

# Access/R produksjonsapparat for folkehelseprofiler: Brukerveiledning

---

(Filversjon 1, 15.4.2015 – men oppdatert med detaljer innimellom)

## Oversikt

Innledning.....	2
Innlesing av filer.....	3
Kubeproduksjon (dvs. all output).....	13
Eksempel på kompleks indikator: INNTULIKHET.....	21
Sammenhengen mellom fil-elementene (til bruk ved f.eks. kval.kontroll).....	21
Statusinformasjon om hver indikator.....	22
Kodeverk og systemverdier.....	22

## Innhold

Innledning.....	2
Access-database, path:.....	2
Se også nederst!.....	3
Begreper.....	3
Arbeidsprinsipp.....	3
Pass på:.....	3
Innlesing av filer.....	3
1. Tabell FILGRUPPER:.....	3
Dersom vi får en ny dataleveranse med samtlige årganger, men nytt format: .....	4
2. Skjema StyrFILGRUPPER:.....	4
Bruk:.....	5
Feltene vi må forholde oss til:.....	5
Tabell 1:.....	5
Tabell 2 (INNLESING): Selve innlesingsspec'en.....	6
Soner:.....	8
Grunnkretser:.....	8
Andre kommuner enn bydels-byene:.....	8
Trengs Reshape av to ting samtidig?.....	8
Innlesingsargumenter (INNLESARG), syntaks:.....	9
Hvis samme inndatafil skal brukes i flere filgrupper (eller flere innlesingsspec'er):.....	9
Bruk av RSYNT- (Statasynt-) punkter:.....	10

3. Kjør innlesing med disse parameterne:.....	10
I RStudio:.....	10
4. Sjekk resultatet: Skjema LOGGskjema.....	11
5a. Faste koder: Endres ikke.....	11
Felter:.....	11
5b. Kodebok:.....	12
Legge inn ny kode: Tabell KODEBOK.....	12
Sjekke hvordan det gikk: Tabell KODEBOK_LOGG_se.....	12
Gjelder dette nå, etter at "versjonering=TRUE" er satt som default? MEN SÅ MÅ FILGRUPPEN LAGES I R!.....	12
SPESIELLE TRIKS OSV.....	13
Gjenbruke en innlesingsspec: Samme innfil i to filgrupper.....	13
Begrense tidsserien i én ORGfil, når to ORGfiler har overlappende data: Bruk KODEBOK.....	13
Kubeproduksjon (dvs. all output).....	13
1. Tabell TNP_PROD.....	14
2. Tabell KUBER.....	14
Felter:.....	14
Fylles ut når kuben skal direktestandardiseres i Stata (brukes i NH):.....	16
Syntaks for Stata_TAB1:.....	17
HVIS INDIKATOREN SKAL VISES I BAROMETERET: Tabell FRISKVIK.....	17
Feltene:.....	17
(Feilmelding ved kjøring:.....	19
3. FOR Å KJØRE:.....	20
4. PROBLEMLØSING/FEILSØKING.....	20
Ta runtimedump:.....	20
Skrive selve filgruppen til en CSV-fil, direkte fra R:.....	20
Ryddde buffer:.....	21
Friske opp globale bakgrunnsdata:.....	21
Uforklarlig prikking av sammenslåtte Geo?.....	21
Eksempel på kompleks indikator: INNTULIKHET.....	21
Sammenhengen mellom fil-elementene (til bruk ved f.eks. kval.kontroll).....	21
Statusinformasjon om hver indikator.....	22
Kodeverk og systemverdier.....	22
Interne koder for de ulike elementene (kolonnene) i en tabell:.....	22
Landbakgrunn, koder i KH_KODER:.....	23
Sivilstand, koder i KH_KODER:.....	23
Utdanning, koder i KH_KODER:.....	23

## Innledning

### Access-database, path:

[\\fhi.no\Felles\Forskningsprosjekter\PDB 2455 - Helseprofiler og til \\_PRODUKSJON\STYRING\KHELSA.mdb](\\fhi.no\Felles\Forskningsprosjekter\PDB 2455 - Helseprofiler og til _PRODUKSJON\STYRING\KHELSA.mdb)

Kåre bemerket (jun-2016) at han med hensikt hadde holdt seg til det gamle Access-filformatet .mdb, og ikke brukt .accdb fra Office 2010. Han mente det var kommet til ytterligere bells and whistles i grensesnittet i det nye formatet, som ville heve brukerterskelen.

## Se også nederst!

Mer generelt om sammenhengen i databasen, og hvordan den brukes (konvensjoner) beskrives etter kapitlene om detaljert bruk.

## Begreper

- Navigasjonsruta til venstre i Accessbildet har de ulike tabeller, skjemaer osv. samlet under deloverskrifter (Access-begrep: «grupper»). Disse nevnes i denne filen først i navnet på det objektet vi skal inn i, slik: «STYRING\Filgrupper».
- Stabled fil: Alle de separate inndatafilene for en indikator (f.eks. alle årgangene) lagt oppå hverandre til én stor fil. DSS. filgruppe.

## Arbeidsprinsipp

- Prosessen har to hovedtrinn:
  - 1) Innlesing av leverte/nedlastede datafiler og produksjon av en «stabled fil».
  - 2) Kubeproduksjon, dvs. all videre behandling av stabled fil.
- Det er bare ett nivå av lagret mellomprodukt: Stabled fil. Denne ligger som R-datafil, og er normalt ikke meningen å behandle utenom systemet. Her ligger alle årganger, men normalt bare telleren (antall).
- Hver gang noe skal leses inn (f.eks. ved ny årgang), leses samtlige tidligere originaldatafiler inn samtidig (det vil si, de som ikke er tatt ut av produksjonen). Stabled fil lages altså alltid på nytt fra bunnen, vi hekter ikke bare siste årgang bakpå en stor samlefil.
- Både for innlesing (dvs. hva som må til for å få lest filene riktig inn) og produksjon av output-filer styres detaljene fra en Access-database. Dermed trenger vi ikke tukle med selve produksjonsscriptene for å gjøre endringer.
- Innlesing settes i gang inne fra Access med en programmert funksjonsknapp.
- Kubeproduksjon (dvs. all output) kjøres direkte i R. (Det kan nok bli endret etter hvert)

## Pass på:

Klikk alltid «Oppdater alt»-knappen for HVER tabell (etc) jeg er inne og redigerer i! Hvis jeg bare bytter til en annen tabell, har jeg opplevd at siste innlagte verdi ikke blir lagret.

OG gå ut av den raden du har redigert i, så den ikke står i redigeringsmodus, FØR du klikker Oppdater.

## Innlesing av filer

Venstremeny i Access: Avsnitt LAG FILGRUPPE – MÅ BRUKE

/home/ybk/Downloads/Bruksanvisning.docx

## 1. Tabell FILGRUPPER:

Én linje per ferdig stablet fil. Det svarer alltid til en unik gruppe inputfiler. Innhold: Grunnleggende definisjoner for en filgruppe. Evt. R-syntakser for spesialbehandling.

(Noe av denne informasjonen brukes også ved kubeproduksjon.)

- Haker av for hvilke av standardvariablene som fins i innfilene.
- Må oppgi: Hva betyr alder=ALLE? F.eks. at i denne filgruppen er det «16\_69».
- TAB1 (-2, -3): Dimensjoner i tillegg til de obligatoriske - hva SKAL de hete i den stablede filen.
- B\_STARTAAR: Når data for Bydeler kan brukes. F.eks 2004 betyr at tallene for bydeler før 2004 ikke skal brukes i KUBE pga. ukomplette, for få etc.. etc..
- DK2020\_STARTAAR: Sett "2020" dersom ORGdata er: inndelt etter kommune (ikke grunnkrets), med gamle kommunekoder, og er satt opp slik at missing betyr null. Det gjør at tall f.o.m. 2020 for delingskommuner får SPVFLAGG == 1, og dermed ikke serieprikkes.
- SY\_PAA\_FILGR
- RSYNT\_SYPAA (Ikke implementert punkt!)
- ValErAarsSnitt
- VAL1navn (-2-, -3-): Tilsv. som TAB.., for verdikolonnene.  
VAL1navn er gjerne telleren.  
VAL2navn er ofte nevneren, hvis den følger med i inputfilene. (SEPARATE FILER: Se nedenfor)  
DERSOM inndatafilen inneholder verdikolonner vi ikke skal bruke (f.eks. ferdige prosenter), bare la være å nevne dem.
- VAL1sumbar (-2-, -3-): Om verdiene kan aggregeres.
- VAL1miss (-2-, -3-): Hvordan håndtere at VAL1 er missing? Skriv inn hvilken kode disse cellene skal ha: **OBS: SJEKK AT JEG HAR FORSTÅTT DETTE RIKTIG!**  
0        hvis det er implisitte nuller  
..        Manglende data
- (...) VERSJONFRA, VERSJONTIL: Datoer for når/om denne spec'en er tatt inn eller ut av systemet.
- FILTER1 (-2, -3): Ikke i bruk i scriptet. Kan brukes til flagging av rader i lista (for sortering etc).
- RSYNT\_PRE\_FGLAGRING: R-syntaks som kjøres før den ferdige stablafilen lagres.  
*F.eks. for å lage Middelfolkemengde fra befolkningsfil: Inndata har folketall 1.jan., R-syntaksen beregner flere versjoner av Middelfolkemengde, og det genereres noen ekstra variabler med disse Midbef-tallene. Stablafile har m.a.o. noen flere kolonner enn vanlige stablafiler.*

### Hvis Teller og Nevner leveres i separate filer,

så leser vi dem inn til separate filgrupper og kobler dem sammen i KUBE-trinnet.

### Dersom vi får en ny dataleveranse med samtlige årganger, men nytt format:

Ikke lag ny linje i tabellen, og ikke endre filgruppenavn - vi bare redigerer eksisterende linje.

MEN da må vi også gå gjennom **samtlige gamle inndatafiler** (i neste verktøy, se neste avsnitt!) og sette «I bruk til»-dato lik i dag.

## 2. Skjema StyrFILGRUPPER:

Viktigste styringsverktøy for innlesingen.

Skjemaet har fire deler:

1. Enkeltfelter øverst, som oppsummerer det som ble lagt inn i forrige avsnitt.
2. En tabell (her kalt Tabell 1) med én rad per unike inndatafil. (Den henter fra "ORIGINALFILERse" under LAG FILGRUPPE – MÅ BRUKE.)  
Gir path og filnavn, noen grunnleggende opplysninger om filen, og forteller hvilken av radene i tabell 2 som skal brukes.
3. En tabell (Tabell 2, ledetekst «INNLESING») med én rad per sett av innlesingsparametre innen samme filgruppe («Innlesingsspec»). (Den henter fra tabell INNLESING)  
Gir de detaljene som gjør at én bestemt inndatafil blir lest inn riktig og evt. reshaped til korrekt («long») format.
4. Et større skjema (ledetekst «StyrFILGRUPPER EnInnles», heading INNLESING), som viser den valgte raden fra tabell 2 så man ser flere detaljer i ett blick.

Når flere inndatafiler er like, kan de i tabell 1 kobles til samme innlesingsspec ved å oppgi navnet på spec'en. Da får man flere rader i tabell 1 for hver rad i tabell 2.

HVIS SAMME INNDATAFIL skal brukes i flere filgrupper: Se nedenfor!

- OG hvis du ikke får lov å skrive inn filnavnet: Sjekk i tabell ORIGINALFILERse om den ligger inne fra før...

Hvis man skal lese inn samme fil flere ganger bare med ulik delID/spec:

1. gå til «ORGINNLESkobl» - her kopierer man FILID på den filen man skal duplisere og legge inn samme filgruppe (med mindre den skal brukes i en annen) og den nye delID/spec. Når man har oppdatert dette, går man tilbake til «styrFILGRUPPER» og oppdaterer, og da skal de nye innlesingene være der.

### Bruk:

Velg riktig filgruppe: Verktøyknapp «Sorter og filtrer/Avansert/Filtrer etter skjema» gir nedtrekksmeny i det valgte feltet nedenfor. Velg riktig filgruppe (En ny filgruppe vil ikke ligge i lista, da skrives navnet), og klikk «Aktiver/deaktiver filter».

### DERSOM ovennevnte klikk ikke fungerer:

Sett markøren i skjemafeltet FILGRUPPE, og klikk knappen Filtrer (stort traktsymbol). Da kommer en rullegardin under feltet, som er bredere (lettere å se FG-navnene) og har avkrysningsbokser. Huk av riktig FG og klikk OK.

For å fylle ut Tabell 2 kan det lønne seg å åpne inndatafilen for å se hvordan den er.

Hvis det blir feilmeldinger og skjæring: Vi fikk det til en gang ved å gå i tabellene i bakgrunnen (ORIGINALFILERse, f.eks. (siden filen allerede var innlest, men uten å skrive inn FILGRUPPE, så den havnet et annet sted i systemet)) og slette hele raden, og begynne på nytt.

### Feltene vi må forholde oss til:

#### Tabell 1:

- LEVERANDØR: Må fylles ut først, ellers nekter Access. Ved utfylling kommer feilmelding. Klikk OK, så lagres det likevel.
- FILGRUPPE: Samme navn som brukt tidligere. TIPS: Hvis du ikke får lov å skrive i feltet, piltast litt mot høyre og venstre, så gir den seg ...
- DELID: Navnet på den innlesingsspec'en som skal brukes (se tabell 2).
- FILNAVN: Må inkludere path fra første nivå under F:\Prosjekter\KommuneHelse\PRODUKSJON\. Det vil i nåværende system si: **ORGDATA\<leverandør>\<tema>\ORG\<profilår>\<filnavn.ext>**
- FORMAT: Datatypen for inndatafilen (selv om den også var med i <filnavn.ext>). OBS: Ikke dss. filtypen: En CSV-fil har format CSV, selv om den heter «data.txt».

ANM: per 26.6.2015 hadde vi IKKE type «DTA»/Stata implementert! Men det er kjøpt å lage.

- LEVERANSEAAR: Det samme som profilår, dvs. det året dataene er (eller kunne vært) med i profilene. For data levert etter at profilene er laget, settes året etter.
- IBRUKFRA: Dokumenterer når denne filen ble innlemmet i systemet. Som default kommer dagens dato ved første kjøring av innlesingsprosessen, men ofte settes den datoen filen ble f.eks. lastet ned fra SSB Statistikkbanken.
- IBRUKTIL: Standardverdi: 01.01.9999. Her settes dagens dato når vi avslører feil e.l. i en fil, som gjør at den ikke skal brukes mer. Da flyttes den filen fysisk fra ... \ORG\... til ... \GML\... (\*\*), og dette faktum rettes i filnavnet i tabell 1. Til-datoen er egentlig første dag denne filen *ikke* skal regnes med lenger.

(\*\*) Dette trengs ikke hvis hele tidsserien er erstattet i den nye leveransen.

*MEN OBS: Jeg lagde ny innlesingsspec for MFR Keisersnitt\_NH. Måtte sette "ibrukt" på forrige årgangsfil for at systemet skulle ta min nye fil.*

#### Tabell 2 (INNLESING): Selve innlesingsspec'en.

(Her vises én enkelt rad fra tabellen STYRING\INNLESING – men ikke alle feltene.)  
(Dette kan kanskje gjøres også i skjemaet 4, nederst i skjermbildet, men det har vi ikke testet.)

(Marie sa først at det kan være greit å få på plass de andre tingene først, og så ta RESHAPE-argumentene etterpå. Men rekkefølgen i prosesseringen er ikke lik rekkefølgen i dette skjemaet, så noen ting høyere opp faller ikke på plass før reshape virker.)

OBS: Feltene GRUNNKRETS, SONER, SKALA\_VAL1 (-2 -3) og KOPI\_KOL vises ikke i skjemaet! Må fylles ut i selve tabellen INNLESING.

- FILGRUPPE: Som før.
- DELID: Nytt navn for hver ny rad.
- INNLESARG: Argumentene til innlesingsfunksjonen, se nedenfor. OBS: Dette slår inn FØR Rsynt1, så ting man skriver her kan påvirke inndataene til Rsynt1.
  - - CSV-fil som er ekte kommaseparert (ikke semikolon): skriv *sep* = ","
- MANHEADER: Kan styre hvilke kolonner som får hvilken header.  
Syntaks: [1,3,6,7:34]=c("GEO","KJONN","ALDER",1986:2013)
- KASTKOLS: Kan spesifisere kolonner som skal droppes.  
Syntaks: c(2,4,5,35,36,37)  
(Begge de to siste: Se f.eks. tabell INNLESING, Filgruppe BEFOLK, ID nr. 9.)  
OBS: Denne parameteren slår inn *etter* RSYNT1. Hvis kolonner lager krøll i RSYNT1, hjelper det ikke å KASTE dem her.
- FYLLTAB: Navn på TAB-kolonner (dvs. dimensjoner) som i datafilen ikke er fullstendig fylt ut. F.eks. når det står et fylkesnavn bare på første linje med data for dette fylket. Syntaks: Anførselstegn rundt, komma (ikke blank) imellom.  
"GEO", "ALDER"
- RESHAPEid (**OBS: Se nedenfor hvis dobbel Reshape!**): De dimensjonskolonnene som utgjør ID, og som allerede er på long-format. Skriv de «ferdige» navnene (GEO osv.) Syntaks: Som FYLLTAB. OBS: Legg også inn «VAL2» (bokstavelig, altså ikke det faktiske inndatavariabelnavnet) hvis den allerede er på Long.
- RESHAPEmeas: Hvis det er flere verdikolonner som skal bli til én, men der de opprinnelige kolonnetitlene er meningsbærende og skal bli til en ny dimensjon. Skriv de faktiske variabelnavnene i innfilen (de blir kategorier), og rams opp alle som skal bli til én kolonne. Double quotes, og komma. (Prøvde å skille mellom to targetkolonner med semikolon, men det gir kræsje og feilmelding "unexpected ';' ".)  
Eks: "avekt\_u1500","avekt\_u2500","avekt\_o4500".  
Dersom det ikke er spesifisert noe her, vil Reshape ta alle variabler/kolonner som ikke er nevnt i de andre feltene på denne raden, og reshape dem til én.
- RESHAPEvar: Dimensjoner som SKAL reshapes – hva de(n) ferdige kolonnen(e) skal hete, hvor de ovennevnte kategoriene skal være verdiene. Eks: FVEKTKATEGORI.
- RESHAPEval: Hva ferdige verdikolonner skal hete. Eks: ANTALL. (Trenger ikke quotes)
- GEO, AAR, KJONN, ALDER: Hva denne kolonnen heter i inndatafilen. Har den ikke noe kolonnehode, skrives bokstaven for Excelkolonnen. Hvis kolonnen først oppstår etter Reshape, skriv navnet på den ferdige kolonnen (samme som i ReshapeVar ovenfor). Også hvis kolonnen lages i RSYNT1, skriv navnet på ferdig kolonne.  
Dersom kolonnen ikke finnes, og ikke trengs: skriv bindestrek - .

Dersom GEO mangler (dvs. for en fil med landstall): Skriv **<0>** . Da gis alle rader landskoden, 0.

Dersom AAR mangler: Skriv **<\$y>** . Da hentes dataårgangen for denne ORGfila fra feltet DEFAAR i tabell ORIGINALFILER (*som m.a.o. må fylles ut*).

Dette er inndata til Rename: Rett før Reshape renames alle variabler som har originalnavn angitt på denne raden, til standardnavnene (VAL1 osv). Så Reshapes, og da oppstår jo noen nye variabler, som vi har sagt i RESHAPEval at skal hete **<noe>**. Så går det en ny runde Rename, der de nyopprettede variablene endres fra «våre» navn til standardnavnene. De endelige utfil-navnene settes på helt til slutt.

OBS FOR GEOKODER: Tenk på Soner (se nedenfor)

**KODE <0> :**

- GEOd2: Hvis det f.eks. er to nivåer, så denne kolonnen må være med for å gi entydig Geo. Kolonnenavnet i inndatafilen.
- UTDANN, SIVST, LANDBAK: Hvis disse ikke fins i filen, skriv « **<ALLE>** ».
- TAB1 -2 -3: Dimensjoner som skal være med i tillegg til de ovennevnte.
  - Hvis variabelen fins i inndatafilen og ikke skal reshapes, skriv variabelnavnet derfra.
  - Hvis den oppstår ved Reshape, skriv hva ferdig reshapet kolonne skal hete (som for GEO osv. ovenfor).
  - Hvis den skal opprettes, og fylles med en standardverdi: skriv verdien slik: **<verdi>**
- VAL1 -2 -3: Verdikolonnene. Som for TAB-kolonnene.
- GRUNNKRETS: Sett lik 1 hvis det er grunnkretskoder i GEO.
- SONER: Se nedenfor.
- SKALA\_VAL1 (\_VAL2, \_VAL3): Dersom en VAL-kolonne leses inn med feil størrelsesorden (f.eks. at «prosent» ikke tolkes riktig). Innleste tall multipliseres med den faktoren som legges inn her. (Ny des-2019)

Feltene lengre til høyre ble ikke forklart i første gjennomgang.

### Soner:

I noen tilfeller må vi hjelpe Geo-dekodingen: (Vet ikke om dette er relevant for NH/fylkesnivå)

Dersom data er levert slik at for de byene der det er bydeler, så får vi bare bydelene og ikke selve kommunetallet, mens for andre kommuner får vi kommunetallet:

(Det vil si at summen av alle geo blir landstallet)

- Da må vi kalle geo-kodene for **Soner**. Og det innebærer å legge inn i feltet INNLESING/SONER (ikke i StyrFILGRUPPER-bildet!) hvor lange geokoder som skal gjøres om til (eller behandles som) sonekoder.

Det er tre tilfeller:

<b>Alle kommuner, PLUSS bydeler:</b>	<b>Bare bydeler for de aktuelle kommunene, vanlig koding:</b>	<b>Noen har seks sifre hele veien:</b>
0104	0104	010400



0214	0214	021400
<b>0301 (NB!)</b>		
030101	030101	030101
030102	030102	030102
0402	0402	040200
= IKKE soner. Skriv ingen ting i SONER-feltet. Da vet systemet (by default) at 4 siffer =K, 6 siffer=B.	= SONER. Angis som «4,6».	=SONER. Angis som «6».

### Grunnkretser:

Grunnkretser har ikke hierarkisk koding som samsvarer med bydelene. De fire sifferne bak kommunenummeret er selve GK, og må oversettes til bydelsnummer via en egen tabell. Derfor må vi si fra til systemet at kodene er GK, ved å sette "1" i feltet INNLESING/ GRUNNKRETS.

### Andre kommuner enn bydels-byene:

Grunnkrets er egentlig irrelevant. Alt bak kommunenummeret klippes av, og data aggregeres til kommuneverdiene, uten noen egen styring/kodebok/tabell. Feltet GRUNNKRETS må settes til 1.

Testet ifm. MFR Keisersnitt-NH: ORGdata hadde kun grunnkretser – inkl. "ukjent kommune", "kjent kommune, men Missing grunnkrets", og "kjent kommune, men ukjent grunnkrets". Da ble tall med kjent kommune tatt med i aggregeringen på fylkesnivå, uavhengig av hva som sto for GK. De med ukjent kommune ble med i landstallet.

### Trengs Reshape av to ting samtidig?

Det må gjøres i to trinn (OG DET FIKK JEG IKKE TIL Å FUNGERE):

- Innlesing: Reshaper til bare én tallkolonne, med en TAB-kolonne som har dobbelt sett kategorier. (ELLER må det være to TAB-kolonner, en for hvert sett?)
- Kubeprod: Bruker det ene settet verdier i (ELLER den ene?) TAB-kolonnen til å opprette en ny tallkolonne med de ønskede verdiene, og setter et filfilter som deretter kaster de nå overflødige radene fra TAB2.

### Eksempel:

Inndata:

År - Geo - Kjønn - Teller\_500 - Teller\_1000 - Nevner\_500 -  
Nevner\_1000

Behovet er altså å lage en kolonne Teller og en Nevner, fordelt på TAB-kolonnen med -500 og -1000.

Innlesing:

Reshape alle fire til én tallkolonne. Hvis det bare er disse fire (som i eksemplet), kan *RESHAPEmeas* stå tom. Det dannes en TAB1 med fire kategorier: T\_500, T\_1000, N\_500, N\_1000.

SE NEDENFOR om videre spec.

(ANM: Det ovenstående fungerte, dvs. laget én tallkolonne og én TAB-kolonne. Men i den videre prosesseringen klarte jeg ikke å få ut riktige tall.)

### **Innlesingsargumenter (INNLESARG), syntaks:**

Skrives uten mellomrom, adskilt av komma.

**Marie og Nora pleier å shoppe funksjoner ved å se på tidligere innlesingsspec'er. Samtlige innlesingsspec'er er listet opp i tabellen HOVEDSTYRING\INNLESING.**

### **Eksempler:**

skip=n	Hopp over n rader øverst i inndatafilen.
sep=","	Filen er kommaseparert.
slettRader=n	Etter innlesing, slett denne/disse radene. Radnummer som i original fil (uten hensyn til Skip).
TomRadSlutt=TRUE	Gjør at inndataene kuttet ved første tomme rad. Hjelper f.eks. mot SSBs kommentarer nederst i filen.
header=FALSE	Hvis øverste rad (f.eks. etter <i>skip</i> ) ikke inneholder var-navn, og dermed ikke skal slettes.

### **Hvis samme inndatafil skal brukes i flere filgrupper (eller flere innlesingsspec'er):**

Systemet nekter i den vanlige prosessen, så vi må sette opp dette i en bakgrunnstabell.

Gå i LAG FILGRUPPE – KAN BRUKE\ORGINNLESkobl – tabellen over koblingene mellom innfiler, spec og filgruppe.

Skriv inn inndatafilens eksisterende FILID på ny rad, sammen med (det nye) filgruppenavnet og DELID. FILID finnes sikrest i (query/tabell)ORIGINALFILERse. (DELID kan være det samme som før, men dette vil ikke automatisk kopiere innlesingsspec'en til den nye filgruppen.)

Klikk Oppdater Alt i skjemavinduet, så dukker filen opp i Tabell 1!

### **For å få brukt innlesingsspec'en om igjen:**

Gå i LAG FILGRUPPE – MÅ BRUKE\INNLESING og kopier spec'en til ny rad. Skriv inn det nye filgruppenavnet, og evt. ny DELID (men den *kan* være den samme som før. DELID er bare unike innen hver filgruppe).

### **Triks for å kopiere mange spec'er:**

Ved innliming av fire kopierte spec'er i tabell INNLESING protesterte Access, «dubletter i et nøkkelfelt» eller noe sånt. De kopierte radene ble automatisk limt inn i en nyopprettet tabell «Innlimingsfeil», og Access sa «Rett opp feilene der, og kopier derfra til target-tabellen». Da fikk jeg skrevet inn nye verdier i alle fire, og kopierte fra «Innlimingsfeil» til INNLESING. Det ble godtatt.

## Bruk av RSYNT- (Statasynt-) punkter:

Se oversikt over aktive punkter i [\\fhi.no\Felles\Forskningsprosjekter\PDB 2455 - Helseprofiler og til\\_PRODUKSJON\DOK\Oversikt over alle RSYNT-punkter.docx](\\fhi.no\Felles\Forskningsprosjekter\PDB 2455 - Helseprofiler og til_PRODUKSJON\DOK\Oversikt over alle RSYNT-punkter.docx)

## 3. Kjør innlesing med disse parameterne:

DOBBELTKLIKK (NB!) «Lag stablet fil»-knappen.

OBS: Egen knapp for 64-bits Access.

Disse knappene er Visual Basic, som Kåre ikke kan. R-versjon er hardkodet – det er en del av path til programmet. Oppdatert sep-17 til vår "frosne" R på Kommunehelsa-disken.

Hvis knappen ikke virker, funker det å kjøre direkte i R. Da kan jeg også se evt. feilmeldinger.

UTDATERT des-2020: Vi er over på versjonshåndtering med Git, som krever at vi kjører fra R/RStudio.

### I RStudio:

Åpne R-prosjekt «khfunctions». Det må være satt opp på forhånd, se egen bruksanvisning ...

Gå i fane Terminal og velg riktig branch i Git (skal som standard være Master eller Main). Sørg for at den er oppdatert (git pull).

Åpne F:\PDB 2455 - Helseprofiler og til\_PRODUKSJON\BIN\ SePaaFil.r

**#NB: Alle kommandoer under krever at linja under kjøres en gang ved oppstart**

```
source('KHfunctions.r')
```

Hvis du får feilmelding: «..... first argument is not an open RODBC channel» må også denne kjøres som oppstart:

```
KHGlobs<-SettGlobs()
```

### *Kommandoer:(De fleste ligger i scriptet)*

FG<-LagFilgruppe("ENEFHIB",versjonert=TRUE) lager versjonert **stablaorg** ("stablet fil"). Her er det foreløpig ikke gjort noen beregninger, det følger i neste trinn.

Denne lagres bare som R-datafil, uten noen csv-kopi, men det kan endres hvis vi vil.

K<-LagKubeDatertCsv(c("NPRKOLS","NPRPSYK","NPRSOMAT","KREFT")) lager datert **kube**, med kopi i CSV (og datert Friskvik) for alle på lista i parentes i kommandoen.

### *Runtime dump av datasettet:*

Enkeltfiler eller stablet fil (avh.av hvilket dumpunkt vi ber om) dumpes ut fra minnet til en lagret fil på ...\\PRODUKSJON\RUNTIMEDUMP\. Enkeltfiler vil ha et løpnummer som skiller dem fra hverandre, mens etter stabling heter filene bare <filgruppenavnet><dumppunktnavnet>.

Mulige punkter: <\\fhi.no\Felles\Prosjekter\KommuneHelsa\PRODUKSJON\DOK\DUMPPUNKTER.docx>

- 1) Sett "dumps" til hvilke punkter og hvilken utfil-type du vil ha.
- 2) Kjør LagFilgruppe med "dumps"-delen av kommandoen inkludert.

Eksempel:

```
dumps=list()           #-> Ingen dump
dumps=list(maKUBE0="CSV",RSYNT2pre=c("CSV","STATA"))
FG<-LagFilgruppe("KEISERSNITT_NH",versjonert=TRUE,dumps=dumps)
```

#### 4. Sjekk resultatet: Skjema LOGGskjema

Dette skjemaet viser dumper og analyser av innlesingen på flere trinn i prosessen, og eventuelle feilmeldinger. Slik er det lett å se på hvilket trinn noe skar seg.

1. Øverst er en oppsummering: Etter litt basisinfo om inndatafilen kommer raden som starter med «INNLES\_OK». **Alle feltene i denne raden skal ha verdien 1** (eller blank, hvis f.eks. den aktuelle variabelen ikke fins).
2. INNLESh viser filen slik den ser ut rett etter rå innlesing. (Det hender at dette er tomt)
3. modINNLESh viser filen etter «skip» og «slettRader» o.l. helt enkle regler.
4. RESHAPEh viser rett etter reshape.
5. FINALh viser sluttresultatet, etter at År og Alder er splittet til «lav» og «høy» og vises som intervaller, det er lagt til en kolonne med FYLKE, og hver VAL-kolonne har fått to flagg:
  - a. «VAL1.f» er årsaken til at det står «NA» dvs. Missing som verdi. Denne tilsvarer (og blir til) SPVflagg.
  - b. «VAL1.a» er antallet rader som er kollapset for å gi tallet i denne raden. Intern QA-variabel.

Innimellom de nevnte tingene er det felter med feilmeldinger fra R, inndelt etter hvor de oppsto.

#### 5a. Faste koder: Endres ikke.

Tabell: KH\_KODER.

Her er en del faste koder for standarddimensjonene definert. De stemmer for det meste med leveransene, men noen ganger matcher ikke inndata denne kodingen. I så fall legger vi inn omkoding fra inndataverdien til KH\_KODER-verdien i tabell KODEBOK (se nedenfor).

#### Felter:

DE L	(som betyr)	KODE, LABEL
---------	-------------	-------------

<b>A</b>	Alder	888_888 Ugyldig, 999_999 Ukjent
<b>Gn</b>	Geonivå	L land, H Hreg, F fylke, K kommune, B bydel, S sone, G grunnkrets, U ugyldig.
<b>K</b>	Kjønn	0 kjønn samlet, 1 menn, 2 kvinner, 8 ugyldig, 9 ukjent.
<b>L</b>	Landbakgrunn	(se kodeliste nederst i filen)
<b>S</b>	Sivilstand	(se kodeliste nederst i filen)
<b>U</b>	Utdanning	(se kodeliste nederst i filen)

## 5b. Kodebok:

Venstremeny i Access: Avsnitt LAG FILGRUPPE – KAN BRUKE.

Det hender filene inneholder verdier som ikke er selvforklarende, og som må spesifiseres. Ulike varianter av blanke, missing osv., non-standard eller skiftende Geokoder, stringverdier med store vs. små bokstaver i ulike årganger ...

OBS: Koding av VALn missing -> 'noe annet' : Se også felt FILGRUPPE/VAL1miss .

### Legge inn ny kode: Tabell KODEBOK

- Identifiser FILGRUPPE, DELID og FELTTYPER (dvs. variabelnavn) koden gjelder for.
- DELID kan oppgis til FELLES. Da gjelder denne raden for alle DELID innen samme FILGRUPPE. Men IKKE sett FELLES hvis det fins flere varianter av samme inn-kode (dvs. som skal bli til én ut-kode). Det kan se ut som en FELLES gjør at andre varianter ikke slår inn?
- Default TYPE er «KB». Noen steder står det «SUB», og da gjøres omkodingen med Regular Expressions, ikke bare ved å bytte ut en fast string. Dersom hele strengen skal omkodes, brukes «KB», men dersom deler av en/flere strenger skal omkodes, brukes «SUB».
- ORGKODE er koden i inndatafil.
- NYKODE er hva den skal kodes om til.

### Regular expressions: Litt ulikt Stata.

- Det er akkurat det som matcher, som byttes ut. Må lage en RE som matcher hele stringen for å få erstattet hele.
- Symboler eller koder for klasser av tegn: \d digit, \s space (inkl. space, tab, newline osv.), og stor bokstav \D og \S betyr «alt som IKKE er» hhv. digit og space.
- Antallsoperator: Det foregående uttrykket matches et antall ganger: {n} nøyaktig n ganger, {n, } n eller flere ganger, {n,m} minst n men max m ganger. Så \d{2} matcher to siffer.
- Erstatt med et av sub-uttrykkene (Stata: regexs(2) ): \n der n er 1-9. Matcher parentes nummer n i det komplette RE. (Brukes f.eks. i «NYKODE»-kolonnen)

### Sjekke hvordan det gikk: Tabell KODEBOK\_LOGG\_se

(i avsnitt LAG FILGRUPPE – MÅ BRUKE)

(OBS: Kjempetabell, det tar mange sekunder å åpne den og filtrere den. Selv det å åpne filter-nedtrekksmenyen tar et par minutter første gang. Vær litt tålmodig.)

Her finner du igjen Filgruppe, Delid, Feltype.

- ORG er koden i inndata.
- KBOMK er koden den skulle kodes til.
- OMK er koden den faktisk ble omkodet til.
- OK flagger de radene der dette ikke stemmer overens (=0), så man kan filtrere fram disse.

**Gjelder dette nå, etter at "versjonering=TRUE" er satt som default?**

**MEN SÅ MÅ FILGRUPPEN LAGES I R!**

Kjøring i Access (med knappen) viser om det ble vellykket, men lager ikke det endelige sluttproduktet som plukkes opp av kubeproduksjonen!

Se under 3)

## **SPESIELLE TRIKS OSV**

**Gjenbruke en innlesingsspec: Samme innfil i to filgrupper**

Det går an, men er litt plundrete. Systemet nekter å bruke samme filnavn om igjen.

- Les FILID fra ORIGINALFILER.
- Gå i ORGINNLESkobl og finn FILID-nummeret der.
- Der kan jeg nå *skrive inn samme Filid* og kopiere resten. Det vil bli en ny KOBLID, og vi skriver altså ny FILGRUPPE.
  - o OBS: DELID kan ikke kopieres her – det er ikke entydige navn.
- Gå i Hovedstyring/INNLESING og kopiér innlesingsspec'en. Skriv inn den nye Filgruppen, og evt. nytt Delid (kan hete det samme).

**Begrense tidsserien i én ORGfil, når to ORGfiler har overlappende data: Bruk KODEBOK**

Eksempel: ORGfiler til ABORT\_NH i 2021.

- Forrige leveranse dekket hele tidsserien 1979-2019.
- Nyeste leveranse har 2016-2020.

Da bruker vi alle årgangene i nyeste fil, for å få med oppdateringer i registeret. De overlappende årene i eldste fil må tas ut før de to filene stables til filgruppe.

- Sørg for at de to filene har hver sin DELID.
- Sett i KODEBOK, på DELID for eldste fil: I variabel AAR skal 2016 (og 2017 osv) kodes til "-" (strek). Da slettes disse årene i innlesingen.

## **Kubeproduksjon (dvs. all output)**

Parameterne defineres først i to Access-tabeller, så må vi ut i R og kjøre et script bitvis.

/home/ybk/Downloads/Bruksanvisning.docx

Venstremeny i Access: Avsnitt LAG KUBE – MÅ BRUKE.

OBS: Kjører du LagKube om igjen fordi Filgruppen er kjørt på nytt, legg merke til om filgruppen leses på nytt!

```
> k<-LagKubeDatertCsv("BARN_S0SHJELP")
```

```
BATCH: 2020-12-18-14-33
```

Hentet FIL BARN\_S0SHJELP fra BUFFER (22903 x 16)

-DVS. GJENBRUK AV FORRIGE (GALE) FILGRUPPE!

i motsetning til slik du ønsker:

```
> k<-LagKubeDatertCsv("BARN_S0SHJELP")
```

```
BATCH: 2020-12-18-14-40
```

Lest inn FIL BARN\_S0SHJELP (22920 x 17), batch=2020-12-18-14-16

BUFFER kan tømmes, se i kap. "Problemløsning/feilsøking".

NB: Kuber som skal direktestandardiseres i Stata (brukes i NH) må lages etter en egen, litt mer kompleks metode: Først lages en "pre"-kube i R-løypa, og deretter kjøres standardiseringsscriptet i Stata. Parametre for begge trinn i prosessen spesifiseres i tabell Kuber som forklart nedenfor.

## 1. Tabell TNP\_PROD

Definere teller og nevner.

Felter:

TNP\_NAVN Vi lager et utfilnavn. Det er ofte lik tellerfil-navnet.

TELLERFIL Filgruppenavnet (se neste punkt).

NEVNERFIL Må gis når nevner ikke er inkludert.

OBS: Disse to ...FIL-parameterne kan peke til **virtuelle** filer definert i tabellen FILFILTRE (venstremeny avsnitt LAG KUBE – KAN BRUKE). I så fall dannes disse runtime. Det er navnet fra feltet FILVERSJON som skrives her.

NYEKOL\_RAD Nye kolonner fra rader: Lag ny kolonne med <navn> ut fra verdiene i <denne variabelen> når <denne tab-kolonnen> har <disse verdiene>.

Syntaks:

Nevner=ANTALL{TAB1=="aperinat\_fodt\_500\_who"|

TAB1=="aperinat\_fodt\_1000\_who"}

ANM: Det ovenstående fungerte ikke. Det lot seg kjøre, men tallene som kom ut var alt for høye (det skjedde en utilsiktet collapse). Kåre sa i forbifarten at «dette er egentlig laget for noe annet», så det er mulig at metoden ikke egner seg for min «reshape wide»-problemstilling.

...

TELLERKOL Hvilken kolonne i tellerfila er den telleren som skal brukes. Se navnet i tabell Styring/FILGRUPPER, avsnitt 1 ovenfor. VAL1navn etc.

NEVNERKOL Tilsvarende. Innfilene til denne delen av prosessen er jo laget av systemet, så det er spesifisert tidligere hva variablene heter. Kan oppgi <navn> fra NYEKOL\_RAD-kommandoen ovenfor!

## 1 b - Eventuelt: Tabell FILFILTRE

Inneholder enda flere mekanismer for å lage teller/nevner under kjøringen. Brukes f.eks. der middelfolkemengde er nevner. Denne er ikke dokumentert fra starten av, så etter at Kåre forsvant ut, baserer bruken seg på erfaring og prøving-og-feiling ...

FILVERSJON Vi lager et navn. Dette brukes i tabell TNP\_PROD som TELLERFIL eller NEVNERFIL.  
ORGFIL Filgruppenavnet.

## 2. Tabell KUBER

Definerer hva som skal beregnes og hvordan utdataene skal se ut. Inneholder felter både for ren R-løypebehandling og for direktestandardisering i Stata (for NH-indikatorer). Stata har noen helt egne felter, men bruker også mange av de vanlige. Der syntaks i feltet er ulik for R- og Statabehandling, er det angitt:

### Felter:

KUBE_NAVN	Vi lager et utfilnavn.
MODUS	NH eller KH. Styrer hvor utfiler havner.
TNP	Skriv TNP_NAVN fra forrige tabell.
Stata_dropp	Feltet er utdatert (nov-2018). Styres nå ved å oppgi kubenavn i dialogen for start av statascript Direktestandardisering (gjelder bare NH). STANDARDISERING I STATA: "1" hvis denne skal utelates fra kjøringen. Trenger bare fylles ut hvis REFVERDI_VP er lik 'D'.
AAR_START	Default 0. Legges inn for å begrense hvilke årganger som blir med.
AAR_STOP	Default 9999. Ditto. Siste år som inkluderes. <b>OBS: Man må enten bruke disse to, eller "AAR" nedenfor. Hvis det står noe i AAR, virker ikke AAR_STOP.</b>
AAR	Enkeltår som skal tas ut! Syntaks: -[2006,2007]. MFR: -[0]
GEOniv	Hvilke som skal kjøres ut. B, K, F, H, L.
KJONN	Ditto.
KJONN_0	Er kategorier som skal brukes senere i prosessen, men som ikke skal vises i kuben. Det kan f.eks. være at vi vil vise andel «høy» av noe og at nevneren skal være «høy» + «lav», da må «lav» også være med litt videre selv om vi ikke vil vise det i kuben. (Skriv inn alle felter som inngår i senere beregning,



	også de som <i>skal</i> vises ut.) <b>Det gjelder alle kolonnene som slutter på «_0».</b>
ALDER	Som KJONN. Tom -> alt som ligger i fila. «ALLE» -> bare aldersgruppen ALLE, som er definert annet sted. Ellers: Må spesifisere alle grupper som skal vises i kuben.
...	<i>OBS: En PRE-fil for direkttestandardisering (Norges helse, se nedenfor) kan ikke inneholde overlappende aldersgrupper (f.eks. 0-4 og 0-19), den må bare ha de gruppene som skal brukes i standardiseringen (som regel streitte 5-årsgrupper 0-4, 5-9 osv.).</i>
(TAB1formel -2 -3	Finnes ikke i løypescriptet)
TAB1 -2 -3	Tom -> alt som ligger i fila. Ellers: Spesifiserer hvilke kategorier fra hver TAB som skal være med, adskilt med komma. Kan angi kategorier som skal <i>droppes</i> i R-løypa: -[R03] SYNTAKS FOR STATA: AARSAK~keep~SELVMORD - kan angi <i>keep</i> eller <i>drop</i> .
TAB1_0 etc.	Se «KJONN_0» ovenfor.
RATESKALA	Bestemmer om rate blir per 100, per 1000 osv.
MOVAV	Bestemmer lengden på glidende snitt-perioden.
REFVERDI_VP	Bestemmer standardisering. P -> Indirekte standardisering (for KH), D -> Direkte standardisering (for NH), V -> ingen std.
REFVERDI	Hvilket geonivå som er ref. OBS ulik syntaks: For KH: Alltid landet i hht. nåvær.beslutning. <i>GEOniv</i> == 'L' For NH: <i>bef1jan2012</i>
NESSTARTUPPEL	Hvilke måltall som skal være med. OBS ulik syntaks: <b>For R-løypa:</b> T teller (rå antall), N nevner, RATE (ikke-std.), MEIS (std.rate), SMR (Norge=100, kan være med selv om ikke std.tall), MT måltall, SN sum nevner, ST sum teller, RN Rate.n (antall år det aktuelle tallet er gj.snitt av – blir lavere enn MOVAV hvis en kommune mangler tall for ett eller flere av årene i perioden). <b>For Stata:</b> antall, crude, std, smr,nevner.
EKSTRAVAR	For eksempel Dekningsgrad.
DIMDROPP	Hvilke dimensjoner skal droppes. UTD, LANDBAK, SIVSTAND, evt KJONN er typisk.
PRIKK_T	Prikkeregler. Dette antallet eller lavere prikkes. (TELLER)
PRIKK_N	Ditto. (NEVNER)
STATTOL_T	Grense for å skjule serier. «Hvis mer enn 50 % av punktene er basert på <dette tallet> eller færre tilfeller, skjules hele serien». (OBS: Feil formulert i metadata fram til feb-2016: "færre enn <dette tallet>".)
OVERKAT_ANO	<b>For R-løypa:</b> Hvilken kategori skal bevares ved naboprikking. «KJONN=1» gjør at Kvinner og Samlet prikkes. Erfaring: Hvis det skal naboprikkes etter en ekstradimensjon (f.eks. INNVKAT, UTDANN), må var-navnet stå her (med angivelse av overgruppe).

**For Stata:** Styres i *Stata\_nabopr....*

VERSJONFRA

VERSJONTIL

Kan brukes til å hindre at en kube kjøres ut, ved å sette «til»-dato før dagens.

(ANM: Jeg fikk ikke til å lage en ny linje her med samme kubnavn, selv om jeg satte «til»-dato på den gamle og «fra»-dato på den nye, og hadde forskjell i «nesstartuppel»-feltet. Måtte ha avvikende kubnavn for at linja skulle bli godtatt.)

SLUTTREDIGER RSYNT-punkt, kan legge inn Statakode. (Ligger ikke helt til slutt i løypa, det skjer flere ting etter dette punktet).

RSYNT\_POSTPROSESS RSYNT-punkt like før ferdig kube lagres. OBS: Variabler i datasettet er fremdeles inkludert hjelpevariabler.

FHP\_geoniv I hvilke FHP brukes kubene? (B, K, F). Kun til info.

### Fylles ut når kubene skal direktestandardiseres i Stata (brukes i NH):

Stata\_spesdim Navnet på spesialdimensjonen, f.eks. diagnose, dødsårsak el.l.

Stata\_KJONN Oppgi alle grupper, også såkalte implisitte undergrupper som ikke skal vises i kubene (de som skal vises angis i KJONN).

Stata\_ALDER Oppgi alle grupper, også såkalte implisitte undergrupper som ikke skal vises i kubene (de som skal vises angis i ALDER). Eks.: Hvis 0-120 og 45+ skal vises, er 0-44 implisitt, og må oppgis her.

Stata\_TAB1 De verdiene av spesialdimensjonen som ønskes vist i kubene. Også verdier som må lages. Syntaks: Se nedenfor

Stata\_ford Statistisk fordeling: "binomial" eller "poisson"

Stata\_antall Hvilke Geonivå skal ha Antall - Tillatt er L, H, F, kombinasjoner av disse (uten skilletegn), eller tomt felt (som gir antall på alle nivå). Så lenge vi ikke naboprikker på GEO, bør kanskje antall begrenses til 'L'.

Stata\_naboprKjonn Rekkefølge for bevaring av nabokategorier, Kjønn.

KJONN~niva1{0,1,2 }.

Avansert syntaks: Kan angi avvikende rekkefølge ved angitte vilkår. KJONN~niva1{0,1,2 (2,0,1\_if\_ICD==C50)

(2,0,1\_if\_ICD==C53) (1,0,2\_if\_ICD==C61)}

Stata\_naboprAlder Rekkefølge for bevaring av nabokategorier, Aldersgrupper.

Alle gruppene ("trekantene" i prikkelogikken) må oppgis.

Komplekst eksempel (Reseptreg.):

ALDER~**niva1**{0\_74,0\_79,75\_79}**niva1**{20\_79,0\_79,0\_19}**niva2**{0\_74,45\_74,0\_44}**niva2**{0\_14,0\_19,15\_19}**niva3**{0\_44,25\_44,15\_24,0\_14}**niva4**{15\_24,15\_19,20\_24}. Hele rekkefølgen har betydning når flere enn tre kategorier (se nedenfor).

Stata\_naboprSpes Rekkefølge for bevaring av nabokategorier, (se de foregående).

**NB, hele rekkefølgen har betydning. I eksempelet "INNVKAT~{0,2, 3,20 }" har «0» høyest prioritet, og «20» har lavest.**

### Syntaks for Stata\_TAB1:

Hvis man bare ønsker et sub-sett av verdiene i en spesialdimensjon, eller man ønsker å avlede en ny verdi fra de eksisterende, angis dette i denne kolonnen.

- Angi ønskede verdier av spesialdimensjonen(e) slik:  
spesialdimnavn~verdi. Eksempel: TRINN~10
- Dersom flere verdier, skill med komma, ingen mellomrom. Eksempel:  
TRINN~10,8
- Dersom flere spesialdimensjoner, skill med # . Ingen mellomrom.  
Eksempel: SPRSMLID~307#TRINN~10
- Dersom ny aggregert verdi (f.x. i tillegg til de eksisterende): Listes opp sammen med de andre verdiene, men etterfulgt av "=" og deretter angivelse av hvilke eksisterende verdier det skal aggregeres fra (skilt med "&") og så "@" og angivelse av *hvordan* teller (T) og nevner (N) skal aggregeres. Tenk nøye gjennom sistnevnte!  
Eksempel på syntaks der man ønsket ny verdi, Vdg\_tot, for alle som har mer enn grunnskole (eksempel hentet fra 2014-filen):  
"Vdg\_tot=Uni\_Høg\_kort&Uni\_Høg\_lang&Videregående@sumT\_meanN"
- Mulige aggregeringsmetoder: sum, mean.

### HVIS INDIKATOREN SKAL VISES I BAROMETERET: Tabell FRISKVIK

Her er parametre for å lage tabellen s.4 og barometerfigurene: både styre hva R-løypa skal slice ut til å vises, og hvordan Statascript skal gjøre beregningene for å produsere dem. Her angis også om indikatoren skal ha setning på side 1.

Ny, separat rad for hver indikator/geonivå/årgang.

OBS: Stata direktetestandardiseringsscript (for NH) sjekker alltid om aktuell kube (KUBE\_NAVN) ligger i tab FRISKVIK med AARGANG==inneværende profilårgang og MODUS=="F". I så fall må hele raden være utfylt, og det slices ut en Friskvikfil.

#### Feltene:

INDIKATOR	gi den et navn.
MODUS	dss. geonivå, F K B
AARGANG	profilårgang
FILNAVN	
KUBE_NAVN	Kuben tallene tas fra
AARh	Hvilken årgang data skal vises
KJONN	Hvilket kjønn skal vises
ALDER	Hvilken aldersgruppe skal vises (hvis fila har fler). <i>[mars-22: tom celle, eller "-", for en fil med flere ald.grp -&gt; alder 0 ble valgt (Indikatorer e0K_NH_LF og e0M...).]</i>

EKSTRA\_TAB      Hvilken verdi fra en ekstradimensjon skal vises. F.eks.:  
 ANTALL\_GANGER=='engangellerflere' & SOES=='0'.  
 Her er ANTALL\_GANGER TAB1 og SOES TAB2.

UTDANN            (tilsv.)  
 SIVST             (tilsv.)  
 LANDBAK          (tilsv.)

ALTERNATIV\_MALTALL    Hvis noe annet enn <> skal brukes. Eks.  
 DEKNINGSGRAD for indikatoren Drikkevann/dekningsgrad.  
*(De følgende trenger ikke fylles ut ved kubekjøring – de brukes av Stata i flatfil-  
 og barometerproduksjonen)*

KOMMENTAR

tema\_bm            Under hvilken temaoverskrift i tabellen?  
 tema\_nn            (tilsv.)  
 temaId             Denne temaoverskriftens nummer  
 LNRtabells4        Radnummer i tabellen side 4  
 ekstrafil\_loep      Dette er en lite brukt parameter, men behøves dersom man  
 har en kombinasjon av

- a) indirekte standardisering, og
- b) fast referanseår

Frem til januar 2019 har dette kun gjelt RFU-tall, fylkesprofiler.

**Forklaring:** Alle statistiske tester, både av bydels-,  
 kommune- og fylkestall, skal være mot landstall samme år.

Ved indirekte standardisering (se pkt. a) betyr dette at det  
 trengs «predteller»<sup>1</sup> basert på de aldersspesifikke ratene på  
 landsnivå samme år. Siden de ordinære Friskvikfilene for  
 fylkesprofiler kjøres ut med fast referanse-år (2012), vil alle  
 predtellere i disse filene være basert på de aldersspesifikke  
 ratene på landsnivå i 2012, hvilket ikke er det som trengs.  
 Situasjonen krever at det kjøres ut ekstra filer med løpende  
 standardisering. Altså: Riktig MEIS<sup>2</sup> må komme fra de  
 ordinære filene standardisert mot fast standard-år (for at MEIS  
 i profilene skal stemme med MEIS i statistikkbanken), mens  
 rett predteller (og følgelig Forholdstall og figurverdier) må  
 komme fra ekstrafilene standardisert mot løpende standard-  
 år.

**I praksis:** Man lager som regel ordinær fil og ekstrafil ved å  
 spesifisere henholdsvis

«GEOniv=='L' & AARI=='2012'», og  
 «GEOniv=='L'»

i parameteren REFVERDI i tabellen KUBER . Navnet man gir  
 ekstrafilen må slutte på «\_L» (som i Løpende), f.eks.

<sup>1</sup> Forventet antall tilfeller i bydelen, kommunen eller fylket

<sup>2</sup> Aldersstandardisert rate

ROYK\_NH\_5\_16\_44\_L. Det er navnet på ekstrafilet som skal inn i dette feltet (ekstrafile\_loep).

varname_bm	Indikatornavn i tabellen
varname_nn	(tilsv.)
Side_1	Skal indikatoren ha setning side 1? (avkrysningsboks) - krever at ogs� nesten er kryssset av.
Side_4	Skal indikatoren v�re med p� side 4? (avkrysningsboks) - Her kan man ta ut igjen en indikator som likevel ikke skal v�re med, uten � m�tte slette all info.
Enhet	Tekst til kolonnen Enhet i tabellen.
periode_bm	Beskriver dataenes tidsperiode til fotnote s.4
periode_nn	(tilsv.)
kreftkorr	<a href="#">Korrigerings for at teller og nevner allerede var periodesummer fra leverand�r (se Kreftreg.) og likevel er blitt ganget opp med antall �r i perioden i R-l�ypen. Eksempel: Dagens (2018) praksis er at Kreftregisteret leverer grunnlagstall i form av 10�rige glidende summer av teller og nevner. Anta at dette i v�r preprosessering er blitt tolket som ett�rige tall, og derfor (feilaktig) ganget opp teller og nevner med 10. Da kan man korrigerer dette ved � dividere igjen med 10, ved � skrive «10» i &lt;kreftkorr&gt;. Skriver man «1» eller ingenting, s� skjer det ikke noen korreksjon.</a>
ekskluderteAar	
FN_alder	Beskriver dataenes aldersgruppe til fotnote s.4
forklaring_bm	Beskriver dataene, til fotnote s.4
forklaring_nn	(tilsv.)
SammeFotnSom	Angi <i>LNRtabells4</i> for indikatorer som skal ha samme fotnote
testtype	Hvilken statistisk test? Chiang, binomial, poisson, ELLER kommentar om at testtype er hardkodet i scriptet.
feilvei	Sett JA hvis x-aksen i barometeret skal snus s� sm� verdier vises som "bedre". Ellers sett NEI.
skalaforsskyv	
fargetPrikk	Sett JA hvis barometersymbolet skal "verdilades" med r�d/gr�nn prikk. Ellers sett NEI for � f� hvit prikk med signifikansmerking.
enFastDesimal	<a href="#">JA, NEI, ingenting. Vanligvis: «NEI» eller ingenting. Da vil tall&gt;10 komme uten desimaler, mens tall lavere enn 10 vil f� gradvis flere desimaler. Alle tall under 100 f�r, p� denne m�ten, samme antall «significant digits». Dette er den konvensjonelle m�ten � presentere tall p�. Skriver man «JA» vil alle tall komme med akkurat �n desimal.</a>
fleksSkala	<a href="#">JA, NEI, n/a</a>
topp10	<a href="#">JA, NEI, n/a</a>
ant_perioder	<a href="#">Brukes for Ungdata</a>
TidsserieURL	URL for tidsserie i KH/NH. Tas med i flatfile for opplasting til SQL-databasen. ( <a href="#">VAR KUTTET i Develop-tabellen</a> )

### **(Feilmelding ved kjøring:**

Når det kjøres ut en Friskvikfil i R-løypa, kommer en feilmelding:

```
*KHFEIL!! FEIL I FRISKVIKFILTER (GEOniv %in% GEOfilter) & ALDER=='0_120' &  
AARh==2017 & KJONN==2 & KMI_KAT=='Overvekt' GIR bare 475 / 487 rader!
```

Denne kan ignoreres: Den skyldes at løypa ble tilpasset antall kommuner+fylker+bydeler+land når det ble utviklet. Nå er dette antallet redusert ved kommunesammenslåinger. )

### **3. FOR Å KJØRE:**

Gå i Rstudio.

Script: SePaaFil.r (samme som vi har åpnet før, se 3.)

Kjøres bitvis.

#### ***#LAGE EN NY KUBE (datert), med kopi i csv***

```
K<-LagKubeDatertCsv("RESEPT_PREstatb_NH")
```

(«K<-» -delen gjør at output går til et objekt i stedet for til skjermen.

Sparer mye skjermutskrift, og gjør det mulig å se på resultatet i R etterpå.)

#### ***For å få ut en runtime dump av datasettet:***

Se beskrivelse ovenfor, under Innlesing "3. Kjør innlesing..."

Eksempel:

```
k<-LagKubeDatertCsv("TEST_SESJON1_NH",dumps=dumps)
```

#### ***For å se på en fil (som i Statas datafilvindu):***

```
FG<-FinnFilT("NPR")
```

```
subset(FG,AARh==2012 & GEO=="0214" & KJONN==1)
```

ELLER/OG:

View(FG) -ja, med stor V.

Kan lage et subset først, da " utvalg <- subset(FG,...) " og View(utvalg)

#### ***UTFILER HAVNER etter hva som er satt i «Modus» - NH eller KH.***

Står i «ferdig»-meldingen etter en vellykket kjøring i R.

```
"F:/Forskningsprosjekter/PDB 2455 - Helseprofiler og til_/PRODUKSJON/  
PRODUKTER/KUBER/NORGESHELSE/DATERT/csv//FODSELSVEKT_NH_2015-06-22-  
11-37.csv"
```

### **4. PROBLEMLØSING/FEILSØKING**

#### **Ta runtime dump:**

Det er ofte nyttig å ta ut runtime dump på ulike steder. Det kan avsløre hvor i prosessen ting går galt, og gi assosiasjoner til hvilken parameter som må endres.

Kommando LagFilgruppe eller LagKube med parameter "dumps", se i SePaaFil.r.  
Se liste over punktene i [\\fhi.no\FellesForskningsprosjekter\PDB 2455 - Helseprofiler og til \\_PRODUKSJON\DOK\DUMPPUNKTER.docx](http://fhi.no/FellesForskningsprosjekter/PDB_2455_-_Helseprofiler_og_til_PRODUKSJON/DOK/DUMPPUNKTER.docx)

### **Skrive selve filgruppen til en CSV-fil, direkte fra R:**

Etter at LagFilgruppe() er kjørt, ligger filgruppen som R-objektet FG. Kan evt. leses fra disken med `FG<-FinnFilt("EIERSTATUS")`.

Eksportkommando (filnavnet trenger ikke eksistere fra før):

```
write.table(FG, file='F:\\Forskningsprosjekter\\ PDB 2455 - Helseprofiler og til _\\  
PRODUKSJON\\RUNTIMEDUMP\\EIERSTATUS_FG.csv', sep=';', na="", row.names =  
FALSE)
```

### **Ryddde buffer:**

Opplevd: R har en buffer hvor resultat-«filer» lagres for gjenbruk. Ved vellykket kjøring ryddes den til slutt, men ved kræsj skjedde det at feil-outputfila ble liggende der. Ved neste kjøring ble ikke riktig filgruppe lest inn, den gamle med feil ble bare kjørt inn om igjen.

Sjekke bufferen:

```
> names(BUFFER)          viser hvilke «filer» som ligger der.
```

Slette ting:

```
> BUFFER$navn <- NULL sletter «navn»
```

```
> BUFFER <- NULL      sletter alt. Koster litt tid, for da må BEF_Gka leses inn  
om igjen
```

(når den trengs).

### **Friske opp globale bakgrunnsdata:**

```
> Khglobs              Leser dem inn på nytt.
```

Men for å være helt sikker, er det lurt å restarte R helt! Og da uten å lagre workspace ...

### **Uforklarlig prikking av sammenslåtte Geo?**

Sjekk at alle VAL-tallene er satt som "sumbar" i tabell FILGRUPPER. Hvis ikke, kan jo ikke tallet for to gamle kommuner summeres til den nye – og da vil prikkerutinen slette resten av tallene på samme linje.

### **Eksempel på kompleks indikator: INNTULIKHET**

Inndata: Måltallene Ginikoeffisient og P90/P10 kommer i hver sine filer, som må ha ulike innlesingsspec. Tallene er ikke aggregerbare.

Ønsket kube: Med de to måltallene i hver sin kolonne.

/home/ybk/Downloads/Bruksanvisning.docx

I Filgruppen har vi én verdikolonne, og en ekstradimensjon som forteller hvilken rad som er Gini og hvilken P9010. Dette må splittes opp i kubeproduksjonen, og dimensjonen droppes.

## Sammenhengen mellom fil-elementene (til bruk ved f.eks. kval.kontroll)

Dersom det ikke har vært endringer i hvordan en filgruppe ser ut, går det an å kjøre ut en ny versjon av fjorårets kube uten å gå den tunge veien til versjonsarkivet etc. Man bare manipulerer hvilken dato de ulike originalfilene er tatt inn og ut av bruk. Men da må man kjenne til fil-sammenhengen ...

I tabellen brukes notasjon: TABELLNAVN - FELTNAVN.

<b>Fil:</b>	<b>Navnet defineres hvor:</b>	<b>Fil som er grunnlag:</b>
Ferdig kubedatafil til statistikkbanken ("kube", output fra R-systemet)	Defineres i tabell KUBER - KUBE_NAVN.	Bygger på en TNP-fil, angitt i KUBER - TNP.
TNP-fil	Defineres i tabell TNP_PROD - TNP_NAVN.	Bygger på en tellerfil, angitt i TNP_PROD - TELLERFIL, og evt. en nevnerfil angitt i - NEVNERFIL.
Tellerfilen = en filgruppe	Defineres normalt i tabell FILGRUPPER - FILGRUPPE, som regel i skjemaet STYRING\StyrFILGRUPPER øverste tabell. (Kan evt. defineres i FILFILTRE, dvs. lages runtime. Det er mer aktuelt for nevnerfilen.)	Bygger på én eller flere inndatafiler (ORGfiler), som angis i - FILNAVN.  Disse har hver sin IBrukFra og IBrukTil-dato.

Både ferdig kubedatafil, TNP-fil og tellerfilen (filgruppen) har ofte SAMME NAVN i Access. Det ødelegger "selvforklarende"-prinsippet, men er greiest i daglig bruk.

## Statusinformasjon om hver indikator

Prosess: Innsamling av data, innlesing av originalfiler -> filgruppe, produksjon av kubedatafiler/Friskvikfiler, kvalitetskontroll av output. Hvor langt hver indikator er kommet, kan leses ??



Tabell KUBESTATUS (KH2020\_KUBESTATUS etc) viser prosessen med kvalitetskontroll av ferdig kube.

## Kodeverk og systemverdier

### Interne koder for de ulike elementene (kolonnene) i en tabell:

Gn = Geonivå, Y = År, osv.

#### Se tabell KH\_DELER

### Landbakgrunn, koder i KH\_KODER:

KODE LABEL

0	ALLE
1	Europa unntatt Tyrkia
2	Afrika
3	Asia med Tyrkia
4	Nord- Amerika
5	Sør- og Mellom- Amerika
6	Oseania
7	Statsløse
8	Uoppgitt
9	Andre
88	Ugyldig
-1	Norge
100	Innvandrere
1B	Europa unntatt Tyrkia_innvand
1C	Europa unntatt Tyrkia_norskf
2B	Afrika_innvand
2C	Afrika_norskf
3B	Asia med Tyrkia_innvand
3C	Asia med Tyrkia_norskf
4B	Nord- Amerika_innvand
4C	Nord- Amerika_norskf
5B	Sør- og Mellom- Amerika_innvand
5C	Sør- og Mellom- Amerika_norskf
6B	Oseania_innvand
6C	Oseania_norskf

### **Sivilstand, koder i KH\_KODER:**

KODE LABEL

0	ALLE
1	Ugift
2	Gift
3	Enke/ enkemann
4	Skilt/ separert
5	Annen
8	Ugyldig
9	Ukjent

### **Utdanning, koder i KH\_KODER:**

KODE LABEL

0	ALLE
1	Grunnskole
2	Videregående skole
3	Universitets/ høyskoleutdanning
4	Annet (ingen/ uoppgitt)
8	Ugyldig
9	Ukjent
23	VidergåendeOgHøyere
123	Oppgitt

(NB: Her bruker vi ofte 3 = Univ/Høy, kort og 4 = Univ/Høy, lang. )