

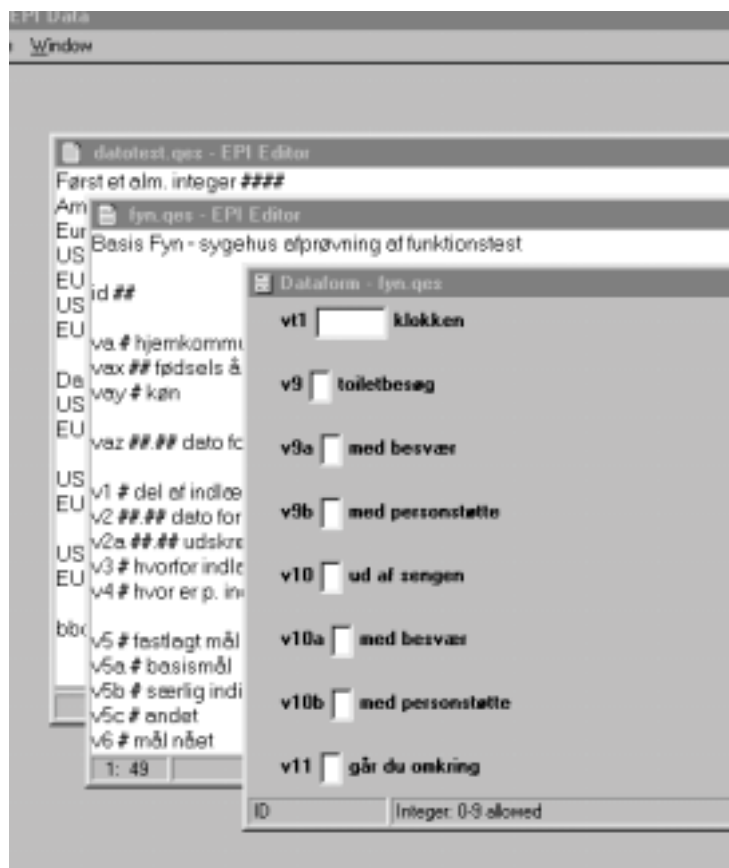
# EpiData

## Introduktion



### Opbygning af skemaer, indtastning og fejlretning af data.

- Formulere spørgsmål i skemaer
- Omsætte spørgsmål til variable, der kan edb bearbejdes
- Indlæse data til edb bearbejdning
- Minimere fejl i data (kontrolleret indtastning)
- Rette data på en måde, der senere kan dokumenteres
- Udarbejde datadokumentation, der opfylder kravene fra udvalget vedr. videnskabelig uredelighed
- Foretage grundlæggende beskrivelser af data
- Introduktion til programmet EpiCalc 2000.
  - Confidensintervaller og statistisk testning ved direkte indtastning
  - Stikprøve størrelse beregning



© Jens M. Lauritsen

Udgave 1.0. (4. Dec. 2000)

**Om denne note:**

1. udgave 2000 (revideret pr 21.nov.2000)

Noten er indlagt på [www.epidata.dk](http://www.epidata.dk) til personligt brug. Noten må printes i sin helhed eller dele deraf til **personligt brug**. Hvis nogen ønsker at bruge hele noten eller dele af noten til undervisning, udgivelse eller distribution skal sådan brug anmeldes skriftlig til forfatteren og Copy-Dan.

Notens indhold svarer til et kursus på 4-8 timer afhængig af deltagernes forudsætning og hvor meget deltagerne har læst på forhånd. Ved 4 timer forudsættes at de første generelle dele er læst forinden.

© Jens M. Lauritsen. 2000. JM.Lauritsen@dadlnet.dk

ISBN: 87-987843-2-3 (trykt udgave)

ISBN: 87-987843-3-1 (elektronisk udgave)

**EpiData** - krav til computeren:

**PC:** Windows 95/98/2000/NT. Processor: 486, pentium eller hurtigere. Ram ingen særlige krav.

**Machintosh:** Mac G3 233 mghz, system 8.1 og 128 Mb RAM, samt emulator RealPC og windows95. (Muligvis også andre kombinationer).

Programmet er tilgængeligt fra internet og koster ikke noget. Se [www.epidata.dk](http://www.epidata.dk)

**Nogle nyttige internet sider om Stata, biostatistik eller hjælpeprogrammer:**

Statistik rutiner og uddybende forklaringer af forskellige analysetyper i statistikprogrammet Stata: ("Resources to Help you Learn and Use Stata") <http://www.oac.ucla.edu/training/stata/>

Introduktionsnote inkl. øvelsesdatasæt til Stata på dansk: [www.bola.suite.dk](http://www.bola.suite.dk)

Instruktionsrutiner og forklaring af analyse af forskellige datatyper, inklusive øvelsesdata. Er baseret på EpiInfo, men principperne er generelt gyldige uanset statistikprogram:

<http://www.sjsu.edu/faculty/gerstman/EpiInfo>

Enkle manualer [http://mkn.co.uk/help/extra/people/Brixton\\_Books](http://mkn.co.uk/help/extra/people/Brixton_Books)

Indtastningsprogram EpiData (gratis) til win95/98/NT/2000: første udgave aug./sept. 2000.

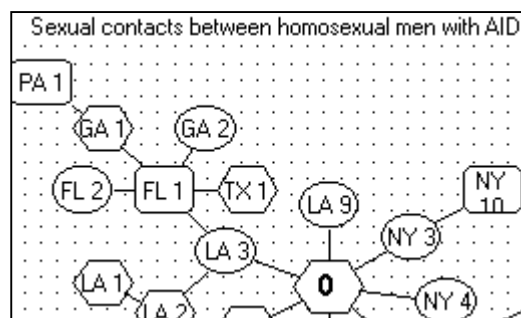
Se [www.epidata.dk](http://www.epidata.dk)

Hjælpeprogrammer (gratis). Blandt andet en nyttig "lommeregner" (Epicalc 2000) og et simpelt program til at tegne flowcharts og diagrammer (EpiGram). Figuren nedenfor er tegnet med EpiGram

<http://www.myatt.demon.co.uk/>

Uddrag af diagram tegnet med EpiGram:

(prikket baggrund findes ikke på udskrifter)



## Indholdsfortegnelse

Forord .....	5
Oversigt – indsamling, oparbejdning og analyse af data. ....	7
Fra ide til skema. ....	8
Hvor mange skemaer ?.....	9
Indhold i det enkelte skema. ....	9
Øvelse 1 Præcision af formuleringer.....	10
Svarmuligheder .....	10
Indsamling af data. ....	10
Fra skema til indtastning.....	11
Antal og typer af variable ?.....	11
Gruppering af data - NEJ !!.....	11
Styringsvariable. ....	11
Kvalitet af en oplysning - “kontrolvariable”.....	11
Dokumentation. ....	14
1. Mulighed for at finde tilbage til originalmateriale.....	14
2. Kvalitetskontrol, dokumentation og arkivering. ....	14
Datadokumentation.....	15
Øvelse 2 Beslut hvilken delmængde af skemaerne du vil bruge til analysen.....	17
Forbehold som skal huskes under dataanalysen. ....	17
2 Konkret indtastning med EpiData .....	19
Øvelse 3 Download af flow chart for EpiData. ....	19
Øvelse 4 Download og Installering af EpiData. ....	19
Øvelse 5 Indstilling af brugen af EpiData. ....	19
Flow Chart for arbejde med EpiData. ....	20
Øvelse 6 Start EpiData uden brug af mus. ....	21
Øvelse 7 Start EpiData .....	21
Øvelse 8 Hvilke dele af flowchartet svarer til "work process" toolbar ? .....	21
Øvelse 9 Arbejd dig igennem EpiTour.....	21
Øvelse 10 Første "reelle" datasæt i EpiData.....	22
Øvelse 11 Opbyg en definition til indtastning af data. ....	23
Øvelse 12 Navngivning af variable. ....	23
Øvelse 13 Navngivning af variable - first word. ....	23
Øvelse 14 Opret en datafil ud fra hdur.qes.....	24
Øvelse 15 Indtastning af data - uden kontrol og backup kopi. ....	24
Øvelse 16 Ny datafil - med simpel range kontrol under indtastning.....	24
Øvelse 17 Indtastning af data - med range kontrol og backup kopi. ....	24
Øvelse 18 Tilføj labels til indtastning. ....	24
Øvelse 19 Indtast igen med labels. ....	25
Øvelse 20 Udvidet kontrol og beregning af temperatur under indtastning. ....	25
Øvelse 21 Afsluttende indtastning med beregning af temperatur. ....	25
Øvelse 22 Vend beregningen om. ....	25
Øvelse 23 Dokumentation med og beregning. ....	25
Øvelse 24 List data. ....	25
Øvelse 25 Simple tabeller for data - codebook. ....	25
Øvelse 26 Sammenlign de to datafiler. ....	26
Øvelse 27 Exportér data til Stata. ....	26
Øvelse 28 Exportér data til Excel eller dbase.....	26
Opbygning af check filer. ....	27

---

Repetition Opbygning af skema og indtastning - dit eget projekt. ....	30
Øvelse 29 Et projekt du arbejder på - hvilken data struktur.....	30
Det enkelte skema udarbejdes. ....	30
Øvelse 30 Et projekt du arbejder på - skitse til et konkret skema. ....	31
Beslutning om faktisk indtastningsmetode.....	31
Øvelse 31 Dit eget projekt - beslut en indtastningsmåde. ....	32
Øvelse 32 Dit eget projekt - første dokument om oparbejdning af data. ....	32
Epicalc 2000 .....	33
Øvelse 33 Hent EpiCalc 2000 fra internet og installér det.....	33
Øvelse 34 Konfidensinterval for en proportion.....	33
Øvelse 35 Stikprøvestørrelse.....	33
Euroqol 5-d.....	34

## Forord

Der findes utallige programmer til analyse af data. Intentionen med EpiData er **ikke** at udvikle et nyt data analyse program, men at videreføre de principper for indtastning af data, som findes i programmet Epi-Info version 6<sup>1</sup>. Principperne er uovertruffet enkle og modsvarer alle krav til indtastning af langt de fleste data. Desværre fungerer EpiInfo ikke uden en del problemer i de nye styresystemer (Windows), specielt i forbindelse med nye netværk. Nutidens brugere har desuden vanskeligt ved at arbejde med et system baseret på "dos" principperne<sup>2</sup>.

EpiData er udviklet ud fra erfaringer med Epi-Info version 6 med tre begrundelser:

1. Der findes ikke noget dataindtastningsværktøj, som samtidig gør det let at "kvalitetssikre" indtastning.
2. At udvikle et let tilgængeligt program der udelukkende fokuserer på indtastning. Samt at sikre finansiering af udviklingen for derefter at distribuere programmet gratis.
3. At indbygge muligheder for at dokumentere indtastningsprocessen og fokusere på datakvalitet. Bl.a. med tilføjelse af labels, logisk konsistenskontrol og udskrifter.

Med EpiData har brugeren et færdigt værktøj til de faser af dataindtastning, der starter med opbygning af skemaer og leder hen til arkivering af data eller distribution af data til samarbejdspartnere. Derefter kan data analyseres i en lang række statistikprogrammer.

Programmet udvikles i et samarbejde mellem Fyns Amt, Initiativ for Ulykkesforebyggelse ved Jens Lauritsen, programmør Michael Bruus og Mark Myatt fra forlaget Brixton Health UK ("http://mkn.co.uk/help/extra/people/Brixton\_Books"). Dokumentation og manualer m.v. er skrevet udenfor Fyns Amt's regi.

Finansieringen af udviklingen er sikret gennem bidrag fra:

- Fyns Amt, sundhedssekretariatet <http://www.fyns-amt.dk>.
- Brixton Health, UK <http://www.brixtonbooks.demon.co.uk>
- Danish Data Archives/ERAS, Denmark <http://www.dda.dk>
- University of Southern Denmark, Faculty of Health - Odense.  
<http://www.sdu.dk/indexE.html>
- Valid International. London UK <http://www.validinternational.org/>
- London School of Hygiene & Tropical Medicine, UK <http://www.lshtm.ac.uk/>
- International Centre for Eye Health, UK <http://www.ucl.ac.uk/ico/ircpb.htm>

Bidrag til den videre udvikling modtages i alle størrelser med tak. Midlerne vil gå til aflønning af programmør, koordineringsomkostninger ved samarbejdet med Brixton Health, udvikling af dokumentation, instruktionsmateriale og præsentation af EpiData ved videnskabelige

---

1 Se fx: "http://www.cdc.gov/epo/epi/epiinfo.htm" eller "http://www.gruk.no/epi-info"

2 CDC (Center for Disease Control) i USA begyndte at udarbejde en windows udgave af Epi-Info kaldet Epi-Info 2000 før 1998. Programmet er netop frigivet i første udgave, men udviklingsstrategien fastholder ikke **det simple**. Bl.a. fylder programmet knap 40 Mb og der er endnu ikke udviklet et modul til sammenligning af dobbeltindtastede data, ligesom flere moduler fortsat er dos programmer. Brugerne opfordres til at afprøve programmet og vurdere om princippet i EpiData eller Epi-Info 2000 er mest tiltalende. EpiData svarer til MakeView og EnterData modulerne i Epi-Info 2000. Data indtastet med EpiData kan senere bruges i Epi-Info 2000, hvis man skulle ændre mening.

møder mv. Se nærmere i hjælpefilen om dette. Anvendelse af midlerne revideres af Dansk Selskab for Samfundsmedicin. Forslag og ønsker til fremtidige versioner kan meddeles [info@epidata.dk](mailto:info@epidata.dk)

Denne note giver en introduktion til opbygning af skemaer, formulering af spørgsmål i skemaer og "oversættelse" af skemaer til variable og indtastningsbilleder.

**Intentionen er:** At læseren efter arbejde med **alle øvelser i kronologisk rækkefølge** er introduceret til principper for gode spørgeskemaer samt principper for datadokumentation. Desuden at brugeren kan omsætte et skema til et antal variable der indtastes på en computer og efterfølgende forklares og dokumenteres. Som en del af indtastningen kan der tilføjes regler for konsistens af data og eller kontrol af gyldige værdier m.v. Dokumentation og datafil kan til sidst samles i en "pakke" der er velegnet til arkivering og videre forsendelse til samarbejdspartnere.

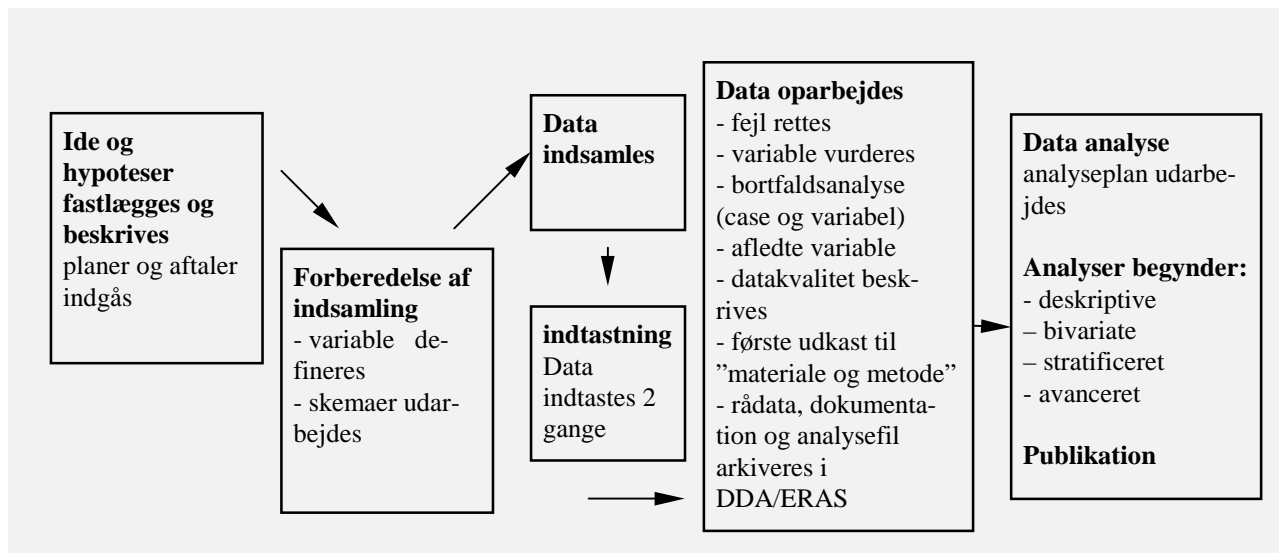
Forslag til rettelser af noten modtages gerne.

Kontakt Jens M. Lauritsen pr e-mail: [JM.Lauritsen@dadlnet.dk](mailto:JM.Lauritsen@dadlnet.dk) eller [info@epidata.dk](mailto:info@epidata.dk)

## Oversigt – indsamling, oparbejdning og analyse af data.

Når der indsamles kvantitative data (spørgeskema, kodede interviewspørgsmål ...) er det meget væsentligt, at sikre god kvalitet af de registrerede data. Både teknisk (indtastning etc.) og indholdsmæssigt. Der findes en lang række af skala'er og afprøvede enkeltspørgsmål, som belyser en bestemt problemstilling. Find disse skala'er eller formuleringer og anvend dem<sup>1</sup>. Et af områderne er generelle helbredsstatusmål. Et af disse anvendes i øvelserne (EuroQol 5-d) og er gengivet bagest i noten. I Danmark vil et udmærket udgangspunkt være det standardspørgeskema, som Statens Institut for Folkesundhed (tidl. Dansk Institut for Klinisk Epidemiologi "www.dike.dk") har udarbejdet. Dette giver mulighed for at sammenligne med andre undersøgelser. Ved oversættelse fra andre sprog af standard skalaer er der en række forhold, som skal sikres<sup>2</sup>

Den samlede proces kan opdeles i disse trin:



Bemærk, at dataopbejdning afsluttes med at første udkast til materiale og evt. også metode skrives. Megen tid og præcision af analysearbejdet vil gå tabt hvis der ikke er god forståelse og logisk konsistens af data **inden** analysen påbegyndes. En gennemgang af hele processen er beskrevet<sup>2</sup>. Oplysninger om arkivering af sundhedsvidenskabelige data kan rekvireres fra Dansk Data Arkiv/Eras ([www.sa.dk/ERAS](http://www.sa.dk/ERAS)).

### Supplerende litteratur:

1. McDowell I and C: Measuring Health: A guide to Rating Scales and Questionnaires. Oxford University Press. 2nd ed: 1995/96. Dirksen A, Christensen E, Jørgensen T, Kampmann JP, Kjær P. Klinisk forskningsmetode, en grundbog. København: Munksgaard; 1996:93-110. Udvalget vedrørende Videnskabelig uredelighed: Vejledning i god forskningspraksis ([www.forskraad.dk/publ/vejl\\_vid\\_praksis/prorap.htm](http://www.forskraad.dk/publ/vejl_vid_praksis/prorap.htm)). Jørgensen PH, Kyvik KO. Registrering og arkivering af sundhedsvidenskabelige data. Ugeskr Læger 1997; 159: 963-4. ([www.dda.dk/eras](http://www.dda.dk/eras)). Stewart AL and Ware JE (Eds): Measuring Functioning and Well-being. The MOS study approach. Duke University Press, 1992 (ISBN 0-8223-1212-3). Modelspørgeskema til undersøgelse af befolkningens sundhed og helbred. Dike, København, 1995.
2. Kvamme, O.J.; Mainz, J.; Helin, A.; Ribacke, M.; Olesen, F.; Hjortdahl, P. Oversættelse av spørreskjema. Et oversett metodeproblem. Nord.Med. 1998; 113: 363-6.
3. Hansen JM, Lauritsen JM. Dataindsamling Og Analyse. Ugeskr Laeger 1999 (juli)

## Fra ide til skema.

Før de konkrete skemaer udarbejdes skal der selvfølgelig findes en ide eller et tema der skal belyses, samt tilhørende formål og delmål. Når disse er beskrevet fastlægges den overordnede datastruktur i projektet. Herunder analyseenhed, antal skemaer og de konkrete spørgsmål.

## Hvad er analyseenheden ?

**Hvad observeres ?** En antal patienter, skoleelever, skoleklasser, en lærer og de elever der undervises, en håndboldspiller eller et håndboldhold. Dette har betydning for valg af statistisk metode. For almindeligt anvendte statistiske metoder er det en forudsætning, at analyseenhederne er indbyrdes uafhængige.

## Eksempel:

- I et studie indgår et håndbold hold med 17 spillere, der træner sammen. Er analyseenheden så et hold eller et antal spillere, hvor nogle egenskaber er knyttet til holdet. Hvis det sidste er tilfældet bør der i analyserne tages højde for en mulig gruppeeffekt (såkaldt klyngeeffekt eller på engelsk clustereffect). Hvis det primært er det sociale sammenhold på håndboldholdet der studeres ville observationsenheden nok være et hold.
- Effekten af et kirurgisk indgreb - hvis kirurgens erfaring eller individuelle dygtighed har stor betydning kan det være nødvendigt at tage højde for det i analysen. Patienterne vil så i nogen grad være knyttet sammen i klynger (grupper) afhængig af kirurgen.

Heldigvis er der i en lang række situationer ingen stærk gruppetilknytning, men vær alligevel opmærksom på skjulte klynge effekter (fx behandler effekt)

**Eksempel:** I en blodbank ønskes det at belyse omfanget af komplikationer ved transfusioner. Hvilken af A..D er den rigtige analyseenhed ?

Donorer: ①②③④⑤⑥⑦⑧⑨⑩

Modtagere: ①②③④⑤⑥⑦⑧⑨⑩ (Patienter)

Mulige analyseenheder:

**A.En transfusion:** ②① ②① ⑦① ⑦① (eksempler)

**B.Et transfusionsdøgn:** ②①① ⑦①① ②⑧⑨⑩ (den samlede mængde blod,  
②②②② ②③③③ ②④③③ givet til hver patient i et døgn)

**C.En patient og samtlige modtagne portioner blod:**

②①②③④④⑤⑥⑦⑧⑨⑩ (Patienten indeholder potentielt antigener fra alle 10 donorer)

**D.En donor og samtlige patienter, der har modtaget blod fra pågældende:**

⑥①②③④⑤⑥⑦⑧⑨⑩ (Alle patienterne indeholder potentielt antigener fra donoren)

Svaret må selvfølgelig være: Det kommer an på arten af komplikationer der har interesse.

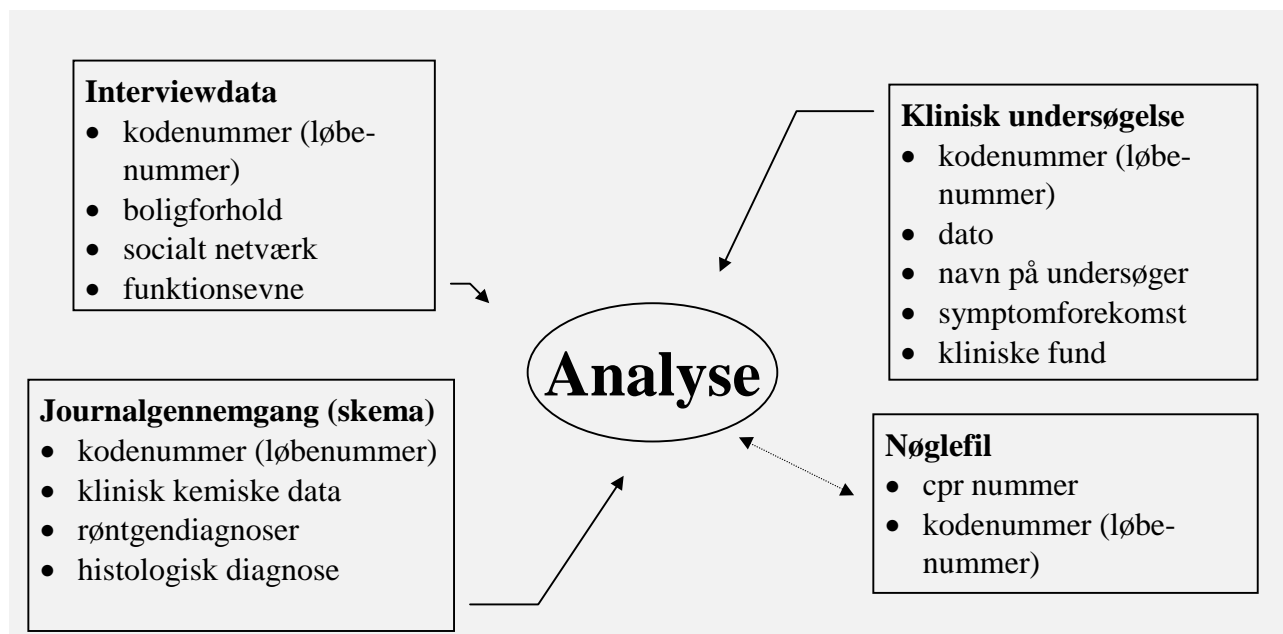


## Hvor mange skemaer ?

Enkle undersøgelser kan gennemføres ud fra ét skema, fx et spørgeskema med 10 spørgsmål til alle respondenter (personer, som har svaret), mens andre har flere typer skemaer.

### Eksempel:

En undersøgelse indeholder: interviewdata, klinisk undersøgelse og et skema til journalgennemgang.



Der opbygges et papirskema og en indtastningsrutine til hvert skema. Senere kobles alle skemaer sammen i data **Analysen** ud fra et fælles id nummer (identitets eller anonymiseringsnummer).

En særlig fil med både identitetsnummer og cpr nummer gemmes særskilt og aflåst efter de regler, som datatilsynet har givet.

### Indhold i det enkelte skema.

For hvert af de ønskede skemaer udarbejdes et første udkast med de forskellige emner, der skal indgå. Det er som regel en god ide, at starte med de spørgsmål, der specifikt vedrører projekts emne. Fødselsdato, køn etc kan indsættes senere i skemaet.

Opdel hvert skema i forskellige sektioner og tænk på om rækkefølgen er både logisk og acceptabel for de der skal svare. Det er nyttigt at anvende filterspørgsmål, især ved interview.

### Eksempel: Filterspørgsmål

<b>10</b>	Har De været i arbejde siden 1. januar ?	<input type="checkbox"/> <sup>1</sup> Ja <input type="checkbox"/> <sup>2</sup> Nej - fortsæt i spørgsmål 28
-----------	--	--

De konkrete formuleringer af enkelte spørgsmål tages fra standardiserede spørgsmål, en valideret skala eller en diagnostisk kode (fx ICD-10 kriterier), se side 5. Pas på med "private" oversættelser og typografiske fejl, som er meget lette at overse. Det er en fordel at skitsere de grafer og tabeller, der skal findes i de planlagte publikationer. Hvis du ikke kan anvise hvilke grafer mm. der skal bruges til sidst, ved du heller ikke hvilke data der skal indsamles.

Alle formuleringer nedenfor er fra DIKE's sygdoms- sundhedsundersøgelse fra 1994. (spørgsmålene indgik i en anden rækkefølge)

### Øvelse 1 Præcision af formuleringer.

Sammenlign formuleringerne i spørgsmål 1 og 2. Vil de give samme svarfordeling?

Helbred	
1. Hvorledes vil du vurdere din nuværende helbredstilstand i almindelighed?	1 <input type="checkbox"/> Virkelig god 2 <input type="checkbox"/> God 3 <input type="checkbox"/> Nogenlunde 4 <input type="checkbox"/> Dårlig 5 <input type="checkbox"/> Meget dårlig
2. Hvordan synes du dit helbred er alt i alt ?	1 <input type="checkbox"/> Fremragende 2 <input type="checkbox"/> Vældig Godt 3 <input type="checkbox"/> Godt 4 <input type="checkbox"/> Mindre Godt 5 <input type="checkbox"/> Dårligt
To andre generelle spørgsmål om helbred og dagligdag er:	
3. Føler du dig frisk nok til at gennemføre det, som du har lyst til at gøre ?	1 <input type="checkbox"/> Ja, for det meste 2 <input type="checkbox"/> Ja, af og til 3 <input type="checkbox"/> Nej, sjældent
4. Er du glad og tilfreds med din tilværelse ?	1 <input type="checkbox"/> Ja, for det meste 2 <input type="checkbox"/> Ja, af og til 3 <input type="checkbox"/> Nej, sjældent

### Svarmuligheder

Det er vigtigt at opbygge hvert spørgsmål så alle personer kan finde ét svar (såkaldt exhaustivt eller udtømmende krav), og samtidig at du har besluttet om der kun må være et svar i hvert spørgsmål (kategorier er gensidigt udelukkende eller exclusive). Endelig om svarmulighederne skal være kategoriale (mand, kvinde), ordinale eller rangordnede (for det meste, af og til, sjældent) eller kontinuerte ( hvor mange meter ?). Datoer er med **EpiData** lette at indtaste og regne på (fx alder ved et bestemt ambulatorie besøg). Tekster og forklaringer kan give gode uddybende forklaringer, som ikke nødvendigvis skal testes ind. I indtastningen kan man nøjes med at angive om der findes en tekst eller ej.

Svarmulighederne omsættes senere i variable med de samme typer. Nogle spørgsmål kodes i én variabel (fx køn) andre gange må der flere til (hvilke af følgende aviser læser du ?).

### Indsamling af data.

Det er nødvendigt at etablere en rutine til at sikre, at alle skemaer indsamles og en opsamlings/rykkerprocedure hvis der mangler information. Dette kan være en omfattende arbejdsopgave, hvor grundighed betaler sig. Det skal være let at overskue (fx ved at udskrive lister) hvem der har svaret eller hvilke journaler der er gennemgået på et givet tidspunkt.

## Fra skema til indtastning.

For hver oplysning der indsamles besluttes, hvordan oplysningen kodes som variable i dataanalysen. En variabel er den mindste informationsenhed i et skema. Variable kaldes også parametre, kovariater, items mm afhængig af faglig baggrund og tradition hos den der udfører undersøgelsen.

## Antal og typer af variable ?.

En oplysning kan fx være et spørgsmål i et spørgeskema, en blodprøveværdi eller en histologisk diagnose for en vævsprøve. Nogle gange kodes et spørgsmål i én variabel (fx køn) andre gange må der flere til (hvilke aviser læser du ?).

Variable er enten grupperede (d.v.s. et afgrænset antal svarmuligheder ja/nej/ved ikke), rangordnede (grupperet, men med en bestemt rækkefølge stor/større/størst), kontinuerte (numeriske), fritekst strenge (“Han tog hans hat og gik hans vej”), logiske (Sand/Falsk - som regel kodet som J/N, Y/N eller 1/0) eller datoer (datoen indtastes).

Variable er af forskellig type. De er alle vigtige for et samlet reproducerbart resultat. Udover de emnemæssigt mest relevante (de der afspejler formålet) er det som hovedregel nødvendigt med nogle baggrundvariable (alder, køn ....), nogle kontrolvariable (til såkaldt confounder kontrol), styringsvariable (dataindsamlings processen) og kvalitetsvariable. Se nærmere nedenfor.

## Gruppering af data - NEJ !!

Det er **meget vigtigt** at **INGEN** information grupperes/summeres på tidspunktet for dataopsamling eller dataindtastning. Det gøres meget lettere og bedre under analysearbejdet.

Indtast fødselsdato - ikke alder, vægt i kg ikke grupperet vægt, sygedage i antal dage - ikke over/under 7 dage. Indtast hele datoen for besøg i et ambulatorie, ikke kun årstal. Hvis du bruger en tidskode (1=start 2= 3mdr 3=6 mdr ...), så tast også datoen for en sådan oplysning ind. Det kan være der er årstidsvariation eller at det senere viser sig, at der var problemer med kalibrering i en bestemt kalender periode.

***Ingen datareduktion  
under indsamling eller  
indtastning af data !!!!***

## Styringsvariable.

Et antal styringsvariable er som regel nødvendige: dato for interview, hvornår er skema retuneret, hvem har kodet tekstinformation, hvem har interviewet .....

## Kvalitet af en oplysning - “kontrolvariable”.

Ofte varierer oplysningers kvalitet med andre kendetegn. Fx må det forventes, at en røntgen diagnose stillet af en speciallæge er mere præcis end hvis diagnosen stilles af yngste reservelæge. Et andet eksempel kunne være dato for en analyse, der er kendt for at have årstidsafhængig variation (Fx holdbarhed i forhold til transport temperatur).

For at bruge præcisionen af en oplysning i analysen kan der inkluderes en eller flere variable, som angiver usikkerheden på hovedvariablen (diagnostikers erfaring eller prøvedato). Senere i dataanalysen kan det undersøges om der er systematiske afvigelser ud fra undergrupperne i disse “kontrolvariable”.

### Navngivning af variable.

De forskellige variable gives navne. Det er nødvendigt for at sikre, at det er de rigtige tabeller der tolkes.

Det er en smagssag om variable skal have numre (v1,v2....v129) eller betegnelser, som antyder indholdet (alder, gender, ACTH, .....). Men i analyser mv er det hurtigere at skrive fx "v1" end "hoejde78". Derfor foreslår jeg at man bruger nummerlignende betegnelser. Navnet skal være entydigt og kan med fordel referere til de spørgsmål, der er i registrerings- /interview- /spørgeskemaer (Fx s1,s2 ... for spørgsmål og L1,L2 ... for laboratoriedata). Anvend højst 8 bogstaver/tal i variabelnavne, **aldrig æøå**. Første tegn i et navn må ikke være et tal **Inavn** er ulovligt, mens **navn1** er ok.

### Forskel mellem uoplyst og irrelevant.

Det er vanskeligt at opstille en fast regel for hvordan dette skal håndteres under indtastningen og i skemaer.

- **Uoplyst:** Uoplyst betyder at oplysningen **ikke** er tilvejebragt. Om det skyldes svar personens manglende lyst til at give svaret eller om det ikke kan fremskaffes er ligegyldigt. Svaret findes ikke tilgængeligt for analysen.
- **Irrelevant:** Irrelevant er et svar, som ikke kan og ikke skal besvares. Fx er antal gennemførte graviditeter uden betydning for mænd (i biologisk somatisk forstand).

Under indtastningen kan det være en fordel at alle uoplyste værdier blot overspringes, dvs tages med en blank (ingen) værdi. Ved et filterspørgsmål springes de irrelevante over, og der indsættes en irrelevant værdi (fx 8) i de spørgsmål der springes over. Den irrelevante værdi kan enten indsættes under indtastningen (automatisk) eller det kan ske i oparbejdningen..

Hvordan skal der kodes hvis et filterspørgsmål er uoplyst ?

#### Eksempel: Filterspørgsmål

<b>10</b>	Har De været i arbejde siden 1. januar ?	<input type="checkbox"/> <sup>1</sup> Ja <input type="checkbox"/> <sup>2</sup> Nej - fortsæt i spørgsmål 28
-----------	--	--

Der kunne kodes uoplyst for spørgsmål 10-27 for alle, som ikke har besvaret ovennævnte spørgsmål, men der kunne også kodes irrelevant. Hvis der kodes irrelevant svarer det til at uoplyst tolkes som et nej. Beslutningen må dokumenteres i undersøgelsen.

### En kodebog

En kodebog viser hvilke variable der findes, hvordan de enkelte variable kodes og hvilken kontrol der udføres under indtastningen. (se næste figur) I kodebogen er hver oplysning omsat til en eller flere variable, der hver er defineret nøje. Hvilken datatype (kontinuert, grupperet, åben tekst ...), hvilke svarkategorier findes, hvilken talkode skal anvendes, hvis variabelen er uoplyst etc. Bredden er det antal cifre eller antal bogstaver der skal bruges, fx 3 ved 999 og 2 ved 99. For kontinuerte variable skal minimum og maksimum anføres. Desuden skal nøglevariable besluttes, d.v.s. variable som skal belyses for hver person.

Der kan som nævnt ovenfor være brug for to typer "uoplyst". Den ene type gælder når en ønsket oplysning ikke findes (*personen svarede ikke, blodprøvesvaret er endnu ikke ankommet etc.*), mens den anden type er "irrelevant spørgsmål", der opstår ved et "filterspørgsmål".

Alt dette vil fremgå af en kodebog, som **EpiData** kan udskrive når indtastningen er defineret.

## Endelig udarbejdelse af skemaer

Når analysevariabler er defineret kan skemaer finpudses og dataindsamlingen begynde. "Pilot-afprøvning" foretages på en gruppe der svarer til den endelige modtagergruppe. Efter pilotfasen gennemgås hvert spørgsmål - spørges der kun om én dimension (éet spørgsmål), er alle mulige svar med i et grupperet spørgsmål ....

## Simpel kodebog for et skema kan udskrives med EpiData (Uddrag)

DATAFILE: C:\EPI6\eqtest\eqtest.rec

Filesize: 3220 bytes  
 Last revision: 10. feb 1999 23:27  
 Number of fields: 19  
 Number of records: 47  
 Checks applied: Yes (Last revision 21. nov 2000 13:29)

Fields in datafile:

No.	Name	Variable label	Fieldtype	Width	Checks	Value labels
1	ALFA	ID	Text	1	Must enter	
2	ID1		Integer	2	Must enter Legal: 0-99	
.....uddrag.....						
6	T2	t2	Integer	2	Legal: 20-65	
7	T3	t3	Integer	1	Legal: 1,2	sex 1: Male 2: Female 9: Unknown
8	T4	t4	Integer	1	Legal: 1,2,3	smoke 1: yes 2: no 3: previous
..... .uddrag.....						
14	TEQ2	teq2	Integer	1	Legal: 1-3	ord 1: no 2: moderate 3: extreme
19	TEQG	teqg	Integer	3	Must enter Legal: 0-100	

## **Dokumentation.**

Dokumentation består af de oplysninger, som i princippet er nødvendige for at gentage en given undersøgelse en anden gang. Det vil sige at det både indeholder projektplaner, beskrivelse af beslutninger i processen, godkendelser fra datatilsyn og samarbejdspartnere, eventuel godkendelse fra videnskabsetisk komite og en række tekniske dokumenter om selve de indtastede data. Der er ingen formelle lovbestemte krav til dokumentation, men i praksis viser erfaringen at det er meget nyttigt at dokumentere processen. Ingen kan huske hvad der er besluttet i en konkret fase af et projekt nogle år senere. Med "datadokumentation" menes den del af dokumentationen, som er nødvendig for at kunne anvende et givet indsamlet datasæt til analyse og beskrivelse.

I forbindelse med sundhedsvidenskabelig forskning er der udarbejdet vejledende retningslinier fra "udvalget vedr. videnskabelig uredelighed" (se side 5). Reglerne er retningslinier og - desværre - udformede i generelle henstillinger. Fx skrives - kvalitetskontrol - uden at specificere dette nærmere. To krav kan nævnes:

### **1. Mulighed for at finde tilbage til originalmateriale.**

Det skal være muligt for enhver oplysning (fx et punkt i en figur) at finde tilbage til originalmaterialet. Det vil sige, at der skal være knyttet et id nummer til alle observationer. Id nummeret følger med i alle udgaver af data og er **entydigt** forbundet med de originale observationer. Originalmaterialet skal opbevares i 10 år. Originalmaterialet omfatter notater (også håndskrevne retteblade mm), spørgeskemaer, analyseskemaer, .....

### **2. Kvalitetskontrol, dokumentation og arkivering.**

Samtidig med ovenstående proces dokumenteres antal fundne fejl under indtastning mv. og de konsekvenser det har haft. Data arkiveres sammen med projektplaner, kopi af anvendte skemaer og den udarbejdede dokumentation. For større projekter med fordel i Dansk Data Arkiv/ERAS. Ved arkivering i ERAS bevarer den person der har afleveret data den fulde kontrol over hvem der må få data udleveret igen. Kontakt ERAS for nærmere oplysning: "<http://www.dda.dk/eras>" eller "[mail@dda.dk](mailto:mail@dda.dk)". Se desuden næste side.

## Datadokumentation.

Datadokumentation er en delmængde af den samlede dokumentation. Det består helt enkelt i en samling af noter (fejlrutiner, rykkerprocedurer, håndskrevne fejllister m.v.) og kopi af de værktøjer (indtastningsfiler m.v.) der er anvendt under indtastning og fejlfinding. Endelig indgår kopi af alle filer der er anvendt undervejs og dokumentationen afsluttes med en liste over filer der er tilgængelige i dataanalysen.

Det vigtigste formål med datadokumentationen er at redegøre for indtastningsprincipper, ændringer foretaget som konsekvens af kvalitetskontrol og at begrunde hvor mange personer, der indgår i den konkrete undersøgelse. Herunder vurdere omfanget af uoplyst og beskrive særlige forhold som der skal tages hensyn til under analyse af data.

Omfanget af dokumentation skal svare til ambitionsniveauet med studiet. For flerårige studier bør der findes omfattende dokumentation. For mindre evalueringer eller kortlægninger fylder det måske kun ½ A4 side.

Som en del af dokumentationen er det en god ide at arbejde med et dokument, hvor beslutninger løbende indskrives.

### Eksempel: Beslutninger under indtastning og oparbejdning af data fra et landbrugsprojekt:

For personen 82809 er det en fejl, at v7a2x er kodet 1, dette er rettet til 0.

Den ene bedrift og tre personer uden information udelukkes:

```
select if (bnr <> 841 and v1 <> 82309 and v1 <> 8299 and v1 <> 2599).
```

Grunden til at de var med i filen var, at de som besøgende har haft en materielskade ulykke.

Disse personer er kodet som bystander og givet 1 time for alle risikotider:

```
V1 V3 TIMER2 RISKÅR1 RISKÅR2 RISKÅR3 RISKÅR4 RISKTOT V7A2 V7A2X
```

1102	1	1	1	1	1	1	1	2	2
2299	1	1	1	1	1	1	1	1	1
3999	1	1	1	1	1	1	1	1	1
7203	1	1	1	1	1	1	1	1	1
7304	1	1	1	1	1	1	1	1	0

Det samlede antal ulykker er sat til ovenstående i variabelen v7a2 og v7a2x, men er sat til 0 i variableerne v7a2b og v7a2bx.

### Brugstype

Denne var uoplyst for brug med nr: 16, 41,252,260,293

Disse kodes indtil videre som ”andet” i brugstype og uoplyst i rentabilitet.

På denne måde er der ikke tvivl om hvad der besluttet. Selve teksten er selvfølgelig noget kryptisk for udenfor stående, men er med til at opfylde kravet "data skal kunne følges fra oprindeligt skema til endelig analysefil". Noterne skal opfattes som et arbejdsdokument.

Nedenfor vises endnu et eksempel med en samlet variabelliste. Eksemplet er tilpasset ud fra en konkret arbejdsmiljøundersøgelse af psykisk helbred og stress indenfor politiet (ref: Bjarne Ibsen. *Politiets Psykiske Arbejdsmiljø. Arbejdsmiljøfondet*). Undersøgelsen blev gennemført som en kombination af en landsdækkende tværsnitsundersøgelse og en opfølgende undersøgelse i Fyns Amt, hvor der skulle udfyldes skemaer 5 gange på et år. Formålet var at nedbringe de psykiske konsekvenser af politiarbejde i forbindelse med voldsomme hændelser (fysisk vold, trusler mod politiet, oplyse forældre om børn's død, skudepisoder mv.).

Ind imellem teksten kommer nogle øvelser, hvor du skal tage beslutninger.

### Eksempel: Datadokumentation for et konkret projekt.

#### Personer:

For at vide hvor mange personer der har været ansat er der udarbejdet en slags råfil med personoplysninger. Grundlaget for denne var den registrering, som har været anvendt til at udsende skemaer efter. Fra filen er anvendt idnr, start og slutdato, samt beregnet varighed af ansættelse. Filen havde desuden en variabel charge (er personen overordnet ?), aktiv (personen skulle modtage skema nr 5) og en variabel der angav køn på basis af navnene (en række personer var benævnt med initialer fx P.K.Hansen og har fået værdien 2 for køn, de fleste af disse vil være mænd). Kodningen er sket ud fra filen: 'Find datoer.do' (vedlagt). **OBS.** *Af anonymitetshensyn har navne og id nr altid været adskilt på tidspunkter, hvor skemaet er blevet håndteret.*

#### Spørgeskemaer:

Skemaer er modtaget fra Esbjerg Centralsygehus udelukkende med id nr på. Indtastningen er sket fortløbende med kontrol af lovlige værdier m.v. ved hjælp af indtastningssystemet Epi-Data. Al indtastning er foretaget af NN.

Alle skemaer er indtastet en gang. De 10 vigtigste variable er kontrolleret mod skemaernes indhold efter indtastning.

#### Besvarelse af de udsendte skemaer:

Det skal nu besluttes hvilken del af skemaerne der skal anvendes ved analyserne. Det er vanskeligt at afgøre hvilke skemaer, der skal indgå i analyserne. Valget indebærer en afvejning af kompleksitet af skemaer i forhold til besvarelsesprocent. For at gøre analyserne gennemskuelige vil jeg foreslå at basere analyserne på alle der har været ansat i 10 mdr. Det er denne gruppe, som har været længst i kontakt med projektet som var et interventionsprojekt. Desuden må det også gælde, at mindst første og sidste skema er afleveret.

#### Analyserne kunne baseres på følgende subpopulationer:

Skemaer:	Besvarelsesprocent:	ansat mindst 3 mdr	ansat mindst 10 mdr
Første	554/674 = 82 %	534/641 = 83 %	423/498 = 85 %
Sidste	487/674 = 72 %	468/641 = 73 %	381/498 = 77 %
Første og sidste	448/674 = 67 %	435/641 = 68 %	359/498 = 72 %
Første og sidste + ét mere	446/674 = 66 %	433/641 = 68 %	357/498 = 72 %
Første og sidste + to mere	424/674 = 63 %	413/641 = 64 %	343/498 = 69 %
Alle fem	355/674 = 53 %	346/641 = 54 %	290/498 = 54 %
Procentandel af de ansatte:	100 %	95 %	74 %



## Øvelse 2 Beslut hvilken delmængde af skemaerne du vil bruge til analysen.

Kig på tabellen og tag en beslutning. Uanset hvilken beslutning du tager bliver der et bortfald. Enten et skemabortfald eller et personbortfald. Udform derefter et afsnit til metodeafsnit, der beskriver dit materiale og bortfaldets størrelse.

Efter en yderligere vurdering af omfanget af uoplyste udformes en tabel med oversigt over bortfald og variable i analysen. Afhængig af de valg der blev truffet ovenfor kunne tabellen se sådan ud:

Materialet for data fra Fyn 1995 bliver derfor:

	Personer I alt	Antal skemaer afleveret	Antal hændelser rapporteret i alt i foregående 3 mdr.	Uoplyst om hændelser
Samtlige ansatte personer	674			
Ingen skemaer afleveret	<u>120</u>			
Samtlige indtastede skemaer	554	1577	240	8
Mere end 19 uoplyste spørgsmål		<u>42</u>	<u>21</u>	<u>6</u>
Herefter tilgængelig	554	1535	219	2
Mindre end 3 mdr's ansættelse i 1995	9	19	2	
<b>Kan indgå i analysen</b>	<b>545</b>	<b>1516</b>	<b>217</b>	<b>2</b>

Udover de indsamlede data fra Fyn beskrevet ovenfor foreligger et landsmateriale fra 1992. Der er herefter tre datafiler til rådighed for alle de afsluttende analyser:

'Fyn alle kvartaler.dta'	Anvendes til dannelse af skalaer alene	1516 records	Variable se bilag
'Fyn 1995.dta'	Hver person én record, dvs analyser på tværs af året.	545 records	Variable se skema nedenfor
'Hele landet.dta'	Analyser for hele landet, bl.a. afsluttende ph.d. analyser	9035 records	Variable se skema nedenfor

**En record eller observation svarer til en person (et skema).**

### Forbehold som skal huskes under dataanalysen.

Da projektets dataindsamling og analyse blev planlagt var det hensigten, at der skulle indsamles de samme data (forløbs data) ved starten af projektet og ved afslutning af hvert af de 4 kvartaler i løbet af 1995. Denne indsamlingsmetode måtte opgives, fordi der var stor modstand mod at besvare skemaet fra nogle afdelinger. Følgende variable er tilgængelige. Andre problemstillinger kan analyseres fra de oprindelige data efter yderligere oparbejdning.

Indhold	Lands- projektet	Fyn				
		startskema	1. kvartal	2.kvartal	3.kvartal	4.kvartal
Arbejde på fyn	sted					
Alder (agegrp er grupperet)	age agegrp			age agegrp		
Køn	gender			gender		
Antal skemaer afleveret i alt				antal		
Skema afleveret i dette kvartal ?		in_0	in_1	in_2	in_3	in_4
Ordens /kriminalpoliti	besk	besk_0				besk_4
Leder (charge)	charge	charge				
Anciennitet i politiet	ancp					
Anciennitet i nuværende position	anc	anc				
Skiftet afdeling siden 1.1.95						anc_4
Skiftende arbejdstider	skift	skift_0				skift_4
Bureaukratisk sagsgang	d3a					
For stor arbejdsmængde	arbmgd	arbmgd				
Støtte og opmuntring fra leder	e10	e10				
Støtte og opmuntring fra kolleger	e11	e11				
Det sociale miljø på arbejdspladsen	i1c	i1c				
Mulighed for råd og hjælp fra leder	e8	e8				
Overvejet at finde andet arbejde	c1a	c1a				
Social kontakt med andre	social	social				
Deltagelse i aktiviteter	aktiv	aktiv				
Generelt selv vurderet helbred	j13a	j13a_0	j13a_1	j13a_2	j13a_3	j13a_4
Glad og tilfreds med tilværelsen	j14a	j14a_0	j14a_1	j14a_2	j14a_3	j14a_4
Tilfreds med arbejdsforhold - alt i alt	L7	L7_0	L7_1	L7_2	L7_3	L7_4
Mindst én uges sygefravær sidste år	j5a					
Mindst 3 dages sygefravær sidste 2 mdr	j4a1					
Mindst 3 dages sygefravær sidste 3 mdr		j4a_0	j4a_1	j4a_2	j4a_3	j4a_4
Stadig påvirket af hændelse fra tidligere	b5y	b5y				
Angivet voldsom hændelse for foregående 12 måneder	b5t	b5t_0				b5t_4
Udsat for voldsom hændelse foregående 3 måneder	b5	b5_0	b5_1	b5_2	b5_3	b5_4
Hændelserne (-en) påvirkede privatlivet	b5f	b5f_0	b5f_1	b5f_2	b5f_3	b5f_4
Er du stresset ?	j8k	j8k_0	j8k_1	j8k_2	j8k_3	j8k_4
Manglende lyst til at gå på arbejde pga hændelse	c5d8	c5d8_0	c5d8_1	c5d8_2	c5d8_3	c5d8_4
Skala: Graden af gener fra hændelse	gener	gener_0	gener_1	gener_2	gener_3	gener_4
Skala: <b>Fysisk</b> velbefindende (4 item)	velbf4	velbf4_0	velbf4_1	velbf4_2	velbf4_3	velbf4_4
Skala: <b>Psykisk</b> velbefindende (4 item)	velpf4	velpf4_0	velpf4_1	velpf4_2	velpf4_3	velpf4_4
Medvirket i intervention						delt

Variable markeret med grå er kun oplyst for personer, der har haft hændelser i de foregående tre måneder.

Variablene (**gener** og **c5d8**) samt (**gener\_0** og **c5d8\_0**) refererer til besvarelsen af spørgsmålet om hændelser i de foregående 12 mdr. I de resterende kvartaler i Fyns undersøgelsen refererer det til hændelser fra de foregående 3 måneder i hvert af kvartalerne.

## 2 Konkret indtastning med EpiData

Før du kan arbejde med EpiData skal du hente en kopi fra internet.

### Øvelse 3 Download af flow chart for EpiData.

Start en internet browser, kig på [www.epidata.dk](http://www.epidata.dk) og find punktet "download". Find det punkt, der viser "flow chart". Klik på "download" ud for dette (**epidata.pdf**) og du vil nu se et flowchart på skærmen. Dette er vist på næste side. Hvis du ikke kan se "pdf" filer skal du også downloade en "adobe pdf reader" - link er angivet på download siden.

### Øvelse 4 Download og Installering af EpiData.

Fortsæt nu med at hente EpiData fra [www.epidata.dk](http://www.epidata.dk). Hent følgende filer:

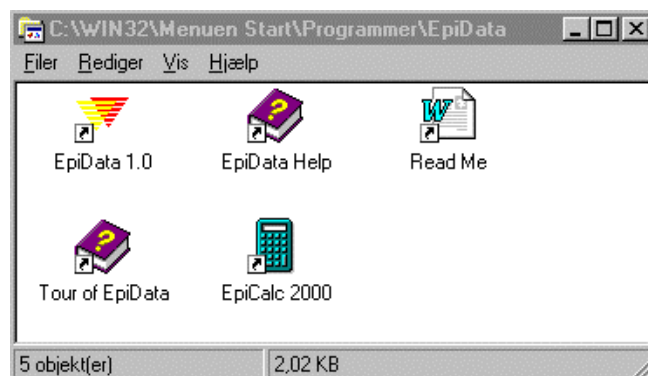
- ◆ De filer der indeholder installations filerne. Der bruges cirka 10 filer, som er pakket sammen i et såkaldt "zip" arkiv<sup>3</sup>. (pr 15.11.2000: **epid12.zip**).
- ◆ PDF filer med dokumentation. EpiTour.pdf og EpiHelp.pdf. Print disse filer og læs dem.
- ◆ Hvis du ikke har et "unzip-udpaknings" program i forvejen, så hent også (UnZip programmet).

Når du henter filerne vil browseren spørge "save file as/gem fil som" eller lignende. Gem fx i mappen c:\temp. Udpak dernæst epid12.zip<sup>4</sup> til fx. c:\temp og køр derefter c:\temp\setup.exe.

Programmet installeres og der oprettes nu en programgruppe, der hedder EpiData.

Når programgruppen er oprettet vil du i de fleste systemer se et skærmbillede som her<sup>5</sup>.

Når det vises kan du udføre øvelse 2<sup>6</sup>:

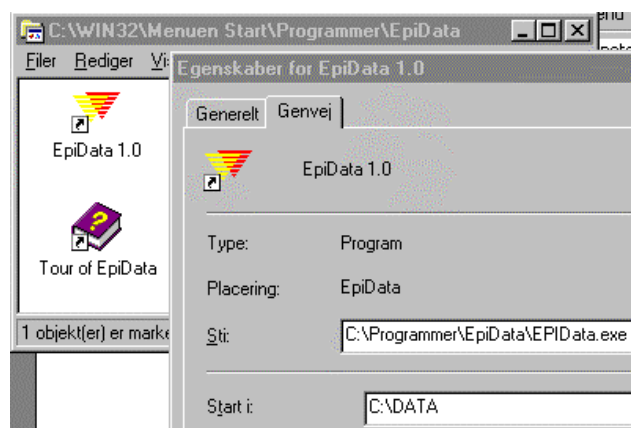


### Øvelse 5 Indstilling af brugen af EpiData.

Du skal fastlægge det sted på computeren, hvor EpiData gemmer data. Se til højre.

Egenskaber ved en ikon kan bl.a. aktiveres ved højre tryk med mus ud for EpiData ikonen og fra tastaturet (tast nr 3 th. for mellemrumstangenten).

Hvis du ikke har et "C:\data" i forvejen må du oprette det.



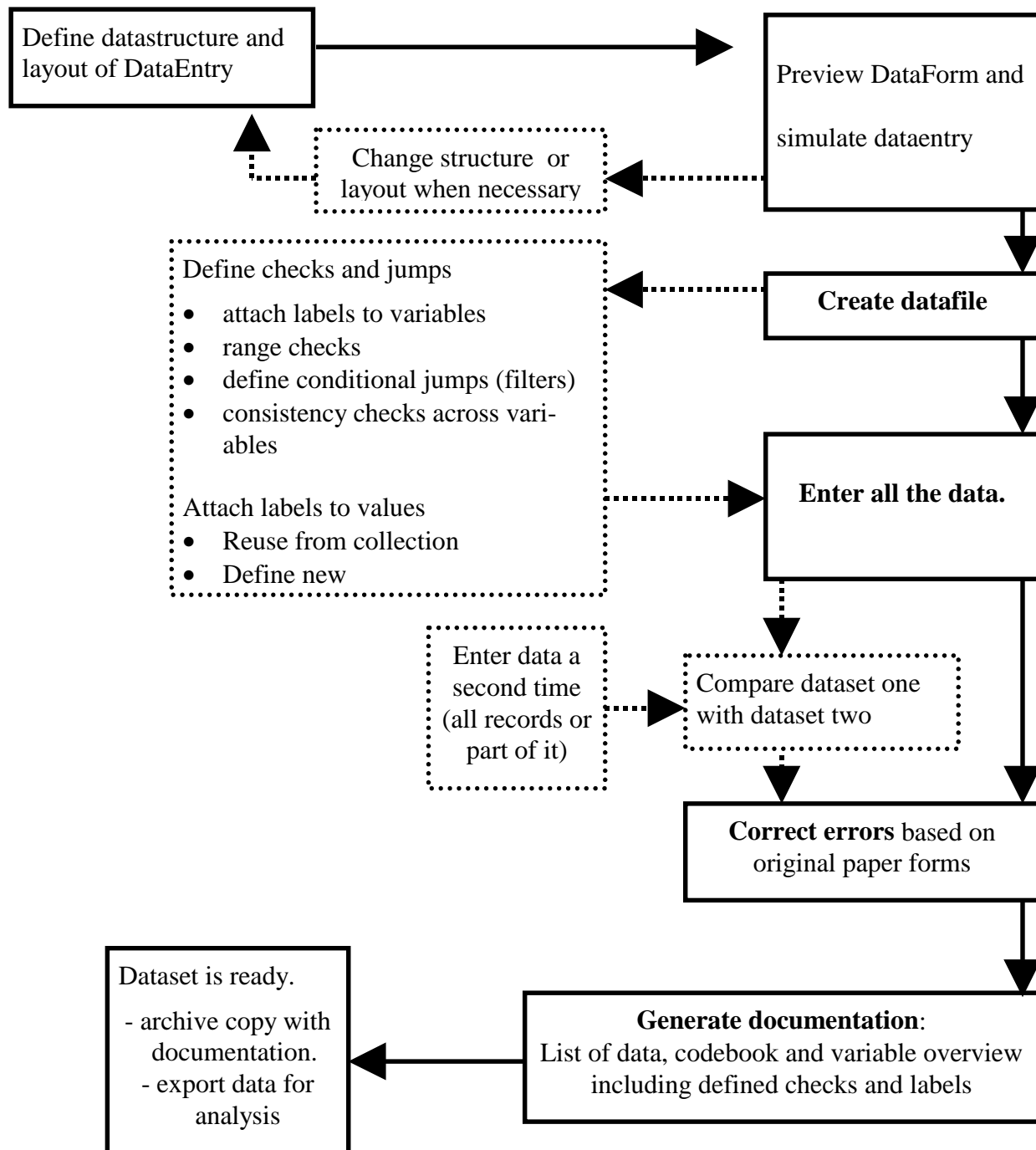
<sup>3</sup> Zip filer og PDF filer forklares nærmere på download delen af [www.epidata.dk](http://www.epidata.dk).

<sup>4</sup> Brug minunzip.exe ved at starte programmet og angive epidata1.zip eller et andet "unzip" program.

<sup>5</sup> Hvis du foretrækker at starte programmet ud fra "skrivebordet" kan du kopiere ikonen til skrivebordet.

<sup>6</sup> Konkrete skærmbilleder og den rækkefølge hvor forskellige emner vises på skærmen varierer meget mellem forskellige udgaver af windows/mac/pc systemer. Det tekniske punkt, som øvelse 2 henviser til er den måde, hvor du fastlægger "Egenskaber for Ikoner"

### Flow Chart for arbejde med EpiData.



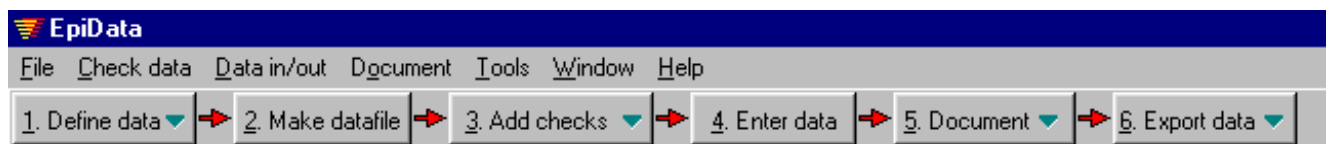
### Øvelse 6 Start EpiData uden brug af mus.

Hvis du med venstre musetast holder "fast" i EpiData ikonen og fører den ned til "Start" oprettes nu en menulinie i Start Menuen. Derefter er det nemt at starte EpiData.



### Øvelse 7 Start EpiData

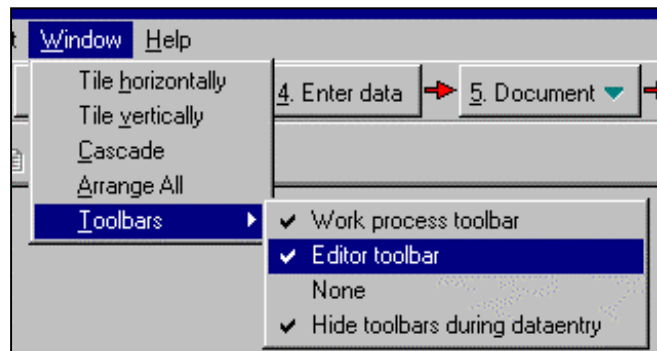
Start programmet og kig på skærbilledet.



Der er en menulinie, samt to såkaldte "ToolBar's".

Prøv at slå disse til og fra ud fra menupunktet "Windows". Se figuren til højre.

Afslut med at vise både "editor" og "Work process" toolbar.



### Øvelse 8 Hvilke dele af flowchartet svarer til "work process" toolbar ?

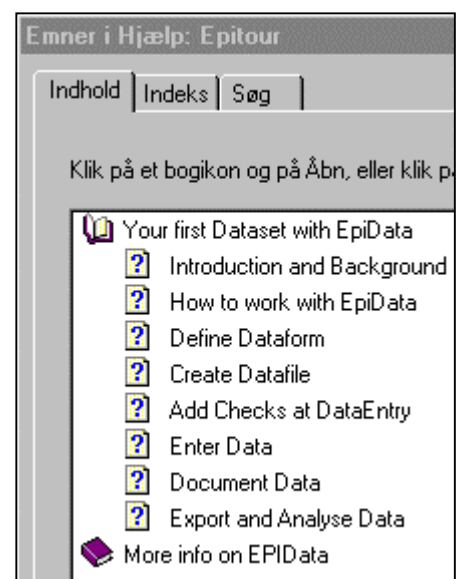
Notér på flow chartet hvor hver "kasse" fra "work process toolbar" foregår.

### Øvelse 9 Arbejd dig igennem EpiTour

EpiTour er en introduktions "rundtur" i programmet. Den er opbygget på samme måde som punkterne i "Work process toolbar".

Følg nu eksemplet i EpiTour fra "Define Dataform" til "Export and Analyse Data".

Specielt under "Add checks at DataEntry" skal du kun kigge på de første punkter om range, legal og labels. Resten kan du vende tilbage til senere.



Når du har arbejdet dig igennem EpiTour ud fra eksemplet First.qes - First.rec etc, er du klar til at prøve med et mere realistisk datasæt. Det er en fordel fra [www.epidata.dk](http://www.epidata.dk) at hente filen epitour.pdf og printe den ud. Det gælder også hjælpefilen.

### Øvelse 10 Første "reelle" datasæt i EpiData.

Varighed af hospitals indlæggelse<sup>7</sup>. Se på disse data og bestem hvilken variabeltype, der skal anvendes for hver variabel. Slå op i hjælpesystemet om variabeltyper.

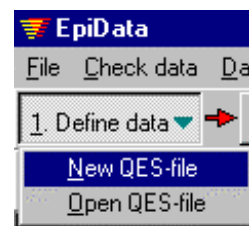
Data repræsenterer en stikprøve fra et større hospitalsstudie af antibiotika anvendelse. Data indeholder følgende variable: DUR (varighed af indlæggelse - dage), AGE (alder i år), SEX (M = mand, K = kvinde), Patientens temperatur (Fahrenheit: tempf. Celcius: tempc), Hvide blodlegemer - antal (WBC: 100/dl), Antibiotika brug under indlæggelsen (AB Y=Ja,N=Nej), bloddyrkning (CULTure 1 = ja, 2 = nej) og afdelingstype (SERvice 1 = medicinsk 2 = kirurgisk).

v1	v2	v3	v3a	v4	v5	v6	v7	v8
AGE	SEX	TEMP (F)	TEMP (C)	WBC	AB	CULT	SERV	DUR
30	K	99,00	37,22	8	N	2	1	5
73	K	98,00	36,67	5	N	1	1	10
40	K	99,00	37,22	12	N	2	2	6
47	K	98,20	36,78	4	N	2	2	11
25	K	98,50	36,94	11	N	2	2	5
82	M	96,80	36,00	6	Y	2	2	14
60	M	99,50	37,50	8	Y	1	1	30
56	K	98,60	37,00	7	N	2	1	11
43	K	98,00	36,67	7	N	2	1	17
50	M	98,00	36,67	12	N	1	2	3
59	K	97,60	36,44	7	N	1	1	9
4	M	97,80	36,56	3	N	2	2	3
22	K	99,50	37,50	11	Y	2	2	8
33	K	98,40	36,89	14	Y	1	2	8
20	K	98,40	36,89	11	N	1	2	5
32	M	99,00	37,22	9	N	2	2	5
36	M	99,20	37,33	6	Y	2	2	7
69	M	98,00	36,67	6	N	2	2	4
47	M	97,00	36,11	5	Y	2	1	3
22	M	98,20	36,78	6	N	2	2	7
11	M	98,20	36,78	10	N	2	2	9
19	M	98,60	37,00	14	Y	2	2	11
67	K	97,60	36,44	4	N	2	1	11
43	K	98,60	37,00	5	N	2	2	9
41	K	98,00	36,67	5	N	2	1	4

<sup>7</sup> Kopieret fra: B.Gerstman: <http://www.sjsu.edu/faculty/gerstman/EpiInfo/Epi6Basics.htm>  
Data source: Townsend et al., 1979, as reported in Rosner, 1990, p. 36. (med tilladelse fra B.Gerstman)

### Øvelse 11 Opbyg en definition til indtastning af data.

Opret en ny qes fil under **1. Define Data** og så snart du har fået en "untitled" oprettet, så gem filen (Ctrl+S eller menu File) under et passende navn, fx "hdur" eller "antibotika og indlæggelse indlæggelse".



Når du har gemt filen, så se efter at der i menulinien øverst står **hdur.qes** (eller hvilket navn du brugte). Det er vigtigt, at endelsen (filtypen) er QES, som er en forkortelse for Questionnaire.

To eksempler på opbygning af dataform ud for eksemplet.

```
id <idnum>
V1 Age ##
V2 Sex #
V3 Temp ##.##
V3a Temp ##.##
V4 WBC ##
V5 AB #
V6 Cult #
V7 Serv #
V8 Dur ##
```

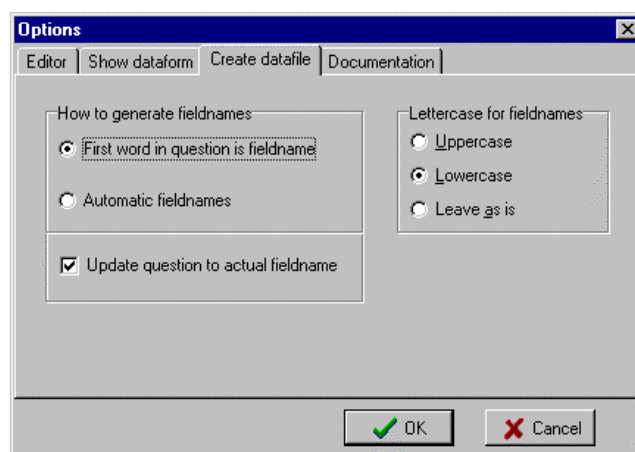
```
id <idnum>
Age ##
Sex #
Tempf ##.##
Tempf ##.##
WBC ##
AB <Y>
Cult #
Serv #
Dur ##
```

Skriv eksemplet til venstre ind i QES filen og tryk dernæst på "Ctrl+T" (Preview dataform). Hvad hedder de forskellige variable ?

### Øvelse 12 Navngivning af variable.

EpiData har to forskellige måder at definere navne på. Det ene er ud fra første ord i linien og det andet "automatic" ud fra de første bogstaver i linien.

Luk "Test Dataform" igen (Ctrl+F4 eller menu File) og find "Options" på menu'en File. Sæt dernæst options under "Create datafile", som vist på figuren til højre. Bemærk at "lettercase" er sat til små bogstaver.



### Øvelse 13 Navngivning af variable - first word.

Prøv nu med "PreView Dataform" igen (Ctrl+T) og se hvad navnene på de forskellige variable bliver. De skal nu være v1....v8.

- ◆ Luk Preview Dataform igen. (Ctrl+F4 - eller menu File). Nu kan datafilen oprettes.
- ◆ Se efter i flow-chart, hvor du er i forhold til work process toolbar

### Øvelse 14 Opret en datafil ud fra hdur.qes

Når datafilen er oprettet skal du sørge for at lukke alle filer for at arbejde videre (menu File, Close all).

### Øvelse 15 Indtastning af data - uden kontrol og backup kopi.

Fortsæt nu direkte til "Enter Data" og indtast alle data uden videre kontrol. Når du har indtastet alle data fortsætter du til "export data" vælg "back-up" og tag en kopi til en diskette (A:) eller et andet sted.

### Øvelse 16 Ny datafil - med simpel range kontrol under indtastning.

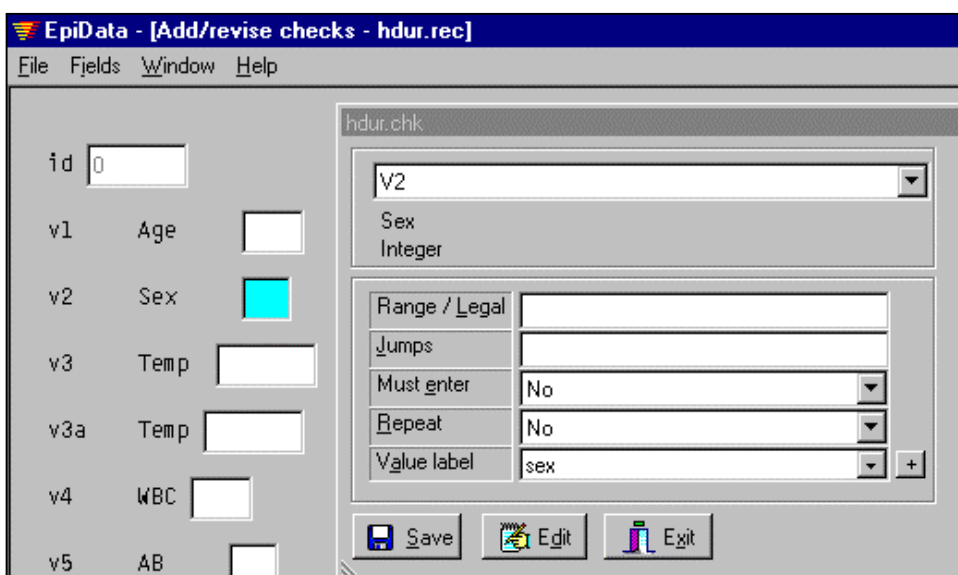
Opret en ny datafil med et nyt navn ud fra samme "QES" fil som ovenfor. Men fortsæt nu med "Add checks". Tilføj lovlige værdier til variablene v2, v5,v6,v7. (alle må kun være 1 og 2).

### Øvelse 17 Indtastning af data - med range kontrol og backup kopi.

Fortsæt direkte til "Enter Data" og indtast de første 5 personer - virker "range" kontrol ?. Afslut "enter data" efter de først 5 personer.

### Øvelse 18 Tilføj labels til indtastning.

Gå tilbage til "Add checks" og giv value label til køn, således at "value label" defineres som 1 "male" og 2 "female".



Det er meget let du skal blot under labels angive "sex". Det er den lille pil til højre for Value Label. Kontrollér evt. værdierne ved at taste "+" når du har valgt label "sex".



### Øvelse 19 Indtast igen med labels.

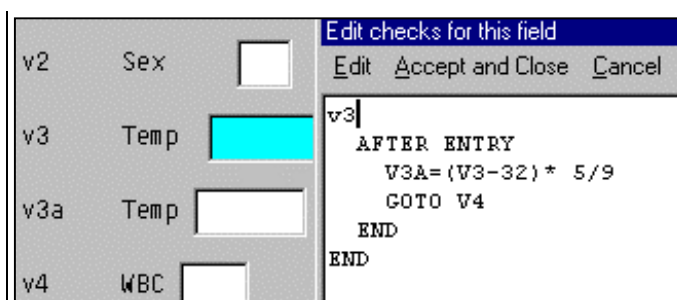
Afslut "add checks", fortsæt med "Enter Data" og indtast de næste 5 personer. Du kan nu under indtastning af køn trykke på "+" (den grå +) eller F9 og få værdierne vist på skærmen. Bemærk, at værdier i en label **ikke** skal være defineret under range/legal.

### Udvidede øvelser om beregning under indtastning, kan evt. springes over.

#### Øvelse 20 Udvidet kontrol og beregning af temperatur under indtastning.

Under "Add checks" kan du nu definere udvidet checks for v3 og v3a. Det gøres ved at angive udvidet kontrol og beregning under indtastning. Opbygningen kan ske i "add checks", punktet "Edit". Se nedenfor.

Tilføj under v3 linierne vist i figuren. Ideen er, at temperaturen i Celcius kan beregnes ud fra den formel og sættes ind i v3a. For at det virker skal du desuden i feltet v3a indsætte en linie, der hedder "NOENTER". Hvilket betyder at variabelen under indtastningen slet ikke tastes, men springes over.



#### Øvelse 21 Afsluttende indtastning med beregning af temperatur.

Tast yderligere 5 skemaer ind mens du ser, at temperaturen beregnes automatisk.

#### Øvelse 22 Vend beregningen om.

Vend nu i "add checks" beregningen om, så det er Fahrenheit temperaturen der beregnes og Celcius der tastes ind. Når du har defineret det korrekte tastes de sidste data ind.

#### Øvelse 23 Dokumentation med og beregning.

Sammenlign den dokumentation der vises under pkt 5 i Work Process Toolbar "Datafile Documentation" ved den fil du lige har indtastet og det der vises ved den simple indtastning i starten "**Indtastning af data - uden kontrol og backup kopi.**"

Hvilke forskelle er der ? Du kan vise de to samtidig side om side ved at udføre "Window Tile Vertically". Huskede du i øvrigt "back-up" undervejs ?

#### Øvelse 24 List data.

List data (Se under Document) i den ene eller begge filerne og læs korrektur. Var der fejl ?

#### Øvelse 25 Simple tabeller for data - codebook.

Codebook data (Se under Document) for den ene eller begge filer. Det er en oversigt over de indtastede data.

**Øvelse 26 Sammenlign de to datafiler.**

Sammenlign de to indtastede datafiler med "validate", som findes i "Tool" menu'en. Var der overensstemmelse ? Hvor stor fejlprocent ? Ret eventuelle fejl.

**Øvelse 27 Exportér data til Stata.**

Filerne er nu klar til export. Udført export til Stata. Læg filerne på en diskette sammen med behørig dokumentation. Hvilke filer skal være med ? Ved eksport til Stata følger labels med.

**Øvelse 28 Exportér data til Excel eller dbase**

Filerne er nu klar til export. Udført export til et andet format end Stata. Læg filerne på en diskette sammen med behørig dokumentation. Hvilke filer skal være med ?

## Opbygning af check filer.

Checkfiler indholder de kommandoer, som definerer kontrol under indtastningen. Princippet for opbygning af check filer uddybes nedenfor, sammen med et forslag til kodning af spørgsmål, hvor man der kan sættes kryds ved flere svar.

### Eksempel 1:

Helbred	
1	Hvorledes vil du vurdere din nuværende helbredstilstand i almindelighed?
	<input type="checkbox"/> Virkelig god <input type="checkbox"/> God <input type="checkbox"/> Nogenlunde <input type="checkbox"/> Dårlig <input type="checkbox"/> Meget dårlig

Her er det let at opbygge kontrol til indtastningen. Der skal blot defineres et indtastningsfelt i QES filen og derefter når datafilen er oprettet i "Add/revise" defineres en label med 5 værdie:

```
label helbred
  1 "Virkelig god"
  2 God
  3 Nogenlunde
  4 Dårlig
  5 "Meget dårlig"
```

end

Den pgl. label knyttes til variabelen, der koder spørgsmålet om helbred.

### Eksempel 2:

I et spørgsmål fra evaluering af en konference ønskes det oplyst, hvilke dele af dagsprogrammet personerne har deltaget i. Derfor stilles dette spørgsmål:

Dagsprogram	
6	Hvilke dele af dagsprogrammet deltog du i ?
	<input type="checkbox"/> Oplysning og udstilling <input type="checkbox"/> Arbejdsulykker og – sikkerhed <input type="checkbox"/> Børneulykker <input type="checkbox"/> Registrering og forebyggelse (+ ældre) <input type="checkbox"/> Forebyggelse af selvmord Sæt kryds ved de dele du deltog i: <input type="checkbox"/> Trafikulykker <input type="checkbox"/> Vold <input type="checkbox"/> Opsamling fra temagrupper i plenum <input type="checkbox"/> foredrag af Torben K.Jensen
	Alternativ udformning af svarkategorier hvor der skal svares i alle linier kunne være sådan: (Her vist med Vold som eksempel).
	Vold ..... <input type="checkbox"/> Ja ..... <input type="checkbox"/> Nej

I programmet var punkterne fra Arbejdsulykker til og med Vold arbejdsgrupper og personerne har derfor for de flestes vedkommende kun deltaget i en af dem. Ved indtastningen kan der let ske det, at man tror der er kryds i trafikulykker, men rent faktisk er det i Forebyggelse af selvmord.

Da personerne kan have deltaget i flere dele skal der oprettes kodefelter for hver af de ni punkter. Men hvis kun én af værdierne fra 1 til 9 tillades i hver underpunkt under tastningen, så er man sikret mod en fejl, hvor et kryds i fx Trafikulykker kodes under selvomrd.

Et godt indtastningsprincip bliver i så fald: (Antag at variablene hedder s7n1, s7n2 .....s7n9)

Variabel	Lovlig værdi for denne
s7n1	1
s7n2	2
s7n3	3
s7n4	4 .....

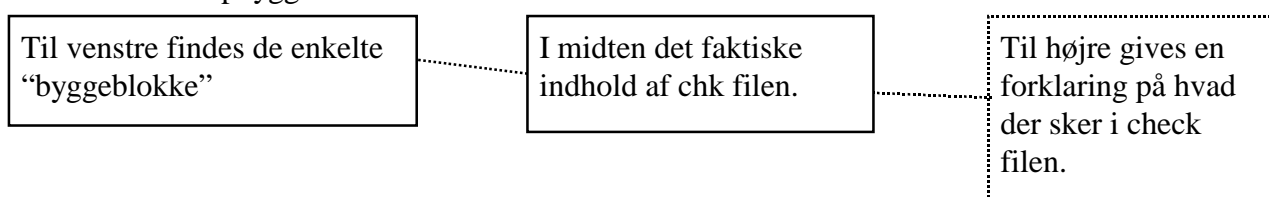
o.s.v for de resterende variable op til s7n9. Hvis man taster et 4 ind i nr 3 felt afvises indtastningen, mens et ikke udfyldt svar tillades. På denne måde sikres mod parallel forskydningsfejl under indtastning. I analysen af data ville man straks omkode 1,2,3.....9 til værdien 1 og et manglende svar ville kodes som 0.

På næste side gennemgås et længere eksempel. Id nummer angives med en “key”, således at programmet gør opmærksom på genindtastning af tidligere anvendt id nummer.

Prøv fx at se på disse muligheder:

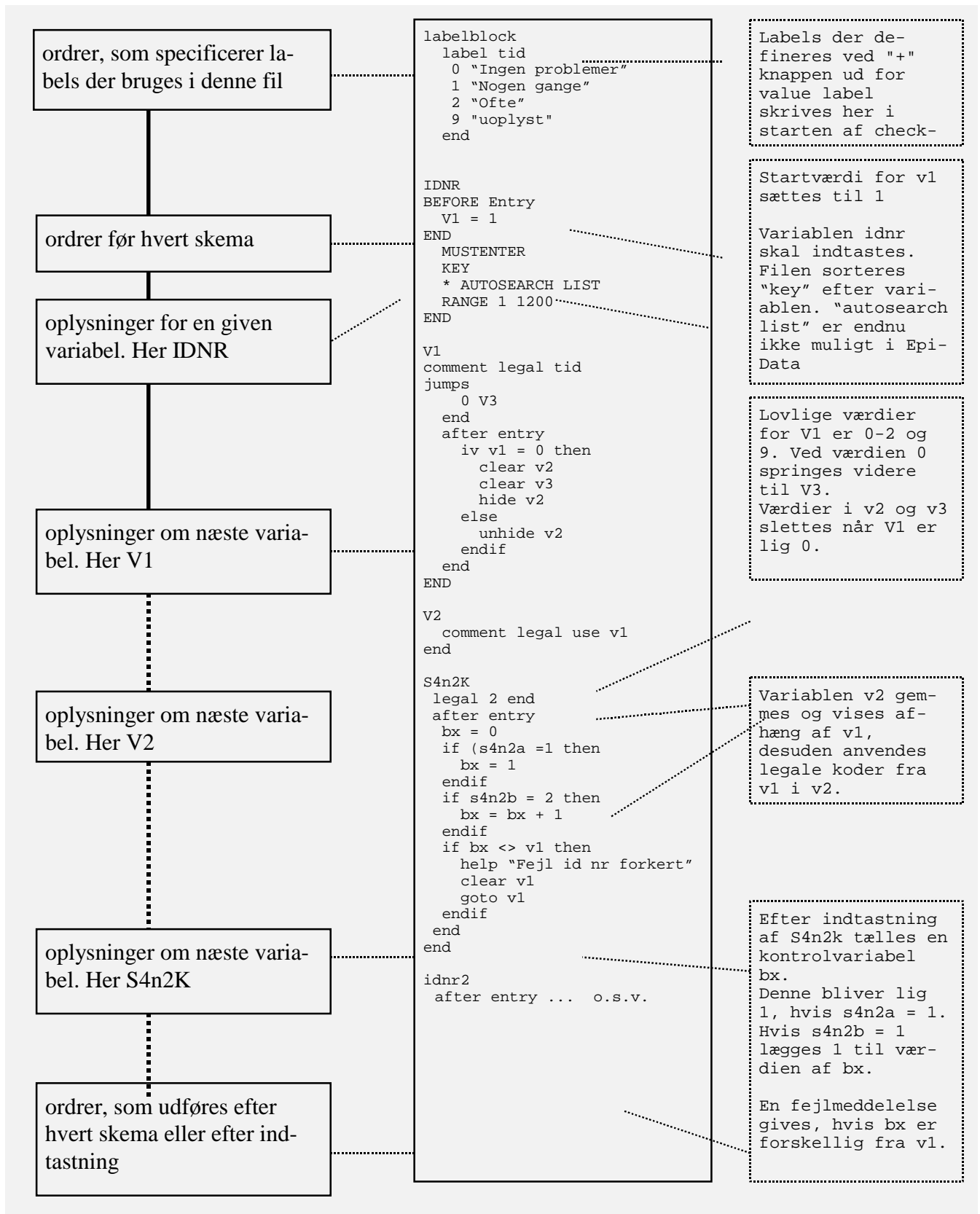
- check ved en beregning (som s4\_2k næste side) at det rigtige antal kryds er indtastet.
- prøv at indlægge forklarende koder med comment legal eller value label
- check at første og sidste indtastning af id nummer er korrekt.
- vis en fejl eller hjælpetekst på skærmen. Find kommandoen i hjælpefilen for EpiData
- gem eller fremhæv nogle spørgsmål afhængig af værdien af andre.
- nulstil variable pga filterspørgsmål

Næste side er opbygget således:



## Principskitse: Opbygning af en check fil

Her er vist principper. De konkrete linier er meget afhængige af det konkrete studie.



## Repetition Opbygning af skema og indtastning - dit eget projekt.

Bemærkningerne ovenfor resumerede arbejdet med at opstille skemaer. Nu repeteres processen ud fra øvelser. **Læs hele afsnittet igennem og udfør øvelserne.**

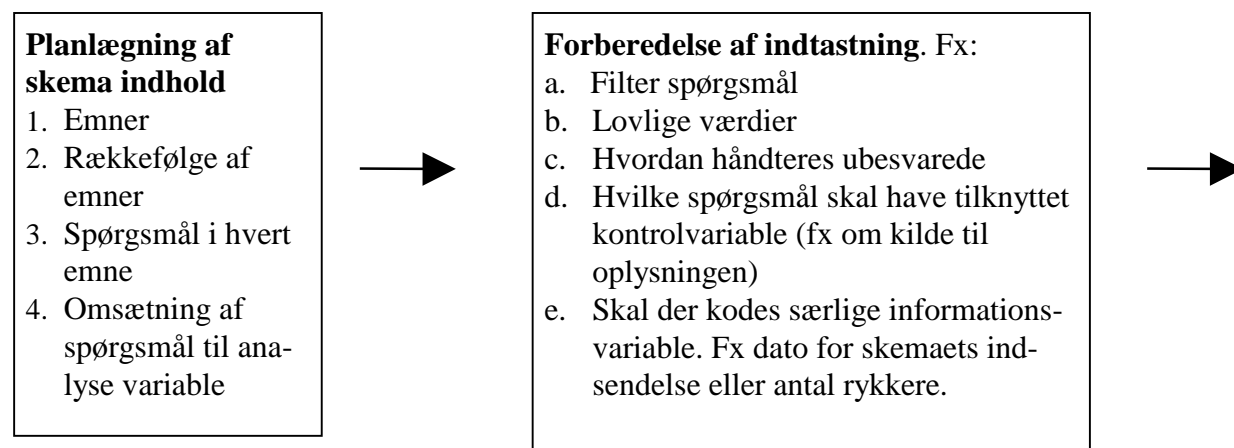
### Øvelse 29 Et projekt du arbejder på - hvilken data struktur.

Tegn en figur, der beskriver de (den) datakilder, som projektet indeholder. Beskriv hvilke emner, der indgår i hver datakilde, hvordan skal oplysningerne indsamles og hvordan skal de forskellige datatyper kobles sammen. Hvordan sikres anonymitet. OBS: (mange projekter har kun en datakilde = et skema)

### Det enkelte skema udarbejdes.

Spørgeskemaet (eller registreringsskemaet) **skal** umiddelbart kunne anvendes som grundlag for indtastning. Ved trykning skal der tages hensyn til, at det er let at bladere i skemaet under indtastning. Det er fx ikke nødvendigvis en god ide at trykke på begge sider. Skemaets skal afpasses til et format, der kan håndteres under indtastningen. Måske er A4 bedre end A5.

For hvert spørgsmål skal det angives direkte i skemaet, hvilken kode, der skal tages for det pågældende spørgsmål, fx ved et lille tal ud for hver svar mulighed. Anvend altid samme koder, fx 1 for ja og 2 for nej igennem hele skemaet. (0 er ikke god at taste hurtigt). Overvej grundigt, om et givet spørgsmål indeholder både udtømmende (exhaustive) og gensidigt udelukkende svarmuligheder (Dvs om alle svarpersoner kan finde et og kun et gyldigt svar). Hvis der kan sættes kryds ved flere underspørgsmål skal der for hvert af underspørgsmålene kodes en variabel. "Ved ikke" bør ikke angives som svarmulighed. Erfaringen viser, at personer som er i stor tvivl alligevel ikke svarer på et spørgsmål.



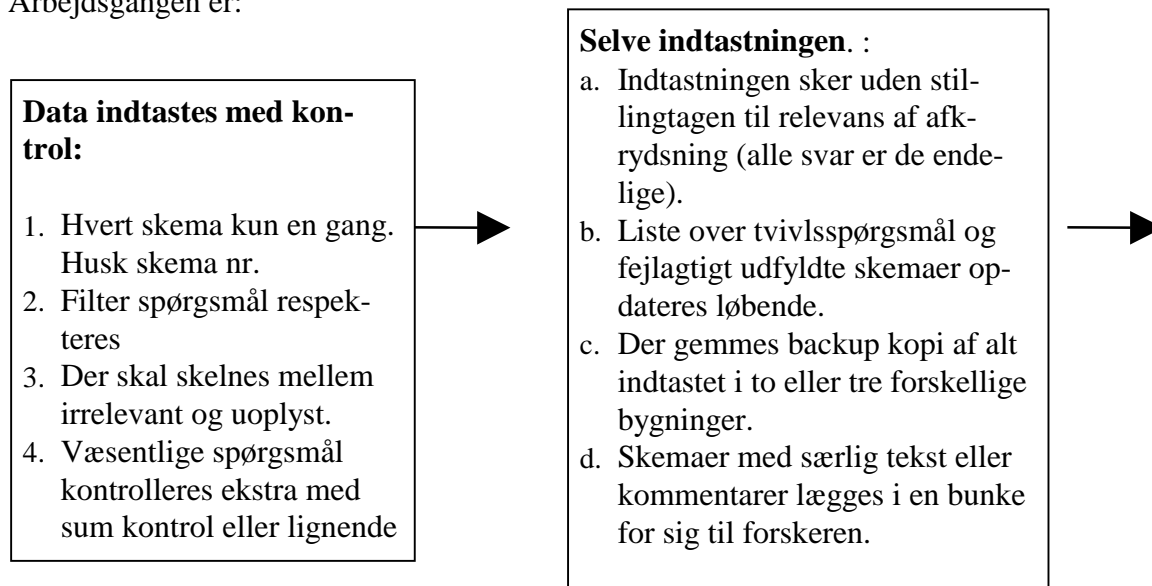
“Tipskupon” formuleringer af spørgsmål er ikke anbefalelsesværdige. Det er min opfattelse, at der bliver for store fejlmuligheder. I en tipskupon kan der på lidt plads stå mange svar, men det er meget let at sætte kryds et forkert sted, samt taste et forkert svar ind.

Søg efter tidligere formuleringer for givne spørgsmål i Dansk Data Arkiv. Spørgsmålene kan kopieres direkte over i tekstbehandlingsanlæg. Internet adresse: (WWW): <http://www.dda.dk>



tastes eller korrekturlæses ?) og lignende. Det giver stor indsigt i mulige fejl og rækkevidden af konkrete spørgsmål, at kontrollere hele processen selv. Selve indtastningen kan derimod varetages af hjælpepersonnel. Som udgangspunkt bør der dobbeltindtastes så meget, at fejlprocent (promille !!) kan dokumenteres.

Arbejdsgangen er:



### Øvelse 31 Dit eget projekt - beslut en indtastningsmåde.

Beslut dig for en indtastningsmetode, herunder hvordan du vil sikre dig, at have sikkerhedskopier af data til enhver tid.

### Øvelse 32 Dit eget projekt - første dokument om oparbejdning af data.

Beskriv den samlede struktur for data i projektet, de enkelte skemaer, de enkelte spørgsmål og hvilke variabler, der udgår fra spørgsmålene. Indtastningsmetoden beskrives i princip.

Beskrivelsen skal udformes som et resumé af øvelse 30-33 på en form, der er velegnet til at indgå som bilag i en rapport eller ph.d. afhandling. Omfang nogle få (én) sider med tekst vedlagt de konkrete skemaer.



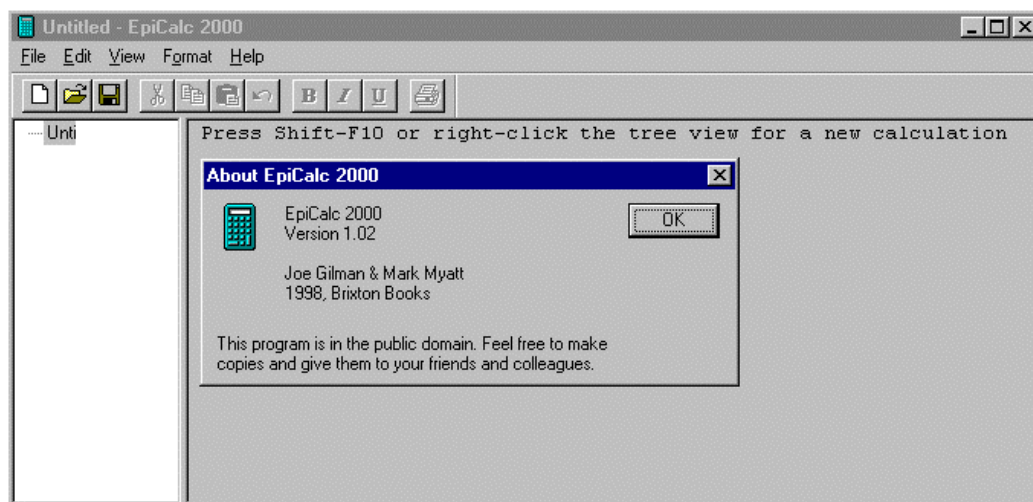
## Epicalc 2000

Hvis der kun skal beregnes et enkelt konfidensinterval, en odds ratio eller en forventet teststyrke ved planlægning af en undersøgelse kan det være nyttigt med et specialprogram. En slags udvidet "lommeregner". EpiCalc 2000 er et sådant program.

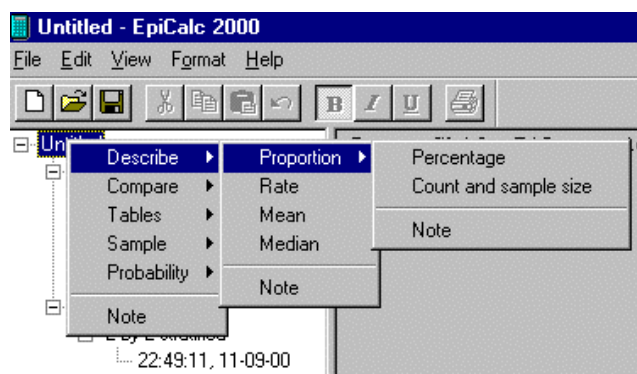
### Øvelse 33 Hent EpiCalc 2000 fra internet og installér det.

Du henter blot filen fra <http://www.myatt.demon.co.uk/>. Størrelsen er cirka 1.54 Mb, så det kan godt være du får brug for en rimelig hurtig internetforbindelse. Efter download skal du installere

ud fra den modtagne fil, som i version 1.02 hedder ec2v102.exe. (peg på filen og klik). Efter installation kan du starte programmet og skærmbilledet ser sådan ud:



For at få den første beregning udført taster du shift+f10 eller højre muse klik, mens du peger på "untitled". Herefter kommer en delmenu frem. Derefter er programmet selvforklarende.



### Øvelse 34 Konfidensinterval for en proportion.

Beregn et 95 % konfidensinterval for en proportion på 0.21 og en stikprøve på 234. Hvad er resultatet og hvordan tolkes det ?

### Øvelse 35 Stikprøvestørrelse.

I en population er 70 % som har et bestemt kendetegn, mens der i referencepopulationen er en hyppighed på 80 %. Hvor stor en stikprøve skal der udtages, hvis nulhypotesen om "ingen forskel" skal forkastes på 5% niveau med en teststyrke på 80 % ?

## Euroqol 5-d<sup>10</sup>

Et af de generelle sundhedsstatusmål hedder Euroqol-5d.

Der er indsamlet øvelsesdata på et møde, hvor deltagerne udfyldte skemaet ved mødets start og igen tre timer senere. En kopi af skemaet findes efterfølgende. I eksempelsamlingen findes der datafiler med navnet eqfirst.rec og eqsecond.rec

Indhold	Kodning	Navn på variabel.	
		Indtastning af data fra mødets:	
id	Identifikation af besvarelsen.	start	slut
alvorlig sygdom	Har du oplevet alvorlig sygdom hos a: dig selv b: i din familie c: ved at passe andre	id	id
		s1a	t1a
		s1b	t1b
		s1c	t1c
alder	angivet antal år ved besvarelse	s2	t2
køn	1=mand 2=kvinde	s3	t3
rygning	1: nuværende ryger, 2: forhenværende ryger 3: aldrig ryger	s4	t4
sundhedsarbejder	arbejder du/har du nogensinde arbejdet i social og sundhedssektoren 1:ja 2:nej	s5	t5
beskæftigelse	Hovedbeskæftigelse er 1: i beskæftigelse /selverhvervende 2:pensioneret ...6.andet	s6	t6
uddannelse	Fortsatte du uddannelse efter folkeskolen 1:ja 2:nej	s7	t7
universitet	Har du en universitetsuddannelse 1:ja 2:nej	s8	t8
euroqol dimensioner	De seks euroqol spørgsmål, se bagest i noten	eq1-eq6	teq1-teq6
euroqol samlet	Aflæsning af visual analog skala (0-100)	eqg	teqg

Data fra første udfyldelse er tastet to gange og findes i filerne: eqfirst.rec og first-xt.rec

Data fra anden udfyldelse findes i filen: eqsecond.rec . Variabelnavne fra eqsecond.rec er som vist ovenfor, blot er der foran navnet sat et "t", dvs at s7 i anden indtastning hedder t7. Id variabelen hedder det samme i begge filer.

Angiv, ved at sætte kryds (sådan ) i én af kasserne i hver gruppe, hvilke udsagn, der bedst beskriver din helbredstilstand i dag.

### Bevægelighed

- Jeg har ingen problemer med at gå omkring   
 Jeg har nogle problemer med at gå omkring   
 Jeg er bundet til sengen

### Personlig pleje

- Jeg har ingen problemer med min personlige pleje .....  
 Jeg har nogle problemer med at vaske mig eller klæde mig på .....  
 Jeg kan ikke vaske mig eller klæde mig på.....

### Sædvanlige aktiviteter (fx. arbejde, studie, husarbejde, familie- eller fritidsaktiviteter)

- Jeg har ingen problemer med at udføre mine sædvanlige aktiviteter .....  
 Jeg har nogle problemer med at udføre mine sædvanlige aktiviteter .....  
 Jeg kan ikke udføre mine sædvanlige aktiviteter .....

### Smertes/ubehag

- Jeg har ingen smerter eller ubehag .....  
 Jeg har moderate smerter eller ubehag .....  
 Jeg har ekstreme smerter eller ubehag .....

### Angst/depression

- Jeg er ikke ængstelig eller deprimeret .....  
 Jeg er moderat ængstelig eller deprimeret .....  
 Jeg er ekstremt ængstelig eller deprimeret .....

Sammenlignet med min helbredstilstand gennem de seneste 12 måneder, er min helbredstilstand i dag:

- Sæt kryds i  
én kasse
- Bedre .....  
 Stort set den samme .....  
 Værre .....

“Nu vil jeg gerne beskrive din helbredstilstand i dag med et enkelt tal.”

For at sige hvor god eller dårlig din helbredstilstand er i dag skal du forestille dig, at vi kunne måle den ligesom med et termometer. Hvor den bedste helbredstilstand du kan forestille dig er 100 og den værste helbredstilstand du kan forestille dig er 0”

“På en sådan skala fra 0 til 100, hvor godt eller dårligt mener du dit eget helbred er i dag ?”

“Sig et tal fra 0 til 100” (Tegn tallet ind i figuren)

Tegn en streg her fra og til det tal, som bedst svarer til din helbredstilstand i dag

OBS, den originale form skal anvendes til konkrete formål, dette er blot en tilnærmet !!

