



Datatillgång och användbarhet

Källor, möjligheter, utmaningar och jämförelser



UPPSALA
UNIVERSITET



Vasa centralsjukhus
Vaasan keskussairaala



Österbottens förbund
Pohjanmaan liitto



UMEÅ UNIVERSITET



KUST-ÖSTERBOTTENS SAMKOMMUN FÖR SOCIAL- OCH PRIMÄRHÄLSOVÅRD

Datatillgång och användbarhet – Källor, möjligheter, utmaningar och jämförelser

En delrapport inom Botnia-Atlanticaprojektet *SOS 1.0 – Simuleringsstödd optimering av servicekedjor inom specialist-, primär- och hemsjukvård.*

Författare: Mats Johansson, Konrad Abramowicz

Foto: Freepik

Region Västerbotten Januari 2020

Innehåll

Abstract	4
Bakgrund och syfte	5
Datamaterial – Källa och innehåll	5
Resultat – Möjliga analyser	7
Primärvården.....	7
Vårdtid och demografi	7
Vårdproduktion.....	8
In- och utskrivningsmönster	10
Specialistvården	15
Integration av primärvårds- och specialistsjukvårdsdata	17
Processen.....	17
Visualisering av vårdkedjor i Västerbotten.....	17
Visualisering av vårdkedjor i Kust-Österbotten	21
Koppling till geografisk information: Sverige	23
Frågor som ej ännu är besvarade utifrån givet data	25
Frågor som är obesvarade oavsett data.....	25
Utmaningar i datatillgång och bearbetning.....	25
Gemensamma utmaningar	25
Sverigefallet	26
Finlandsfallet	27
Diskussion & Slutsats	28

Abstract

The aim of this report is to shed light on the current access to healthcare data within regions of study in Sweden and Finland. We list the sources of the data at the levels of primary and specialist healthcare. We illustrate what possible questions of relevance to the SOS 1.0 project can be answered with help of these data sources. Further, we discuss integration of the two data sources and present the possibilities that it opens in the analysis of the healthcare chains for individual patients. Finally, by adding information from external data source we present how information about patients post code can be used to find and analyze patients travel patterns which provides us with further input on the quality of care offered by health providers.

At each level, we utilize the information from Swedish and Finish healthcare agencies to visualize the similarities and differences between the two systems. We are also clearly distinguishing what conclusions that can and cannot be drawn based on the dataset analyzed.

Bakgrund och syfte

Målet med projektet i stort är att

”skapa ett proof-of-concept som visar att det är möjligt att simulera och optimera gränsöverskridande vårdkedjor och lägga grunden för att etablera metoden som ett stöd till en mer processbaserad hälso- och sjukvård.”

Simulering är baserat på exekveringar av matematiska modeller och en av de viktigaste delarna i en modell är kvalitén på det data man använder. Därav är det av högsta vikt att man inventerar det datamaterial som finns och visar på hur man kan bearbeta och iordningställa detta data för att fungera till en simuleringsmodell.

Då modellen och tillhörande data ska sträcka sig över gränserna är det också av intresse och vikt att se vad det ursprungsdata kan säga oss om en patients vårdresa och hur den beror på egenskaper som diagnos, kön och demografi samt hemvist i syfte att kartlägga resmönster och enheters beläggning.

I denna studie har datakällor kartlagts. Ett urval av dessa datakällor har använts för att bygga analyser som ger en bild av verkligheten utifrån frågeställningar insamlade i projektet. Syftet med detta är framförallt att visa vad datamaterialet skulle kunna säga och inte i sig att göra analyser för att besvara faktiska sakfrågor.

Datamaterial – Källa och innehåll

Vårdssystemet är en komplex verksamhet med många instanser som tillsammans ska behandla en patient på bästa sätt. För att kunna samverka effektivt och tillsammans bli mer än de enskilda delarna tillsammans krävs en väl utvecklad kommunikation av information. Detta utbyte av information har under lång tid stegvist effektiviserats av datasystem som ersatt de analoga kommunikationsvägar och pappersjournaler som förut var standard. Då behandlingar också blir mer och mer avancerade finns ständigt ett ökat behov av att följa en stor mängd information om patienten, både ur medicinska men också logistiska perspektiv. Denna information är väldigt utspridd och fragmenterad i olika system, både på grund av utveckling där nya system hela tiden tillkommer men också exempelvis i den diversifierade flora av leverantörer av medicinteknisk utrustning som ofta har egna system vilka är kopplade mot applikationen men inte nödvändigtvis kommunicerar på ett standardiserat sätt med andra system. Det finns ett antal system som försöker att bli det heltäckande vårdinformationssystemet för att komma bort från majoriteten av detta men dessa system är trots den aggressiva marknadsföringen inte så heltäckande som de vill göra gällande ifall kunden inte helt går in i leverantörens ekosystem och en så kallad lock-in sker. I projektets kvalitativa del, se separat rapport, så har det dock tydligt framkommit att detta är en av vårdens stora utmaningar då mängden system börjar göra det komplicerat att arbeta samt att administration tar mer och mer av vårdpersonalens tid.

I projektet SOS 1.0 har en inventering skett av vårdinformationssystem använda av de tillhörande regionerna Region Västerbotten, Vasa Sjukvårdsdistrikt samt Närpes stad. Denna inventering har framförallt skett utifrån projektets utgångspunkt att ta fram en kravspecifikation för ett modelleringsbaserat beslutsstödsystem samt att implementera en proof-of-concept på detta. Denna inventering är därför en riktad inventering med tonvikt på system som kan bidra till detta.

De system som är mest lämpliga att bidra till projektets målsättning är patientinformationssystemet där de logistiska delarna som samlar övergripande information om patienters vårdvistelser samt journalsystemen använts i aggregerad form.

Patientdatasystemen samlar in data som kan vara av intresse för myndigheter som t.ex. Socialstyrelsen i Sverige. Här framgår bland annat patientens grundläggande data som kön, ålder, hemkommun men också övergripande information om inskrivningen som vilka datum in/utskrivning skett, vilken huvudsaklig diagnos som sattes samt vilken typ av behandling som patienten genomgick. Kravbilden för denna typ av data är liknande mellan länderna och då detta är aggregerade system (eg. de samlar data från underliggande system) så är datamängden från detta system väldigt lik länderna mellan. All data som projektet bearbetat har varit pseudoanonymiserad där namn ej framgått och personnummer varit ersatt med en slumpmässig kod.

Den information som system ger möjliggör en grov bild av patienters resa genom systemet från inskrivning till klinikbyten och sedermera utskrivning. Dock så är denna bild i vissa fall alltför grov då patienter med liknande demografiska förutsättningar och samma diagnos kan ha två vitt skilda behandlingstider vilket antingen är slump eller har underliggande orsaker varav det sista är det mest sannolika. För att ta reda på detta måste man studera patienternas journaler vilka till stor del består av fri text skriven på olika sätt. Texten är idag bättre än historiskt då läkare idag måste skriva så att patienten kan läsa och förstå sin journal men det är fortfarande svårt att därifrån maskinellt extrahera egenskaper som kan besvara ovanstående fråga. Här finns en enorm utvecklingspotential inom informationshantering så att datainmatningen blir mer kvalitetssäkrad och också enklare att föra in samt mer maskinellt läsbar för djupare patientanalyser. Visserligen kan man resonera att det idag finns system med artificiell intelligens som kan läsa fritext men de är ännu inte på den nivå att de kan användas på ett effektivt sätt, detta är dock något att undersöka i framtida studier.

I kommande kapitel där visualiseringar visar på möjligheter i det data som analyserats inom ramen för detta projekt härrör från datamaterial från patientdataregister under perioden 2016–2018. Datamaterialets struktur i sin helhet kan ses i bilaga A.

Datamaterialet är hämtat från två specialistsjukhus (Lycksele samt Vasa) och från två Kliniker i Primärvården (Storuman sjukstuga och Närpes HVC sjukhus).

Tabell 1: Beskrivning av dataset		
Klinik	Datapunkter	Typ
Närpes HVC	1 973	Primärvårdsklinik
Storuman Sjukstuga	1 000	Primärvårdsklinik
Vasa Centralsjukhus	23 568	Specialistvårdsklinik
Lycksele Lasarett	22 164	Specialistvårdsklinik

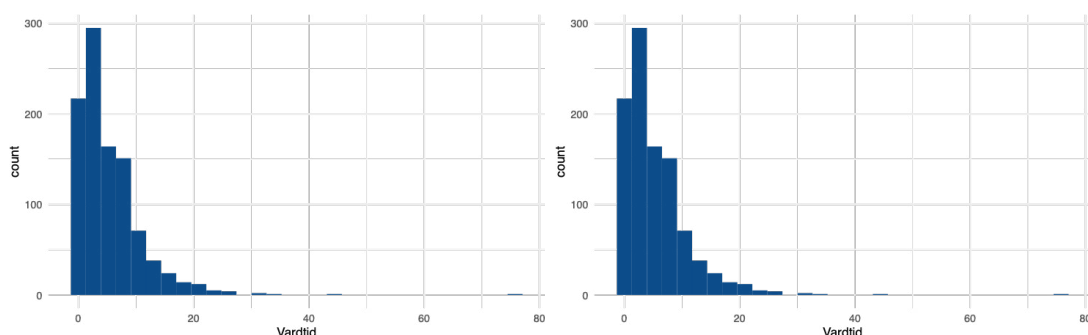
Resultat – Möjliga analyser

Detta avsnitt kommer att behandla de möjliga analyser och frågeställningar som datamaterialet skulle kunna besvara. Då detta är ett proof-of-concept så ges retoriska frågeställningar och observationer men inga slutsatser eller djupare analyser då detta i många fall kräver kompletterande data och en djupare analys i varje fall, vilket inte varit i linje med målet för detta projekt. De visualiseringar som följer är gjorda datadrivet från materialet och alltså inte, ifall det inte explicit beskrivs, manuellt justerat från det ursprungliga datamaterialet.

Primärvården

Vårdtid och demografi

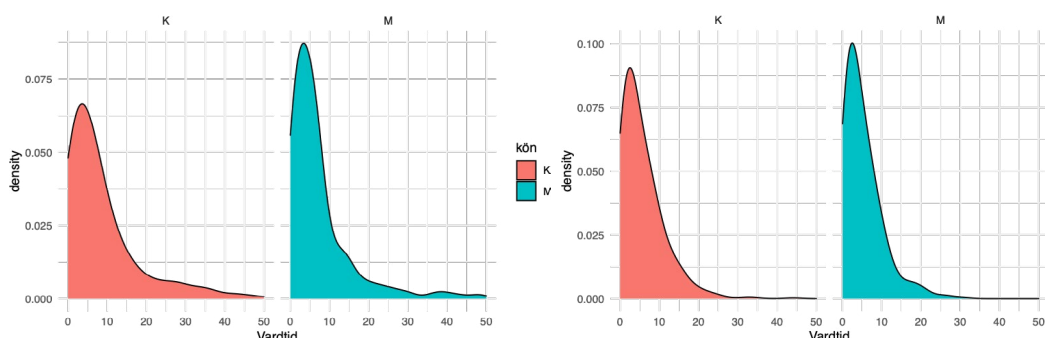
En av de stora frågorna som kom upp till ytan tidigt var tiden patienter spenderar inom primärvården och vårdtid relaterat till klinik och olika demografiska egenskaper.



Figur 1 - Jmf. Vårdtid i dagar. Storuman (Vä) och Närpes (Hö)

Från figur 1 kan det ses att vårdtiden i Närpes har ett fåtal patienter som stannar väldigt länge, vilket inte kan ses i Storuman. Från detta kan man se att försiktighet får iakttas när man jämför medelvårdtider. Ytterligare kan man observera att Närpes har en högre andel patienter som stannar mindre än en natt. Detta är till stor del patienter som kommer in för ett antibiotika-dropp, de skrivs in men tar inte upp en vårdplats.

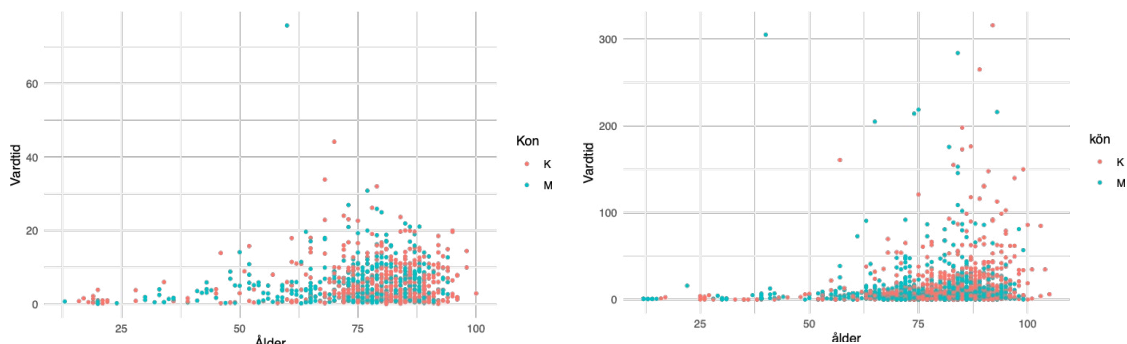
Vårdtiden i sig är en indikator som är användbar som en benchmark men säger inte mycket om vårdepisoden på denna högt aggregerade nivå. Människor och vård är otroligt komplext och det finns otaliga egenskaper som påverkar att vårdtiden blir just som den är. Då vårdtiden utan någon som helst urskiljning inte ger så mycket kunskap görs försök att hitta mönster med demografiska egenskaper och börjar med kön. Detta är fortfarande på en väldigt aggregerad nivå men kan ge en känsla för skillnaden i kön, ifall man vill ha en ännu mer rättvisande bild kan bilden förfinas till jämförbara patientkategorier.



Figur 2 - Jmf. Vårdtid i dagar med hänseende på kön. Storuman (Vä) och Närpes (Hö)

I figur 2 kan det ses en likhet mellan länderna men en olikhet mellan könen där vårdtiden är mer koncentrerad till kortare vårdtider för män och något längre för kvinnor.

Ifall ålder läggs till som en ytterligare dimension kan det kanske ge mer information kring ifall dessa demografiska faktorer påverkar vårdtiden.



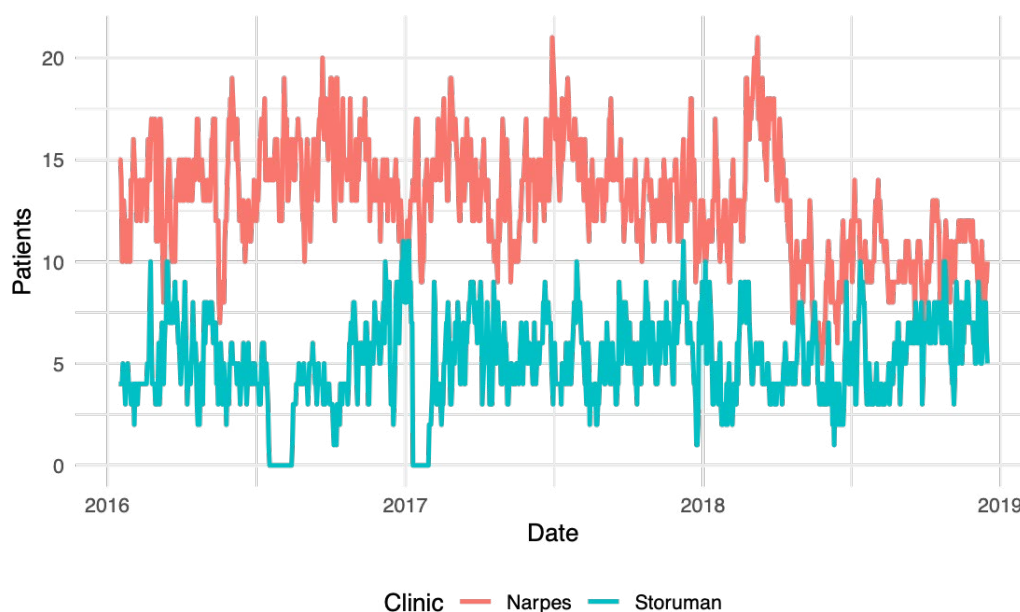
Figur 3 - Jmf. Vårdtid i dagar med hänseende på kön (färgkodat) och ålder. Storuman (Vä) och Närpes (Hö)

I figur 3 ses en tydlig skillnad i vårdtid för olika åldersgrupper men inte en lika framträdande könsskillnad i jämförelse med ålder. Dock så ser det rent visuellt ut som att äldre kvinnor vårdas något längre på båda enheterna.

Vårdproduktion

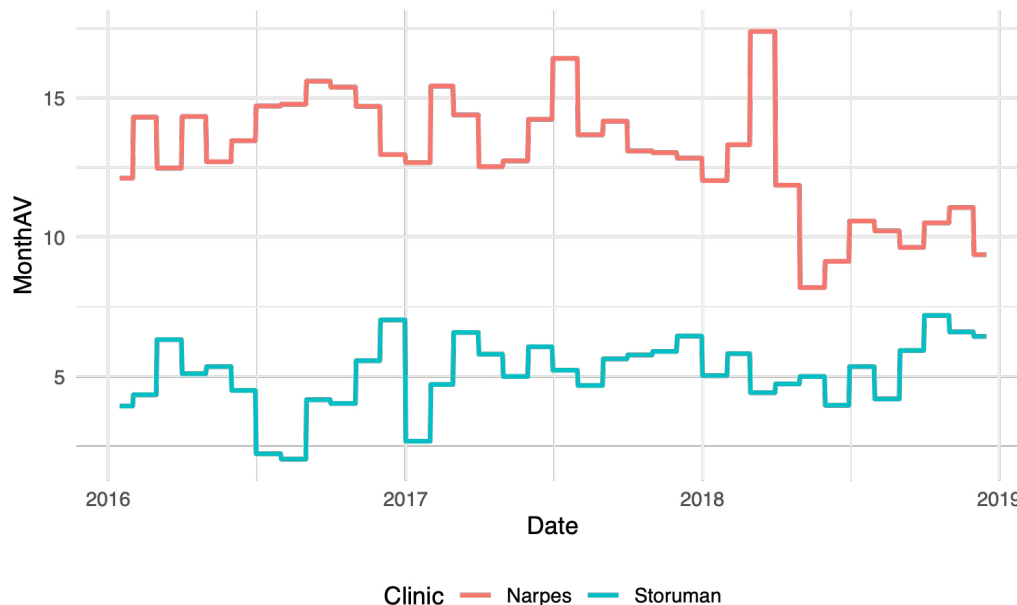
I detta avsnitt lämnas patientens perspektiv och vårdproduktionen under tidsperioden undersöks. Här har korrigeringar skett då de första två veckorna av datamaterialet har exkluderats från visualiseringarna då urvalet av datamaterialet annars felaktigt skulle indikerat att vårdenheterna var tomma eller väldigt lågt belagda under denna period. Då datamaterialet begränsar sig till slutenvården är det denna vårdproduktion som åsyftas nedan.

Vårdproduktionen för de bägge enheterna kan ses nedan i figur 4 där datamaterialet utan ytterligare aggregering visas som vårdproduktionen för varje dag.



Figur 4 – Vårdproduktion 2016-2018

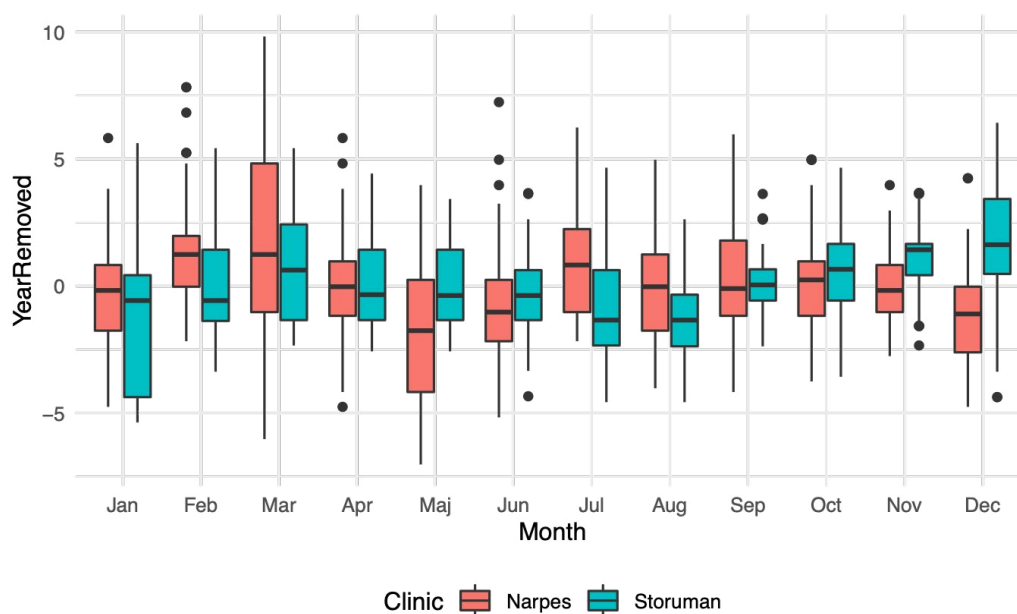
Då figur 4 visar data per dag så kan det lätt bli rörigt ifall läsaren vill se det på en större skala som i detta fall över tre år. Detta kan åtgärdas genom att t.ex. gruppera data i månader som nedan i figur 5. Viktigt dock att notera att varje förenkling medför förlust av upplösning, i nedanstående fall förloras variationen i dagar.



Figur 5 – Vårdproduktion 2016-2018, grupperat per månad

Utifrån figur 5 kan man notera en spik på våren 2018 i Narpes samt därefter en lägre produktion på grund av ombyggnationer vilket resulterade i en minskning med 5 tillgängliga vårdplatser. samtidigt som man kan notera en nedgång sommaren 2016 samt januari 2017 i Storuman.

En möjlig fråga att belysa är huruvida månader eller veckodagar skiljer sig åt i vårdproduktion och detta visualiseras nedan i figur 6 samt figur 7.

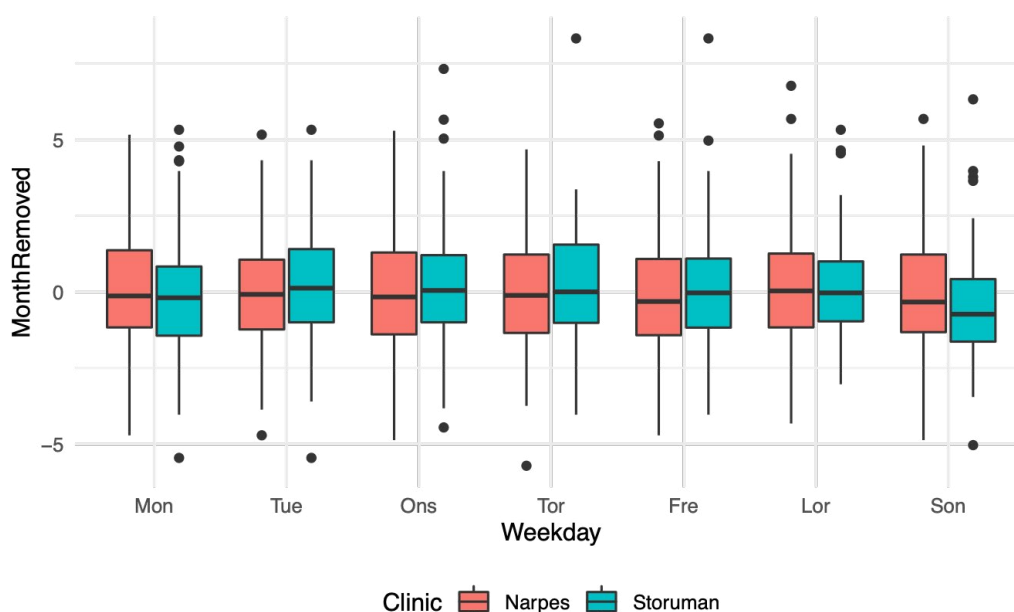


Figur 6 – Vårdproduktion 2016-2018, månadsvariation med årseffekt borttagen

I figur 6 ovan ses en ny typ av figur där de färgade blocken innehåller 50 % av datamaterialet och där den tjockare linjen i blocken representerar medianen (mättet där hälften av datapunkterna faller ovan respektive under). Denna figur som kallas för en box-plot kan användas för att visualisera ett datas egenskaper som medianvärde och spridning. I denna figur är årsvariationer bortplockade då dessa skulle ge missvisande resultat.

Ur ovanstående kan man se att vårdproduktionen håller sig relativt jämn över månaderna men med en förhöjd variation i början på året med januari i Storuman och mars i Närpes som sticker ut. Även maj i Närpes har förhöjd variation.

Gällande frågan ifall veckodagar påverkar vårdproduktionen undersöks i figur 7 nedan då även månadsvariationerna tagits bort för att isolera och visualisera veckodagsvariationerna.



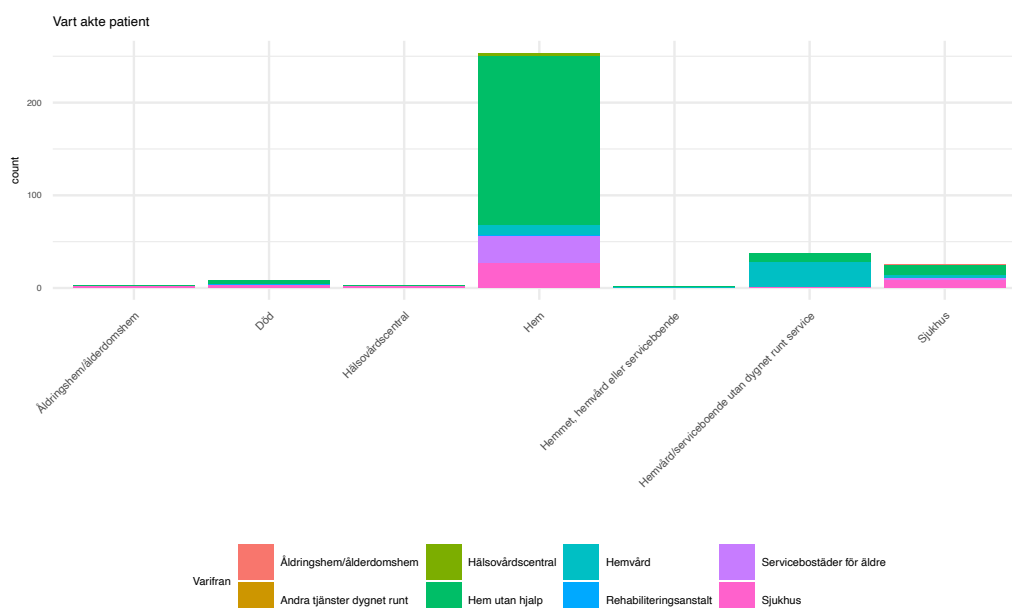
Figur 7 – Vårdproduktion 2016-2018, veckodagsvariation med årseffekt och månadseffekt borttagen

Utifrån det figur 7 visar så ses inga betydande variationer beroende på veckodag på enheterna. Eventuellt att en något lägre produktion kan skönjas på söndagen i Storuman men skillnaden är väldigt liten.

Vårdproduktionsfigurer är ett viktigt verktyg i att kunna kapacitetsplanera, både på en större skala med anställningar men också på en mindre skala med t.ex. hur stor helgbemanningen ska vara. Att se på detta datadrivet kan visa på mönster och egenskaper som man inte tänkt på. Denna typ av figurer går även att förfinas med t.ex. att bygga ett vårdtyngdsmått som då beskriver veckovariationer av vårdtyngd istället för bara antal.

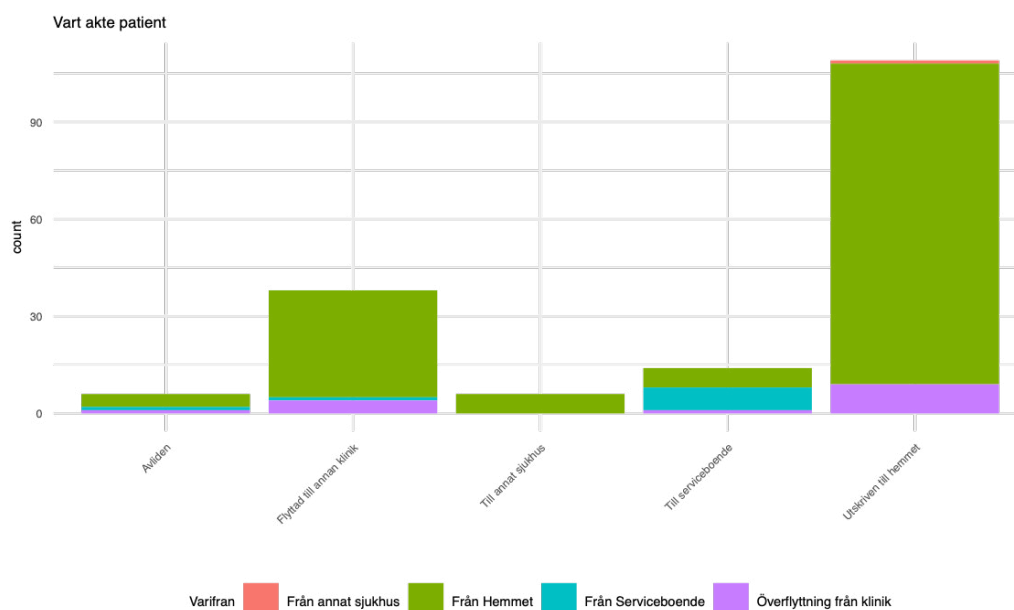
In- och utskrivningsmönster

En av frågorna som har dykt upp under projektet är hur patientflödenas struktur har sett ut i primärvården. Med särskilt fokus på patienter som stannar kort tid (maximalt 1 dag) samt de med längre liggtider (över 1 dag). Nedan är detta visualiserat genom staplar och färger med de koder för in- och utskrivning som finns i datamaterialet.



Figur 8 – Inskrivningar (Färgkod) och Utskrivningar (Staplar) Närpes HVC för korta besök

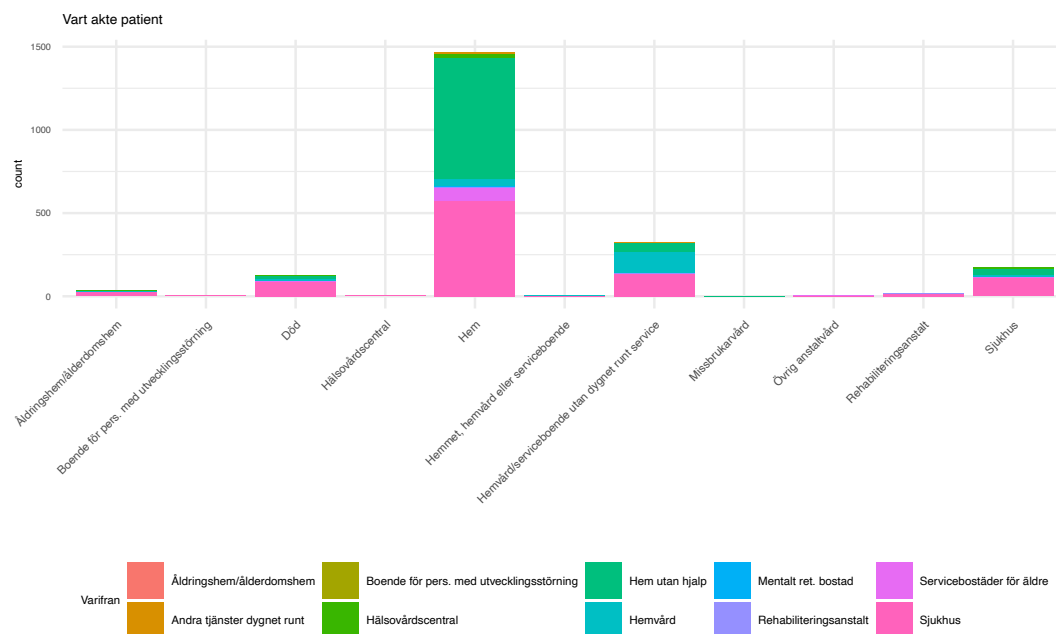
I figur 8 kan ses att de flesta av patienterna som stannar kortare än en dag kommer hemifrån utan hjälpinsatser och skrivs även ut hem utan hjälp. En mindre del av de som skrivs hem utan hjälp på under en dag har kommit från sjukhus eller servicebostad. Ett fåtal av patienterna skrivs åter till sjukhus inom en dag från ankomst från sjukhus.



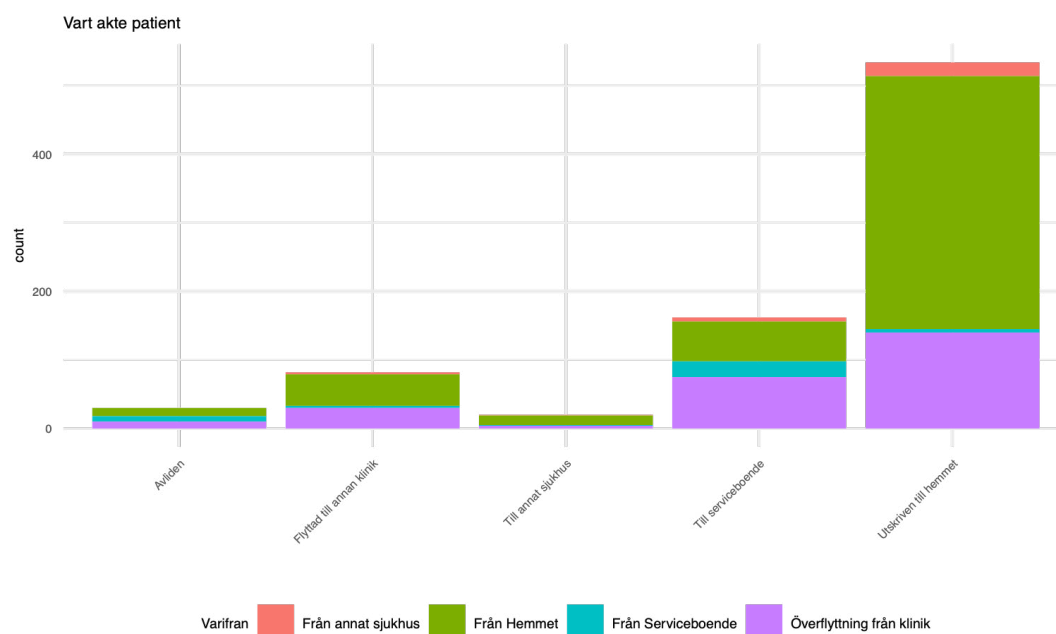
Figur 9 – Inskrivningar (Färgkod) och Utskrivningar (Staplar) Storuman Sjukstuga för korta besök

I figur 9 ses att de allra flesta skrivs hem alternativt till annan klinik. Vi kan se att de flesta patienter vilka kommer till sjukstugan för korta besök kommer från hemmet och endast en mindre del från andra kliniker i regionen.

För att kunna göra ovanstående rättvisa så visas här nedan i figur 10 och figur 11 fördelningen av alla patienter oavsett vårdtid för jämförelse.



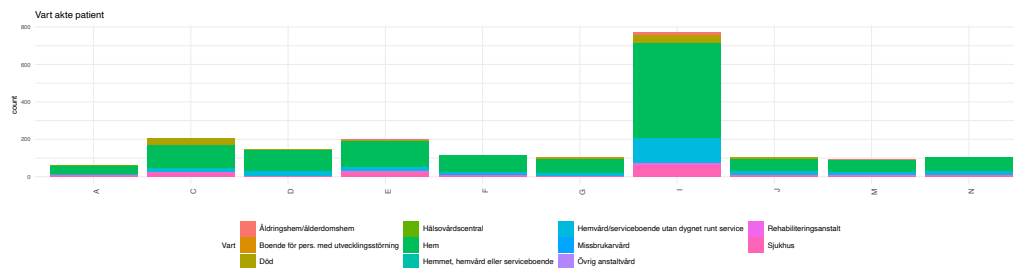
Figur 10 – Inskrivningar (Färgkod) och Utskrivningar (Staplar) Närpes HVC



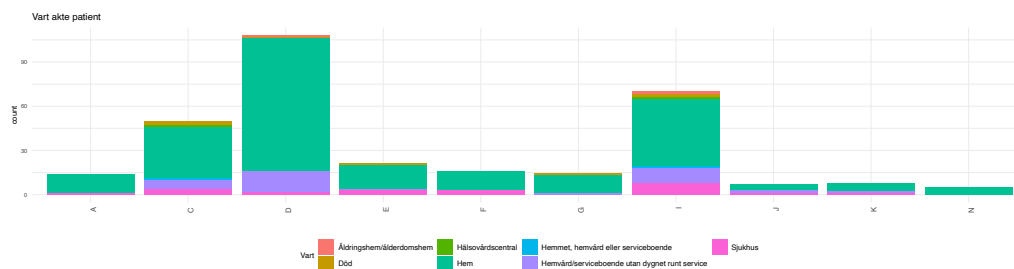
Figur 11 – Inskrivningar (Färgkod) och Utskrivningar (Staplar) Storuman Sjukstuga

Av vikt för att förstå ovan så kan diagnoskategori visualiseras mot utskrivningskod för att se ifall det finns korrelation mellan diagnos och utskrivning. Diagnoskategorierna är enligt ICD-101 begränsat till kategoribokstaven. De tre största i Närpes är C (Maligna tumörer), D (Benigna eller okända tumörer), M (Skelett) samt I (Hjärta-Kärl).

1 ICD-10-SE, <http://icd.internetmedicin.se/>

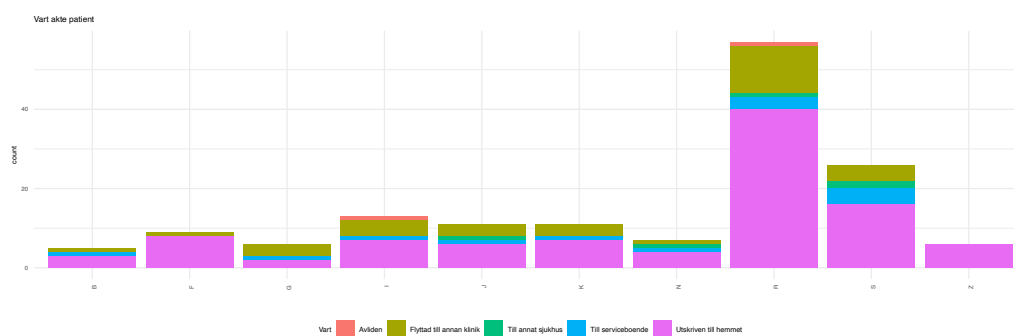


Figur 12 – Utskrivningar (Färgkod) mot diagnoskategori Närpes HVC för korta besök (10 största diagnoskategorierna)

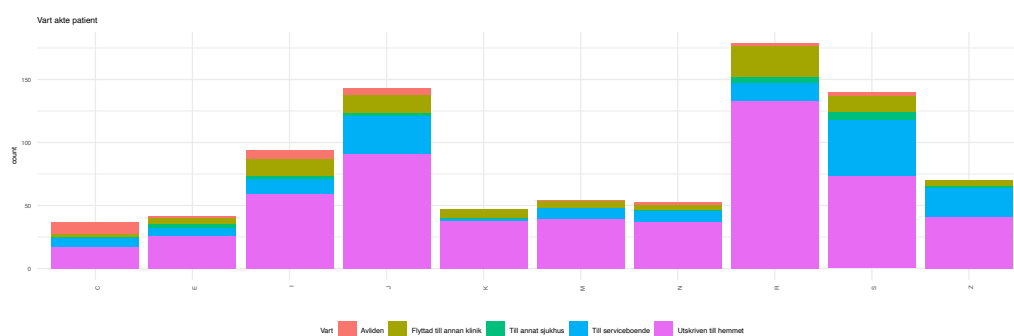


Figur 13 – Utskrivningar (Färgkod) mot diagnoskategori Närpes HVC (10 största diagnoskategorierna)

Och i figur 14 och 15 nedan detsamma för Storuman sjukstuga. Här är de stora kategorierna I (Hjärta Kärl), J (Luftvägar), R (Symptomdiagnoser), S (Trauma).

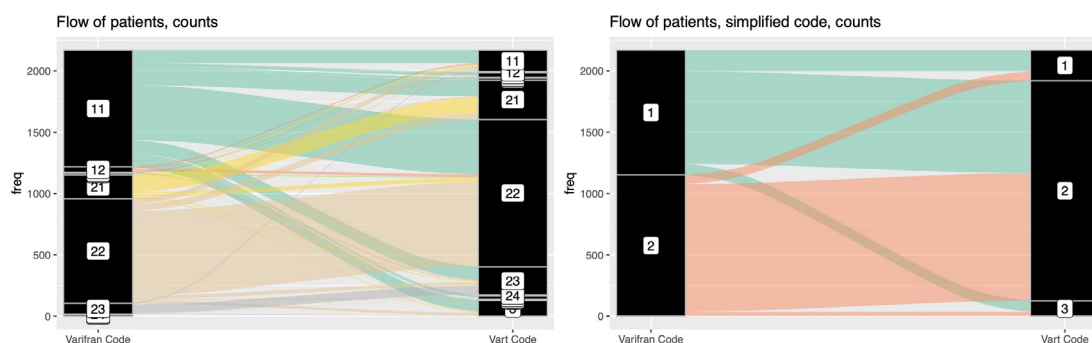


Figur 14 –Utskrivningar (Färgkod) mot diagnoskategori Storuman Sjukstuga för korta besök

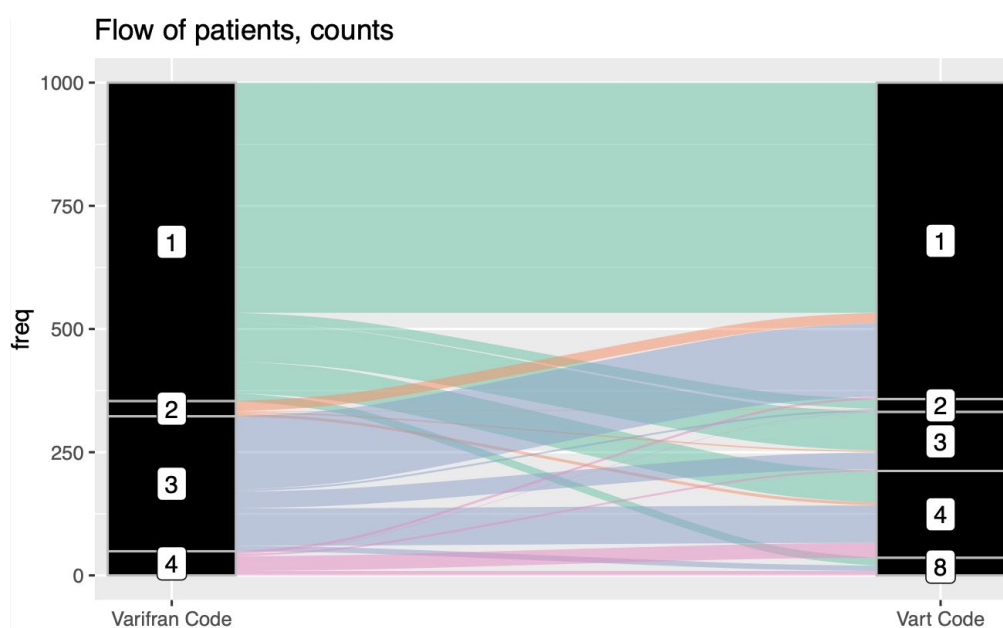


Figur 15 –Utskrivningar (Färgkod) mot diagnoskategori Storuman Sjukstuga

Ett annat sätt att visualisera dessa patienters resa, eller snarare fördelningen av patienter som rör sig i systemet kan visualiseras med ett s.k. Sankey diagram där flödet av patienter ses mellan olika in- och utskrivningskoder. I Närpes fall presenteras både det fullständiga och förenklade flödet då den första siffran i Finlands system ger huvudkategorin vilket endast registreras ibland istället för den fördjupade tvåsiffriga koden. Här är koden 1 och dess underkoder anstaltsvård (Slutenvård) medan koden 2 betyder hemförlovning eller vård i hemmamiljö, kod 3 betyder att patienten avlidit. I det svenska datamaterialet är det tvärtom där 1 är hemgång och 2- 4 är behandling inom sjukvården. Motsvarande kod för att patienten avlidit är 8.



Figur 16 – Flöde mellan ut- och inskrivningskoder Närpes HVC. Förenklat flöde till höger.



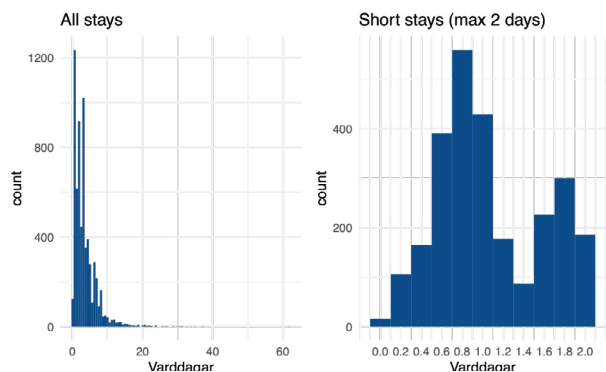
Figur 17 – Flöde mellan ut- och inskrivningskoder Storuman Sjukstuga.

Dessa figurer är användbara för att följa patientflöden och få en intuitiv känsla för vilka patienter som kommer in och färdas vidare till olika enheter och var majoriteten av enhetens inskrivningar och utskrivningar går till. Detta kan bland annat användas för att prioritera resurstilldelning eller vilka processer som ger mest att förbättra.

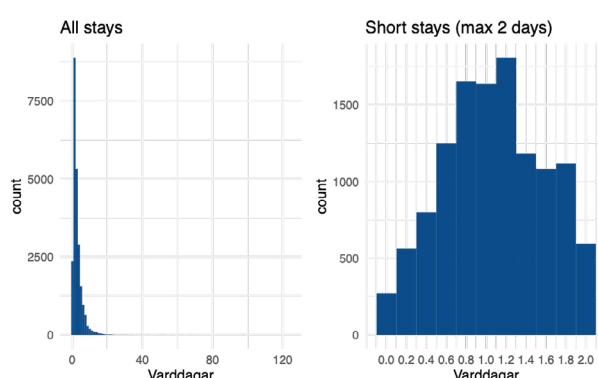
Specialistvården

Analyserna i specialistvården utgår från datamängder från Vasa Centralsjukhus (Avdelningar T1, T2, A2) och Lycksele lasarett (Kirurgkliniken och Medicinkliniken). Data har använts både från det allmänna patientunderlaget samt inriktat på de kirurgiska samt ortopediska kliniker och i de mest specifika fall på kategorierna M16, M17 samt S72.

Vårdtiderna inom specialistvården som visas i figur 18 och figur 19 är fördelade med majoriteten korta besök och en svans av något längre besök. Vid sidan av dessa har en visualisering gjorts av fördelningen hos de patienter som stannar maximalt två dagar.

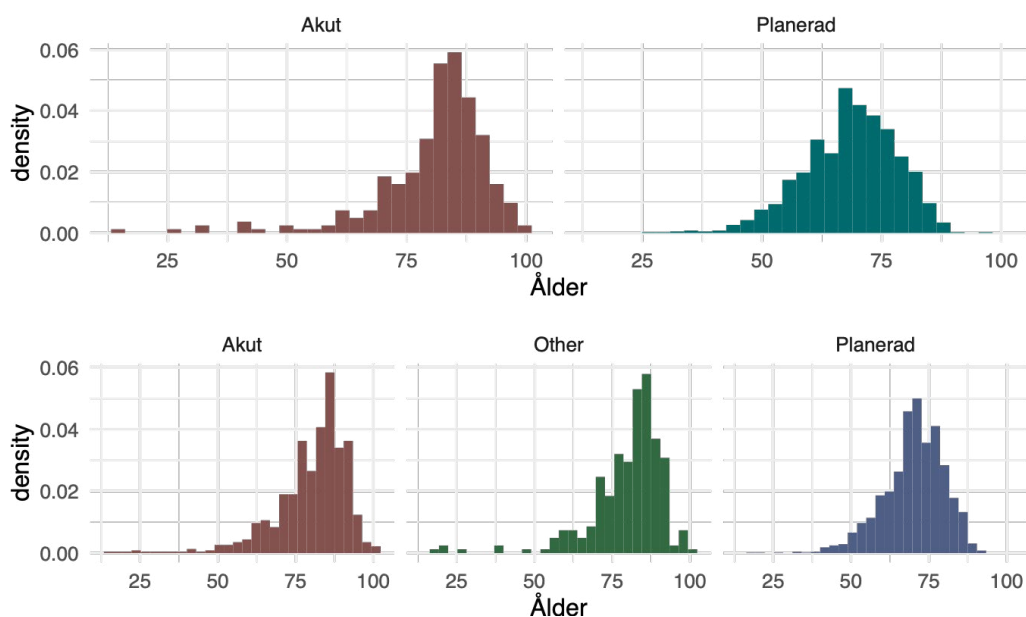


Figur 18 – Vårdtid per inskrivningstillfälle Lycksele



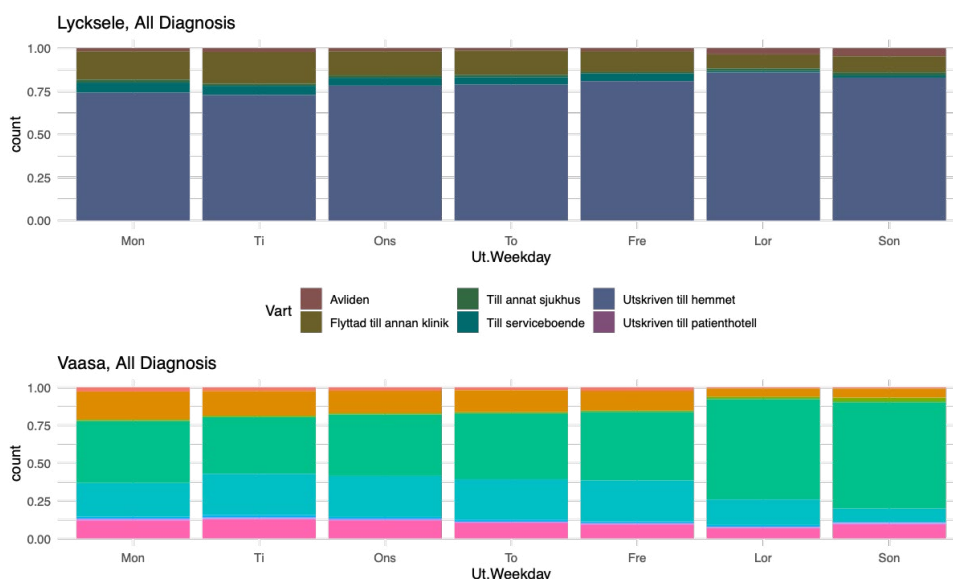
Figur 19 – Vårdtid per inskrivningstillfälle Vasa

Utöver vårdtid så visualiseras också patienternas demografi mot ankomsttyp i figur 20 för att se åldersfördelningen i patientpopulationen. Ankomsttypen tas med då dessa två populationer skiljer sig åt markant i många fall. Akut är de som akut skrivits in på en avdelning medan planerad är planerade ingrepp där patienten kommit vid en på förhand bestämd tid.



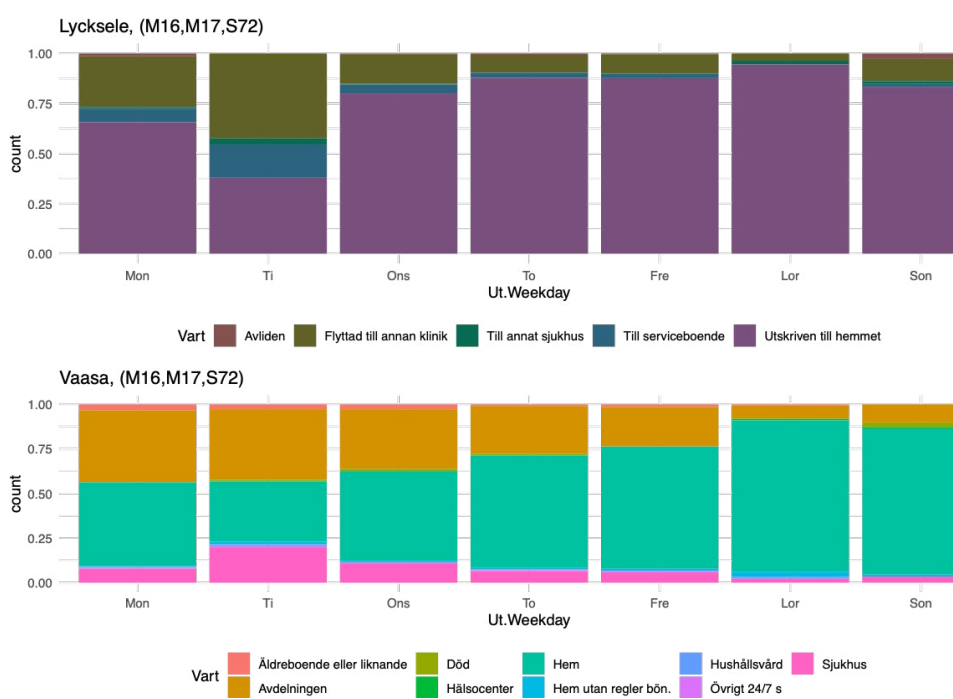
Figur 20 – Fördelning av patienter efter ålder och ankomsttyp, Lycksele ovan och Vasa nedan.

Analogt med fallet för primärvården kan också här patientflöden visualiseras. Här har inte frågat om de som stannar korta eller längre perioder dykt upp utan snarare avsmalning med hänseende till de utvalda diagnoserna M16, M17 samt S72 samt vilken effekt utskrivningsdagen har på utskrivningsplatsen.



Figur 21 – Fördelning av patienters utskrivningsplats mot veckodag

Från figur 21 kan utläsas att för båda inrättningarna ökar andelen utskrivning till hemmet med veckan. Liknande mönster syns i de utvalda patientgrupperna i figur 22 men där tisdagen är den dag där färst människor skrivs till hemmet.



Figur 22 – Fördelning av patienters utskrivningsplats mot veckodag för utvalda diagnoser

Denna information är ofta självklar för de som dagligen arbetar med produktionen men kan vara nyttig för de som jobbar med operativ planering eller chefer på högre nivåer.

Integration av primärvårds- och specialistsjukvårdsdata

Eftersom all datapunkt är utrustad med en unik identifierare för varje patient, kan en patientresa genom sjukvården förenas. På så sätt kan vi skapa vårdkedjor för patienter som går över gränserna eller primär- och specialistsektorn för hälso- och sjukvård. I detta avsnitt integreras data från en primär och en specialistenhet. I datamängden hittas individer som finns i båda enheterna. Genom att titta på deras in- och ut-koder, diagnos, liksom in- och ut-tider på båda enheterna, kan vi sedan avgöra om deras besök i båda enheterna skapas som en del av samma process.

För att undvika problemet med ofullständig vårdperiod, begränsas studien till de perioder som börjar med ankomsten hemifrån och slutar i retur hem (eller död).

Ytterligare avsmalning görs med att endast fokusera på projektets utvalda diagnoser av intresse (M16, M17, S72). Med detta fokus blir datamängderna väldigt begränsade. Detta begränsar möjligheten att analysera data ur befolkningsperspektivet, men möjliggör individanalyser (analyser av enstaka vårdperioder).

En del av studierna som presenteras nedan kommer att använda den totala datamängden och inte enbart de som passerat både en specifik specialistvårdavdelning samt en specifik primärvårdsavdelning. Anledningen till detta är att dessa datamängder blir väldigt avsmalnade med enbart ett fåtal patienter i förhållande till den totala mängden och därför inte representativ för hela populationen.

Processen

Processen att länka ihop datamängderna grundar sig i det krypterade personnumret som möjliggör att kunna följa patienten genom flera vårdenheter. Två definitioner skapas

Vårdepisod - Period av vård där patienten ej har längre uppehåll i vårdperioden än 30 dagar.

Vårdperiod - Period från att patienten inkommer hemifrån till vården, flyttas mellan kliniker och till slut skrivs hem, alltså en hel period där patienten kontinuerligt är inskriven.

Vårdperioden tas fram genom att det krypterade personnumret matchas mellan kliniker och en automatisk analys av tiden mellan utskrivning och nästa i tiden förekommande inskrivning, är denna tid tillräckligt låg så anses det tillsammans med matchande in- samt utskrivningskoder utgöra bevis för att patienten har flyttats direkt mellan klinikerna.

Utifrån denna kedja kan data tas fram som t.ex. antal dagar som patienten under sin kedja spenderat i specialistvård kontra primärvård samt hur diagnoskodning varierat under patientens resa (något som denna rapports metod inte är beroende av men som skulle förenkla).

Denna analys är mycket mer omfattande i den svenska regionen då datamaterialet för den finska motsvarigheten är mindre och inte lika rikt (hela regionens förflyttningar i Sverige till skillnad från ett flöde i Finland).

Visualisering av vårdkedjor i Västerbotten

I Västerbotten finns specialistvård och primärvård under samma huvudman och därför finns data tillgängligt för somatiska kliniker i Västerbotten men som någon gång varit patienter i Lycksele. Vi följer här 18 895 patienter genom deras vårdresor.

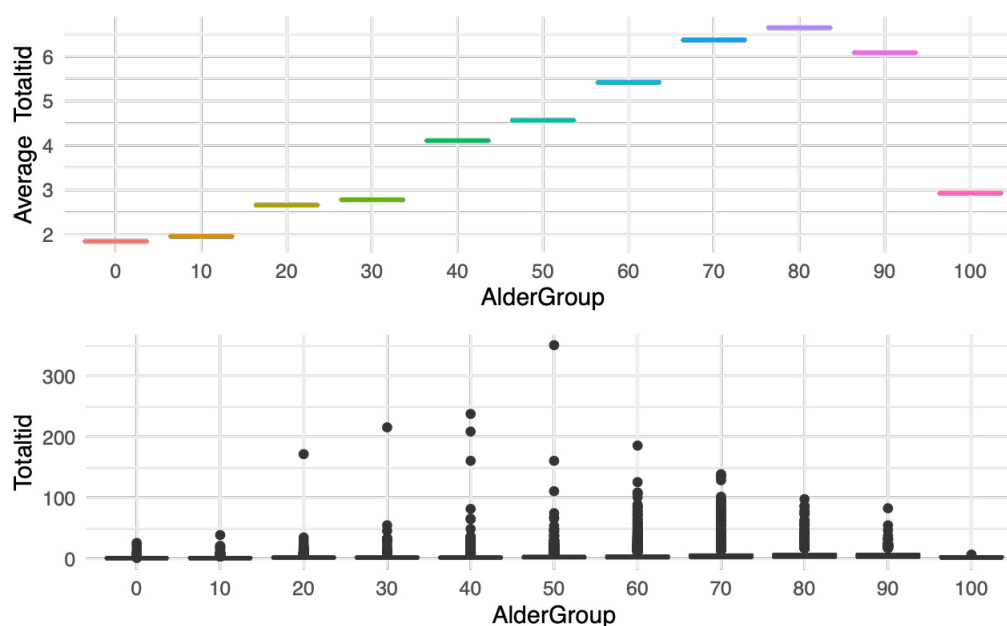
De visualiseringar som presenteras i detta avsnitt refererar till vårdkedjor som beskrivs i förra avsnittet och inte som tidigare enskilda inläggningar.

Datamaterialet är fördelat enligt

Åldersgrupper										
0	10	20	30	40	50	60	70	80	90	100
140	438	1472	1322	943	1738	3256	4723	3904	945	14

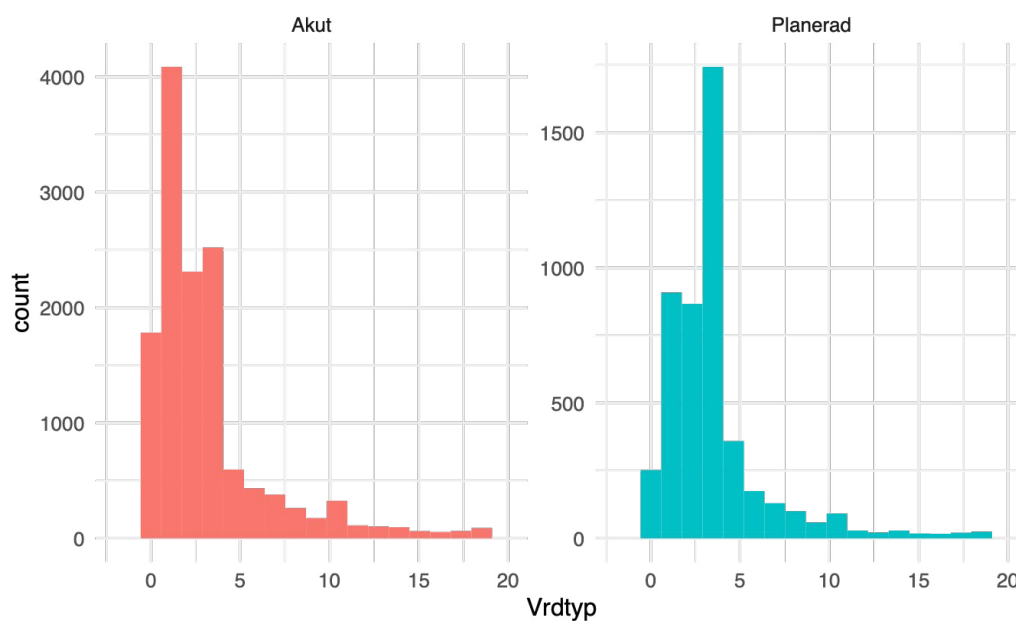
Kön		Vårdtyp		Hemmaregion - Postnummerområde				
K	M	Akut	Planerad	90	91	92	93	Other
10464	8431	13925	4970					

Initialt, för att få ett grepp om populationen så visualiseras patienternas ålder vid initiala inläggandet mot den totala tiden de vårdas i både primär- och specialistvård i en vårdperiod (sammanhängande inläggningar för en patient).



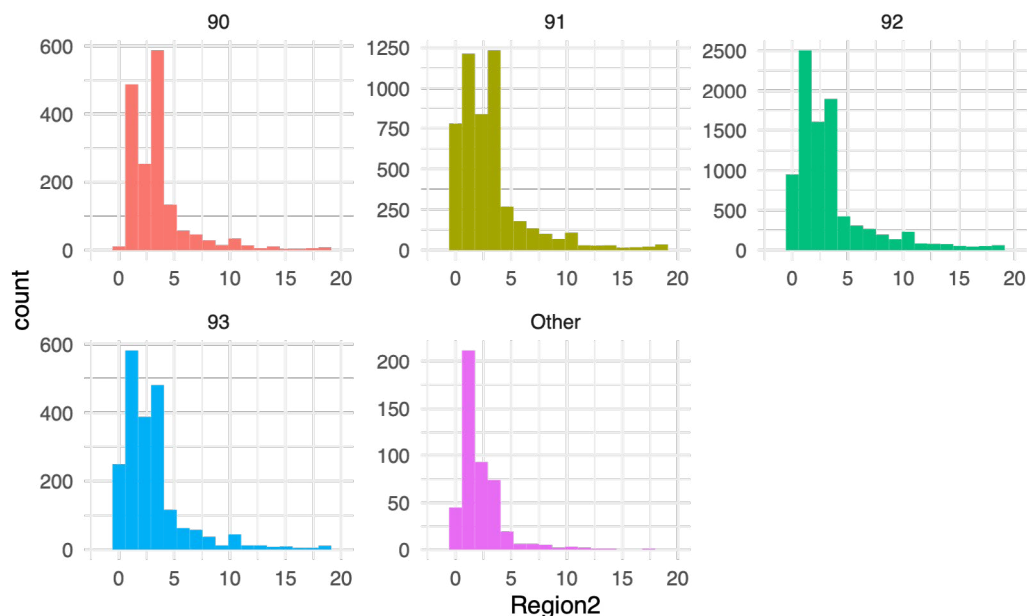
Figur 23 – Vårdtid som funktion av åldersgrupp, medeltal ovan och faktiska nedan

Då det tidigare lades fram att det var skillnad i mönster mellan elektiva patienter med planerade vårdtillfällen och akuta patienter så visualiseras här i figur 24 vårdtiden uppdelad per Inskrivningstyp (Akut/Planerad). I figuren syns bland annat att planerade patienter generellt har en högre inskrivningstid än akuta patienter.



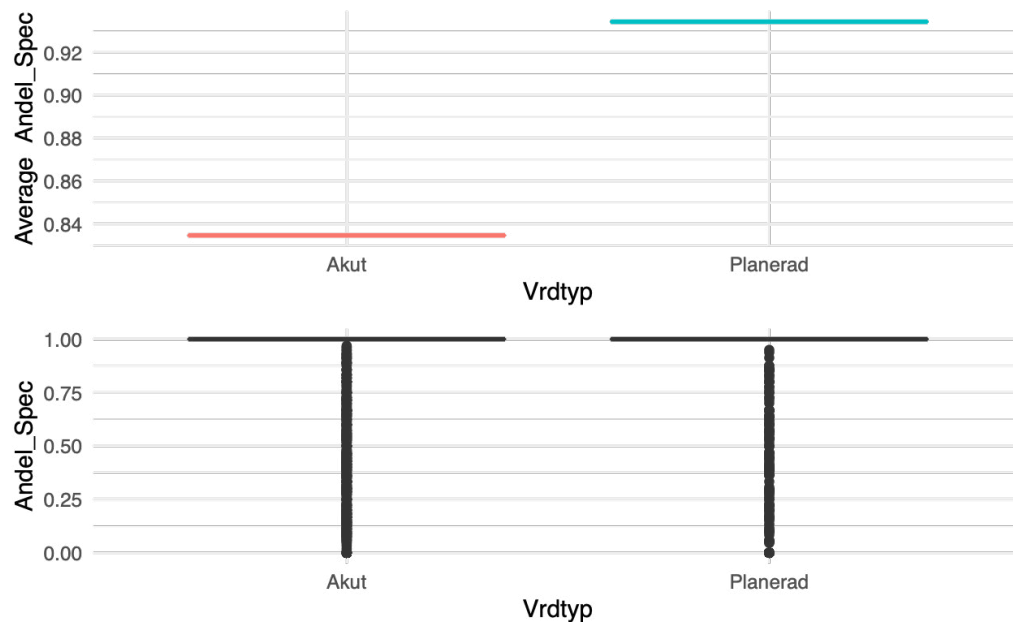
Figur 24 – Vårdtid som funktion av Inskrivningstyp

Det kan också vara intressant att undersöka ifall en patients hemvist påverkar patientens inskrivningstid vilken man då också kan visualisera (se figur 25) med en viss förenkling för att få tillräckligt stora grupper av patienter. Denna gruppering är gjort på de första två siffrorna i postnummerområdet.

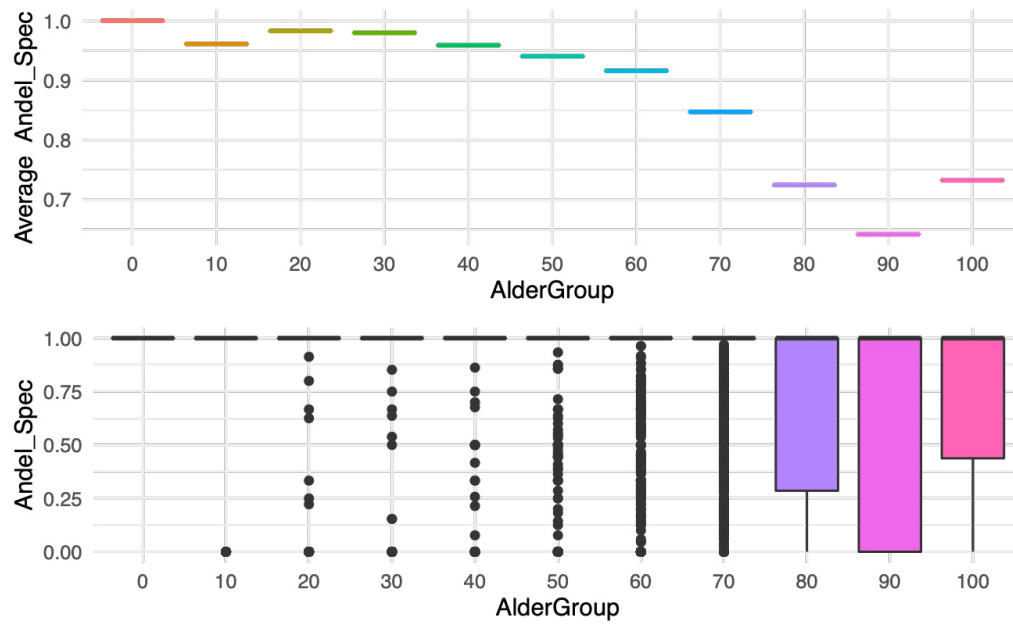


Figur 25 – Vårdtid som funktion av patientens hemvist

Utifrån att datamaterialet har både specialistvård och primärvård kan förhållande mellan de två undersökas för att se ifall det finns mönster i förhållandet av inskrivningstider mot t.ex. ålder eller inskrivningstyp.



Figur 26 – Andelen av vårdtiden spenderad i Specialistvården mot Inskrivningstyp



Figur 27 – Andelen av vårdtiden spenderad i Specialistvården mot åldersgrupp

Visualisering av vårdkedjor i Kust-Österbotten

I Österbotten finns datamaterialet hos två huvudmän då kommunerna ansvarar för primärvården. I datamaterialet för Österbotten finns 17 483 vårdperioder.

Information about the data

```
table(chain$AlderGroup)
```

```
##  
## 10 20 30 40 50 60 70 80 90 100  
## 329 830 987 1322 1951 3350 4394 3390 915 15
```

```
table(chain$Kon)
```

```
##  
## M K  
## 9321 8162
```

```
table(chain$Vrdtyp)
```

```
##  
## A P  
## 9964 7519
```

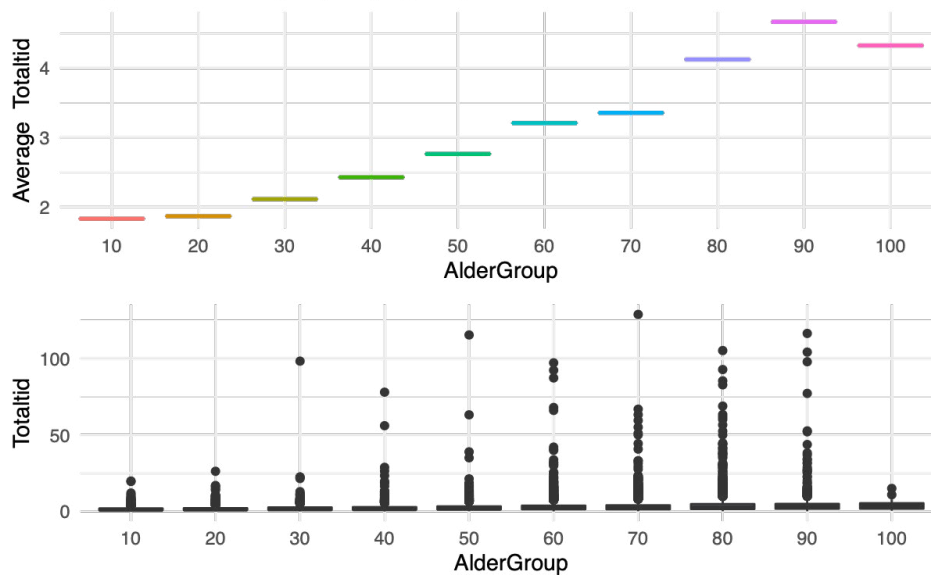
```
table(chain$Region2)
```

```
##  
## 0  
## 17483
```

Figur 28 – Beskrivning av de finska vårdkedjorna

Analogt med det svenska fallet visualiseras total vårdtid i en vårdperiod mot åldersgrupper.

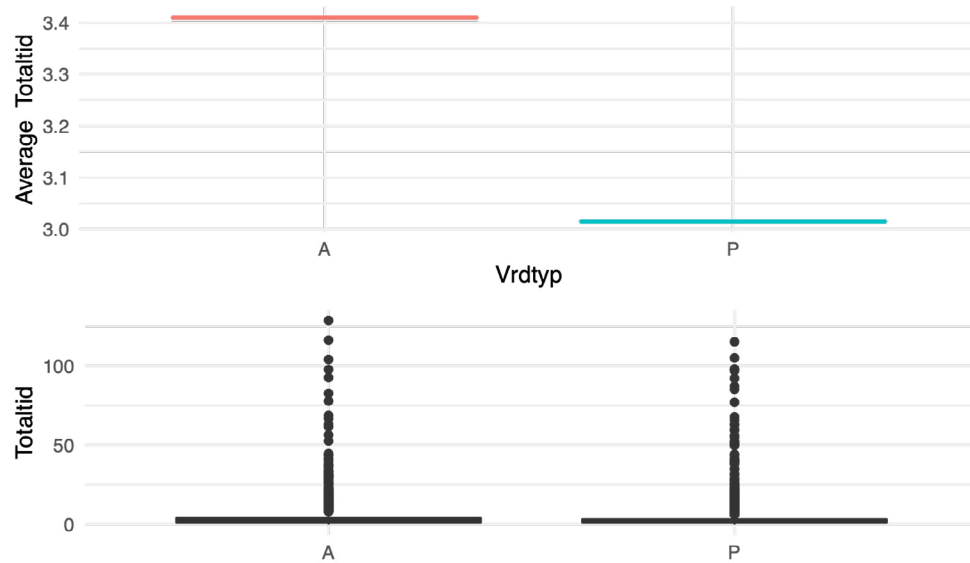
Total lenght of vardkadjä by Age group



Figur 29 – Vårdtid som funktion av åldersgrupp, medeltal ovan och faktiska nedan

Då ankomsttyp/vårdtyp är väldigt särskiljande i datamängden presenteras här också den totala vårdtiden mot vårdtypen (A = Akut besök, P = Planerat besök).

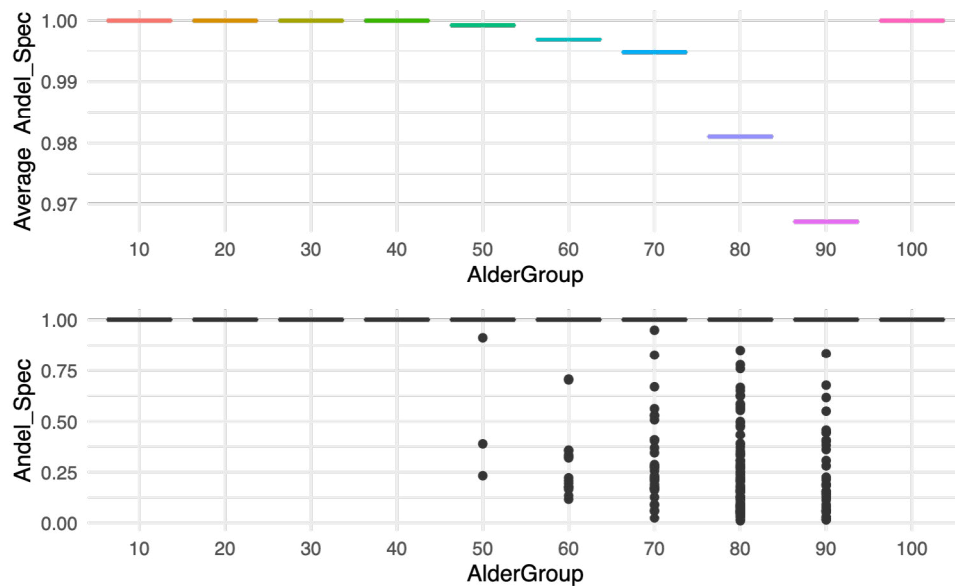
Total lenght of vardkadja by Vardtyp



Figur 30 – Vårdtid som funktion av Inskrivningstyp

Det är också av intresse här att se hur stor andel av vårdperioden som spenderas i specialistvården och här kan man se att den absoluta majoriteten upp till 70 år spenderar sin tid nästan uteslutande i specialistvården medan patienter med ökande ålder har något större andel på primärvårdsenheten.

Proportion of Specialistvard dagar by Age group



Figur 31 – Vårdtid som funktion av patientens hemvist

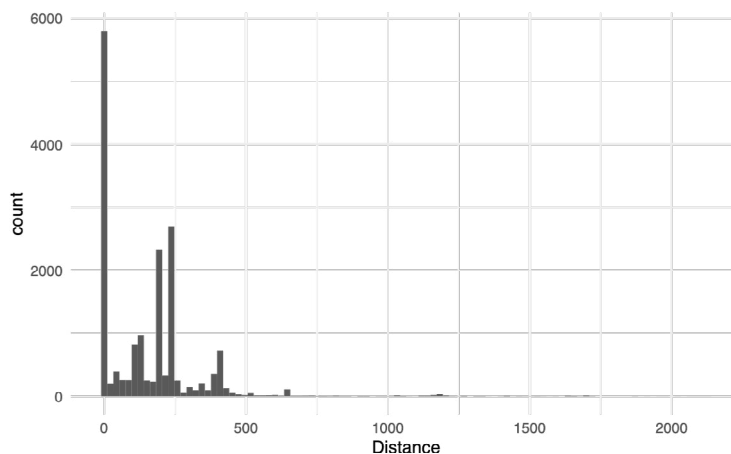
Koppling till geografisk information: Sverige

Den tillgängliga datainformation kan utökas med ytterligare information från externa källor. Ett exempel på detta är att använda informationen om patientpostnummer, tillsammans med geografiska databaser i samband med patientens kedjor. Genom att kartlägga hem, samt primära och specialiserade vårdenheter, kan en proxy till det fysiska avståndet hittas som patienten behöver göra för att få de fullständiga vårdprocesserna. Förhållandet mellan dem och de enskilda faktorerna, som hem kommun, sjukhusvistelse typ eller ålder på resesträckan kan sedan studeras. En sådan aspekt, även om den inte nödvändigtvis är direkt relaterad till resultatet av den sjukvård som tillhandahålls, kan vara en viktig del av patientens erfarenhet.

Nedan har visualiseringar gjorts över patienters tänkta reseavstånd till, inom och från vården. I denna grundläggande analys räknas fågelvägen mellan platserna. Patientens hem är förenklat till den vägda medelpunkten av givet postnummer. Patienter som ej har rest eller de som har reseavstånd över 500 km har tagits bort då det finns externa patienter som enligt ovanstående regler får en osannolikt lång resa (t.ex. skidgäster som skrivits in vid ortopedavdelningen men är hemmahörande i södra delen av Sverige och därför har mycket långt från sitt hempostnummer till Lycksele).

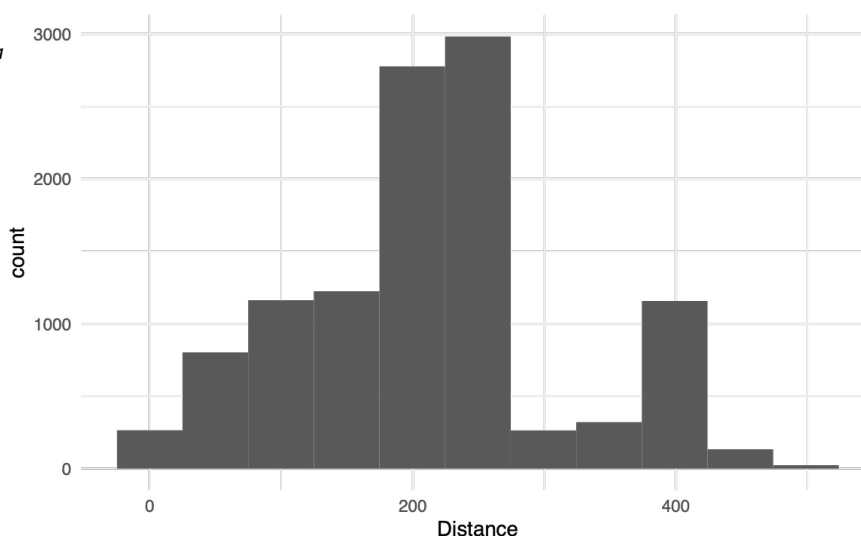
För att se att dessa förenklingar ändå fångar majoriteten av resorna så finns den fullständiga visualiseringen med som jämförelse, se figur 32.

I figur 33 nedan ses resmönstret där majoriteten av de som reser hamnar strax ovan 200km under en vårdperiod.



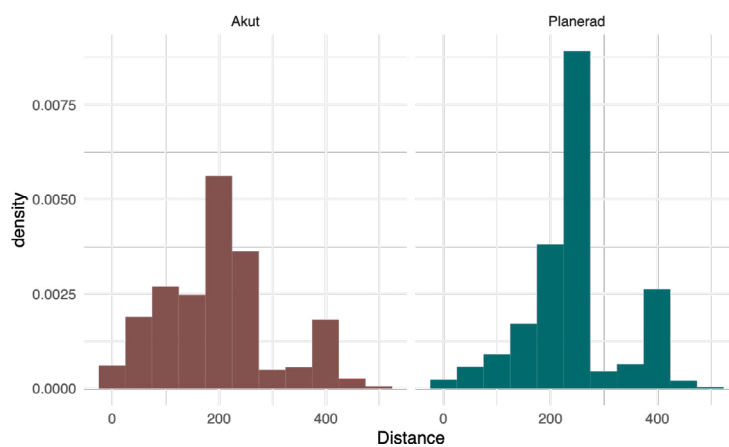
Figur 32 – Fullständig visualisering av restid i kilometer inklusive de som ej reser och långväga

Figur 33 – Visualisering av restider i kilometer med långväga och icke-resande borttaget

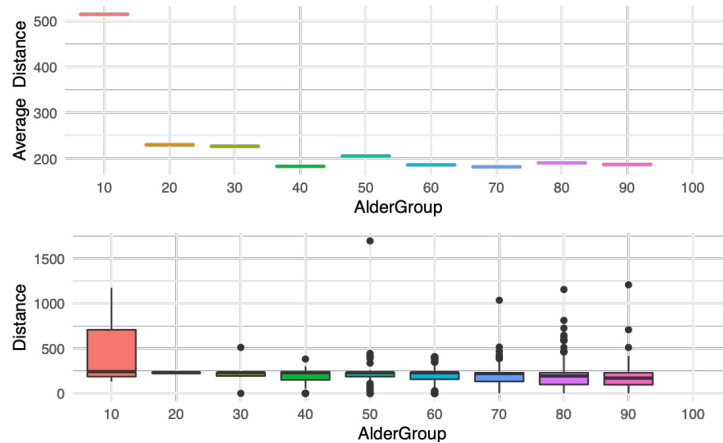


Det kan också visualiseras mot inskrivningstyp (figur 34) samt ålder (figur 35).

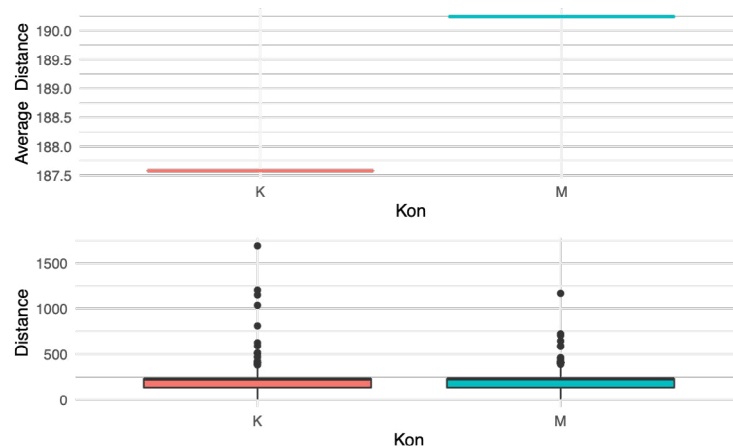
Figur 34 – Visualisering av restider i kilometer uppdelat per vårdtyp



Figur 35 – Visualisering av restider i kilometer uppdelat per åldersgrupp



Figur 35 – Visualisering av restider i kilometer uppdelat per kön



Frågor som ej ännu är besvarade utifrån givet data

Det är viktigt med en medvetenhet om vilken befolkning som analyseras, och vilka slutsatser som dras. Oavsett om det gäller enskilda vårdenheter eller kedjor måste det finnas en tydlig bild av vilka data som används vid analys. Det är osäkert och eventuellt farligt att extrapolera resultat från ett observerat prov till att gälla hela befolkningen.

I det tillgängliga datamaterialet fanns flera inbyggda begränsningar när det gäller ursprung och omfattning av data. På svenska sidan har endast patienter observerats som någon gång tagits in på Lycksele sjukhus eller enheter i dess region. Detta ger dock inte all information om patientflödet i regionen eftersom det finns en stor grupp av patienter som exkluderas, men deras närvaro i det verkliga systemet kan fortfarande påverka mönstren för inkluderade individer.

Situationen är ännu mer angelägen i datamaterialet från Finland där tillgång till information endast är från två enheter. Att här skapa kedjor genom att koppla data resulterar i betydligt färre datapunkter vilket gör det svårare att dra statistiska slutsatser utöver vad som gäller för den observerade populationen. Med detta sagt sådana datamängder fortfarande användas för att hitta mönster som ses i populationen och eventuellt leta efter deras ursprung.

Det skulle vara av högsta intresse att analysera de fullständiga datamängderna från båda regionerna vilket skulle ge oss en bredare och mer fullständig bild av patientflödena då de huvudsakligen rör sig inom regionerna.

Frågor som är obesvarade oavsett data

Det är viktigt att förstå att de observerade data, oavsett vid vilken upplösning, bara är en bild av historien. Med verktyg för modellering och slutledning kan generalisering av resultaten till den generella populationen ske, men detta tillvägagångssätt har också begränsningar. Dynamiken i vårdrutinerna och vårdsystemen samt personalens interna rutiner och preferenser kan och kommer att påverka beslut som återspeglas i patienternas vårdkedjor men vilket är mycket svårt att fånga. Vidare påverkas inflödet av patienter också av flera aspekter av samhället som är i ständig förändring.

Det är därför av högsta vikt att använda modellering som en gång validerats och justerat för observerade data, den kommer då att kunna fånga framtida variation. Inom modelleringsramverket kan flera scenarier av patientens inflöden och systembeteenden modelleras och förse oss med heltäckande bilder av flödena

Utmaningar i datatillgång och bearbetning

Gemensamma utmaningar

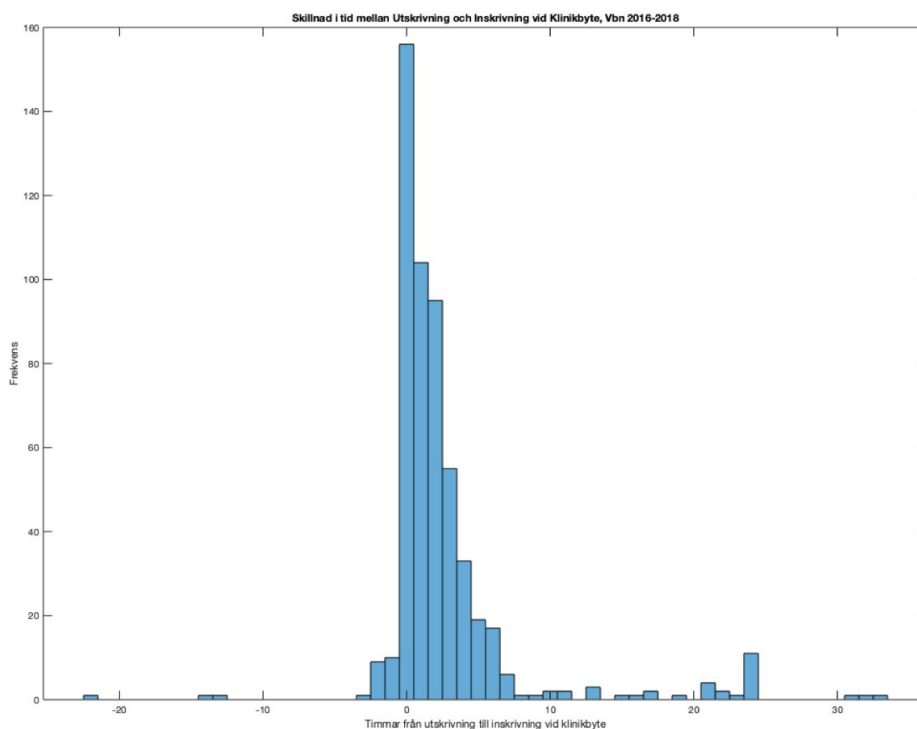
Modellering och simulering av denna typ är en ny företeelse inom vården vilket gör att även dataprocessen är det. Frånvaro av en standardiserad och kvalitetssäkrad procedur för utlämning av logistikdata har under projektets gång utvecklats hos de givande parterna vilket gjort att datauttagens komplexitet och framförallt tidsåtgång varierat till det positiva mot slutet av projektiden.

För att bättre förstå och kunna dra slutsatser av den information som lyckas framställas ur givna data så skulle det vara bra för framtida projekt att ha resurser på plats med stor domänkunskap som med kort notis skulle kunna värdera framkomna resultat.

Sverige-fallet

De svenska data som erhöles var av relativt bra kvalité men en del parametrar försvårade analyserna.

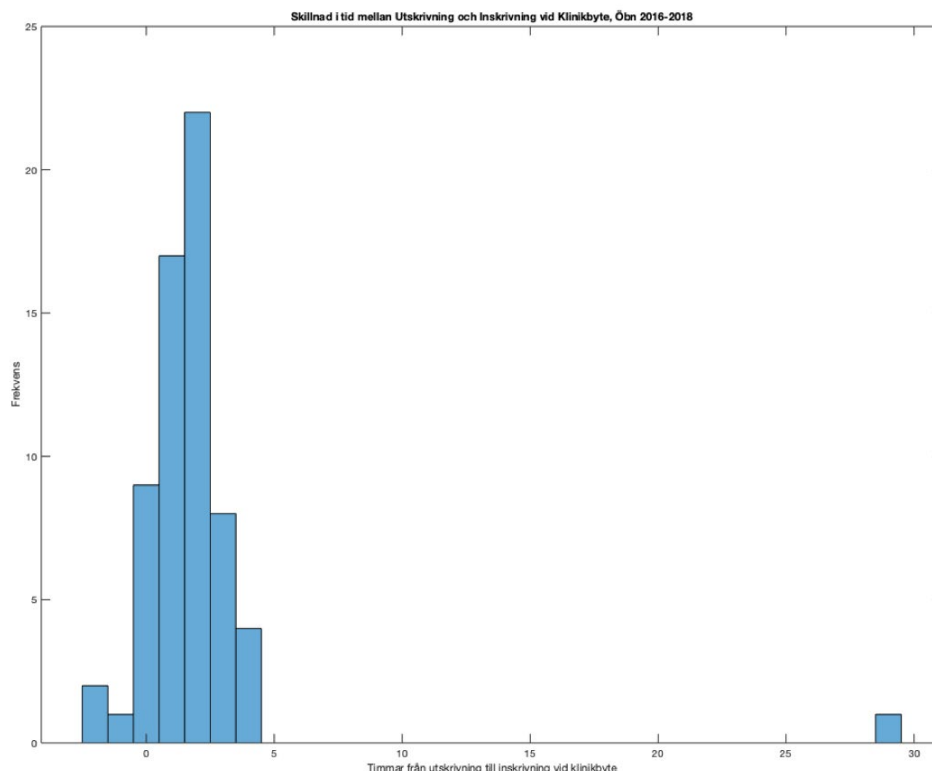
1. För återbesök så finns både varianterna "Ring och boka en tid" eller att en tid bokas direkt. Detta gör skillnad i systemet då den första registreras som akut och den senare som planerad, trots att det borde vara likriktat då den faktiska sökorsaken är lika.
2. Koder för ut och inskrivning är ibland godtyckligt inskrivna (kan ses viss korrelation mellan utskrivande enhet och mönster i utskrivningskod), dessa borde vara förinställda beroende på platsen patienten skrivs till. T.ex. finns det två koder vilka kan användas för utskrivning till en annan klinik och dessa är 2 – Till annat sjukhus samt 3 – Flyttad till annan klinik. Inom Västerbotten används dessa bägge om vartannat vilket gör det svårt att göra matchningar. Vissa mönster kan ses i att olika kliniker använder dessa olika. I matchningen som gjorts i denna studie har detta undersökts och matchning har gjorts när övriga parametrar stämmer.
3. Det borde finnas en standard för i vilka fall och hur man anger komplikationsdiagnoser. I vissa inskrivningar finns den ursprungliga huvuddiagnosen som huvuddiagnos och komplikationerna som bidiagnos, i vissa tvärtom och i andra fall har huvuddiagnosen fallit bort helt. Detta gör det svårt att identifiera vårdkomplikationer och återinskrivning.
4. Det finns variationer i in- och utskrivning (Vissa skrivs naturligt in med en resetid mellan men vissa skrivs in före de skrivs ut). Ett exempel som togs bort från figuren nedan men dock är intressant är en patient som under en och samma period om tre dagar hade två inskrivningar på samma klinik (Vilhelmina sjukstuga, mars 2018)



Finlandsfallet

Finland och Vasaregionen från vilket projektets data kommer ifrån, har ett externt företag som ansvarar för lagring och hantering av data. Datamaterial måste därför beställs därifrån och detta har i projektet bidragit till en ökad väntetid och ett hinder i kommunikationen. Detta är en av anledningarna till att analysresultaten skiljer sig mellan länderna.

1. Huvuddiagnoser på första inskrivning och vidare inskrivning är ej överensstämmande i datamaterialet. T.ex. vid första besök i Vasa får patienten S721 (Trauma), när patienten sen skrivs vidare till Närpes får patienten huvuddiagnos E039 (Sköldkörtelproblem).
2. Vi har inte kunnat göra en analys av resmönster då vi skulle behöva patienternas postnummer. Idag finns endast hemkommun i datamaterialet från Vasa och denna är endast Närpes för alla patienter i urvalet. Därför kan inte avstånd räknas ut från hemmet till HVC och hem igen utan endast avstånd för interna klinikbyten t.ex. mellan Närpes och Vasa.
3. I likhet med det svenska datamaterialet finns det patienter som är dubbelinskrivna samt patienter som vid förflyttningar inte skrivs in direkt. Detta kan var en produkt av restid men 12 timmars restid som vissa glapp har varit är osannolikt.
 - a. I de utvalda diagnoserna (M16/M17/S72) så rör det sig om mellan 1- 2 timmar (3 tillfällen på 121 förflyttningar eller 2,4%)
 - b. Ungefär 1,5% av alla som har kedjor (patienter som är inskrivna på fler än en klinik under sin vårdresa) är vid någon tid dubbelinskriven. De är i snitt dubbelinskrivna i 3 timmar (minst 1 minut och max 12,5 timmar)



Diskussion & Slutsats

Mycket kan göras med det datamaterial som samlas in redan idag, utmaningen är att hitta strukturer för att kunna möjliggöra datadrivna analyser i nära realtid. Utmaningar för detta är både tekniska, juridiska och kulturella. Dessa utmaningar är universella för att kunna använda data i nära realtid för till exempel operativ planering och utveckling oavsett ifall man använder modellering, simulering eller bara visualisering.

Just möjligheten för analyserna att vara datadrivna och kontinuerligt uppdaterade möjliggör aktuella lägesbilder och att de analyser man gör utgår ifrån läget nu och inte för några år sedan då en så komplex och omfattande verksamhet som vården är rörlig, vilket vi ser i vårt data med t.ex. ombyggnationer och omorganisation av kliniker.

För att möjliggöra detta är det viktigt att det finns en helhetstanke från början med datainsamling till slutet med analyser och presentation.

Dagens datainsamling är inte utformad med användning av datadriven analys utan som en digitalisering av de tidigare analoga systemen och nyttjar därför inte kraften i det datadrivna optimalt. Då modellering idag är mer tillgängligt och kraftfullt kommer förhoppningsvis nästa generation av vårdinformationssystem att på ett bra sätt ha det integrerat i sin kärna.

Ovanstående gjorde att det tog tid att få rätt datamaterial och få datamaterialet på rätt sätt. Processen för att få datamaterialet var lång och ganska osäker då det inte var en fastställd process utan var tvungen att förankras och kontrolleras. I Västerbotten har en process för att skapa en uttagsprocess initierats av just denna anledning då detta projekt ej var det första att ställa en liknande fråga. I Österbotten har man en större vana av att arbeta med datamaterial men inte för just denna tillämpning.

En sak som framkommit i projektet är också att det krävs att de som utför dataanalysen och den efterföljande modelleringen har en stor domänkunskap och en nära kommunikation med verksamheten då det finns väldigt mycket kunskap i produktionen som kan ha stor påverkan på flödena men som inte är enkelt läsbart ur datamaterialet.

Ytterligare svårigheter tillstötte när det kommunala datamaterialet på svenska sidan hölls tillbaka då det fanns juridiska oklarheter eller osäkerheter hos de handläggare på kommunen som ansvarade för datamaterialet. Något som skulle underlätta framtida studier är enklare tillgängliga juridiska anvisningar och information för de som handhar datamaterial och kanske inte har tiden eller möjligheten att sätta sig in i reglerna kring datahantering.

Ytterligare ett ämne man skulle behöva belysa är också användandet av integritetskänsliga data i analyser gjorde i nära realtid så att det blir en bra hantering där integritetsskyddet behålls och analyserna inte riskerar att hänga ut enskilda individer eller känsliga uppgifter. Här behövs en hög nivå av de som handhar datamaterialet och de som skriver programvarorna.

Slutsatserna är alltså att det behövs juridiskt, tekniskt och kulturellt arbete för att facilitera en datadriven approach till beslutsstöd där de som utarbetar datamodellerna har en hög kompetens inom statistik men arbetar i nära anslutning till verksamheterna.

