**DRAFT – This is not a final version of the DEPiCTS data requirements or data management strategy document.**

**DEPiCTS – Data Requirements, Data Management**

Version 1

June 14, 2017

# Acronyms

SGS = Stroke Guidance System

EDW = Island Health Enterprise Data Warehouse

ACVS = Acute Cerebrovascular Syndrome

TIA = Transient Ischemic Attack

ED = Emergency Department

CTAS = Canadian Triage …………….

NACRS = National Ambulatory Care Reporting System

# Research Cohorts

There are two relevant groups of subjects:

1. SRAU Cohort [or whatever you name it] - Patients who have been seen in the Stroke Rapid Assessment Unit (SRAU) – these are the patients for whom detailed data are collected and analyzed – see Section 3 below entitled “Data Requirements – Patients Seen in SRAU”.
2. All patients with records in the EDW – in order to generate estimates of base rates for ACVS and related conditions – see Section 4 below entitled “Data Requirements - Data collected on all patients in EDW”. These data are only used to generate aggregate level statistics re: incidence/prevalence of ACVS and related conditions.

There are an estimated XXXXX patients who will be included in the SRAU Cohort.

## Note re: dates; look-back and look-forward period.

In the DEPiCTS research data files, in order to meet the requirements of Safe Harbor de-identification around dates, date will be expressed in terms of days before/after a date that has been set as Day 0. The term “coarsened” is used in the tables below that have been subject to this protection. This method is more conservative than the Safe Harbor requirements, while preserving fully the sequence of events and the duration that separates events.

For the purpose of this research, Day 0 is set at Day of Registration in the SRAU. All dates in the research data file to be generated will be expressed in terms of duration from Day 0.

For data from the EDW (e.g,. Encounters) attached to the SRAU Cohort, the lookback period (from day of registration in the SRAU is set at [I suggest you go back as far as we have data so the lookback period would be set by the earliest day in which data were registered in the EDW).

For data from the EDW attached to the SRAU Cohort, the lookforward period will be set as XXXXX from Date of Registration in the SRAU (Day 0) [[ or will it be date of Discharge???]]. This value will be expressed in the file used by researchers in terms of elapsed days.

## ICD10 Diagnostic Criteria for deeming a case to be an ACVS or ACVS-relevant case

Haemmorhage

* I60 Suarachnoid haemorrhage
* I61 Intracerebral haemorrhage
* I62 Subdural/extradural haemorrhage

Ischemic Stroke

* I63 Cerebral Infarction due to embolism/thrombosis
* I64 Stroke, not otherwise specified

Transient Ischemic Attack (TIA)

* G45.0 Vertebro-basilar artery syndrome
* G45.1 Carotid artery syndrome (hemispheric)
* G45.2 Multiple and bilateral precerebral artery syndromes
* G45.3 Amaurosis fugax
* G45.8 Other transient cerebral ischaemic attacks and related syndromes
* G45.9 Transient cerebral ischaemic attack, unspecified
* H34.0 Transient retinal artery occlusion

# Data Requirements – Patients Seen in SRAU

|  |  |  |
| --- | --- | --- |
| **Data Category or Element** | **Use – Legacy** | **Use - New** |
| **From Stroke Guidance System (SGS)** | | |
| Demographics, Referral Background, Various Dates Associated With Demographics & Referral Background | | |
| MRN [Linking Variable - masked in research data file] | X | X |
| Regional ID (PHN) [Linking Variable] | X |  |
| Gender | X | X |
| Date of Birth | X | Coarsened |
| Age Group (in blocks of how many years????) | X | X |
| Case Management within SRAU | | |
| Referral Form – Clinical Information | | |
| Time of ACVS-relevant clinical event (e.g., TIA, stroke) | X | X |
| Referral form signs, symptoms (e.g., amnesia; anxiety; confusion; diplopia; headache, etc.) (Yes/No) | X | X |
| Etc |  |  |
| Etc |  |  |

# Data Management

1. Person-level data for a cohort whose membership is defined in terms of registrations to the SRAU – approximately 12,000 persons.
2. Data from any and all data sources will be indexed by PHN in order to facilitate data linkage.
3. Data residing in sources outside of the EDW but within the perimeter of the Island Health firewall will be copied directly into the appropriately designated location in the EDW by a designated EDW staff person. Any data residing outside the Island Health firewall will be transmitted via Kiteworks secure email to a designated EDW staff person.
4. All data linkage will be performed by the EDW staff person inside the EDW.
5. All dates will be expressed in terms of duration (days) from Day 0, which is set by the patient registration date into the SRAU.
6. All data elements will be reviewed against the Safe Harbor list of identifiers; all Safe Harbor identifiers will be masked or deleted. Keys to link the de-identified data to PNHs will be retained by the EDW staff and will not be disclosed to the researchers.
7. Researchers will only have access to de-identified data.

RE: US Regulations – see HIPAA Privacy Rule – *US Code of Federal Regulations (CFR): 45 CFR, Part 160 and Subparts A and E of Part 165.*

## Safe Harbor Identifiers

The following material is copied from Malin (2012) ***Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule (***[***https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#safeharborguidance***](https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#safeharborguidance)***) .* For a .pdf version, see** [**https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/De-identification/hhs\_deid\_guidance.pdf**](https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/De-identification/hhs_deid_guidance.pdf)

Guidance on Satisfying the Safe Harbor Method

In §164.514(b), the Safe Harbor method for de-identification is defined as follows:

(2)(i) The following identifiers of the individual or of relatives, employers, or household members of the individual, are removed:

(A) Names

(B) All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code, and their equivalent geocodes, except for the initial three digits of the ZIP code if, according to the current publicly available data from the Bureau of the Census:  
(1) The geographic unit formed by combining all ZIP codes with the same three initial digits contains more than 20,000 people; and  
(2) The initial three digits of a ZIP code for all such geographic units containing 20,000 or fewer people is changed to 000

(C) All elements of dates (except year) for dates that are directly related to an individual, including birth date, admission date, discharge date, death date, and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older

(D) Telephone numbers

(L) Vehicle identifiers and serial numbers, including license plate numbers

(E) Fax numbers

(M) Device identifiers and serial numbers

(F) Email addresses

(N) Web Universal Resource Locators (URLs)

(G) Social security numbers

(O) Internet Protocol (IP) addresses

(H) Medical record numbers

(P) Biometric identifiers, including finger and voice prints

(I) Health plan beneficiary numbers

(Q) Full-face photographs and any comparable images

(J) Account numbers

(R) Any other unique identifying number, characteristic, or code, except as permitted by paragraph (c) of this section; and

(K) Certificate/license numbers

(ii) The covered entity does not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information