

Erfaringer med valideringsstudie for SOReg-N

Steg 1 Studiedesign

Registerleder, registerkoordinator og statistikere ved Nasjonalt servicemiljø for medisinske kvalitetsregistre region Vest avtalte et møte for studiedesign. På dette møte gjennomgikk møtedeltakere flere punkter:

- Er det overkommelig for registeret å besøke alle sykehus? Hvis ikke, skal det trekkes tilfeldige sykehus, eller skal man velge et sykehus i hver region, eller et sykehus i hver størrelsesorden? Er det andre «typer» sykehus eller avdelinger som registeret bør dekke over i valideringsstudien (lokalt/regionalt, barne- og voksenavdelinger, akutt/langvarig osv.)?
 - For SOReg-N ble planen å besøke alle sykehus som leverer data til registeret. Det ville ikke være tid til å besøke alle sykehus i løpet av et år, men en ville bestrebe seg til å få i stand en rullerende besøksplan som ville medføre at alle sykehus ville få besøk minst hvert andre år. Inkludert i besøkene skulle en også ha samtaler og erfaringsutveksling angående det praktiske ved registrering i registeret (utfordringer, spørsmål, ...). Tidligere erfaringer hadde vist at dette kunne være nyttig for å forbedre datakvaliteten, dekningsgraden og innregistreringsløsningen.
- Kan det være gunstig med en pilot ved ett sykehus?
 - SOReg-N planla en pilot ved et sykehus ikke langt ifra registerets lokaliteter, Voss sjukehus.
- Skal man validere alle pasienter, eller skal man ta et utvalg?
 - For SOReg-N var det ikke realistisk å se på alle pasienter. Registerleder og koordinator forhørte seg med statistikerne på møte hvor mange pasienter som burde være med. Statistikerne opplyste om den statistiske usikkerhet for ulike scenarier. Man kom frem til at

hovedfokus skulle ligge på validitet for de enkelte variablene for landet sett under ett. Antall pasienter var avhengig av hvor mye tid registerkoordinator hadde på sykehuset de besøkte. Konklusjonen var at statistikerne skulle lage valideringsfiler med alle pasienter for hvert sykehus, men sette dem i tilfeldig rekkefølge. Registerkoordinator kunne da jobbe gjennom listen så langt hun kom i løpet av besøket, og således få med «så mange som mulig» pasienter.

- Skal man validere alle variabler? Skal man ta et tilfeldig utvalg variabler, eller konsentrere seg om variabler sentrale for registeret (typisk kvalitetsindikatorer)? Er det noen variabler som er uaktuelt å validere, f.eks PROM-data?
 - SOReg-N kom frem til at det var aktuelt å validere de sentrale variablene i registeret, dvs. de variablene som var med i en egen datadump laget for årsrapporter. Men for å få opp bredden i variabler uten at dette gikk utover antall pasienter undersøkt, ville man ikke validere alle variabler for hver pasient. Registeret ble enige med statistikerne om å trekke et overkommelig antall, tilfeldige variabler per operasjon (operasjon = omtrent det samme som pasient, men noen få pasienter hadde hatt reoperasjoner). Planen var i utgangspunktet 7 tilfeldige variabler per operasjon.

Etter å ha samlet inn data for flere sykehus, skulle en undersøke om det var noen variabler som utmerket seg med særlig dårlig datakvalitet. En skulle i så fall gjennomføre tiltak for å få opp datakvaliteten på disse.

Steg 2: Tilrettelegging av data

Etter enighet fra møte arbeidet statistikerne med å hente et tilfeldig uttrekk variabler fra registeret. De strukturerte et SPSS-datasett for validering i statistikkprogrammet R. Datasettet inneholdt følgende variabler/kolonner:

- Indeksvariabler som entydig identifiserte en operasjon, eks. Pasient-ID, forløps-ID, operasjons-ID, operasjonsdato, kjønn, alder og fødselsdato. Operasjonsdato, kjønn, alder og fødselsdato var ikke

nødvendig for å entydige identifisere operasjonen (de kan avledes av pasient-ID, forløps-ID og operasjons-ID), men var med for å gjøre det enklere og raskere å finne frem pasienten i pasientjournalen.

- En kolonne for variabel-ID. Denne beskrev hvilke av de tilfeldig uttrukne variablene som skulle sjekkes for pasienten. Hver pasient hadde 7 rader i datasettet, én for hver av de 7 tilfeldige variablene. Rekkefølgene på radene var basert på en «naturlig» rekkefølge (for eksempel kom baseline-variabler før oppfølgingsvariabler), slik at det skulle være enklere og raskere å finne frem til informasjonen i pasientjournalen.
- En kolonne for tallverdi og en kolonne for datoverdi. Hvis variabelen hadde en tallverdi, ble denne lagt i tallverdikolonnen. Hvis variabelen hadde en datoverdi, ble denne lagt i datoverdikolonnen. Kolonnenavnene hadde prefikset «reg_» («reg_tal» og «reg_dato») for å indikere at det var registerets verdier som lå i disse kolonnene.
- En tilsvarende kolonne for tallverdi og datoverdi som skulle fylles ut fra pasientjournalen. Disse var tomme i utgangspunktet, og hadde prefiks «epj_» («epj_tal» og «epj_dato») for å indikere at disse kolonnene var til verdiene fra journalen.
- Datasettet ble sendt til en SPSS-fil. Statistikerne genererte en egen SPSS-fil for hvert sykehus. Sykehusets navn var med i navnet på SPSS-fila og filene ble lagt i en mappe med navn lik datoen til uthenting av datagrunnlaget (format: ÅÅÅÅ-MM-DD).

Målet med denne strukturen var at det skulle være enkelt for registerkoordinator å legge inn verdiene hun fant i pasientjournal i de tomme «epj_»-kolonnene. Et annet viktig aspekt var at SPSS-formatet sikrer dataintegritet (eks. kan ikke datokolonner inneholde tekst, tall eller noe annet enn datoer) og er egnet for videre databehandling. Dataformat som Excel-regneark var av denne grunn uaktuelle. Registerkoordinator hadde erfaring med SPSS og dette var ansett som et egnet verktøy. (En «ekte» database hadde også vært et egnet verktøy, men det hadde krevet mye mer ressurser å sette opp.)

Steg 3: Validering

Registerkoordinator reiste til Voss sjukehus med tilrettelagt datasett for gjeldene sjukehus. Her fylte hun inn de tilsvarende verdiene for en variabel

funnet i pasientjournalen. Samtidig dokumenterte hun feil slik at de kunne rettes opp i registeret.

I mellomtiden, utarbeidet statistikerne et skript i SPSS for enkel analyse av valideringsdatasettet etter ferdig utfyllelse. Målet var at registerleder og koordinator selv kunne kjøre skriptet i SPSS for å få resultater etter et besøk. I denne sammenheng var det dialog mellom registerleder og registerkoordinator, og statistikerne om hvilke analyser som var ønsket. Det ble enighet om følgende resultater:

- Prosentandelen verdier i registeret som var lik verdiene i pasientjournalen
- Prosentandelen manglende verdier i registeret som var utfylt i pasientjournalen
- De to ovennevnte analysene per variabel, for å identifisere om noen variabler var typisk ikke-besvarte eller feil-utfylte
- Prosentandelen manglende verdier i pasientjournalen som var utfylt i registeret.

Når det for en kontinuerlig variabel eller en datovariabel ikke var samsvar mellom registeret og pasientjournalen (men noe var registrert begge steder), ville det også være aktuelt å regne ut hvor stor diskrepansen var (eks. differansen i registrert kroppshøyde mellom de to datakildene).

Steg 4: Erfaringer

Etter valideringsbesøket til Voss sjukehus, ble noen endringer gjort basert på erfaringer. En erfaring var at å finne mange variabler for en pasient var lettere enn å åpne ny journal for å finne verdiene til en ny pasient. På bakgrunn av dette økte en antall tilfeldige valgte variabler til 10 variabler per pasient.

Videre viste det seg at mange av variablene var uaktuelle for enkelte pasienter. Det var ikke aktuelt å se på 1-års-vekt til pasienter som nylig hadde hatt sin operasjon og 1-års-oppfølging ikke var aktuelt enda. På bakgrunn av dette kunne statistikere legge til i datasettet at variabler som gjaldt målinger på 6-ukers og 1-års oppfølging bare dukket opp for pasienter som var aktuelle for disse oppfølgningene.