC SCORE = D + O + (10 – C) = /20		

Modified/Combined from: [1]**PHQ-2**: Hajek VE et al. Brief assessment of cognitive impairment in patients with stroke. Arch Phys Med Rehabil. 1989 Feb;70(2):114-7. [2]**STOP**: Chung F et al. STOP questionnaire: A tool to screen patients for obstructive sleep apnea. Anesthesiology. 2008 May;108(5):812-21. 74. [3]**MoCA**: Nasreddine ZS et al. The montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. J Am Geriatr Soc. 2005 Apr;53(4):695-9. MoCA copyright of the adapted version: Z. Nasreddine MD. Adapted by Swartz et al., 2013.

Supplementary Appendix 3 – DOC Case Report Form

DOC 2.0 CASE REPORT FORM Ver. Aug 13 2014

Initial Visit Date: _____

❑ 1st visit to Clinic
❑ Assessed at F/U for OBI Eligibility

Age at Visit: _____

Date of Referral: _____

Source of Referral (check only one):
❑ Emergency Department **Institution:** _____
❑ Post-admission Inpatient Follow-up
❑ Community/ Family Physician (GP)
❑ Other Subspecialties **Specify:**
□ Ophthalmology □ Orthopedics □ Cardiology □ Obstetrics
□ Neurosurgery □ Vascular Surgery □ Neurology □ Surgeon
□ Nephrology □ GIM □ Other: _____

Reason for Referral:
Date of event: _____
❑ New/Recent (≤180 days) ❑ Remote > 180 days
Type of event (check only one):
❑ Ischemic Stroke ❑ Hemorrhagic ❑ TIA/Minor Stroke
❑ Sinus Venus Thrombosis ❑ Asymptomatic Carotid Artery Disease
❑ Asymptomatic Abnormal Imaging ❑ Other: _____ ❑ UTD
Symptoms (check all that apply):
□ Left □ Right □ Bilateral □ UTD □ Not Applicable
□ Weakness □ Speech □ Sensory □ Monocular Visual Symptoms
□ Diplopia □ Vertigo □ Headaches/Migraines
□ Other: _____

Vascular Risk Factors (check all that apply):
❑ Previous Stroke **Specify:**
□ Ischemic □ ICH □ SAH □ IVH □ UTD
Most recent occurred: □ ≤ 3months □ > 3months □ UTD
❑ Previous TIA
Most recent occurred: □ ≤ 3months □ > 3months □ UTD
❑ Hypertension (HTN)
❑ Diabetes (DM) **Specify:** □ Type I □ Type II □ UTD
Specify End-organ damage: □ Yes □ No □ UTD
❑ Dyslipidemia/Hyperlipidemia
❑ Family History of Vascular Disease
❑ Carotid Stenosis (Extra Cranial)
□ **Right Side Specify:** □ Symptomatic □ Asymptomatic
□ Trace (<30%) □ Mild (30-49%) □ Moderate (50-69%)
□ Severe (70-99%) □ Occluded (100%)
□ **Left Side Specify:** □ Symptomatic □ Asymptomatic
□ Trace (<30%) □ Mild (30-49%) □ Moderate (50-69%)
□ Severe (70-99%) □ Occluded (100%)
□ **UTD**
❑ Symptomatic Intracranial Arterial Disease (e.g. Moya-moya, Atherosclerosis, Vasculitis, RCVS, Fibromuscular Dysplasia)
❑ Peripheral Vascular Disease (PVD)
❑ Carotid Endarterectomy (CEA)/ Stenting **Specify:**
□ Right □ Left □ Both □ UTD
❑ Deep Venous Thrombosis (DVT) or Pulmonary Embolism(PE)

Smoker Status (check only one):
❑ Life-long non-smoker ❑ Current non-smoker (hist. unknown)
❑ Non-smoker (including quit >5yrs) ❑ Reformed smoker (quit > 1 mo to 5 years) ❑ Current (including quit <30 days) ❑ UTD

Cardiac Disease (check all that apply):
❑ Angina ❑ Prior MI ❑ Prior CABG (triple bypass)
❑ Prior PTCA (angioplasty) ❑ Congestive Heart Failure (CHF)
❑ Valve Replacement ❑ Atrial Fibrillation (A-fib)/Flutter
❑ Significant Murmur (grade ≥3/6 for systolic, any diastolic)
❑ Left Ventricular Hypertrophy (LVH)
❑ Other Cardiac Source of Emboli **Specify:**
□ Poor LV function (low EF) □ Abnormal Aortic or Mitral Valve (Valvular Heart Disease) □ Cardiomyopathy □ PFO or other Shunt □ Intracardiac Thrombosis
❑ CAD type not specified

Others (check all that apply):
❑ Obesity
❑ Prothrombotic State
❑ Known Pregnancy **Specify:**
□ N/A □ 1st trimester □ 2nd trimester □ 3rd trimester
❑ Post-partum **Specify:**
□ N/A □ <6 weeks □ 6 – 12 weeks
❑ Known Obstructive Sleep Apnea (OSA) **Specify:**
□ Using CPAP □ Not using CPAP □ Using other therapy: _____
❑ History of Migraines

Alcohol Consumption (check only one):
❑ Never ❑ Not currently (hist. unknown) ❑ Rare/Social
❑ < 2 drinks/day ❑ > 2 drinks/day ❑ UTD

Systemic Disease (check all that apply):
❑ Dementia
❑ Chronic Renal disease **Specify any of:**
□ Mild (Cr < 256) □ Moderate - Severe (Cr ≥ 256) □ Dialysis
□ Transplantation
❑ Cancer **Specify:**
□ Breast □ Lung □ Leukemia □ Lymphoma □ Metastatic □ UTD
□ Other: _____
❑ COPD/Asthma
❑ Connective Tissue Disease
❑ Peptic Ulcer Disease (PUD)
❑ Liver Disease **Specify:**
□ Mild (Cirrhosis without PHT, Chronic Hepatitis) □ Moderate □ Severe (Cirrhosis with PHT +/- Variceal Bleeding) □ UTD
❑ Hemiplegia or Paraplegia
❑ AIDS
❑ Autoimmune Disease **Specify:**
□ CNS (MS, Vasculitis etc.) □ Rheumatological (Lupus/ SLE, RA etc.)
□ Other: _____

Most Responsible Diagnosis after Initial Visit:
❑ Ischemic Stroke **Specify:**
□ Definite Stroke □ Possible/Query Stroke
❑ Hemorrhage
□ ICH □ SAH □ IVH □ SDH
❑ TIA **Specify:**
□ Definite TIA □ Possible/Query TIA
❑ Sinovenous Thrombosis
❑ Asymptomatic Carotid Artery Disease
❑ Abnormal CT/MRI scan
❑ Other Vascular: _____
❑ Other Non-Vascular: Specify
□ BPPV/Vertigo □ Migraine/Aura □ Seizure/Aura
□ Other: _____