

2.2.3 Laatuvalidointi tuotannossa ja

2.2.5 Käytettävyystestaus

AUTOVALIDATE-TUTKIMUS

Raportti validoinnin tuloksista
Potilastietojen retrospektiivinen haku,
manuaalinen vs. Insite-alustan avulla

Table of contents

Sisällysluettelo

1	AUTOVALIDATE-TUTKIMUKSEN TARKOITUS	3
2	KOHORTIN MUODOSTUS KRITEERISTÖ.....	4
3	INSITE TULOKSET SUHTEESSA MUIHIN KOHORTIN MUODOSTUKSEN TAPOIHIN- RESULTS.....	5
4	TESTAUS ASETELMA JA TESTAUKSEN TULOKSET	24
5	KÄYTTÖKELPOISUUS KLIINIKON NÄKÖKULMASTA JA SYY INSITE IMPLEMENTOINTIIN.....	29
6	MANUAALINEN R-KIELELLÄ TOTEUTETTU POIMINTASKRIPTI.	33

1 AUTOVALIDATE-TUTKIMUKSEN TARKOITUS

Autovalidate-tutkimus perustuu Fibstroke-tutkimuksessa kerättyyn kardiologiseen potilasaineistoon vuosilta 2003-2012, sekä lisäanalyysiin, jossa TYKSin tietokannoista tehdyistä poimminnoista selvitetään tietyin kriteerein potilastapahtumat vuosilta 2013-2017.

Tutkimuksen tavoitteet

Tutkimuksen tavoitteena on selvittää seuraavat kysymykset:

1. Miten hyvin InSite-ohjelmistolla löydetään TYKS:in tietokannasta potilaat, jotka oli manuaalisesti haettu Fibstroke-tutkimukseen vuosilta 2003-2012
2. Miten hyvin InSite-ohjelmistolla löydetty FibStroke-tutkimuksen potilaiden tiedot vastaavat potilaiden terveyttä, sairautta ja interventioita koskevia tietoja FibStroke aineistossa
3. Ovatko aineistojen välillä mahdollisesti todettavat eroavaisuudet lähdeaineiston puutteita vai InSite-ohjelman toiminnallisuuteen liittyviä ilmiöitä
4. Mitä haasteita InSite ohjelmiston käyttämiseen klinikon kannalta liittyy käytettävyyden tai palvelun käyttäytymisen kannalta

Tutkimusaineisto

FibStroke tutkimuksessa on tunnistettu retrospektiivisesti sairaskertomuksista potilaat, joilla on sekä eteisvärinädiagnoosi että aivoverenkiertohäiriödiagnoosi (stroke/TIA) milloin tahansa 10-vuotistutkimusajanjakson (vuodet 2003-2012) aikana. Tutkimukseen otettiin mukaan vain potilaat, jotka ovat asuneet sairaalan alueella aivohalvauksen ilmetessä ja vähintään vuoden ajan sen jälkeen, jotta kattavat tiedot komplikaatioista voidaan kerätä. Myös lääkitystiedot ja riskitekijöitä CHA2DS2-VASc ja HAS-BLED arvioitiin.

2 KOHORTIN MUODOSTUS KRITEERISTÖ

1. Aikuisia (>18 vuotiaita) 1.1.2003 diagnoosihetkeltä alkaen
 2. Potilaalla Diagnosoitu eteisvärinä → Atrial Fibrillation (AF) or atrial flutter I48
 3. Yksi tai useampi seuraavista aivoverenkierron häiriöistä → Disturbances in cerebral blood flow **during 01.01.2003-31.12.2012:**
 - a) Aivoinfarkti → Ischemic stroke, I60.0-I60.9, I61.0 - I61.9, I62.0 - I62.9, I63.0 - I63.9, I65.0-I65.9, I66.0-I66.9, I69.0 - I69.9 **or**
 - b) Ohimenevä aivoverenkiertohäiriö → TIA, G45.0 - G45.9, G46.0 - G46.9 **or**
 - c) kallonsis. Verenvuoto → Intracranial bleeding, S06.0- S06.9
- ➔ Kuinka monella potilaalla oli **eteisvärinä AF diagnoosi ENNEN ensimmäistä diagnoosia a, b tai c.** ("ENNEN" määritelty, ainakin 1 päivä ennen diagnoosia a, b tai c).

3 INSITE TULOKSET SUHTEESSA MUIHIN KOHORTIN MUODOSTUKSEN TAPOIHIN- RESULTS

	Insite search	Fibstroke study original manual search	BC REQUEST
Number of patients who suffered from stroke, TIA or Intracranial bleeding during 2003-2013 and who had ever been diagnosed of AF (see Fig 1)	6230	2920	10 934
Number of events (stroke, TIA or/and intracranial bleeding) in total during 2003-2012 (as some had more than one event).	9114	3762	
AF diagnosed before the event. Please see explanation in the end of this document why the last values (4493 and 2177) are not comparable with each other.	4493	2177	

Manual search -R script (see chapter 6)

	Manual R-script (see chapter 6)	comments
Number of patients who suffered from stroke, TIA or Intracranial bleeding during 2003-2013 and who had ever been diagnosed of AF (see Fig 1)	3950	Observe that this result is closer to the result from the Fibstroke study done by manual search
Patients who are over 18 years and suffered i48 during 01.01.2003-31.12.2012	23 919	Observe that this cohort count is 23 865 throughout Insite

Fig 1. Number of adult patients who suffered from stroke, TIA or/and Intracranial bleeding during 2003-2013 and who had ever been diagnosed of AF. Number of patients 6230 in Insite.

localhost:8080/app/cohort/studies/study/builder/import?cohortStudyId=8&_csrf=5d16bacd-dff6-4bea-a00a-2bde5eab317a

Suggested Sites Imported From IE laadullinen tutkimus Sote-asiakastietojen Moodle Login Tohtorikoulutus Suomi.fi-tunnistus Strategies to improve Strategies to improve

InSite Studies Cohort Analysis Recruiting Studies NiinaN Notification

Cohort Builder

Save Cohort & Explore Patients Save Cohort

Recalculate results Show matches

Choose cohort: 1 016 137 IN COHORT

Age 18 or more on 1.1.2003 (1 016 137 patients)

Reference date: 31 DEC 2012

FILTERS Filters Add filter

1. first diagnosis of Atrial fibrillation and flutter 23 865 IN COHORT
2. the following events:
 - first diagnosis of Nontraumatic subarachnoid hemorrhage
 - or first diagnosis of Nontraumatic intracerebral hemorrhage
 - or first diagnosis of Other and unspecified nontraumatic intracranial hemorrhage
 - or first diagnosis of Cerebral infarction
 - or first diagnosis of Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction
 - or first diagnosis of Occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction
 - or first diagnosis of Sequelae of cerebrovascular disease
 - or first diagnosis of Vascular syndromes of brain in cerebrovascular diseases
 - or first diagnosis of Intracranial injury
 - or first diagnosis of Transient cerebral ischemic attacks and related syndromes

happened at most 10 Years before now

6 230 IN COHORT

Fig 2. Number of events occurred during 2003-2012. To get number of each diagnose, each code was run separately one by one, ie by changing the order of the filters. The example in this picture below is for i63.

localhost:8080/app/cohort/studies/study/builder/import?cohortStudyId=8&_csrf=5d16bacd-dff6-4bea-a00a-2bde5eab317a

sted Sites Imported From IE laadullinen tutkimus Sote-asiakastietojen Moodle Login Tohtorikoulutus Suomi.fi-tunnistus Strategies to improv Strategies to improv

InSite Studies Cohort Analysis Recruiting Studies NiinaN

< Cohort Studies autovalidate Cohort builder

Choose cohort: 1 016 137 IN COHORT

Age 18 or more on 1.1.2003 (1 016 137 patients)

Reference date: 31 DEC 2012

FILTERS Filters + Add filter

- the following events: 23 865 IN COHORT
 first diagnosis of Atrial fibrillation and flutter
 happened at most 100 Years before now
- ABSTRACT i63 3 879 IN COHORT
 the following events:
 first diagnosis of Cerebral infarction due to thrombosis of precerebral arteries
 or first diagnosis of Cerebral infarction due to embolism of precerebral arteries
 or first diagnosis of Cerebral infarction due to unspecified occlusion or stenosis of precerebral arteries
 or first diagnosis of Cerebral infarction due to thrombosis of cerebral arteries
 or first diagnosis of Cerebral infarction due to embolism of cerebral arteries
 or first diagnosis of Cerebral infarction due to unspecified occlusion or stenosis of cerebral arteries
 or first diagnosis of Cerebral infarction due to cerebral venous thrombosis, nonpyogenic
 or first diagnosis of Other cerebral infarction
 or first diagnosis of Cerebral infarction, unspecified
 happened at most 10 Years before now
- ABSTRACT g46 31 IN COHORT

The results of each dg runned separately the same way as in picture above:

Stroke	first dg of i69 (sequelae of cerebrovascular disease) happened at most 10 years ago	first dg of i66 (occl and stenosis of cerebral arteries, not resulting in cerebral infarction) happened at most 10 years ago	first dg of i65 (occl and stenosis of pre-cerebral arteries, not resulting in cerebral infarction) happened at most 10 years ago	first dg of i63 (cerebral infarction) happened at most 10 years ago	first dg of i62 (other and unspecified nontraumatic intracranial hemorrhage)	first dg of i61 (nontraumatic intracerebral hemorrhage) happened at most 10 years ago	first dg of i60 (non-traumatic sub-arachnoid hemorrhage) happened at most 10 y ago (sama jos "any")
	1941	19	168	3879	131	444	105
TIA	first dg of G45 (transient ischemic cerebral attacks and related syndromes) happened at most 10 years ago	first dg of G46 (vascular syndromes of brain in cerebrovascular diseases) happened at most 10 years ago					
	1432	80					
Intracranial bleeding	first dg of S06 (intracranial injury) happened at						

	most 10 years ago						
	915						

Table 2. Total amount of Stroke, TIA and intracranial bleeding 9114. Sum calculated manually together from the table)

Patient groups were also tested separately ie only stroke patients with AF, only TIA patients with AF and only Intracranial bleeding patients with AF. The results between Insite and manual search are listed in table below.

	Insite search	Fibstroke study original manual search	BC RQUEST
Number of patients who suffered from stroke during 2003-2013 and who had ever been diagnosed of AF	4924	1979	
Number of patients who suffered from TIA during 2003-2013 and who had ever been diagnosed of AF	1502	493	
Number of patients who suffered from Intracranial bleeding during 2003-2013 and who had ever been diagnosed of AF	915	341	

Table 3. Number of patients with AF and stroke during 2003-2012, number of patients with AF and TIA during 2003-2012 and number of patients with AF and intracranial bleeding during 2003-2012.

Testing whether first dg of AF happened before or after the first dg of stroke, TIA or intracranial bleeding:

- The same population than in the first run (Fig 1=AF + Stroke, TIA or Intracranial bleeding), but added with filter “time constraint: X before Y”.
- The aim to get events in which AF was diagnosed at least 1 day before the event (either stroke, TIA or Intracranial bleeding).
- Insite found 4493 patients, who’s first dg of AF was before their first dg of stroke, TIA or intracranial bleeding event. For reference, Fibstroke manual search found 2177 patients.
- However, this is not comparable with the Fibstroke study value, because Fibstroke recorded number of **events** (not patients) in which AF was at least 1 day before the event. Insite searched for **patients** whos first AF dg was at least 1 day before stroke, TIA or IC bleeding.

Fig 3. Testing patients who had their first dg of AF before their first dg of Stroke, TIA or Intracranial bleeding

localhost:8080/app/cohort/studies/study/builder/import?cohortStudyId=8&_csrf=d0b2ac1f-39b0-4aef-901e-cf2982861248

ted Sites Imported From IE laadullinen tutkimus Sote-asiakastietojen Moodle Login Tohtorikoulutus Suomi.fi-tunnistus Strategies to improve

Studies Cohort Analysis Recruiting Studies NiinaN

Cohort Builder

SETTINGS

Choose cohort: 1 016 137 IN COHORT

Age 18 or more on 1.1.2003 (1 016 137 patients)

Reference date: 31 DEC 2012

FILTERS

1. first diagnosis of Atrial fibrillation and flutter 23 865 IN COHORT

2. the following events:
first diagnosis of Nontraumatic subarachnoid hemorrhage
or first diagnosis of Nontraumatic intracerebral hemorrhage
or first diagnosis of Other and unspecified nontraumatic intracranial hemorrhage
or first diagnosis of Cerebral infarction
or first diagnosis of Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction
or first diagnosis of Occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction
or first diagnosis of Sequelae of cerebrovascular disease
or first diagnosis of Vascular syndromes of brain in cerebrovascular diseases
or first diagnosis of Intracranial injury
or first diagnosis of Transient cerebral ischemic attacks and related syndromes
happened at most 10 Years before now 6 230 IN COHORT

3. the following events:
first diagnosis of Nontraumatic subarachnoid hemorrhage
or first diagnosis of Nontraumatic intracerebral hemorrhage
or first diagnosis of Other and unspecified nontraumatic intracranial hemorrhage
or first diagnosis of Cerebral infarction
or first diagnosis of Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction
or first diagnosis of Occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction
or first diagnosis of Sequelae of cerebrovascular disease
or first diagnosis of Transient cerebral ischemic attacks and related syndromes
or first diagnosis of Vascular syndromes of brain in cerebrovascular diseases
or first diagnosis of Intracranial injury
happened at least 1 Days After these events:
first diagnosis of Atrial fibrillation and flutter 4 493 IN COHORT

Stroke, TIA and IC bleeding sub-populations were also tested in 3 subgroups:

Fig 3a. Testing patients who had their first dg of AF before their first dg of Stroke

localhost:8080/app/cohort/studies/study/builder/import?cohortStudyId=8&csrf=d0b2ac1f-39b0-4aef-901e-cf2982861248

Suggested Sites Imported From IE laadullinen tutkimus Sote-asiakastietojen Moodle Login Tohtorikoulutus Suomi.fi-tunnistus Strategies to improve

InSite Studies Cohort Analysis Recruiting Studies NiinaN

Cohort Studies autovalidate Cohort builder

Choose cohort: 1 016 137 IN COHORT Reference date: 31 DEC 2012

Age 18 or more on 1.1.2003 (1 016 137 patients)

FILTERS Filters Add filter

1. first diagnosis of Atrial fibrillation and flutter 23 865 IN COHORT
2. the following events:
 - first diagnosis of Nontraumatic subarachnoid hemorrhage
 - or first diagnosis of Nontraumatic intracerebral hemorrhage
 - or first diagnosis of Other and unspecified nontraumatic intracranial hemorrhage
 - or first diagnosis of Cerebral infarction
 - or first diagnosis of Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction
 - or first diagnosis of Occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction
 - or first diagnosis of Sequelae of cerebrovascular disease
 happened at most 10 Years before now 4 924 IN COHORT
3. **ABSTRACT** AF + Stroke 3 620 IN COHORT

the following events:

 - first diagnosis of Sequelae of cerebrovascular disease
 - or first diagnosis of Occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction
 - or first diagnosis of Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction
 - or first diagnosis of Cerebral infarction
 - or first diagnosis of Unspecified nontraumatic intracranial hemorrhage
 - or first diagnosis of Nontraumatic intracerebral hemorrhage
 - or first diagnosis of Nontraumatic subarachnoid hemorrhage
 happened at least 1 Days After these events:
 - first diagnosis of Atrial fibrillation and flutter

Fig 3b. Testing patients who had their first dg of AF before their first dg of TIA

localhost:8080/app/cohort/studies/study/builder/import?cohortStudyId=8&_csrf=d0b2ac1f-39b0-4aef-901e-cf2982861248

sted Sites Imported From IE laadullinen tutkimus Sote-asiakastietojen Moodle Login Tohtorikoulutus Suomi.fi-tunnistus Strategies to improv

InSite Studies Cohort Analysis Recruiting Studies NiinaN

Cohort Studies autovalidate Cohort builder

Cohort Builder

Save Cohort & Explore Patients Save Cohort

Recalculate results Show match

SETTINGS

Choose cohort:

Age 18 or more on 1.1.2003 (1 016 137 patients)

1 016 137 IN COHORT

Reference date:

31 DEC 2012

FILTERS Filters Add filter

1. first diagnosis of Atrial fibrillation and flutter 23 865 IN COHORT
2. the following events:
first diagnosis of Transient cerebral ischemic attacks and related syndromes
or first diagnosis of Vascular syndromes of brain in cerebrovascular diseases
happened at most 10 Years before now 1 502 IN COHORT
3. **ABSTRACT** TIA AFTER AF 923 IN COHORT
the following events:
first diagnosis of Transient cerebral ischemic attacks and related syndromes
or first diagnosis of Vascular syndromes of brain in cerebrovascular diseases
happened at least 1 Days After these events:
first diagnosis of Atrial fibrillation and flutter

Fig 3c. Testing patients who had their first dg of AF before their first dg of Intracranial bleeding

The screenshot shows the Cohort Builder interface. The URL bar displays: `localhost:8080/app/cohort/studies/study/builder/import?cohortStudyId=8&_csrf=d0b2ac1f-39b0-4aef-901e-cf2982861248`. The interface includes a top navigation bar with 'InSite Studies', 'Cohort Analysis', and 'Recruiting Studies'. The main section is titled 'Cohort Builder' and shows a 'Choose cohort:' dropdown set to 'Age 18 or more on 1.1.2003 (1 016 137 patients)' with '1 016 137 IN COHORT' displayed. A 'Reference date:' is set to '31 DEC 2012'. Under the 'FILTERS' section, three filters are listed:

1. the following events:
first diagnosis of Atrial fibrillation and flutter
happened at most 100 Years before now
23 865 IN COHORT
2. the following events:
first diagnosis of Intracranial injury
happened at most 10 Years before now
915 IN COHORT
3. **ABSTRACT** IC bleeding AFTER AF
excluding HF
the following events:
first diagnosis of Intracranial injury
happened at least 1 Days After these events:
first diagnosis of Atrial fibrillation and flutter
613 IN COHORT

A blue arrow points from the URL bar to the 'IC bleeding AFTER AF' filter in the third filter entry.

Patients with AF + Stroke	4924		
AF before stroke	AF before stroke	AF after stroke	AF after stroke
at least 30 days before	at least 1 day before	at least 1 day after	at least 30 days after
3056	3620 (Fig 3a)	1285	972

	<i>For reference, Fib-stroke value 1475</i>		
--	--	--	--

Patients with AF + TIA	1502		
<i>AF before TIA</i>	<i>AF before TIA</i>	<i>AF after TIA</i>	<i>AF after TIA</i>
<i>at least 30 days before</i>	<i>at least 1 day before</i>	<i>at least 1 day after</i>	<i>at least 30 days after</i>
788	923 (Fig 3b) For reference, Fib-stroke value 369	479	431

Patients with AF + IC bleeding	915		
<i>AF before IC bleeding</i>	<i>AF before IC bleeding</i>	<i>AF after IC bleeding</i>	<i>AF after IC bleeding</i>
<i>at least 30 days before</i>	<i>at least 1 day before</i>	<i>at least 1 day after</i>	<i>at least 30 days after</i>
574	613 (Fig 3c) For reference, Fibstroke value 323	287	248

Table 4. Stroke at least 1 day before AF, TIA at least 1 day before AF and Intracranial bleeding at least 1 day before AF. The same test was used to identify patients who had got their first dg of AF at least 1 day **after** the event. They all were also tested with run “first dg of AF at least 30 days before” and “30 days after” the event. These results are also presented in the table above.

CHA2DS2-VASc score

CHA2DS2-VASc ja HAS-BLED –scores help to choose suitable anticoagulation treatment by estimating thrombosis and bleeding risks. Risk factor filtering was done for patient population who's first dg of AF occurred at least 1 day before stroke. Same test was done to sub-populations of TIA and intracranial bleeding patients. To get number of each diagnose, each diagnosis code was run separately one by one, ie by changing the order of the filters. The example in Fig 4a. below is for diabetes (E10 and E11).

Fig 4a. CHA2DS2-VASc, AF + stroke

localhost:8080/app/cohort/studies/study/builder/import?cohortStudyId=8&_csrf=aa28b7c8-0e66-442d-8595-fff2d21e7a4c

Imported From IE laadullinen tutkimus Sote-asiakastietojen Moodle Login Tohtorikoulutus Suomi.fi-tunnistus Strategies to improve

Cohort Builder

Choose cohort: Age 18 or more on 11/2022 (1 016 137 patients) 1 016 137 IN COHORT Reference date: 31 DEC 2012

FILTERS

Filter	Count	Status
1. diagnosis of Atrial Fibrillation and Rutter	23 865	IN COHORT
2. the following events: diagnosis of Nontraumatic subarachnoid hemorrhage diagnosis of Nontraumatic intracerebral hemorrhage diagnosis of Other and unspecified nontraumatic intracerebral hemorrhage diagnosis of Cerebral infarction diagnosis of Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction diagnosis of Occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction diagnosis of Sequelae of cerebrovascular disease happened at most 10 years before now	4 924	IN COHORT
3. ABSTRACT of Stroke	3 620	IN COHORT
4. the following events: diagnosis of Type 1 diabetes mellitus diagnosis of Type 2 diabetes mellitus happened at most 1 day before these events diagnosis of Atrial Fibrillation and Rutter	533	IN COHORT
5. the following events: diagnosis of Heart failure happened at most 1 day before these events diagnosis of Atrial Fibrillation and Rutter	106	IN COHORT
6. the following events: diagnosis of Essential (primary) hypertension happened at most 1 day before these events diagnosis of Atrial Fibrillation and Rutter	60	IN COHORT
7. the following events: diagnosis of Atherosclerosis diagnosis of Aortic aneurysm and dissection happened at most 1 day before these events diagnosis of Atrial Fibrillation and Rutter	4	IN COHORT
8. any demographic record of Female	2	IN COHORT
9. How or why to 75 years old	2	IN COHORT
10. ABSTRACT of stroke or TIA	0	IN COHORT

Fig 4b. CHA2DS2 ja AF+ TIA (example in Fig 4b is for heart failure)

localhost:8080/app/cohort/studies/study/builder/import?cohortStudyId=8&_csrf=aa28b7c8-0e66-442d-8595-fff2d21e7a4c

sted Sites Imported From IE laadullinen tutkimus Sote-asiakastietojen Moodle Login Tohtorikoulutus Suomi.fi-tunnistus Strategies to improv

Site Studies Cohort Analysis Recruiting Studies NiinaN

Cohort Studies > autovalidate > Cohort builder

Choose cohort: 1 016 137 IN COHORT Reference date: 31 DEC 2012

Age 18 or more on 1.1.2003 (1 016 137 patients)

FILTERS

1. first diagnosis of Atrial fibrillation and flutter	23 865 IN COHORT
2. the following events: first diagnosis of Vascular syndromes of brain in cerebrovascular diseases or first diagnosis of Transient cerebral ischemic attacks and related syndromes happened at most 10 Years before now	1 502 IN COHORT
3. ABSTRACT AF + TIA	923 IN COHORT
4. the following events: first diagnosis of Heart failure happened at most 1 Days Before these events: first diagnosis of Atrial fibrillation and flutter	89 IN COHORT
5. the following events: first diagnosis of Atherosclerosis or first diagnosis of Aortic aneurysm and dissection happened at most 1 Days Before these events: first diagnosis of Atrial fibrillation and flutter	2 IN COHORT
6. any demographic record of Female	0 IN COHORT
7. the following events: first diagnosis of Essential (primary) hypertension happened at most 1 Days Before these events: first diagnosis of Atrial fibrillation and flutter	0 IN COHORT
8. the following events: first diagnosis of Type 1 diabetes mellitus or first diagnosis of Type 2 diabetes mellitus happened at most 1 Days Before these events: first diagnosis of Atrial fibrillation and flutter	0 IN COHORT
9. More or equals to 75 Years old	0 IN COHORT

Fig 4c. CHA2DS2 ja AF + Intracranial bleeding (example in Fig 4c. is for diabetes)

localhost:8080/app/cohort/studies/study/builder/import?cohortStudyId=8&_csrf=aa28b7c8-0e66-442d-8595-fff2d21e7a4c

sted Sites Imported From IE laadullinen tutkimus Sote-asiakastietojen Moodle Login Tohtorikoulutus Suomi.fi-tunnistus Strategies to improve

Site Studies Cohort Analysis Recruiting Studies NiinaM

Cohort Studies autovaihtaja Cohort builder

Choose cohort: 1 016 137 IN COHORT Reference date: 31 DEC 2012

Age 18 or more on 1.1.2003 (1 016 137 patients)

FILTERS Filters + Add filter

1.	first diagnosis of Atrial fibrillation and flutter	23 865 IN COHORT
2.	ABSTRACT AF + IC bleeding	613 IN COHORT
3.	the following events: first diagnosis of Type 1 diabetes mellitus or first diagnosis of Type 2 diabetes mellitus happened at most 1 Days Before these events: first diagnosis of Atrial fibrillation and flutter	76 IN COHORT
4.	the following events: first diagnosis of Atherosclerosis or first diagnosis of Aortic aneurysm and dissection happened at most 1 Days Before these events: first diagnosis of Atrial fibrillation and flutter	2 IN COHORT
5.	the following events: first diagnosis of Heart failure happened at most 1 Days Before these events: first diagnosis of Atrial fibrillation and flutter	1 IN COHORT
6.	any demographic record of Female	0 IN COHORT
7.	the following events: first diagnosis of Essential (primary) hypertension happened at most 1 Days Before these events: first diagnosis of Atrial fibrillation and flutter	0 IN COHORT
8.	More or equals to 75 Years old	0 IN COHORT
9.	ABSTRACT history of stroke or TIA	0 IN COHORT

	AF + Stroke	AF + TIA	AF + IC blee- ding
	*)1d	1d	1d
CHA2DS2-VASc score	3620	923	613
Heart failure (i50) **)	536	89	75
Hypertension >(i10)	1424	303	183
Diabetes (E10 and E11)	533	119	76
Vascular disease (i70-i71)	76	14	10
female	1923	428	242
history of Stroke or TIA	na	na	na
age >75	na	na	na

Table 5. Number of patients who had certain risk factor diagnosed at the time of their first AF.

*) first dg of AF at least 1 d before first dg of stroke

**) These dg existed already at the time of first AF (at least 1 day before AF).

For example: Number of patients who got AF and later on stroke, had heart failure diagnosed already at the time of their first AF. There were 536 such patients.

	AF + Stroke	AF + TIA	AF + IC blee- ding
PAIRED RUNS	4924	1502	915
The stroke risk factor	*)1d	1d	1d
CHA2DS2-VASc score **)	3620	923	613
Hypertension + Heart failure	233	39	19
Hypertension + Diabetes	328	69	43
Hypertension + vascular disease	41	4	0
Hypertension + female	844	167	90
Diabetes + vascular disease	28	4	2

Diabetes + female	279	54	35
Diabetes + heart failure	106	17	10
Vascular disease + female	33	5	1
Vascular disease + heart failure	15	2	3
heart failure + female	304	54	37
	Fig 4a.	Fig 4b.	Fig 4c.

Table 6. Number of patients who had certain risk factors diagnosed at the time of their first AF. For example: Number of patients who got AF and later on stroke, had heart failure AND Hypertension diagnosed already at the time of their first AF. There were 233 such patients.

Medication

It was tested, what medication was in use after the patient had got his first dg of AF. Medication runs were done for 3 sub-populations: stroke, TIA and Intracranial bleeding. Medication filters were tested one by one, ie by changing the order of the filters.

Fig 6a. Medication for AF + stroke. Example in this Fig 6a. is for novel anticoagulants.

localhost:8080/app/cohort/studies/study/builder/import?cohortStudyId=8&_csrf=d0b2ac1f-39b0-4aef-901e-cf2982861248

Imported From IE laadullinen tutkimus Sote-asiakastietojen Moodle Login Tohtorikoulutus Suomi.fi-tunnistus Strategies to improv

InSite Studies Cohort Analysis Recruiting Studies NiinaN

Cohort Studies autovalidate Cohort builder

FILTERS Filters + Add fil

1.	the following events: first diagnosis of Atrial fibrillation and flutter happened at most 100 Years before now	23 865 IN COHORT
2.	ABSTRACT AF before Stroke	3 620 IN COHORT
3.	the following events: first administration of dabigatran etexilate or first administration of rivaroxaban or first administration of apixaban or first administration of rivaroxaban happened at least 1 Days After these events: first diagnosis of Atrial fibrillation and flutter	63 IN COHORT
4.	the following events: first administration of Vitamin K antagonists or first administration of warfarin happened at least 1 Days After these events: first diagnosis of Atrial fibrillation and flutter	29 IN COHORT
5.	the following events: any administration of acetylsalicylic acid or any administration of acetylsalicylic acid happened at least 1 Days After these events: first diagnosis of Atrial fibrillation and flutter	1 ^o IN COHORT (OUTDATED)
6.	the following events: first administration of clopidogrel or first administration of ticagrelor or first administration of prasugrel happened at least 1 Days After these events: first diagnosis of Atrial fibrillation and flutter	0 ^o IN COHORT (OUTDATED)

6b. Medication AF+ TIA. Example in this Fig 6b. is for varfarin.

localhost:8080/app/cohort/studies/study/builder/import?cohortStudyId=8&_csrf=d0b2ac1f-39b0-4aef-901e-cf2982861248

sted Sites Imported From IE laadullinen tutkimus Sote-asiakastietojen Moodle Login Tohtorikoulutus Suomi.fi-tunnistus Strategies to improv

Site Studies Cohort Analysis Recruiting Studies NiinaN

Cohort Builder

SETTINGS

Choose cohort: 1 016 137 IN COHORT

Age 18 or more on 1.1.2003 (1 016 137 patients)

Reference date: 31 DEC 2012

Filters + Add filter

1. the following events: first diagnosis of Atrial fibrillation and flutter happened at most 100 Years before now	23 865 IN COHORT
2. ABSTRACT AF before TIA	923 IN COHORT
3. the following events: first administration of Vitamin K antagonists or first administration of warfarin happened at least 1 Days After these events: first diagnosis of Atrial fibrillation and flutter	448 IN COHORT
4. the following events: first administration of dabigatran etexilate or first administration of rivaroxaban or first administration of apixaban or first administration of rivaroxaban happened at least 1 Days After these events: first diagnosis of Atrial fibrillation and flutter	15 IN COHORT
5. the following events: any administration of acetylsalicylic acid or any administration of acetylsalicylic acid happened at least 1 Days After these events: first diagnosis of Atrial fibrillation and flutter	5 IN COHORT
6. the following events: first administration of clopidogrel or first administration of ticagrelor or first administration of prasugrel happened at least 1 Days After these events: first diagnosis of Atrial fibrillation and flutter	0 IN COHORT

6c. Medication AF-IC bleeding. Example in this Fig 6c. is for novel anticoagulants.

localhost:8080/app/cohort/studies/study/builder/import?cohortStudyId=8&_csrf=d0b2ac1f-39b0-4aef-901e-cf2982861248

ested Sites Imported From IE laadullinen tutkimus Sote-asiakastietojen Moodle Login Tohtorikoulutus Suomi.fi-tunnistus Strategies to improv

nSite Studies Cohort Analysis Recruiting Studies MiinaN

Cohort Builder

Save Cohort & Explore Patients Save Cohort

Recalculate results Show match

Choose cohort: 1 016 137 IN COHORT


Reference date: 31 DEC 2012

Age 18 or more on 1.1.2003 (1 016 137 patients)

FILTERS

Filters Add filter

- the following events:
first diagnosis of Atrial fibrillation and flutter
happened at most 100 Years before now 23 865 IN COHORT
- ABSTRACT AF before IC bleeding 613 IN COHORT
- the following events:
first administration of clopidogrel
or first administration of ticagrelor
or first administration of prasugrel
happened at least 1 Days After these events:
first diagnosis of Atrial fibrillation and flutter 24 IN COHORT
- the following events:
any administration of acetylsalicylic acid
or any administration of acetylsalicylic acid
happened at least 1 Days After these events:
first diagnosis of Atrial fibrillation and flutter 15 IN COHORT
- the following events:
first administration of dabigatran etexilate
or first administration of rivaroxaban
or first administration of apixaban
or first administration of rivaroxaban
happened at least 1 Days After these events:
first diagnosis of Atrial fibrillation and flutter 1 IN COHORT
- the following events:
first administration of Vitamin K antagonists
or first administration of warfarin
happened at least 1 Days After these events:
first diagnosis of Atrial fibrillation and flutter 1 IN COHORT



	AF + Stroke at least 1 day after		AF + TIA at least 1 day after		AF + IC bleeding at least 1 d after	
	4924		1502		915	
	*) 1d	1d	1d	1d	1d	1d
	3620		923		613	
	medication in use at least 1 day after AF	medication in use at least 1 month after AF	medication in use at least 1 day after AF	medication in use at least 1 month after AF	medication in use at least 1 day after AF	medication in use at least 1 month after AF
ASA **)	596	526	177	163	119	115
Varfarin, K-vitamin antagonist	1288	1137	→ 448	419	217	208
Novel anticoagulants (diabigatran, rivaroxaban, apixaban)	→ 63	55	33	26	6	6
Antithrombotic medicines: clopidogrel, prasugrel or ticagrelor	123	55	40	36	→ 24	23
Antithromb + ASA	83	74	25	24	15	14

Table 6. Number of patients who got ASA, varfarin, novel anticoagulants, antithrombotics or antithromb+ASA after their first dg of AF. For example: 63 patients got novel anticoagulants after AF and later got stroke. This test was also run for all sub-populations with additional time constraint to see what medication was in use at least 1 month after the AF.

4 TESTAUS ASETELMA JA TESTAUKSEN TULOKSET

Testaus on toteutettu ajalla Q1-Q2 2018. Insite mahdollistaa tutkimuskohortin muodostuksen selainkäyttöliittymän avulla. Yllämainitussa tutkimusasetelmassamme kohortin muodostus ei toiminut erityisen luotettavasti. Insiten haku liittää ymmärryksemme mukaan potilaita, joita ei poimintaan kuuluisi kriteeristö huomioiden olla luettu osaksi kohorttia. Pyysimme insitelta selvitystä tästä, vaan emme ole saanut vastausta vielä raportin kirjoitushetkellä vaikkakin olemme toimittaneet omat skriptimme heidän nähtäville.

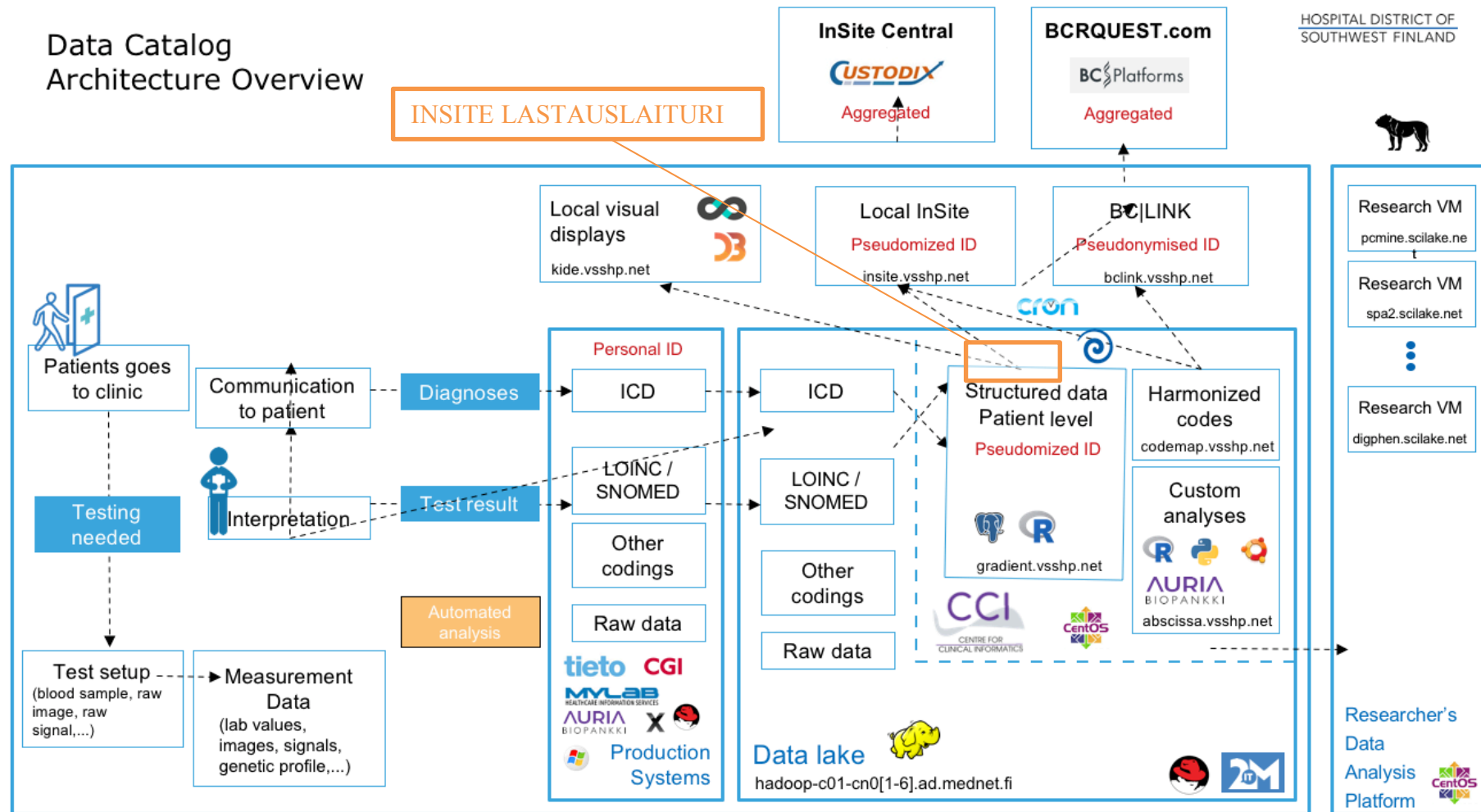
Insiten perusidea on, että sen avulla voi saada selville, paljonko tietyt kriteerit omaavia potilaita on sairaalassa. Tähän tarkoitukseen se soveltuukin hyvin. Layout näyttää selkeästi potilasmäärän vähenemisen sitä mukaa kun kriteereitä lisätään. Insite listaa potilasmääriä, Fibstroke-tutkimuksessa oli listattu myös mm. aivotapahtumien lukumääriä per potilas, minkälainen haku on Insitellä mahdotonta tai ainakin hyvin hankalaa.

Fibstroke-tutkimuksen ottaminen vertailututkimukseksi ei ehkä ollut paras vaihtoehto validoinnin kannalta, sillä aineistoon/väitöskirjaan ei ollut kirjattu kriteereitä tai tietoja potilaista, joita ei oltu kelpuutettu mukaan manuaalisen haun perusteella. Myöskään tietopoiminnan toteutuksen skriptejä ei ole tallella organisaation ja prosessin muutossyistä. Nykyään tietoluovutuksen prosessiin kuuluu tallentaa sekä SQL/R-ajon skriptit, että tutkimusluvut samaan kansioon. Tämän takia suora luotettava vertailu on mahdotonta. Myöskään ei ollut mahdollista määritellä samaa aikakriteeriä kuin Fibstroksessa sisällyttää sellaiset potilaat, jotka olivat TYKS:in asiakkaita vähintään vuoden aivotapahtuman jälkeen.

Insitellä saatu kohdepopulaation määrä oli kaksinkertainen Fibstroke-tutkimuksen hakuihin. Huomattavaa oli kuitenkin se, että potilasmäärät olivan johdonmukaisesti lähes kaikissa testeissä manuaalisella haulla 40-48 % pienemmät kuin Insitellä haettuna. Ei ole olemassa yhtä selittävää tekijää suuresta erosta. Tälläisiä syitä voisi olla mm. se, että väitöskirjassa on käytetty jotain mainitsematonta kriteeriä alkuperäisen kohortin siivoamiseen ("data curatointiin"). Tietysti myös tuolloin hyödynnetystä SAS-järjestelmästä tehty poiminta on voitu toteuttaa hieman eri tavoin. Kokemustemme mukaan ja manuaaliseen uusintapoimintaan verrattuna vaikuttaa kuitenkin siltä, että Insite scripti poimii mukaan potilaita, joiden ei kriteerien perusteella tulisi kohortissa näkyä. Mainittakoon vielä, että tiedustelimme useampaan kertaan Insitelta poimintalogiikkaa kun tunnistimme, että insite ottaa hakuihinsa mukaan myös potilaita, jotka oli rajattu halutun aikakriteerin ulkopuolelle, eli potilaita jotka eivät pitäisi tulla mukaan mikäli potilaspoiminta suoritettaisiin SQL-skriptin tai R-skriptin avulla suoraan kannasta.

Insitella tehtyjen lääkitys CHA2DS2-VASc ajojen tuloksia emme ole ennättänyt verrata lainkaan Fibstroksen manuaaliseen aineistoon vielä tämän raportin kirjoitus hetkellä.

Data Catalog Architecture Overview



Testausasetelma teknisen arkkitehtuurin kulmasta katseltuna

YHTEISTYÖ

Oheisen kysymyksen olemme toimittaneet Insite-tiimille useampaan kertaan, katso kuvakaappauksen punaisella merkitty laatikko. Saimme 12.6 validoinnin ajojemme oikeellisuudelle, mutta tarkempi selvitystyö lienee vielä kirjoitushetkellä 19.06.2018 kesken.

(ensimmäinen maili **Lähetäjä:** Varjonen Juha-Matti

Lähetetty: maanantai 30. huhtikuuta 2018 9.20

Vastaanottaja: 'laura.meloni@custodix.com' <laura.meloni@custodix.com>

Kopio: Blomster Juuso <Juuso.Blomster@tyks.fi>; 'Niina Nieminen' <niinie@utu.fi>

Aihe: Autovalidate study/University of Turku)

Age 18 or more on 1.1.2003 (1 016 137 patients) IN COHORT 31 DEC 2012

FILTERS Filters: + Add filter

- the following events:
any diagnosis of Atrial fibrillation and flutter
happened at most 100 Years before now 23 865
IN COHORT
- the following events:
first diagnosis of Intracranial injury
happened at most 10 Years before now 915
IN COHORT
- ABSTRACT** IC bleeding AFTER AF
excluding HF 613
IN COHORT
the following events:
first diagnosis of Intracranial injury
happened at least 1 Days After these events:
first diagnosis of Atrial fibrillation and flutter

Oheinen filttointi Insite laustauslaiturilta (* katso arkkitehtuurikuva yllä) osoittaa, että Insite haku sisällyttää virheellisesti hakuun potilaita, joilla aikakriteeri ei täyty. Käytimme SQL kyselyjä suoraan kantaan ja saimme poikkeavat tulokset, vaan emme selvitystä Insitelta (538 patients versus 613 patient).

First S06 (Intracranial injury) happened at least 1 Days After first I48 (Atrial fibrillation and flutter).

Ohessa esim. 2 potilasta, jotka mielestämme eivät täytä kohortin poimintakriteeriä (tarvittaessa koko lista on saatavissa ~70 patients)

```
-- Patient_ID = zVxoHpWVkh2s+1VrEcXmJA==
SELECT patient_id, date, diagnosis_code, is_main_diagnosis, is_cause_diagnosis,
```

```

        location, update_time_stage, update_time_ehr4cr, diagnosis_id,
        visit_id
FROM ehr4cr_insite.diagnosis where patient_id = 'zVxoHpWVkh2s+lVrEcXmJA==' and
diagnosis_code ~* 'I48|S06'
ORDER BY date

-- "zVxoHpWVkh2s+lVrEcXmJA==" ; "2008-06-04" ; "S06.6"
-- "zVxoHpWVkh2s+lVrEcXmJA==" ; "2008-06-04" ; "S06.0"
-- "zVxoHpWVkh2s+lVrEcXmJA==" ; "2008-06-05" ; "S06.6"
-- "zVxoHpWVkh2s+lVrEcXmJA==" ; "2008-06-05" ; "I48"

-- =====

-- Patient_ID = zjZe10GEuEsqeyXkWz3URw==
SELECT patient_id, date, diagnosis_code, is_main_diagnosis, is_cause_diagnosis,
        location, update_time_stage, update_time_ehr4cr, diagnosis_id,
        visit_id
FROM ehr4cr_insite.diagnosis where patient_id = 'zjZe10GEuEsqeyXkWz3URw==' and diagnosis_code ~* 'I48|S06'
ORDER BY date

-- "zjZe10GEuEsqeyXkWz3URw==" ; "2008-02-01" ; "S06.5"
-- "zjZe10GEuEsqeyXkWz3URw==" ; "2008-02-01" ; "S06.5"
-- "zjZe10GEuEsqeyXkWz3URw==" ; "2008-02-23" ; "S06.5"
-- "zjZe10GEuEsqeyXkWz3URw==" ; "2008-02-23" ; "S06.5"
-- "zjZe10GEuEsqeyXkWz3URw==" ; "2008-02-25" ; "S06.5"
-- "zjZe10GEuEsqeyXkWz3URw==" ; "2008-02-25" ; "I48"

```

First intracranial bleeding (S06.6) kirjattu 4.6.2008 ja first AF (i48) kirjattu jälkeen 5.6.2008.
Tämä potilas ei olisi kuulunut olla mukana Insiten listalla

First intracranial bleeding (S06.6) kirjattu 1.2.2008 ja first AF (i48) kirjattu jälkeen 25.2.2008.
Tämäkään potilas ei olisi kuulunut olla mukana Insiten listalla. Näitä löytyi noin 70 muutakin kyseisestä testiajosta

Mikä on Insiten kohortinrakentaja skriptin logiikka? Olemmeko käyttäneet Insitea oikein?

5 KÄYTTÖKELPOISUUS KLIINIKON NÄKÖKULMASTA JA SYY INSITE IMPLEMENTOINTIIN

Insitea käytettäessä oli vaikea hahmottaa Insite termejä “at most”(maksimissaan) ja “at least”(ainakin) ja Custodixin toimittama guideline ohje auttoi osittain, mutta jätti silti tulkinnanvaraa kommentojen oikeaoppiselle käytölle. Ei myöskään ollut päivänselvää, miten inkluusio- ja eksklusio-kriteerit erotettiin toisistaan. Monessa kohdassa inhimillisen virheen riski on suuri, sillä käyttöliittymä ei ole vielä tarpeeksi selkeä. Vaikka itse tutkimuksessa oli käytännössä vain muutama kriteeri, niistä piti rakentaa Insiteen monta filtteriä (välikohorttia), jotta haluttu kriteeristö saataisiin ajettua.

Tämä oli aikaa vievää, ohjeet puutteelliset ja vaati tarkastuksia Custodixin asiantuntijoiden avulla. Koska olimme epävarmoja siitä, ovatko kriteerit algoritmisesti oikein niin tarkistutimme asian Custodixin asiantuntijoilla ja pyysimme saada nähdä poimintaskriptit. Validointitutkimuksen kokeimuksen perusteella Insite soveltuu parhaiten uusien tutkimusten suunnittelutyökaluksi sopivien potilaspopulaatioiden ja heitä hoitavien sairaaloiden löytämiseksi. Insite antaa kuvan siitä, että potilasmäärät ovat suuremmat kuin manuaalisessa poiminnassa. Potilasmäärät saadaan kuitenkin edelleen tarkemmin manuaalisella poiminnalla, jossa myös laadunseuranta mahdollistuu. Insitellä ei pystytä havaitsemaan diagnoosien ja lääkkeitöiden onnistumisen trendejä tai analysoimaan dataa potilastapahtumien aikajanaa ajatellen. Kliinisen tutkimuksen näkökulmasta Insitea voidaan käyttää alustavaan arvioon, siitä onko tietyn diagnoosin saaneita potilaita sairaalan potilaina ja alustava arvio siitä, kuinka monta potilasta kohorttiin voitaisiin parhaimmillaan saada kyseisestä sairaalasta. Tähän nimenomaiseen tarkoitukseen Insite on alun perin kehitettykin. Sairaanhoidopiiriin kliinisen tutkimuksen katalyyttinä se voi toimia, mikäli se saavuttaa tutkijoita, jotka eivät muuten hoksaisi tiedustella meiltä aineistojamme omiin tutkimuksiinsa. Näin se toimisi yhden palveluporttaalin pisteenä useampaan sairaalaan ja tarjoaisi VSSHPille oivan ikkunan kliinisen tutkimuksen maailmanmarkkinoille.

Suomella on mahdollisuus nousta kansainvälisesti merkittäväksi rekisteri ja kliinisen tiedon hyödyntäjäksi, sekä maailmalla laadustaan tunnisteutuksi tutkimuspartneriksi. Globaalit lääkefirmat ovat kiinnostuneita laadukkaasta tietovarannon hyödyntämisen mahdollisuuksista päätöksenteon tukena. Useista rekistereistä yhdisteltyä tietoa voidaan hyödyntää muun muassa tutkimuspotilaiden identifioinnissa, satunnaistettujen lääketutkimusten tulosten validoinnissa sekä uusien hoitojen kustannusvaikuttavuuden arvioinnissa. Lääketeollisuuden tutkijat voisivat jo nykyisen tietovarantomme avulla pyrkiä osoittamaan tiettyjen lääkkeiden tehokkuutta suhteessa vanhempiin hoitomuotoihin ja lääkityksiin. Harvinaissairauksiin liittyvät tutkimukset, hyödyttävät merkittävästi potilasta ja juuri niissä laajat ja laadullisesti tasokkaat tietovarannot ovat erityisen tärkeitä kustannustehokkuus näkökulmasta.

Kansainvälisessä mittakaavassa Suomessa on erinomaiset edellytykset kliinisen tutkimusten suorittamiseen maassamme. Meillä on kansakuntana tarjota kehittynyt ja luottamusta nauttiva terveydenhuoltojärjestelmä, mielenkiintoinen geeniperimä, laadukas rekisteridata jo vuosikymmenten ajalta, yksilöllinen henkilöturvatus ja teknistä kyvykkyyttä. Lisäksi tutkimuslainsäädäntöämme ollaan modernisoimassa, mikä osaltaan voisi helpottaa ja keventää Suomessa toteutetun tutkimuksen kustannusrakennetta. Lakiuudistuksista suurimpana valmistelussa on sosiaali- ja terveys-tietojen toisiokäyttöön liittyvää uutta lainsäädäntöä, jonka myötä eri rekistereistä ja lähteistä noudetun tiedon (RWD, real world data) hyödyntämismahdollisuuksia tulisi entisestään laajentamaan.

Sairaanhoitopiiri jossa on paikallinen saatavuuspalvelu , esim. Insite-link implementointi on tehty näkyy tutkimuksen maailmanmarkkinoilla ja mahdollistaa tutkimusvolyymin kasvun. Insite tukee kliinisen tutkimuksista feasibility-kysymysten selvittelyn prosessia ja vastaa kysymykseen, että miten paljon tutkimus asetelmaan sopivia, tietynlaisia kriteereitä omaavia potilaita, kyseisen sairaanhoitopiirin alueelta löytyy. Suurin etu sairaanhoitopiirille kuitenkin syntyy Insite implementoinnissa siitä kun Insiteä hyödyntävät suuret tutkimusorganisaatiot ja lääkeyritykset suunnittelevat tutkimustensa sijoittamista eri maihin ja sairaaloihin tekemällä ajoja koko Insite-verkostoon saavat tietoonsa myös kyseisen sairaanhoitopiirin potilaskohortin. Mikäli tutkijaorganisaatiot ja lääkeyhtiöt tunnistavat tämän portaalin hyödyt laajemmin, on sen ulkopuolelle jääminen myös riski siitä, että tutkimuksia ei saada kotiutettua Suomalaisiin sairaanhoitopiireihin niin tehokkaasti kuin olisi mahdollista. Toki etuna on, että poimintoja ts. kohortin muodostuksen voisi suorittaa muutkin kuin asiaan vihkiytyneet tekniset osaajat, jotka nykyään tekevät kohortti poimintoja manuaalisin keinoin. Juuri tuosta syystä Suomen sairaaloiden mukanaolo olisi tärkeää, sillä Insite toimii näyteikkunana kliinisen tutkimuksen maailmanmarkkinoilla. Parhaassa tapauksessa , lääkeyritykset tunnistaisivat Suomen yhdeksi maaksi, jossa tutkimusta toteutettaisiin. Toki parhaan hyödyn realisoimiseksi tiivis yhteistyö ja käyttöliittymän laajempi validointi ainakin tässä alkuvaiheessa Custodixin kanssa loisi paremmat menestysedellytykset. Tiiviimpi yhteistyö syksyä kohden olisi myös meidän testaaajien toivomus.

Varsinais-Suomen Sairaanhoitopiiriin ,VSSH:n, tutkimuspalveluista kerrottiin, että suurin osa VSSH:lle nykyiseltään saapuvista kansainvälisistä kliinisistä tutkimuksista saapuu meille yleistysti kahdesta vaihtoehdoisesta syystä:

1) Tutkimus tehdään kansallisesti kaikissa yliopistollisissa sairaaloissa ja sen takia myös meillä VSSH:llä.

2) Kliinikan johtava lääkäri tai muu maailmalla luottamusta herättänyt on tutkimusmyönteinen ja saa jalkautettua kansainvälisiä tutkimuksia myös meidän sairaanhoitopiirissämme toteutettavaksi.

Toki on hyvä pitää mielessä, että lääkefirmat puntaroivat VSSH yhteistyössä sopivan yhteistyökliinikon lisäksi myös tutkimuksen hintaa, nopeuttamme rekrytoida potilaita tutkimukseen, sekä kohortin sopivuutta ja datan laatua suhteessa tutkimusasetelmaansa.

- - -

Notes on validation runs:

	Kriteeri	Ajot:	Tulos/kommentti
1	TYKS potilastietojärjestelmään kirjatut potilaat, jotka ≥ 18 vuotta ajankohtana 1.1.2003	Ajo 1: -kaikki potilaat kohortista -Filter: more or equals to 18 (age constraint) -Tässä ajossa: ref date 1.1.2003	Sama tulos käytetäänkö ajoa 1 tai 2.
		Ajo 2: luotiin sub-cohort: Age 18 or more on 1.1.2003	
2	Asuinpaikka VSSH:n alue aivotapahtuman aikana ja 1 vuosi tämän jälkeen		Ei mahdollista testata

3	Eteisvärinä-diagnoosi (milloin tahansa historiassa)	-First dg of i48 happened at most 100 years from now	Aikamääreellä ei näytä olevan merkitystä
		-first dg of i48 happened at most 10 years from now	
		-“Last”/“first”/“any” dg of i48, tuottaako saman tuloksen	Kyllä
4	Aivoverenkiertohäiriön diagnoosi (Stroke, TIA tai Intracranial bleeding during 1.1.2003-31.12.2012)	Ylläolevien filtterien lisäksi: -first dg of i60-i69 (excl i64,i67-i68) happened at most 10 years ago OR -first dg of G45-G46 happened at most 10 years ago OR -firstdg of S06 happened at most 10 years ago	
	AF ennen Strokea	Ylläolevien filterien lisäksi: -Following events (7 kohtaa, välissä OR) happened at least 1 days after these events: AF -Following events (7 kohtaa) happened at least 30 days after these events: AF	Ajettu vertailun vuoksi kahdella aikamääreellä, 1pv ja 1 kuukausi AF jälkeen.
	AF ennen TIAa	-Following events (first dg of <i>transient cerebral ischemic attacks and related syndromes</i> OR first dg of <i>vascular syndromes of brain in cerebrovascular diseases</i>) happened at least 1 days after these events: AF -Following events (first dg of <i>transient cerebral ischemic attacks and related syndromes</i> OR first dg of	Ajettu vertailun vuoksi kahdella aikamääreellä, 1pv ja 1 kuukausi AF jälkeen.

		<i>vascular syndromes of brain in cerebrovascular diseases</i>) happened at least 1 month after these events: AF	
	AF ennen Intracranial bleeding	-Following events (first dg of <i>intracranial injury</i>) happened at least 1 days after these events: AF -Following events (first dg of <i>intracranial injury</i>) happened at least 30 days after these events: AF	Ajettu vertailun vuoksi kahdella aikamääreellä, 1pv ja 1 kuukausi AF jälkeen.

6

MANUAALINEN R-KIELELLÄ TOTEUTETTU POIMINTASKRIPTI

Alla oleva skripti on vain pääkriteerit täyttävä ja voi sisältää virheitä. Skripti on tehty helpottamaan Insite ajojen validointia ja vastaa kuitenkin kysymykseen montako täysi-ikäistä potilasta etäisvärinädiagnoosin ensin saaneita vuosilta 2003- 2012 löytyy kullekin ala ICD-kohortille.

```
# install.packages("data.table")
# install.packages("SnowballC")
# install.packages("wordcloud")

# TÄMÄKIN EHKÄ

# sudo yum install libcurl-devel.x86_64

# install.packages("webutils")

# TÄMÄKIN EHKÄ

# sudo yum install libxml2-devel

# install.packages("xml2")
# install.packages("tm")
# install.packages("gridExtra")
# install.packages("openxlsx")
# install.packages("DBI")
# install.packages("ggplot2")

# yum install libpqxx.x86_64
```

```
# yum install libpqxx-devel.x86_64
```

```
# install.packages("RPostgreSQL")
# install.packages("RJDBC")
# install.packages("lubridate")
```

```
# install.packages("readxl")
```

```
# -----
# lataa n?m? aina kun allaolevat paketit on päivittyneet
# t?m?n j?lkeen k?ynnist? R uudelleen
```

```
install.packages("/home/antti/mnt/local/R/RtoolsKTP_0.5.tar.gz", repos = NULL, type = "source")
install.packages("/home/antti/mnt/local/R/poiminta_1.1.tar.gz", repos = NULL, type = "source")
install.packages("//home/antti/mnt/local/R/poiminta_1.1.tar.gz", repos = NULL, type = "source")
```

```
# -----
# lataa n?m? aina kun R-k?ynnistet??n uudelleen:
```

```
options(stringsAsFactors = F)
options(java.parameters = "-Xmx8000m")
library(RtoolsKTP)
library(poiminta)
library(data.table)
# library(RODBC)
library(RPostgreSQL)
```

```
Sys.setenv(R_ZIPCMD= "/usr/bin/zip")
```

```

library(stringr)
library(openxlsx)
library(plyr)
library(ggplot2)
library(gridExtra)
library(RJDBC)
# install.packages("lubridate")
library(lubridate)
library(foreign)
Sys.setenv(TZ = 'EST')

# ODBC-yhteyden muodostus
# con_pg <- odbcConnect(dsn = "PostgreSQL35W",
#                       pwd=readLines("M:/DATA/Data/Apu3.txt", warn = F))

# Vaihtoehtoinen PostgreSQL drv yhteydenmuodostus

drv <- dbDriver("PostgreSQL")
con_pg <- dbConnect(drv = drv, dbname = "ktp", user = "ktp", host = "gradient.vssh.net", port = 5432,
  password = readLines("/home/antti/.ssh/apu3.txt"))
con_pg_r <- dbConnect(drv = drv, dbname = "research", user = "ktp", host = "gradient.vssh.net", port =
  5432, password = readLines("/home/antti/.ssh/apu3.txt"))

# Tutkimusnumero
# 1. T***/2013(as in projektirekisteri http://atas41/projektirekisteri/default.aspx )
# 2. If only statistics etc KTP***/2017
this_Tutkimusnumero <- "T140/2017"

setwd_by_research(this_Tutkimusnumero,
  dir_path = c(tutkimukset = "/home/antti/mnt/ktp/Tutkimukset",

```

```

    toimeksiannot = "/home/antti/mnt/ktp/Toimeksiannot",
    ktp_kansio = "/home/antti/mnt/ktp"))

# setwd_by_research(this_Tutkimusnumero,
#                   dir_path = c(tutkimukset = "/mnt/pinta/KTP/Tutkimukset",
#                               toimeksiannot = "/mnt/pinta/KTP/Toimeksiannot",
#                               ktp_kansio = "/mnt/pinta/KTP"))

getwd()

Kohortin luominen
```{r eval=FALSE}

time points within data must be (check from researcher or tutkimuslupa)
this_min_project_date <- as.Date("2003-01-01")
this_max_project_date <- as.Date("2012-12-31")
this_date_values <- seq(this_min_project_date, this_max_project_date, by = "day")

#-----

this_diagnoosi_list_a <- c("^I60.[0-9]", "^I61.[0-9]", "^I62.[0-9]", "^I63.[0-9]", "^I65.[0-9]", "^I66.[0-9]",
 "^I69.[0-9]", "^G45.[0-9]", "^G46.[0-9]", "^S06.[0-9]", "AF", "^I48")

ladataan kaikki syöpä diagnoosin saaneet potilaat
diagnoosit_a <- load_diagnoosit_pg(con_pg,
 diagnoosi = this_diagnoosi_list,
 date_start = this_min_project_date,
 date_end = this_max_project_date,
 sql = F)

```

```
lasketaan potilaan ikä tapahtumahetkellä
```

```
diagnoosit_a$IKA_TMP_HETKI <- as.integer((as.Date(diagnoosit_a$TAPAHTUMA_AIKALEIMA)-
 diagnoosit_a$HENKILOTUNNUS, birthday = T)$birthday)/365.25)
diagnoosit_18_ja_yli_a <- diagnoosit_a[IKA_TMP_HETKI >= 18]
```

```
Haetaan yksittäiset henkilötunnukset
```

```
ensi_diag <- diagnoosit[,.SD[which.min(TAPAHTUMA_AIKALEIMA)],list(HENKILOTUNNUS)]
ensi_diag <- diagnoosit_18_ja_yli_a[!grepl("I48",DIAGNOOSI,ignore.case = T)]
```

```
ensi_diag_i48 <- diagnoosit_18_ja_yli_a[DIAGNOOSI == "I48",.SD[which.min(DGN_PVM)],list(HENKILOTUNNUS)]
```

```
testausta <- merge(ensi_diag_i48[,list(HENKILOTUNNUS,MIN_DGN_PVM_I48 = as.Date(DGN_PVM),DIAGNOOSI_I48 =
 DIAGNOOSI)],
 ensi_diag[,list(HENKILOTUNNUS,MIN_DGN_PVM_MUU = as.Date(DGN_PVM),DIAGNOOSI_MUU = DIAG-
 NOOSI)],
 by = "HENKILOTUNNUS",
 all = F)
```

```
diag_tulos <- testausta[DIAGNOOSI_I48 == "I48"][!is.na(MIN_DGN_PVM_I48)][MIN_DGN_PVM_I48 <
 MIN_DGN_PVM_MUU]
```

```
Ladataan kohortti_hetuun minimi aikaleima
```

```
diag_loppullinen <- diag_tulos[,.N,DIAGNOOSI_MUU][order(-N)]
```

```
create_kohortti_pg(con_pg = con_pg ,
```

```

 register_schema = "luovutusrekisteri",
 hetu_list = kohortti_hetu,
 this_Tutkimusnumero = this_Tutkimusnumero
 #,delete_already_exist = TRUE
 #,in_v_asiakas = TRUE
)

```

kohortti (3375)

```

#####
#####

this_kohortti_id_fk <- as.integer(3375)

this_data_to_write <- NULL

this_data_to_write$diag_loppullinen <- diag_loppullinen

if(! "deliverables" %in% list.files()) dir.create("deliverables")

create xlsx file. As input is list of data.tables --> every data.table will go on it's own excel sheet:
xlsx_create(this_data_to_write,
 file_name = paste0("deliverables/", Sys.Date(), "_", gsub("/", "_", this_Tutkimusnumero), "_poiminta_data.xlsx"))

```

MANUAALINEN POIMINTASKRIPTIN TULOS

DIAGNOOSI\_MUU      N

I69.3	3950
I63.4	3425
I63.9	2522
S06.5	785
I63.3	736
G45.1	733
G45.9	496
I63.5	420
S06.0	417
I61.0	399
G45.0	309
I65.2	255
I62.0	225
I69.1	195
I63.1	176
I61.9	154
G45.3	145
I63.0	140
I61.1	104
S06.3	97
I63.8	96
I63.2	78
G45.4	68
I60.9	67
I69.0	61
I61.4	49
I60.1	41
S06.9	33
S06.6	32
I69.2	32
I69.4	26

G46.7*I67.9	26
I60.0	24
G46.4*I67.9	23
G45.8	22
I61.3	20
S06.2	20
I62.9	19
I65.1	18
I60.2	17
S06.4	17
I61.5	17
I69.8	14
S06.8	14
G46.3*I67.9	10
I65.3	10
I60.5	10
I66.9	9
I60.3	9
I60.8	9
I61.8	7
I61.2	7
I60.4	7
I63.6	6
I65.0	6
S06.7	6
I61.6	5
G46.5*I67.9	5
I65.9	4
I67.9+G46.4	4
I67.9+G46.3	3
I66.0	3



I62.1	3
I67.9+G46.8	2
G46.8*I67.9	2
I66.8	2
G46.4*	2
G46.3*	1
I66.3	1
I66.1	1
G46.2*I66.2	1
G46.0*I66.0	1
S06.1	1
I67.9+G46.7	1