

11. oktober 2007
j.nr. 1.2002.82
MD / Iny



Notat

Sundheds- og sygelighedsundersøgelsen 2005 – stikprøvekonstruktion, opnåelse, vægtning samt bortfald og konsekvenser af dette

Michael Davidsen
Statens Institut for Folkesundhed
Syddansk Universitet
E-mail: md@si-folkesundhed.dk

Indholdsfortegnelse

1. Indledning	4
2. Udvælgelse af stikprøven til SUSY-2005	4
2.1 Den kommunale og regionale struktur	5
2.2 Konstruktion af den samlede stikprøve	7
2.2.1 Krav	7
2.2.2 Stikprøvestørrelse baseret på forventet deltagelsesprocent	7
2.3. Delstikprøver	8
2.3.1 Reinterviewstikprøven	8
2.3.2 Den regionalt supplerende stikprøve	10
2.4 Selvadministrerede spørgeskemaer	11
3. Opnåelse	12
3.1 Opnåelse i forhold til SUSY-1994	14
4. Bortfald	15
4.1 Metode	16
4.2 Bortfaldsanalyse for i-bortfald (bortfald efter interview)	19
4.3 Bortfaldsanalyse for s-bortfald (det selvadministrerede skema)	21
4.3.1 s-bortfald blandt inviterede	21
4.3.2 s-bortfald blandt interviewede	23
5. Repræsentativitet	29
6. Vægtning	31
6.1 Vægtning for børn	33
6.2 Lidt statistisk baggrund	34
7. Bortfalds betydning for forekomst af udvalgte indikatorer	35
7.1 Diskussion	41
8. Dataindsamlingsmetoder og deres betydning	42
8.1 Metoder	42
8.2 Betydningen af dataindsamlingsmetode	43
8.2.1 Overgang af interviewerens spørgeskemaudfyldelse fra papir til computer (konvertering fra PAPI til CAPI)	43
8.2.2 Betydning af den årstid hvori data indsamles	44
8.2.3 Betydning af supplerings med telefoninterview (CATI)	45
8.2.4 Betydning af interviewere	45
Resumé	46

9. Referencer	48
10. Bilag.....	49
10.1 Invitation af tidligere inviterede (SUSY-1994 og SUSY-2000)	49
10.2 'Associated paper' om årstidsvariation.....	50
10.3 Paper om effekt af telefon-interview.....	66

1. Indledning

Dette notat er et baggrundsnotat, der redegør for design og analyse-mæssige forhold omkring sundheds- og sygelighedsundersøgelsen 2005. Denne undersøgelse er den fjerde generelle sundheds- og sygelighedsundersøgelse, som Statens Institut for Folkesundhed (SIF) har udført – de tidligere undersøgelser er udført i 1987 (Rasmussen et.al. 1988), 1994 (Kjøller et.al. 1995) og 2000 (Kjøller et.al. 2002). Resultaterne fra sundheds- og sygelighedsundersøgelsen år 2005 (SUSY-2005) er beskrevet i Ekholm et.al. 2006 og nærværende notat er en uddybning af metode- og materialebeskrivelsen i denne afrapportering.

Baggrunden for notatet er dels et ønske om at samle det metodemæssige arbejde, der blev lavet inden og i forbindelse med undersøgelsen, og dels at give en bedre beskrivelse af undersøgelsens bortfald og konsekvenser heraf. Dette hænger sammen med, at bortfaldet har været stigende de senere år og at det derfor er vurderet, at der bør gøres mere ud af dette aspekt.

I notatet gøres der først rede for konstruktionen af stikprøven (størrelse og udvælgelse af personer - afsnit 2), dernæst gennemgås opnåelsen i undersøgelsen og dele heraf (afsnit 3). Forskellige typer af bortfald gennemgås i afsnit 4, efterfulgt af en gennemgang af repræsentativitet (afsnit 5). I afsnit 6 omtales den vægtning, der er foretaget i SUSY-2005 og begrundelsen herfor, og i afsnit 7 gives en vurdering af effekten af bortfald i forhold til forekomsten af nogle centrale indikatorer for sundhed og sygelighed. I afsnit 8 gives et kort resumé af arbejde udført omkring dataindsamlingsmetoder og konsekvenser heraf, et arbejde, der primært blev gennemført inden undersøgelsens start i maj 2005. I afsnit 10 (bilag) vises de artikler og referencer, der kort er resumeret i afsnit 8.

2. Udvalgelse af stikprøven til SUSY-2005

Alle sundheds- og sygelighedsundersøgelser er gennemført på baggrund af en stikprøve af voksne danskere (alder 16 år og derover). Stikprøven er lavet ved at invitere et tilfældigt udsnit af voksne danskere til at deltage i undersøgelsen. I dette afsnit beskrives det, hvordan stikprøven til SUSY-2005 blev konstrueret. Der redegøres for de krav, der blev stillet til stikprøven, hvordan stikprøven konkret blev afgrænset (herunder forhold omkring forskerbeskyttelse) samt de forskellige delstikprøver, undersøgelsen er sammensat af.

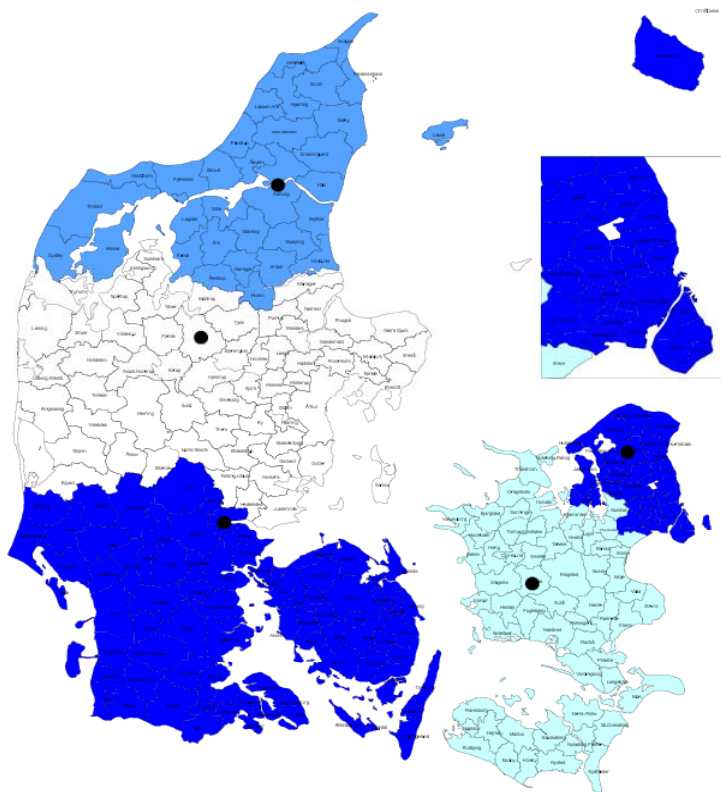
Den konkrete tekniske udvælgelse af stikprøven til SUSY-2005 blev forestået af SFI-SURVEY med godkendelse fra SIF.

2.1 Den kommunale og regionale struktur

Designet af SUSY-2005 blev påbegyndt i efteråret 2004. Den kommunale strukturreform med ikrafttrædelse 1. januar 2007 har spillet en central rolle ved konstruktionen af stikprøven til SUSY-2005. Som bekendt overtog regionerne som følge af strukturreformen ansvaret for sygehusområdet, mens ansvaret for forebyggelse og genoptræning blev placeret i kommunerne. Da SUSY-undersøgelserne primært har været brugt indenfor forebyggelse og planlægning og langt mindre indenfor sygehusbehandling, blev en stikprøveplan baseret på udvælgelse fra kommunerne diskuteret. En sådan stikprøve var det ikke muligt at implementere, primært fordi loven ikke var vedtaget, da stikprøven skulle konstrueres. Derfor blev der i stedet fokuseret på de nye regioner som naturlig geografisk enhed.

Figur 2.1 viser den regionsinddeling, der var kendt pr. 1. marts 2005, hvor den endelige stikprøveplan til SUSY-2005 blev fastlagt. Der er 5 regioner: Region Hovedstaden, Region Sjælland, Region Syddanmark, Region Midtjylland og Region Nordjylland. Ved den konkrete definition af hver region var SIF nødt til at tage udgangspunkt i de kommuner, der eksisterede før 1. januar 2007. Regionsinddelingen baseret på disse kommuner fremgår af tabel 2.1.

Figur 2.1: Regionsinddeling pr. 1. marts 2005.



Tabel 2.1: Regionsopdeling baseret på danske kommuner 2005.

Region Hovedstaden

København, Frederiksberg, Ballerup, Brøndby, Dragør, Gentofte, Gladsaxe, Glostrup, Herlev, Albertslund, Hvidovre, Høje-Taastrup, Ledøje-Smørum, Lyngby-Tårnbæk, Rødovre, Søllerød, Ishøj, Tårnby, Vallensbæk, Værløse, Allerød, Birkerød, Farum, Fredensborg-Humlebæk, Frederikssund, Frederiksværk, Græsted-Gilleleje, Helsingør, Helsingør, Hillerød, Hundested, Hørsholm, Jægerspris, Karlebo, Skibby, Skævinge, Slangerup, Stenløse, Ølstykke, Bornholm

Region Sjælland

Bramsnæs, Greve, Gundsø, Hvalsø, Køge, Lejre, Ramsø, Roskilde, Skovbo, Solrød, Vallø, Bjergsted, Dianalund, Dragsholm, Fuglebjerg, Gørlev, Hashøj, Haslev, Holbæk, Hvidebæk, Høng, Jernløse, Kalundborg, Korsør, Nykøbing-Rørvig, Ringsted, Skælskør, Slagelse, Sorø, Stenlille, Svinninge, Tornved, Trundholm, Tølløse, Fakse, Fladså, Holeby, Holmegaard, HøjrebyLangebæk, Maribo, Møn, Nakskov, Nykøbing F, Nysted, Næstved, Nørre Alslev, Præstø, Ravnsborg, Rudbjerg, Rødby, Rønnede, Sakskøbing, Stevns, Stubbekøbing, Suså, Sydfalster, Vordingborg,

Region Syddanmark

Assens, Bogense, Broby, Egebjerg, Ejby, Faaborg, Glamsbjerg, Gudme, Hårby, Kerteminde, Langeskov, Marstal, Middelfart, Munkebo, Nyborg, Nørre-Åby, Odense, Otterup, Ringe, Rudkøbing, Ryslinge, Svendborg, Sydlangeland, Søndersø, Tommerup, Tranekær, Ullerslev, Vissenbjerg, Ærøskøbing, Ørbæk, Årlev, Årup, Augustenborg, Bov, Bredebro, Broager, Christiansfeld, Gram, Gråsten, Haderslev, Højer, Lundtoft, Løgumkloster, Nordborg, Nørre-Rangstrup, Rødding, Rødekro, Skærbæk, Sundeved, Sydals, Sønderborg, Tinglev, Tønder, Vojens, Åbenrå, Billund, Blåbjerg, Blåvandshuk, Bramming, Brørup, Esbjerg, Fanø, Grindsted, Helle, Holsted, Ribe, Varde, Vejen, Ølgod, Børkop, Egtved, Fredericia, Give, Jelling, Kolding, Lunderskov, Vamdrup, Vejle,

Region Midtjylland

Brædstrup, Gedved, Hedensted, Horsens, Juelsminde, Nørre-Snedede, Tørring-Uldum, Aulum-Haderup, Brande, Egvad, Herning, Holmsland, Holstebro, Ikast, Lemvig, Ringkjøbing, Skjern, Struer, Thyborøn-Harboøre, Thyholm, Trehøje, Ulfborg-Vemb, Videbæk, Vinderup, Aaskov, Ebeltoft, Galten, Gjern, Grenaa, Hadsten, Hammel, Hinnerup, Hørning, Langå, Midtdjurs, Nørhald, Nørre Djurs, Odder, Purhus, Randers, Rosenholm, Rougsø, Ry, Rønde, Samsø, Silkeborg, Skanderborg, Sønderhald, Them, Århus, Bjerringbro, Fjends, Hvorslev, Karup, Kjellerup, Møldrup, Sallingsund, Skive, Spøttrup, Sundsøre, Tjele, Viborg,

Region Nordjylland

Mariager, Hanstholm, Morsø, Sydthy, Thisted, Ålestrup, Arden, Brovst, Brønderslev, Dronninglund, Farsø, Fjerritslev, Frederikshavn, Hadsund, Hals, Hirtshals, Hjørring, Hobro, Læsø, Løgstør, Løkken-Vrå, Nibe, Nørager, Pandrup, Sejlflod, Sindal, Skagen, Skørping, Støvring, Sæby, Åbybro, Aalborg, Års

2.2 Konstruktion af den samlede stikprøve

2.2.1 Krav

Inden den samlede stikprøve til SUSY-2005 blev valgt, blev der defineret en række krav til den. Disse kan opdeles i krav, som har været gældende for alle SUSY-undersøgelser og nye krav til SUSY-2005.

Generelle krav til SUSY-undersøgelser:

- stikprøven skal omfatte voksne danskere – dvs. danske statsborgere alder 16 og derover
- udvælgelsen af personer skal være baseret på tilfældig udvælgelse.

Som ved alle tidligere SUSY-undersøgelser skulle dataindsamlingen foregå ved et personligt interview i respondenternes hjem (besøgsinterview).

De nye krav var:

- Stikprøven skulle være af en sådan størrelse at der blev sikret cirka lige så mange interview som i SUSY-2000 – her blev 16.688 personer interviewet
- Undersøgelsen skulle kunne bruges af de nye regioner
- Der skulle gennemføres et tredje interview af personer fra SUSY-1994,

Som følge af a. og b. ovenfor blev det valgt, at der skulle sigtes mod en stikprøve, der indeholdt 15.000 gennemførte besøgsinterview, 3.000 i hver region.

Den konkrete udvælgelse af stikprøven og størrelsen på forskellige dele af den, vil blive gennemgået nedenfor.

I forbindelse med konstruktionen af stikprøven er forhold omkring forskerbeskyttelse blevet inddraget. I Danmark er det muligt ved flytning at afkrydse et felt på flytteblanketten, der omhandler beskyttelse mod forskerhenvendelse - afkrydningsfeltet står sammen med bl.a. 'ønsker ikke reklamer'. Flytteblanketten er tilgængelig på Internettet. I 2006 havde ca. 10% af den danske befolkning benyttet sig af denne mulighed, fortrinsvist blandt de unge og lavere socialgrupper (upublicerede opgørelser fra Danmarks Statistik).

2.2.2 Stikprøvestørrelse baseret på forventet deltagelsesprocent

Da ikke alle inviterede personer ønsker at deltage i en undersøgelse, er det nødvendigt at invitere flere personer end dem, der forventes gennemført interview med. Derfor blev der inden undersøgelsens start fastlagt en andel, der forventedes at gennemføre et interview.

Som i andre surveys (undersøgelser) er der i sundheds- og sygelighedsundersøgelserne sket et fald i deltagelsesprocenten. I SUSY-1987 var deltagelsesprocenten 79,9 %, i SUSY-1994 77,2 % og i SUSY-2000 74,2 % (Davidsen et al. 2004 - UfL). Derfor blev der i SUSY-2005 fastsat en forventet deltagelsesprocent på 70 %.

I undersøgelserne i 1987, 1991 og 2000 har deltagelsesprocenten i hovedstadsområdet været lavere end i det øvrige Danmark. I SUSY-2000 var deltagelsesprocenten i København og Frederiksberg Kommune 64 %, i Københavns Amt 67 %, mens den i alle andre amter var 70 % eller derover. På den baggrund blev der arbejdet med en deltagelsesprocent i region Hovedstaden på mindst 64 %, mens den i de øvrige regioner skulle være mindst 70 %.

På denne baggrund blev der valgt en stikprøvestørrelse på 21.832 personer. Jf. afsnit 2.1 sigtedes der mod 3.000 gennemførte interviews pr. region. Fra Region Hovedstaden blev der derfor valgt $4.688 = 3.000 / 0,64$ personer, og fra de øvrige 4 regioner $4.286 = 3.000 / 0,70$ personer. Det skal bemærkes, at denne beregning af stikprøvestørrelse ikke er helt hensigtsmæssig, idet den nødvendigvis måtte resultere i en for lav samlet deltagelsesprocent ($100 * (5 * 3.000 / 21832) = 68,7 \%$). I tabel 2.2 ses en oversigt over stikprøvestørrelse og udregning heraf.

2.3. Delstikprøver

For at sikre et tredje interview af personer fra SUSY-1994, blev stikprøven opdelt i to delstikprøver – reinterviewstikprøven og den regionalt supplerende stikprøve.

2.3.1 Reinterviewstikprøven

Reinterviewstikprøven i SUSY-2005 består af følgende to dele:

- alle personer inviteret til geninterviewstikprøven SUSY-2000 og i live 18. marts 2005
- et supplement af unge, der pr. 1. januar 2005 var mellem 16 og 20 år gamle.

2.3.1.1 Geninterviewstikprøven SUSY-2000

I SUSY-2000 bestod den samlede stikprøve af tre delstikprøver (Kjøller 2002), hvoraf en var geninterviewstikprøven. Denne bestod af 5.912 personer og var sammensat på følgende måde:

- alle inviterede til SUSY-1994, som var i live år 2000 (n=5316)
- en stikprøve af unge alder 16-22 år (n=486)

- en stikprøve af personer, der opnåede dansk statsborgerskab mellem 1994 og 2000 (n=110).

Ideen med at sammensætte geninterviewstikprøven på denne måde er at gøre den repræsentativ for den voksne danske befolkning i år 2000. Stikprøven bliver så at sige suppleret op nedefra (med stikprøven af unge) efterhånden som de ældre udgår. Denne ide blev gentaget i år 2005.

I forbindelse med opdatering af alle SUSY-undersøgelser i CPR-registeret, er det konstateret, at der findes personer fra SUSY-1994, der var i live i år 2000, men som ikke blev inviteret til SUSY-2000. Da det drejer sig om ca. 130 personer, er det vurderet, at der ikke udelukkende er tale om problemer med e- og immigration, men også om en reel fejl i forbindelse med SUSY-2000. Derfor er 62 personer, der opfylder nedenstående tre krav, også blevet betragtet som en del af geninterviewstikprøven. Kravene var:

- de var inviteret til SUSY-1994
- de var ikke inviteret til SUSY-2000
- de var i live pr. 1. september 2003

Personerne benævnes nedenfor 'nye geninterviewede'. Det skal bemærkes, at 'nye geninterviewede' ikke har specielle karakteristika mht. køn, alder eller region.

Den samlede geninterviewstikprøve fra SUSY-2000 består således af $5.912+62=5.974$ personer. Der blev indhentet status-oplysninger fra CPR-registeret for disse personer. I alt 517 (8,7%) havde markeret, at de ikke ønskede henvendelse fra forskere (forskerbeskyttelse jf. afsnit 2.2.1), 5.021 (84,1%) var i live, 349 (5,8%) var døde, 76 (1,3%) var emigreret, mens 11 havde andre koder.

I både SUSY-1994 og SUSY-2000 er alle, der blev interviewet, blevet spurgt, om SIF måtte opsøge dem igen. Denne oplysning er sammen med oplysningen om forskerbeskyttelse blevet brugt i forbindelse med invitationen af personer til SUSY-2005. En detaljeret oversigt findes i bilag 10.1 - det generelle princip har været, at hvis man på et tidspunkt har angivet, at SIF gerne måtte kontakte en, blev man inviteret til SUSY-2005. I alt blev 5.377 personer fra geninterviewstikprøven i SUSY-2000 inviteret til SUSY-2005. Se figur 2.2 for en grafisk illustration af gennemført reinterview af geninterviewstikprøven i SUSY-2000.

Tabel 2.2: Udregning af stikprøvestørrelse til SUSY-2005

Region	Forventet antal interviews	Forventet svarprocent	Antal inviterede
Hovedstaden	3.000	64	4.688
Sjælland	3.000	70	4.286
Syddanmark	3.000	70	4.286
Midtjylland	3.000	70	4.286
Nordjylland	3.000	70	4.286
Total	15.000		21.832

2.3.1.2 Supplement af unge

Stikprøven af unge i alderen 16-20 år blev valgt på følgende måde:

I følge Danmarks Statistik udgjorde de 16-20-årige 6,8 % af den voksne danske befolkning pr. 1. januar 2005. Med en forventet samlet størrelse af reinterviewstikprøven på 6.000 personer (den samlede stikprøvestørrelse af SUSY-1994), skulle der således vælges $6.000 * 0,068 = 410$ unge i alderen 16-20 år til stikprøven.

Udvælgelse af unge i alderen 16-20 år blev foretaget på en sådan måde, at kun personer, der ikke ønskede forskerbeskyttelse, blev inviteret.

Den samlede reinterviewstikprøve 2005 består således af 5.377 fra geninterviewstikprøven i SUSY-2000 plus 410 unge i alderen 16-20 år, dvs. i alt 5.787 personer.

2.3.2 Den regionalt supplerende stikprøve

Som anført i afsnit 2.2 skulle der gennemføres 3.000 interview i hver af de 5 regioner. Da reinterviewstikprøven ikke indeholdt tilstrækkeligt med personer, var det nødvendigt med en ekstra delstikprøve. Denne benævnes *den regionalt supplerende stikprøve*.

I tabel 2.3 ses udregningen af, hvor mange personer der efter udvælgelse af personer til reinterviewstikprøven skal indgå i den regionalt supplerende stikprøve i hver region og i alt

Tabellen viser, at der eksempelvis i region Hovedstaden skal inviteres i alt 4.688 personer (afsnit 2.2). I den konstruerede reinterviewstikprøve findes der 1.585 personer fra regionen. Der skal således inviteres $4.688 - 1.585 = 3.103$ personer fra regionen til den regionalt supplerende stikprøve.

Tabel 2.3: Udregning af antal inviterede personer til den regionalt supplerende stikprøve til SUSY-2005.

Region	Inviteres	Fra re-interview stikprøven	Inviteres til den regionalt supplerende stikprøve
Hovedstaden	4.688	1.585	3.103
Sjælland	4.286	857	3.429
Syddanmark	4.286	1.381	2.905
Midtjylland	4.286	1.324	2.962
Nordjylland	4.286	640	3.646
I alt	21.832	5.787	16.045

I alt består den regionalt supplerende stikprøve således af 16.045 personer. Stikprøven blev valgt, så personer, der ønskede forskerbeskyttelse, blev ekskluderet inden stikprøven konkret blev valgt. For detaljer herom henvises til notatets forfatter.

2.4 Selvadministrerede spørgeskemaer

Som ved undersøgelsen i 1994 og 2000 skulle der i SUSY-2005 udleveres et selvadministreret spørgeskema ved afslutningen af det personlige interview. Det blev besluttet, at reinterviewstikprøven skulle have et spørgeskema med nogen specifikke spørgsmål. For at sikre andre brugere af undersøgelsen, en tilstrækkelig stor stikprøve til tværsnitsanalyser, blev det også besluttet at dele den samlede stikprøve i 2 lige store dele med $21.832/2=10.916$ personer i hver del.

Alle 5.787 personer i reinterviewstikprøven blev valgt til at få udleveret spørgeskema 2. De øvrige 16.455 personer blev tilfældigt opdelt i en gruppe på 10.916, der skulle have spørgeskema 1, og en gruppe på 6.039, der ligesom personer i reinterviewstikprøven fik udleveret skema 2.

3. Opnåelse

I afsnit 2 blev det gennemgået, hvordan stikprøven og dens dele er konstrueret. I dette afsnit vil der blive fokuseret på, hvad der kom ud af den konkrete dataindsamling i form af hvor mange personer, der gennemførte et interview hhv. returnerede et selvadministreret spørgeskema.

I afsnit 4 er der fokuseret på bortfald, dvs. de personer, der ikke opnås svar fra. Betydningen af den anvendte dataindsamlingmetode er belyst i afsnit 8.

Den konkrete dataindsamling foregik mellem 1. maj 2005 og 1. marts 2006. I forhold til det personlige interview anses en person, der helt eller delvist gennemfører et interview for at være en svarperson. Der er altså opnået interview med vedkommende, interviewet er gennemført.

I forhold til returnering af det selvadministrerede spørgeskema anses en person, der returnerer et spørgeskema i helt eller delvist udfyldt stand for at have besvaret det selvadministrerede spørgeskema.

Blandt de 21.832 personer udvalgt til stikprøven i SUSY-2005 blev der gennemført interview med 14.566 personer (66,7 %).

I alt modtog 13.977 personer et selvadministreret spørgeskema, og 11.238 returnerede det i helt eller delvist udfyldt stand. I forhold til hele stikprøven på 21.832 har 51,5 % returneret et selvadministreret spørgeskema, mens det i forhold til de 14.566, der gennemførte et interview og dermed havde mulighed for at udfylde skemaet, betyder at 77,2 % har returneret skemaet. Der blev returneret 5.686 spørgeskemaer af type 1 og 5.552 af type 2 (se afsnit 2.4).

I tabel 3.1 ses opnåelsesprocenter fordelt på regioner. Det ses, at der er gennemført over 3.000 interviews i region Syddanmark, hvor der også ses den højeste opnåelse ved interview (72,1 %). Der er gennemført lidt under 3.000 interviews i de øvrige regioner og opnåelsen er lavest i region Hovedstaden (62,4 %). Andelen, der har returneret det selvadministrerede spørgeskema blandt de interviewede, er højst i region Syddanmark og lavest i Region Hovedstaden.

Tabel 3.1: Opnåelse mht. interview og selvadministreret spørgeskema i SUSY-2005

Region	Interview		Selv-administreret spørgeskema		
	Antal	Svarprocent	Antal	Procent - interview	Procent - inviterede
Hovedstaden	2.923	62,4	2.248	76,9	48,0
Sjælland	2.818	65,7	2.156	76,5	50,3
Syddanmark	3.089	72,1	2.421	78,4	56,5
Midtjylland	2.874	67,1	2.260	78,6	52,7
Nordjylland	2.862	66,8	2.153	75,2	50,2
I alt	14.566	66,7	11.238	77,2	51,5

I tabel 3.2 ses responsraten for både interview og selvadministreret spørgeskema fordelt på de to stikprøver samt på typen af planlagt udleveret selvadministreret spørgeskema. Det ses, at svarprocenten ved interview er lidt højere i reinterviewstikprøven (67,2 %) end i den regionalt supplerende (66,5 %). Det ses også, at stikprøven af nye unge (alder 16-20) har den laveste opnåelse for returnering af det selvadministrerede spørgeskema.

Tabel 3.2: Opnåelse i SUSY-2005 fordelt på stikprøver

Stikprøve	Inviteret		Interview		Selvadministreret		
	n		n	%	n	% af inviterede	% af interviewede
Regionalt supplerende	16.045		10.675	66,5	8.214	51,2	76,9
Reinterview	5.787		3.891	67,2	3.024	52,3	77,7
- Unge alder 16-20	410		268	65,4	167	40,7	62,3
- Geninterview	5.377		3.623	67,4	2.857	53,1	78,9
- Inviteret til SUSY-1994	4.825		3.262	67,6	2.591	53,7	79,4
- Unge alder 16-21	433		292	67,4	215	49,7	73,6
- Opnået statsborgerskab mellem 1994 og 2000	67		41	61,2	28	41,8	68,3
- Ny geninterview	52		28	53,8	23	44,2	82,1
Selvadministreret skema							
Spørgeskema 1	10.916		7.291	66,8	5.686	52,1	78,0
Spørgeskema 2	10.916		7.275	66,6	5.552	50,9	76,3

I tabel 3.3 ses en oversigt over de begrundelser der er givet for ikke at deltage i SUSY-2005 fordelt på regioner. Det ses, at den laveste andel af nægttere findes i Region Syddanmark, mens den højeste andel af ej trufne er i Region Hovedstaden.

Tabel 3.3: Grunde til bortfald fra interview fordelt på regioner. SUSY-2005

Region	Interview gennemført	Nægter	Ej hjemme - hospital	Ej truffet	Andet
Hovedstaden	62,4	22,4	1,0	8,6	5,6
Sjælland	65,8	23,0	0,8	5,0	5,5
Syddanmark	72,1	19,3	0,6	2,9	5,1
Midtjylland	67,1	23,2	0,4	3,9	5,4
Nordjylland	66,8	24,4	0,6	3,2	5,0
Total	66,7	22,5	0,7	4,8	5,3

3.1 Opnåelse i forhold til SUSY-1994

I SUSY-undersøgelserne er der som noget unikt indarbejdet et geninterview af stikprøven fra SUSY-1994, som er suppleret både i år 2000 og 2005. For en nærmere forklaring se afsnit 2.3.1.

I dette afsnit vil det blive opgjort, hvor mange gennemførte interviews, der findes på de samme personer, når der

- tages udgangspunkt i at følge stikprøven fra SUSY-1994. Personer kan være interviewet 0-3 gange (i 1994, 2000 og 2005).
- tages udgangspunkt i geninterviewstikprøven fra SUSY-2000. Personer kan være interviewet 0-2 gange (i år 2000 og 2005).

I tabel 3.4 ses antal interview for personer i SUSY-1994 (n=5.995) fordelt på alder i 1994. Det ses, at 2.598 personer (43,3 %) har gennemført tre interviews mens 762 ikke har gennemført noget interview overhovedet. Det ses også, at andelen, der har gennemført tre interview, er højest blandt de personer, der i 1994 var mellem 25 og 44 år, mens andelen, der ingen interview har gennemført, er højest blandt personer på 65 år eller derover. Det kan af tabellen ikke ses, hvilke undersøgelser, personer der har fået et eller to interview, har deltaget i, ligesom antallet af dødsfald ikke kan ses.

Tabel 3.4: Antal gennemførte interviews 1994-2000-2005 for personer inviteret til SUSY-1994

Alder 1994		Antal gennemførte interview				I alt
		0	1	2	3	
16-24	n	52	149	283	424	908
	%	5,7	16,4	31,2	46,7	
25-44	n	208	305	504	1.154	2.171
	%	9,6	14,0	23,2	53,2	
45-64	n	225	306	364	814	1.709
	%	13,2	17,9	21,3	47,6	
65+	n	277	463	261	206	1.207
	%	22,9	38,4	21,6	17,1	
I alt	n	762	1.223	1.412	2.598	5.995
	%	12,7	20,4	23,6	43,3	

I tabel 3.5 ses en tilsvarende opgørelse med udgangspunkt i geninterviewstikprøven fra SUSY-2000 (n=5.912).

Tabel 3.5: Antal gennemførte interviews 2000-2005 for personer inviteret til geninterviewstikprøven i SUSY-2000

Alder 2000		Antal interviews			
		0	1	2	I alt
16-24	n	118	239	415	772
	%	15,3	31,0	53,8	
25-44	n	347	570	1.234	2.151
	%	16,1	26,5	57,4	
45-64	n	357	454	1.069	1.880
	%	19,0	24,1	56,9	
65+	n	290	408	411	1.109
	%	26,1	36,8	37,1	
I alt	n	1.112	1.671	3.129	5.912
	%	18,8	28,3	52,9	

Det ses af tabel 3.5, at i alt 3129 personer (52,9 %) har deltaget både i SUSY-2000 og SUSY-2005, mens 1.112 ikke har deltaget i nogen af undersøgelseerne. Som i tabel 3.4 ses andelen, der har deltaget 0 gange, at være højest blandt personer i alder 65 og derover i år 2000.

4. Bortfald

I afsnit 3 blev opnåelsen i SUSY-2005 gennemgået i form af opgørelser af, hvor mange personer, der gennemførte et interview hhv. returnerede det selvadministrerede spørgeskema, altså deltog i undersøgelsen. I dette afsnit vil der blive fokuseret på bortfald – de personer, der *ikke* deltog i forskellige dele af undersøgelsen. Afsnittet vil indeholde en række analyser til belysning af bortfaldet i forhold til interview og returnering af det selvadministrerede spørgeskema.

Der har i de senere år været en tendens til at bortfaldet stiger i mange interview- og spørgeskemabaserede surveys. Dette giver anledning til bekymring af to grunde: dels betyder færre deltagere mindre sikkerhed i de statistiske opgørelser, og dels kan dette bortfald betyde en skævvridning af den population, der rent faktisk bliver undersøgt. En sådan skævvridning vil man typisk frygte, når der er mistanke om, at bortfaldet især sker i visse grupper.

I dette afsnit fokuseres på, om der kan påvises en skævvridning i SUSY-2005. I afsnit 5 vil stikprøven og dele af denne blive sammenlignet med den danske befolkning, som stikprøven gerne skal repræsentere, og i afsnit 7 vil konsekvenser af bortfaldet i forhold til interview blive belyst for nogle indikatorer for sundhed og sygelighed.

Der findes ingen præcise angivelser for, hvornår et bortfald er stort. SIF har vurderet at et bortfald på 30 % eller derunder ikke er stort. SIF må derfor konstatere, at bortfaldet efter interview i SUSY-2005 på 33,3% (=100-66,7, afsnit 3) ikke kan anses for at være tilfredsstillende. Dette er baggrunden for de gennemførte bortfaldsanalyser.

4.1 Metode

I sundheds- og sygelighedsundersøgelserne er der to typer af bortfald (jf. også afsnit 2 og 3):

- bortfald i forhold til det personlige interview. Dette kaldes nedenfor i-bortfald og er belyst i afsnit 4.2.
- bortfald i forhold til det selvadministrerede spørgeskema. Dette kaldes nedenfor s-bortfald og er belyst i afsnit 4.3

For at sammenligne bortfaldet i forskellige undergrupper, er det nødvendigt at have ens oplysninger blandt personer med og uden bortfald.

I forhold til i-bortfald er der kun få socio-demografiske og designmæssige oplysninger, der er kendt for både interviewede og ikke-interviewede.

I forhold til s-bortfald er der to måder at opgøre dette på:

- a) i forhold til alle inviterede (afsnit 4.3.1),
- b) i forhold til alle interviewede (afsnit 4.3.2).

Det første er relevant, idet det er den undersøgte population, der basalt set skal repræsentere den danske befolkning; den anden fordi kun personer, der gennemfører interview, får udleveret et selvadministreret spørgeskema, og fordi der i denne situation er mulighed for at opgøre bortfald i forhold til en række oplysninger fra det gennemførte interview.

I tabel 4.1 vises en oversigt over de variable, der er brugt til at beskrive i-bortfald og s-bortfald med.

Tabel 4.1: Variable brugt til at beskrive bortfald SUSY-2005

Oplysningstype	Variabel	Forklaring
Oplysninger kendt for alle inviterede og i den danske befolkning	gender	Køn fra CPR-nummer
	alder00	Alder ved sidste kontakt i 5 grupper
	kqnalder	Kombineret køn og alder00
	civilst	Civiltilstand - ægteskabelig stilling rent juridisk
	region	Region i Danmark
	stikpr	Stikprøve
Oplysninger kendt for alle interviewede	SAMLIV1	Samlivsstatus
	socio	Socioøkonomisk placering svarende til DST 1997
	iscod	Kombineret skole- og erhvervsuddannelse
	erhvakt2	Erhvervsaktivitet i 4 grupper
	hjem_born	119. Antal hjemmeboende børn 15 år eller under
	Helbred	20. Nuværende helbredstilstand i almindelighed
	Frisknok	21. Føler De dem frisk nok til at gennemføre det De har lyst til
	Stresset	22. Føler De dem stresset i Deres dagligdag
	Langsyg	23. Langvarig sygdom, eftervirkning, handicap eller lidelse
	Med_reg	25. Bruger regelmæssigt el. til stadighed medicin
	Ulykke	32. Udsat for ulykker udenfor arbejdstid seneste år
	GENER	34. Smerte eller ubehag inden for de sidste 14 dage
	Nervos	34. Generet af ængstelse, nervøstitet, uro og angst
	Nedtrykt	34. Generet af nedtrykthed, deprimeret, ulykkelig
	Umul_dgl_g	35. Sygdom eller lidelse gjort det umuligt at udføre daglige gøremål
	laege_kont	41. Haft kontakt med læge sidste tre måneder pga. gener
	abeh_kont	42. Kontakt med andre behandlere sidste tre måneder
	alternativ	50. Nogensinde benyttet en af 10 alternative behandlingsformer
	fam_ven	56. Kombination af 'Hvor ofte træffer De' familie hhv. venner og bekendte. Den mest hyppige
	fysakt	83. Fysisk aktivitet i fritiden det sidste år.
	DRUKUGE2	Antal genstande sidste uge
	DRUKSST	Overskrider SST's genstandgrænse
	alkhyp	93. Hvor ofte drikker De alkohol
	rygeadfr	Ryge-adfærd - baseret på cigaretrykning
ryg_dagl	98. Ryger De dagligt	
BMI	Body Mass Index (vægt/højde ²)	
fodt_dk	193. Er De født i Danmark	
egen_indk	199. Hvor stor var husstandens bruttoindkomst i 2004	

Tabel 4.1 er opdelt i de variable (oplysninger), der er kendt både for alle inviterede og i den danske befolkning (via oplysninger fra Danmarks Statistik (se www.statistikbanken.dk), oplysninger der er kendt for alle inviterede i kraft af designet af undersøgelsen (se afsnit 2) og oplysninger, der kun er kendt for personer, der gennemførte interviewet.

Søjlen 'Variabel' angiver de navne, der vil blive brugt i tabel 4.3, 4.6 og 4.7, hvoraf også de brugte kategoriseringer vil fremgå.

Søjlen 'Forklaring' giver en forklaring af variabelen, og der er her så vidt muligt inddraget spørgsmålsnummeret i interviewspørgeskemaet (se www.si-folkesundhed.dk). Fx kommer oplysningen om antal hjemmeboende børn (hjemb_born) fra sp. 119 i interviewet, som lyder 'Hvor mange hjemmeboende børn under 16 år har De?'

Da der er mange spørgsmål og emner i interviewet, har det været nødvendigt at vælge en række faktorer til at belyse s-bortfaldet blandt interviewede med. Variablene tabel 4.1 er valgt ud fra følgende kriterier:

- de skal omfatte centrale indikatorer i SUSY-2005
- de skal omfatte et bredt spektrum af indikatorer. Der er mange emnemæssige områder i SUSY, og det har været hensigten at dække disse områder overordnet uden at gå i detaljer. Fx er der inddraget oplysninger om alternativ behandling men ikke om de enkelte typer af alternativ behandling.
- der skal være en formodning om, at variabelen har betydning for re-turnering af det selvadministrerede skema.

De statistiske analyser fokuserer som nævnt på at beskrive bortfaldet for variablene i tabel 4.1. Derfor er der lavet logistiske regressionsanalyser, hvor effekten af relevante variable fra tabel 4.1 vurderes, idet der kontrolleres for køn og alder (variablen `kqn_alder`). Der udregnes odds-ratios (OR) med 95 % CI for at se, hvilke kategorier, der adskiller sig fra hvilke og for at vurdere, om der er effekt af variabelen, når der kontrolleres for køn og alder. Det har ikke været hensigten at vurdere, hvilke der er vigtigst og der er derfor ikke lavet en egentlig multipel analyse med udvælgelse af de signifikante variable, når der kontrolleres for samtlige resten af variablene.

4.2 Bortfaldsanalyse for i-bortfald (bortfald efter interview)

I alt har 14.566 personer ud af 21.832 inviterede gennemført et personligt interview. Dette svarer til et bortfald på 7.266 personer, dvs. et i-bortfald på 33,3 %.

I tabel 4.2 vises udviklingen af i-bortfald i sundheds- og sygelighedsundersøgelserne fra 1994 til 2005 fordelt på køn og alder.

Tabel 4.2: Bortfaldsprocent (interview) 1994^a og frem fordelt på køn og alder

Køn	Alder	1994		2000		2005	
		n	%	n	%	n	%
Mænd	16-24	474	20,7	1.465	24,0	1.123	36,4
	25-44	1.102	20,1	3.852	26,4	3.638	35,1
	45-64	819	25,0	3.897	26,5	3.928	33,6
	65-79	395	23,3	1.390	22,2	1.603	28,5
	80+	100	38,0	402	27,6	444	44,8
	I alt		2.890	22,6	11.006	25,6	10.736
Kvinder	16-24	434	16,1	1.376	22,0	1.093	34,1
	25-44	1.069	15,8	3.817	21,8	3.471	28,9
	45-64	890	20,8	3.783	25,9	3.866	30,7
	65-79	515	31,8	1.735	32,4	1.835	34,5
	80+	197	43,7	767	38,6	831	49,6
	I alt		3.105	21,7	11.478	25,9	11.096
I alt		5.995	22,2	22.484	25,8	21.832	33,3

Det ses af tabel 4.2 at den tilsyneladende jævne stigning i i-bortfald fra 22,2 % 1994 over 25,8 % i 2000 til 33,3 % i 2005 dækker over temmelig forskellige mønstre indenfor aldersgrupper for de 2 køn. Det er bemærkelsesværdigt, at der især er sket en stigning for aldersgrupperne 16-24 år og 25-44 år, både for mænd og kvinder, samt at i-bortfaldet blandt mænd har været ret ens henover køns- og aldersgrupper i år 2000.

I tabel 4.3 præsenteres en analyse af i-bortfaldet fordelt på de variable i tabel 4.1, der er kendt for alle inviterede.

Tabel 4.3: Bortfaldsanalyse for i-bortfald - SUSY-2005.

Variabel	Gruppe	n	Procent	OR	lower OR	upper OR	p
Total		21.832	33,3				
gender	Mand	10.736	34,1	1,10	1,04	1,16	0,0009
	Kvinde	11.096	32,5	1,00	.	.	.
alder00	16-24	2.216	35,3	1,15	1,04	1,27	<.0001
	25-44	7.109	32,0	0,99	0,93	1,07	.
	45-64	7.794	32,1	1,00	.	.	.
	65-79	3.438	31,7	0,98	0,90	1,07	.
	80+	1.275	47,9	1,97	1,75	2,22	.
kqnalder	M 16-24	1.123	36,4	1,29	1,13	1,49	<.0001
	M 25-44	3.638	35,1	1,22	1,11	1,34	.
	M 45-64	3.928	33,6	1,14	1,04	1,26	.
	M 65-79	1.603	28,5	0,90	0,79	1,02	.
	M 80+	444	44,8	1,84	1,50	2,24	.
	K 16-24	1.093	34,1	1,17	1,02	1,35	.
	K 25-44	3.471	28,9	0,92	0,83	1,01	.
	K 45-64	3.866	30,7	1,00	.	.	.
	K 65-79	1.835	34,5	1,19	1,06	1,34	.
	K 80+	831	49,6	2,22	1,91	2,59	.
region	Hovedstaden	4.688	37,6	1,21	1,11	1,32	<.0001
	Sjælland	4.286	34,3	1,06	0,97	1,16	.
	Syddanmark	4.286	27,9	0,78	0,72	0,86	.
	Midtjylland	4.286	32,9	1,00	0,91	1,09	.
	Nordjylland	4.286	33,2	1,00	.	.	.
civilst	Gift	11.343	28,5	1,00	.	.	<.0001
	Enke	1.772	39,2	1,13	1,00	1,29	.
	Fraskilt	2.018	37,5	1,48	1,34	1,64	.
	Ugift	6.699	38,6	1,80	1,66	1,95	.
stik_pr	Region sup	16.045	33,5	1,00	.	.	0,34
	Reinterview	5.787	32,8	0,97	0,91	1,03	.
sadm_typ	A	10.916	33,2	1,00	.	.	0,83
	B	10.916	33,4	1,01	0,95	1,06	.

For køn (gender jf. tabel 4.1) ses, at der er inviteret 10.736 mænd og 11.096 kvinder. Blandt mænd er i-bortfaldet 34,1 %, mens det blandt kvinder er 32,5 %. Odds-ratio (OR) viser, at mænd efter kontrol for alder har 1,10 gange højere odds for bortfald end kvinder, og at 95 % konfidensintervallet for OR er (1,04-1,16) (lower, upper). Denne forskel er statistisk signifikant på niveau 5 %, p-værdien er $p=0,0009$.

Det ses, at den højeste bortfaldsprocent ligger blandt personer i alder 80 og derover for både mænd og kvinder, samt at også personer i alder 16-24 har et øget i-bortfald.

Som i tidligere SUSY-undersøgelser findes at det højeste i-bortfald ses i Region Hovedstaden og blandt fraskilte, ugifte og enker.

Der synes ikke at være forskel på stikprøverne, og på hvilket selvadministreret skema, folk skulle have fået udleveret. Det ville have været meget overraskende, hvis vi havde fundet forskel på type af udleveret selvadministreret skema.

Der er foretaget en multivariat bortfaldsanalyse i SUSY-2005 rapporten (Ekholm et.al. 2006).

4.3 Bortfaldsanalyse for s-bortfald (det selvadministrerede skema)

I alt accepterede 13.977 personer (96.0 % af interviewede, afsnit 3) at få udleveret et selvadministreret spørgeskema. Der er i det følgende derfor ikke analyseret på personer, der ikke ønskede et skema udleveret, men udelukkende på personer, der ikke returnerede et skema.

4.3.1 s-bortfald blandt inviterede

I alt besvarede 11.238 personer det selvadministrerede spørgeskema og således har 21.832-11.238=10.594 ikke returneret skemaet. Dette svarer til et s-bortfald på 48,5 %.

I tabel 4.4 ses bortfaldet for det selvadministrerede spørgeskema for de SUSY-undersøgelser, hvor der er blevet udleveret et sådant.

Tabel 4.4: Udvikling i s-bortfald 1994-2005 blandt inviterede

Køn	Alder	1994		2000		2005	
		n	%	n	%	n	%
Mænd	16-24	474	35,2	1.465	43,7	1.123	64,4
	25-44	1.102	29,5	3.852	39,5	3.638	54,5
	45-64	819	32,4	3.897	35,8	3.928	48,0
	65-79	395	33,4	1.390	31,9	1.603	41,2
	80+	100	59,0	402	44,5	444	62,6
	I alt	2.890	32,8	11.006	37,9	10.736	51,5
Kvinder	16-24	434	22,8	1.376	32,4	1.093	48,4
	25-44	1.069	23,8	3.817	29,9	3.471	42,8
	45-64	890	28,1	3.783	32,3	3.866	40,3
	65-79	515	44,3	1.735	42,2	1.835	48,0
	80+	197	67,5	767	63,5	831	73,8
	I alt	3.105	31,0	11.478	35,1	11.096	45,7
I alt	I alt	5.995	31,9	22.484	36,5	21.832	48,5

Af tabel 4.4. ses, at s-bortfaldet er steget markant fra 31,9 % i 1994 til 48,5 % i 2005. Der ses især en markant stigning blandt 16-24 og 25-44-årige både blandt mænd og kvinder – blandt mænd er bortfaldsprocenten hos de 16-24-årige i 2005 næsten på højde med den konstant høje bortfaldsprocent blandt de ældste kvinder.

I tabel 4.5 er s-bortfaldet opgjort fordelt på de variable, der er kendt for alle inviterede (jf. tabel 4.1).

Tabel 4.5: Bortfaldsanalyse for s-bortfald for alle inviterede

Variabel	Gruppe	n	Procent	OR	lower OR	upper OR	p
Total		21.832	48,5				
gender	Mand	10.736	51,5	1,31	1,24	1,38	<.0001
	Kvinde	11.096	45,7	1,00	.	.	.
alder00	16-24	2.216	56,5	1,64	1,49	1,81	<.0001
	25-44	7.109	48,8	1,20	1,13	1,28	.
	45-64	7.794	44,2	1,00	.	.	.
	65-79	3.438	44,8	1,04	0,95	1,12	.
	80+	1.275	69,9	3,07	2,70	3,49	.
kqnalder	M 16-24	1.123	64,4	2,68	2,33	3,07	<.0001
	M 25-44	3.638	54,5	1,77	1,62	1,94	.
	M 45-64	3.928	48,0	1,37	1,25	1,50	.
	M 65-79	1.603	41,2	1,04	0,92	1,17	.
	M 80+	444	62,6	2,48	2,03	3,04	.
	K 16-24	1.093	48,4	1,39	1,21	1,59	.
	K 25-44	3.471	42,8	1,11	1,01	1,22	.
	K 45-64	3.866	40,3	1,00	.	.	.
	K 65-79	1.835	48,0	1,37	1,22	1,53	.
	K 80+	831	73,8	4,17	3,52	4,92	.
region	Hovedstaden	4.688	52,0	1,09	1,01	1,19	<.0001
	Sjælland	4.286	49,7	1,02	0,94	1,11	.
	Syddanmark	4.286	43,5	0,78	0,72	0,85	.
	Midtjylland	4.286	47,3	0,91	0,84	0,99	.
	Nordjylland	4.286	49,8	1,00	.	.	.
civilst	Gift	11.343	41,4	1,00	.	.	<.0001
	Enke	1.772	58,6	1,34	1,19	1,52	.
	Fraskilt	2.018	52,9	1,60	1,45	1,76	.
	Ugift	6.699	56,6	1,78	1,65	1,92	.
stikpr	Region sup	16.045	48,8	1,00	.	.	0,10
	Reinterview	5.787	47,7	0,95	0,89	1,01	.
sadm_typ	A	10.916	47,9	1,00	.	.	0,11
	B	10.916	49,1	1,05	0,99	1,10	.

Det ses af tabel 4.5, at bortfaldsprocenten er højere for mænd end for kvinder. Det ses også, at mønsteret mht. alder er temmelig forskelligt for mænd og kvinder – for mænd ligger bortfaldsprocenten meget højt blandt de 16-24 årige, falder gradvist til og med alder 79 og stiger for de 80+ årige, mens den for kvinder ligger relativt højt blandt de 16-24 årige, falder til alder 64 og stiger så igen ved

alder 65-79 og stiger dramatisk for de 80+ årige, hvor kun lidt over en fjerdedel af alle inviterede har returneret et spørgeskema.

Det ses også, at bortfaldsprocenten er højest i Region Hovedstaden og lavest i Region Syddanmark og Midtjylland, samt at den er lavest blandt de gifte.

Det ses også, at der ingen forskel er på s-bortfaldet for de designmæssige karakteristika omkring stikprøve og type af selvadministreret skema.

4.3.2 s-bortfald blandt interviewede

Som omtalt i de to tidligere afsnit har 14.566 gennemført et interview og 11.238 returneret et selvadministreret spørgeskema. Således har 3.322 gennemført interview men ikke returneret et selvadministreret spørgeskema, hvilket giver et s-bortfald blandt interviewede på 22,8 %.

I tabel 4.6 ses s-bortfaldet blandt interviewede for det selvadministrerede spørgeskema fordelt på køn og alder.

Af tabel 4.6. ses, at s-bortfaldet er steget fra 12,5 % i 1994 til 22,8 % i 2005. Som i tabel 4.4 ses der især en markant stigning blandt 16-24 og 25-44 årige mænd og kvinder og blandt de 45-64 årige mænd blandt mænd returnerede kun lidt over en fjerdedel det udleverede skema i 2005.

Tabel 4.6: Udvikling i s-bortfald 1994-2005 blandt interviewede

Køn	Alder	1994		2000		2005	
		n	%	n	%	n	%
Mænd	16-24	376	18	1.114	26	714	44
	25-44	881	12	2.835	18	2.362	30
	45-64	614	10	2.865	13	2.609	22
	65-79	303	13	1.081	12	1.146	18
	80+	62	34	291	23	245	32
	I alt	2.236	13	8.186	17	7.076	26
Kvinder	16-24	364	8	1.073	13	720	22
	25-44	900	9	2.984	10	2.469	20
	45-64	705	9	2.802	9	2.680	14
	65-79	351	18	1.172	14	1.202	21
	80+	111	42	471	41	419	48
	I alt	2.431	12	8.502	12	7.490	19
I alt	I alt	4.667	13	16.688	14	14.566	23

I tabel 4.7 er s-bortfaldet beskrevet for samtlige variable nævnt i tabel 4.1.

Tabel 4.7: Bortfaldsanalyse for s-bortfald blandt interviewede i SUSY-2005

Variabel	Gruppe	n	Procent	OR	lower OR	upper OR	p
Total		14.566	22,8				
gender	Mand	7.076	26,4	1,54	1,42	1,66	<.0001
	Kvinde	7.490	19,5	1,00			
alder00	16-24	1.434	32,8	2,27	1,99	2,59	<.0001
	25-44	4.831	24,6	1,52	1,38	1,67	
	45-64	5.289	17,8	1,00			
	65-79	2.348	19,2	1,10	0,97	1,25	
	80+	664	42,2	3,61	3,04	4,28	
kqnalder	M 16-24	714	44,0	4,87	4,05	5,85	<.0001
	M 25-44	2.362	29,8	2,64	2,29	3,04	
	M 45-64	2.609	21,7	1,72	1,49	1,99	
	M 65-79	1.146	17,7	1,34	1,11	1,61	
	M 80+	245	32,2	2,95	2,21	3,94	
	K 16-24	720	21,7	1,72	1,39	2,11	
	K 25-44	2.469	19,6	1,51	1,30	1,75	
	K 45-64	2.680	13,9	1,00			
	K 65-79	1.202	20,5	1,60	1,34	1,92	
	K 80+	419	48,0	5,72	4,59	7,13	
region	Region Hovedstaden	2.923	23,1	0,91	0,80	1,03	0,00
	Region Sjælland	2.818	23,5	0,97	0,86	1,10	
	Region Syddanmark	3.089	21,6	0,83	0,74	0,94	
	Region Midtjylland	2.874	21,4	0,82	0,72	0,93	
	Region Nordjylland	2.862	24,8	1,00			
civilst	Gift	8.114	18,1	1,00			<.0001
	Enke	1.078	31,9	1,59	1,33	1,90	
	Fraskilt	1.261	24,7	1,54	1,34	1,78	
	Ugift	4.113	29,3	1,45	1,29	1,62	
stikpr	Region sup	10.675	23,1	1,00			0,15
	Reinterview	3.891	22,3	0,94	0,86	1,02	
sadm_typ	1	7.291	22,0	1,00			0,02
	2	7.275	23,7	1,10	1,01	1,19	
samliv1	Gift	8.026	17,9	1,00			<.0001
	Samlevende	2.260	25,6	1,37	1,22	1,54	
	Enlig,t.sep,sk.	903	25,8	1,66	1,42	1,96	
	Enlig, enke	1.011	32,9	1,68	1,40	2,02	
	Ugift	2.361	31,6	1,60	1,41	1,82	
isced	lav <10 år	2.295	31,7	1,49	1,32	1,70	<.0001
	mel 10 år	802	30,4	1,19	0,99	1,42	
	mel 11-12 år	3.309	24,4	1,00			
	lang 13-14 år	4.535	20,8	0,77	0,69	0,87	
	lang 15+ år	3.193	15,0	0,54	0,48	0,62	
	skoleelev	256	28,5	0,82	0,61	1,12	

Tabel 4.7: Bortfaldsanalyse for s-bortfald blandt interviewede i SUSY-2005

Variabel	Gruppe	n	Procent	OR	lower OR	upper OR	p
socio	Selvst. med ansatte	438	31,3	1,47	1,18	1,84	<.0001
	Selvst. uden ansatte	426	29,1	1,47	1,16	1,85	
	Medhjælpende ægtefæl	55	29,1	1,95	1,07	3,54	
	Topledere	467	19,9	0,81	0,64	1,04	
	Løn højeste niveau	1.231	14,2	0,59	0,49	0,71	
	Løn mellemste niveau	1.982	17,9	0,79	0,68	0,91	
	Løn laveste niveau	2.975	22,9	1,00			
	Andre lønmodtagere	643	27,5	1,24	1,02	1,51	
	Arbejdsløs	509	27,5	1,34	1,08	1,67	
	Uddannelsessøgende	1.374	27,5	0,81	0,67	0,97	
	Førtidspension	589	27,7	1,75	1,42	2,15	
	Folkepension	2.898	24,4	0,94	0,68	1,30	
	Efterlønsmodtagere	521	11,1	0,58	0,43	0,78	
	Andre uden for arbej	360	26,9	1,41	1,10	1,82	
	erhvakt2	Erhverv 16-64	7.969	21,6	1,00		
hjemb_born	0	10.595	22,6	1,00			0,08
	1	1.637	21,3	0,99	0,86	1,14	
	2	1.767	24,8	1,15	1,00	1,32	
	3	567	26,5	1,23	1,00	1,52	
Helbred	Virkelig god	5.010	20,1	0,87	0,80	0,96	<.0001
	God	6.486	21,7	1,00			
	Nogenlunde	2.222	29,0	1,60	1,43	1,79	
	Dårlig	839	31,5	1,92	1,63	2,26	
frisknok	Ja, for det meste	11.690	20,9	1,00			<.0001
	Ja, af og til	1.705	30,0	1,72	1,53	1,94	
	Nej (næsten aldrig)	1.156	32,2	2,00	1,74	2,29	
Stresset	Ja, ofte	1.228	23,7	1,14	0,99	1,32	0,19
	Ja, af og til	4.867	22,8	1,04	0,95	1,14	
	Nej (næsten aldrig)	8.453	22,7	1,00			
Langsyg	Ja	5.813	23,4	1,14	1,05	1,24	0,001
	Nej	8.728	22,4	1,00			
Med_reg	Ja	5.755	22,2	1,03	0,94	1,13	0,46
	Nej	8.800	23,2	1,00			
Ulykke	Ja	1.425	24,9	0,98	0,86	1,12	0,76
	Nej	13.118	22,5	1,00			
gener	Nej, ingen gener	2.962	23,0	1,00			0,91
	Ja, gener	11.586	22,7	1,01	0,91	1,11	
Nervos	Ja	1.230	26,1	1,30	1,14	1,49	0,00
	Nej	13.307	22,4	1,00			
Nedtrykt	Ja	1.237	27,0	1,35	1,18	1,55	<.0001
	Nej	13.297	22,3	1,00			
Umul_dgl_g	Ja	1.943	26,4	1,32	1,18	1,48	<.0001
	Nej	12.591	22,2	1,00			

Tabel 4.7: Bortfaldsanalyse for s-bortfald blandt interviewede i SUSY-2005

Variabel	Gruppe	n	Procent	OR	lower OR	upper OR	p
laege_kont	Nej	7.568	22,9	1,00			0,40
	Ja	6.973	22,6	1,04	0,95	1,12	
abeh_kont	Nej	8.131	25,2	1,00			<.0001
	Ja	6.404	19,6	0,78	0,72	0,85	
alternativ	Nej	8.056	25,9	1,00			<.0001
	Ja	6.474	18,8	0,74	0,68	0,81	
fysakt	Stillesid	1.908	36,7	1,00			<.0001
	Moderat	8.724	20,1	0,50	0,45	0,56	
	Hård	3.787	21,0	0,43	0,38	0,49	
DRUKUGE2	kr	3.305	28,4	1,54	1,40	1,70	<.0001
	1-14	8.095	20,0	1,00			
	15-21	1.410	22,1	1,05	0,91	1,21	
	22+	1.616	24,8	1,15	1,01	1,31	
DRUKSST	Overskr. gens.grænse	2.054	23,0	1,01	0,90	1,13	0,86
	Overskr. ej gens.græ	12.372	22,6	1,00			
alkhyp	Drikker ikke	769	37,7	2,51	2,11	2,98	<.0001
	<1 gang pr måned	2.102	26,2	1,49	1,31	1,69	
	1-3 gange pr måned	3.352	22,8	1,17	1,04	1,31	
	1-2 gange pr uge	4.174	20,3	1,00			
	3-4 gange pr uge	1.701	19,5	0,99	0,85	1,14	
	5-7 gange pr uge	2.402	20,7	1,13	0,99	1,29	
ryg_dagl	Nej	10.177	19,7	1,00			<.0001
	Ja	4.354	29,5	1,83	1,68	1,99	
RYGEADFR	ryger-øj cigt	983	22,5	1,15	0,97	1,36	<.0001
	ryger 1-4 cigt	431	25,1	1,29	1,02	1,63	
	ryger 5-14 cigt	1.349	26,9	1,56	1,36	1,79	
	ryger 15+ cigt	2.434	32,2	2,02	1,81	2,26	
	aldrig røget	5.774	20,8	1,00			
	tidl røget	3.557	17,4	0,90	0,80	1,01	
BMI	10-<20 BMI	1.120	25,0	1,19	1,02	1,39	0,00
	20-<25 BMI	6.816	21,8	1,00			
	25-<30 BMI	4.847	22,3	1,07	0,97	1,17	
	30-<66 BMI	1.704	24,8	1,31	1,15	1,48	
fodt_dk	Ja	14.074	22,3	1,00			<.0001
	Nej	459	33,8	1,69	1,38	2,07	
egen_indk	<50.000	853	32,7	1,69	1,37	2,07	<.0001
	50.000 - 99.999	1.442	26,7	1,51	1,29	1,78	
	100.000 - 199.999	3.737	22,4	1,31	1,16	1,48	
	200.000 - 299.999	3.759	19,0	1,00			
	300.000 - 399.999	2.377	19,6	0,95	0,83	1,08	
	400.000 - 499.999	859	19,8	0,89	0,73	1,08	
	500.000 - 599.999	361	22,2	1,01	0,78	1,32	
	600.000 - 699.999	196	24,5	1,14	0,81	1,60	
	700.000 -	308	22,7	1,06	0,80	1,41	

Tabel 4.7: Bortfaldsanalyse for s-bortfald blandt interviewede i SUSY-2005

Variabel	Gruppe	n	Procent	OR	lower OR	upper OR	p
hus_indk	<100.000	684	36,5	2,36	1,91	2,92	<.0001
	100.000 - 199.999	1.909	26,5	1,74	1,47	2,06	
	200.000 - 299.999	1.849	20,2	1,24	1,05	1,47	
	300.000 - 399.999	1.671	20,6	1,21	1,02	1,43	
	400.000 - 499.999	1.634	19,3	1,07	0,90	1,27	
	500.000 - 599.999	1.899	18,9	1,00			
	600.000 - 699.999	1.310	19,1	1,03	0,86	1,23	
	700.000 -	836	16,1	0,85	0,68	1,06	
	800.000 - 899.999	450	18,9	1,02	0,78	1,33	
	900.000 - 999.999	265	20,0	1,10	0,79	1,52	
	1.000.000 -	520	20,8	1,15	0,90	1,47	

Udvalgte resultater i tabel 4.7 er kommenteret ud fra en overordnet gruppering af variablene i følgende grupper taget fra Ekholm et al. (2006): socio-demografiske oplysninger (gender, alder00, kqnalder, region, civilst, samliv1, isced, socio, erhvakt2, hjemb_born, fodt_dk, egen_indk, hus_indk), helbredsrelateret livskvalitet (helbred, frisknok), sundhedsadfærd (livsstil – stresset, fysakt, drukuge2, druksst, alkhyp, ryg_dagl, rygeadfr, BMI), sygelighed (langsyg, ulykke, gener, nervos, nedtrykt), sygdomskonsekvenser (med_reg, umul_dgl_g, laege_kont, abeh_kont, alternativ) og stikprøvekarakteristika (stikpr, sadm_typ).

Generelt ses det, at det største s-bortfald findes i grupper, der antalsmæssigt er relativt små.

Socio-demografiske oplysninger

Den højeste bortfaldsprocent forekommer blandt mænd i alder 16-24 (44,0 %) og blandt de ældste mænd (32,2 %) og kvinder (48,0 %).

Det ses, at

- jo lavere uddannelse des større s-bortfald – OR for mindre end 10 års uddannelse er 1,49 (95 % CI: 1,32-1,70) mens OR for dem med mindst 15 års uddannelse er 0,54 (95 % CI: 0,48-0,62)
- bortfaldet er større blandt selvstændige end blandt erhvervsaktive og blandt ikke-erhvervsaktive højst blandt førtidspensionister
- interviewede, der ikke er født i Danmark har større bortfald end personer, der er (OR=1,31, 95 % CI: 1,15-1,48).
- der synes ikke at være sammenhæng mellem s-bortfald og antal hjemmeboende børn, når der kontrolleres for køn og alder (p=0,08).
- jo højere indkomst jo lavere synes s-bortfaldet at være.

Helbredsrelateret livskvalitet

Det ses, at jo dårligere helbred den interviewede rapporterer, jo højere er s-bortfaldet – OR for personer med virkelig godt selv vurderet helbred er 0,87 (95

% CI: 0,80-0,96), mens OR for personer med dårligt helbred er 1,92 (95 % CI: 1,63-2,26).

- interviewede, der oplyser ofte at være stressede, synes efter kontrol for køn og alder ikke at have et højere s-bortfald end ikke-stressede ($p=0,19$).

Sundhedsadfærd

Det ses, at

- interviewede der er storrygere (ryger 15 eller flere cigaretter om dagen) har højere s-bortfald end de øvrige rygegrupper (OR=2,02).
- personer med BMI over 30 (svært overvægtige) har højere s-bortfald end de øvrige BMI-grupper (OR=1,31).

- interviewede, der ikke har drukket den sidste uge har et højere s-bortfald (OR=1,54, 95 % CI: 1,40-1,70) end personer, der oplyser at have drukket den sidste uge.
- interviewede med stillesiddende fysisk aktivitet har højere s-bortfald end de øvrige grupper (OR for ikke-stillesiddende aktivitet er omkring 0,44).

Sygelighed

Interviewede, der har været nedtrykte eller ængstelige samt interviewede med en langvarig sygdom har større s-bortfald end personer, der ikke har disse karakteristika.

Sygdomskonsekvenser

Det fremgår, at

- interviewede der er begrænset i deres daglige aktiviteter grundet sygdom har større s-bortfald (OR=1,32, 95 % CI: 1,18-1,48) end personer, der ikke er.
- personer, der gør brug af alternativ behandling, har lavere s-bortfald (OR=0,74, 95 % CI: 0,68-0,81) end personer, der ikke gør.

Således synes tendensen at være, at det højeste s-bortfald findes blandt de grupper, der må anses som 'svagest'.

5. Repræsentativitet

Formålet med sundheds- og sygelighedsundersøgelserne er at beskrive forskellige indikatorer for sundhed og sygelighed i den voksne danske befolkning. Det er derfor vigtigt, at stikprøven ligner den danske befolkning så meget som muligt. Det er muligt at undersøge dette på udvalgte socio-demografiske oplysninger, som findes både for hele den danske befolkning og for alle inviterede til SUSY-2005 (jf. tabel 4.1). I dette afsnit vil der blive redegjort for, i hvor høj grad stikprøven og dens forskellige delstikprøver kan siges at ligne den befolkning, den skal repræsentere, nemlig voksne danskere.

Repræsentativiteten vil blive belyst i forhold til oplysninger om køn, alder og civilstand, idet oplysningen om region skal behandles specielt, jf. afsnit 7. Repræsentativiteten mht. disse socio-demografiske oplysninger er vist i tabel 5.1.

Tabel 5.1: Repræsentativitet i forhold til socio-demografiske variable og dele af stikprøven i SUSY-2005

Køn	Alder	Voksne danskere	Inviterede		Interview		Selvadministreret spørgeskema		
			Alle	Reinter-view	Alle	Reinter-view	Begge	A	B
Mænd	16-24	6,3	5,1	5,6	4,8	5,2	3,5	3,5	3,4
	25-44	17,9	17,0	18,0	16,6	18,0	15,2	14,7	15,6
	45-64	16,7	17,7	17,1	17,6	17,0	17,8	18,2	17,4
	65-79	6,4	7,2	6,6	7,8	6,9	8,3	8,4	8,1
	80+	1,7	2,0	1,9	1,6	1,7	1,5	1,3	1,6
	I alt	49,0	48,9	49,3	48,4	48,9	46,2	46,2	46,1
Kvinder	16-24	6,1	5,1	5,1	5,0	5,1	5,1	5,0	5,1
	25-44	17,5	16,2	16,7	17,3	18,1	18,1	17,9	18,2
	45-64	16,6	17,6	17,7	18,3	17,9	20,3	20,1	20,5
	65-79	7,4	8,3	7,6	8,2	7,2	8,5	8,8	8,0
	80+	3,4	3,9	3,7	2,9	2,8	2,0	2,0	2,0
	I alt	51,0	51,1	50,7	51,6	51,1	53,8	53,8	53,9
Civilstand	Gift	49,9	51,1	50,8	54,9	54,1	58,0	58,1	57,9
	Enke	7,6	8,0	7,6	7,3	6,8	6,4	6,5	6,3
	Skilt	9,5	9,4	9,3	8,8	8,6	8,6	8,7	8,7
	Ugift	32,9	31,5	32,3	29,1	30,5	26,9	26,7	27,2

Af tabel 5.1 ses, at blandt voksne i Danmark er 6,3 % mænd mellem 16 og 24 år. Det tilsvarende tal i den samlede stikprøve er 5,1 %. Denne lidt lavere forekomst må sandsynligvis tilskrives problemer omkring forskerbeskyttelse (se afsnit 2.3). Det ses også, at andelen af 16-24-årige blandt de interviewede er 4,8 %, hvilket er en konsekvens af det forøgede i-bortfald i denne aldersgruppe (se tabel 4.3).

Det ideelle ville være, at alle tal for stikprøven (inviterede) og de forskellige undersøgelsespopulationer (interview og selvadministreret spørgeskema) lignede tallene for voksne danskere. Det ses, at dette primært grundet bortfald ikke helt er tilfældet – generelt er andelen af 45-64-årige lidt større end ønsket for både mænd og kvinder, andelen af gifte er højere, og andelen af ugifte er lavere, især mht. til det selvadministrerede spørgeskema. Dette er en konsekvens af bortfaldet som beskrevet i afsnit 4.2 og 4.3.1.

6. Vægtning

I sundheds- og sygelighedsundersøgelsen 2005 er der, som i SUSY-2000, brugt vægtning. Formålet med denne vægtning er at søge at tage højde for den regionale ubalance, der er indført i kraft af den måde, stikprøven er udvalgt på (afsnit 2.2)

Begrundelsen for vægtningen er, at der ved konstruktionen af stikprøven er indført en designmæssig skævvridning af den regionale fordeling, som der skal rettes op på (afsnit 2.2). Da stikprøven er tilfældigt udtrukket, og der ikke er nogen designmæssige forskelle mht. øvrige socio-demografiske variable, indgår disse ikke i en vægtning.

Tabel 6.1: Udregning af vægte til brug ved samlet interview i SUSY-2005

Danmark									
Region	n	%	% inviterede	p (‰)	Vægt inviteret	Interview	Nyt antal	Vægt interview	Vægtet %
Hovedstaden	1.323.117	30,5	21,5	3,5	1,42	2.923	4.153,3	1,43	28,6
Sjælland	645.728	14,9	19,6	6,6	0,76	2.818	2.137,4	0,76	14,7
Syddanmark	942.077	21,7	19,6	4,5	1,11	3.089	3.418,2	1,11	23,5
Midtjylland	961.590	22,2	19,6	4,5	1,13	2.874	3.246,2	1,13	22,4
Nordjylland	464.036	10,7	19,6	9,2	0,55	2.862	1.560,0	0,55	10,7
I alt	4.336.548			5,0		14.566	14.515,1		

Af tabel 6.1 fremgår, at 30,5 % af voksne danskere bor i Region Hovedstaden og 10,7 % i Region Nordjylland. De tilsvarende tal for inviterede til SUSY-2005 er 21,5 % i Region Hovedstaden og 19,6 % i de øvrige regioner. Der er således markant forskellig regional fordeling i stikprøven og blandt voksne danskere. Af søjlen 'p (‰)' ses sandsynligheden i promille for, at en person i en region bliver udtrukket – i region Nordjylland er udvalgssandsynligheden $1000 \cdot 4286 / 464036 = 9,2 \text{ ‰}$ (dvs. 0,92 %), mens den i Hovedstaden er 3,5 ‰. Det ses også, at udvalgssandsynligheden i den samlede stikprøve er 5,0 ‰. Således har personer i Hovedstadsregionen mindre sandsynlighed for at blive udtrukket, end hvad man skulle forvente i forhold til regionens størrelse på landsplan, mens personer i Region Nordjylland har større chance end regionens størrelse tilsiger. For at sikre at personer i stikprøven tæller i forhold til regionens størrelse, skal hver person i Hovedstaden således 'fylde' $5,0 / 3,5 = 1,42$, mens det tilsvarende tal i region Nordjylland er $5,0 / 9,2 = 0,55$. På denne led vægter personer i forskellige regioner forskelligt, og de udregnede forhold kaldes vægte. De her beskrevne vægte kan bruges, når der laves analyser for alle inviterede.

Antallet af interviewede fremgår af søjlen 'Interview'. Multipliseres vægten for inviterede med antal interviewede for derved at finde den vægtede regionale

stikprøvestørrelse fås tallene i søjlen 'Nyt antal'. Summeres tallene i søjlen 'Nyt antal' fås tallet 14.515,1, hvilket ikke svarer til det samlede antal interviewede (14566). For at få en vægt for interviewede personer, der sikrer, at det vægtede antal svarer til den samlede stikprøve, må vægtene for inviterede derfor korrigeres med en faktor $14566/14515,1 = 1,004$. De heraf fremkomne vægte (=1,004 * vægt inviteret) fremgår af søjlen 'Vægt interview'. Vægten for Region Nordjylland er således 0,55, mens den for Hovedstaden er 1,43. Forskellen i forhold til vægtene for inviterede er således minimale.

Af søjlen 'Vægtet %' fremgår, at den vægtede regionale fordeling blandt interviewede ikke svarer helt til den procentvise fordeling blandt voksne danskere. Dette skyldes, at der er forskelligt bortfald i de forskellige regioner (se tabel 4.3). SIF har ved konstruktionen af vægte fundet det mest korrekt at tage udgangspunkt i alle inviterede til SUSY-2005.

Det skal nævnes, at standard statistisk analyse ville bruge vægten 1 i alle regioner.

Det ovenfor gennemgåede princip for vægtning er gennemført for delstikprøver af den samlede stikprøve, idet der dog ikke er lavet vægtning for reinterviewstikprøven, da den er baseret på en helt tilfældigt udvalgt stikprøve blandt voksne danskere)

For de selvadministrerede spørgeskemaer er der taget udgangspunkt i interviewede (og ikke inviterede) – dvs. søjlen '% inviteret' er erstattet med '% interviewede' og søjlen 'Interview' med 'Returneret selvadministreret skema' i tabel 6.1.

I tabel 6.2 vises vægte for det selvadministrerede spørgeskema, både totalt og for hver af de to spørgeskematyper (jf. afsnit 2.4).

Tabel 6.2: Vægte for det selvadministrerede spørgeskema i SUSY-2005

Region	Type af selvadministreret skema		
	Begge	1	2
Hovedstaden	1,52	1,69	1,38
Sjælland	0,77	0,73	0,81
Syddanmark	1,02	1,08	0,97
Midtjylland	1,12	1,18	1,07
Nordjylland	0,54	0,46	0,65

Det er ikke overraskende at vægten i stikprøve 2 for fx Region Hovedstaden er nærmere 1 end vægten for stikprøve 1, idet stikprøve 2 indeholder reinterview-

viewstikprøven, hvor den regionale fordeling er mere lig den på landsplan end i den regionalt supplerende stikprøve.

6.1 Vægtning for børn

I SUSY-2005 er der ved interviewet blevet spurgt om antallet af hjemmeboende børn i alder 0-15 år, og for de fire ældste af disse børn er interviewpersonen blevet stillet en række spørgsmål om hvert af børnenes sundhed og sygelighed. Således er der spurgt til børn gennem svar fra interviewpersonen. Det fremgår ikke, om de afgivne svar vedrører biologiske børn eller ej. På helt tilsvarende måde er der blevet spurgt til børns sundhed og sygelighed i SUSY-1994 og SUSY-2000.

I SUSY undersøgelserne er børnene valgt via interviewpersonen. Således har børn med en enlig forsørger kun halv så stor chance for at 'indgå' i SUSY-2005 (og øvrige SUSY-undersøgelser) som børn med en ikke-enlig forsørger. Det betyder, at ud over den i sidste afsnit beskrevne regionale forskel i sandsynlighed for at blive udvalgt, er der for børn en yderligere designmæssig forskel i forhold til, om der er tale om, at interviewpersonen er enlig eller ej.

Som det fremgår af tabel 4.1 er oplysningen om region kendt for alle inviterede (og for alle danske statsborgere), mens oplysningen om hjemmeboende børn kun er kendt for de interviewede. Det har ved konstruktionen af vægte således været nødvendigt at tage udgangspunkt i interviewoplysninger.

Til opgørelse af, om en person i stikprøven kan anses som enlig forsørger er brugt oplysningen om samlivsstatus (variablen SAMLIV1 i tabel 4.1). En tilsvarende oplysning findes ikke i Danmarks Statistiks opgørelser (www.statistikbanken.dk). Der findes opgørelser baseret på den såkaldte 'c-type', som er et mål for, om en person kan anses for at være enlig eller ej. Denne oplysning er brugt til at konstruere tabel 6.3.

Af tabel 6.3 fremgår fordelingen af antal børn i alder 0-15 år i Danmark og i SUSY-2005 fordelt på region og på, om personen kan anses for at være enlig eller ej samt den heraf udledte udregning af vægte for børn.

Det ses, at der i år 2005 var knap 1,1 mill. børn i alder 0-15 år i Danmark og af dem er der opnået svar for 6.962 (0,64 % eller 6,4 promille). Det ses, at udvalgssandsynligheden for børn med en ikke-enlig forsørger generelt er mere end dobbelt så stor som for børn med en enlig forsørger. Udregningen af vægten for børn er den samlede udvalgssandsynlighed (6,4 divideret med udvalgssandsynligheden for hver region og type af forsørger).

Tabel 6.3: Vægtning for børn i alder 0-15 år i SUSY-2005.

Region	Enlig forsørger	Danmark		Interviewede		Udv. ssh (‰)	Vægt børn
		n	%	n	%		
Hovedstaden	Nej	247.121	22,8	1.148	16,5	4,6	1,38
	Ja	62.306	5,8	119	1,7	1,9	3,37
Sjælland	Nej	134.987	12,5	1.126	16,2	8,3	0,77
	Ja	27.386	2,5	86	1,2	3,1	2,05
Syddanmark	Nej	204.332	18,9	1.468	21,1	7,2	0,89
	Ja	37.375	3,5	115	1,7	3,1	2,09
Midtjylland	Nej	216.492	20,0	1.360	19,5	6,3	1,02
	Ja	37.055	3,4	86	1,2	2,3	2,77
Nordjylland	Nej	99.023	9,1	1.351	19,4	13,6	0,47
	Ja	16.870	1,6	103	1,5	6,1	1,05
I alt		1.082.947		6.962		6,4	

6.2. Lidt statistisk baggrund

Statistisk formelt stammer vægtningstankegangen fra den såkaldte survey-statistik, hvor bogen af Särndal et.al. (1992) står som en milepæl. Designet er et såkaldt stratificeret sample og den ovenfor omtalte vægtning svarer til en vægtning baseret på de forskellige udvalgssandsynligheder i de forskellige regioner (strata) med en passende renormering.

De beskrivende analyser, der er foretaget i (Ekholm et.al. 2006) og (Kjøller et.al. 2002) er lavet med statistikprogrammet SAS, idet de i tabel 6.1 anførte vægte er brugt i en såkaldt WEIGHT-sætning. Dette er formelt set en approksimation til surveystatistikken forstået på den måde, at alle forekomster er upåvirket af dette, men at der kan forekomme minimale afvigelser i forhold til variansberegninger.

Det skal også nævnes, at de herved fundne estimater formelt set ikke er Horwitz-Thomson estimatoren, men at der for alle praktiske formåls vedkommende ikke er forskel. Dette gælder dog ikke for analyserne af børn, hvor den anvendte vægtning præcis svarer til Horwitz-Thompson estimatoren.

Der er meget litteratur på dette område; interesserede er velkomne til at henvende sig til dette notats forfatter.

Det brugte princip for vægtning er tidligere beskrevet i (Davidsen og Kjøller 2002).

7. Bortfalds betydning for forekomst af udvalgte indikatorer

I de forrige afsnit er bortfaldet i SUSY-2005 beskrevet, både i forhold til interview og returnering af det selvadministrerede spørgeskema samt i forhold til repræsentativiteten. Baggrunden for dette er at give mulighed for at vurdere bias (skævheder) i stikprøven, som kan påvirke de mål for sundhed og sygelighed, som SIF publicerer. I dette afsnit vil det blive forsøgt at vurdere bortfaldets betydning for forekomsten af centrale indikatorer for sundhed og sygelighed.

Der findes flere metoder til at vurdere effekten af bortfaldet. De har alle kraftige begrænsninger, idet det i sagens natur er svært at udtale sig om det, man netop ikke ved noget om.

Den mest anvendte metode baserer sig på begreberne 'missing completely at random', 'missing at random' og 'not missing at random', som tager udgangspunkt i antagelser om bortfaldsmekanismen. Denne metode kræver som regel en grundig modellering ud fra en konkret problemstilling og vil ikke blive forfulgt yderligere i dette notat.

En anden metode, som finder anvendelse indenfor surveystatistik (afsnit 6.2), hedder 'calibrating for non-response' (Särndal et.al. 2005). Denne metode tager udgangspunkt i, at man især i de nordiske lande har gode socio-demografiske oplysninger ikke blot om sin stikprøve, men i mange situationer (som her) også for den population, man ønsker at udtale sig om.

Metoden kan illustreres ved at se på tabel 5.1: fordelingen af køn, alder og civilstand (samt region) er kendt for alle voksne danskere og for alle inviterede (og dermed interviewede). På grund af bortfald svarer fordelingen af disse socio-demografiske variable blandt interviewede og blandt voksne danskere ikke til hinanden. 'Calibrating for non-response' beregner forekomst (og andre statistiske mål) ved at kalibrere for den kendte forskel i sammensætning mht. disse variable.

I dette notat vil effekten af non-response på 69 udvalgte indikatorer for sundhed og sygelighed fra interviewet blive belyst med kalibrering. De 69 indikatorer er taget fra (Ekholm et.al. 2006) og omfatter alle, der er oplyst for hele interviewpopulationen. Ved kalibreringen inddrages oplysninger om køn/alder, civilstand og region. I princippet er det muligt at inddrage andre socio-demografiske oplysninger kendt i Danmarks Statistiks databank (www.statistikbanken.dk), men disse oplysninger findes ikke for personerne i SUSY-2005, hvilket er nødvendigt for at kunne kalibrere.

For indikatorer fra det selvadministrerede spørgeskema er det også muligt at kalibrere for non-respons i forhold til de i tabel 4.1 nævnte oplysninger fra interview. Dette kræver dog en nærmere vurdering af, hvilke oplysninger, der er relevante i forhold til en specifik valgt indikator fra det selvadministrerede spørgeskema og ligger derfor udenfor dette notats rammer.

Beregningerne er foretaget med programmet CLAN (Andersson ref).

For at illustrere metoden vises i tabel 7.1 forekomsten totalt for otte udvalgte indikatorer dels som de fremstår i SUSY-2005 rapporten (Ekholm et.al. 2006) og dels efter kalibrering for bortfald. Ud over forekomst præsenteres 95 % konfidensinterval for denne.

De valgte indikatorer fra interviewet er som følger, idet der nedenfor henvises til variabelnavne fra tabel 4.1:

- Godt selv vurderet helbred (variablen 'Helbred' grupperet som virkelig godt eller godt vs. resten).
- Langvarig sygdom (variablen 'langsyg' – ja/nej).
- Aktivitetsbegrænsninger (variablen 'Umul_dgl_g')
- Svær overvægt (om variablen BMI er ≥ 30)
- Overskridelse af genstandsgrænsen (variablen DRUKSST)
- Stillesiddende fysisk aktivitet (om der er svaret 'stillesiddende' til variablen 'fysakt')
- Storrygning (om der ryges 15 eller flere cigaretter om dagen).

Tabel 7.1: Forekomst af udvalgte indikatorer som de fremstår i SUSY-2005 og efter kalibrering

Indikator	Forekomst SUSY-rapport			Kalibreret forekomst		
	%	N ^a	Ø ^b	%	N ^a	Ø ^b
Godt selv vurderet helbred	79,4	78,7	80,1	79,9	79,2	80,5
Langvarig sygdom	39,8	39,0	40,6	39,1	38,3	40,0
Aktivitetsbegrænsninger	13,2	12,6	13,8	13,2	12,6	13,8
Fedme	11,3	10,7	11,8	11,1	10,5	11,6
Overskridelse af genstandsgrænsen	14,3	13,7	14,9	14,3	13,7	14,9
Stillesiddende fysisk aktivitet	12,9	12,3	13,4	13,1	12,5	13,6
Storrygning	16,6	16,0	17,3	16,9	16,3	17,5

^a: nedre 95% konfidensinterval

^b: øvre 95% konfidensinterval

Af tabel 7.1 ses, at den publicerede forekomst af storrygning er 16,6 %, og at et 95 % konfidensinterval er (16,0 – 17,3). Det ses også, at når der kalibreres for bortfald i forhold til køn/alder, civilstand og region fås en forekomst på 16,9 % (95 % CI: 16,3 – 17,5).

Af tabel 7.1 ses at den største forskel mellem ukalibreret (SUSY-rapport) og kalibreret (justeret for bortfald) forekomst er 0,7 %. Det er lidt overraskende, at forekomsten af langvarig sygdom falder, når der kalibreres, idet bortfaldet for de ældre er det højeste, og her er forekomsten af langvarig sygdom også højest. Det ses også, at konfidensintervallet for den i SUSY-2005 rapporten angivne forekomst i alle tilfælde omfatter den kalibrerede forekomst og vica-versa.

Til at supplere angivelsen af forekomst, angiver SIF nogle gange antallet af danskere. Det er også muligt at kalibrere dette antal for bortfald. På denne måde kan de små forskelle i tabel 7.1 sættes ind i dette perspektiv.

Tabel 7.2: Antal danskere (i 1.000) med udvalgte indikatorer ud fra SUSY-2005 og efter kalibrering.

Indikator	SUSY-rapport			Kalibreret forekomst		
	n	N ^a	Ø ^b	n	N ^a	Ø ^b
Godt selv vurderet helbred	3.444	3.399	3.490	3.469	3.440	3.498
Langvarig sygdom	1.726	1.685	1.766	1.701	1.665	1.736
Aktivitetsbegrænsninger	573	547	598	572	547	597
Fedme	489	465	512	480	457	503
Overskridelse af genstandsgrænsen	622	595	649	622	596	648
Stillesiddende fysisk aktivitet	558	533	583	567	542	592
Storrygning	721	693	749	734	706	761

^a: nedre 95% konfidensinterval

^b: øvre 95% konfidensinterval

Det ses af tabel 7.2, at det ud fra SUSY estimeres, at der er 558.000 voksne danskere, der har stillesiddende fysisk aktivitet i fritiden. Dette tal justeres til 567.000, når der kalibreres for bortfaldet i forhold til køn/alder, civilstand og region. Det ses også, at konfidensgrænserne ikke varierer meget i deres bredde med og uden brug af kalibrering.

I tabel 7.1 kan forskellen mellem kalibreret og ukalibreret (SUSY-rapport) forekomst for fx godt/virkeligt godt selv vurderet helbred illustreres som følger:

Forekomsten i SUSY-rapporten er 79,4 %, mens forekomsten kalibreret for køn/alder, civilstand og region er 79,9 %. Den absolutte difference i forekomst er således

$$\text{Difference} = 79,9 \% - 79,4\% = 0,4 \%$$

$$\text{Den relative difference er } 0,4 \% / 79,4 \% = 0,5.$$

De i tabel 7.1 angivne konfidensintervaller for forekomsterne er baseret på deres standardafvigelse (SE). SE for den ukalibrede forekomst er 0,35, mens den er 0,34 for den kalibrede. Da kalibrering forventes at reducere SE (Andersson 1998) udregnes variansreduktionen – dvs. reduktionen i usikkerhed - ved kalibrering: Variansreduktionen er

$$\text{Variansreduktion} = -(100*(0,34-0,35)/0,35) = 2.7 \%$$

I tabel 7.3 er ovenstående tre mål opgjort for 69 indikatorer taget fra SUSY-2005 rapporten (Ekholm et.al. 2006). Disse indikatorer omfatter flere af de i tabel 4.1 nævnte.

Tabel 7.3: Betydning af kalibrering opgjort ud fra 69 indikatorer fra SUSY-2005

Gruppe	Mål	Mean	Min	Q10	Median	Q90	Max
I alt	Difference	-0,1	-2,4	-0,9	-0,1	0,4	1,3
	Relativ difference	-0,6	-5,7	-3,6	-0,7	3,1	5,5
	Variansreduktion	2,1	-4,8	-1,5	1,2	8,4	20,6
Kvinder	Difference	-0,1	-1,5	-0,5	-0,1	0,3	0,7
	Relativ difference	-0,1	-3,3	-1,9	-0,4	2,8	4,6
	Variansreduktion	1,2	-5,2	-1,8	0,6	7,1	19,6
Mænd	Difference	-0,2	-3,2	-1,2	-0,1	0,6	1,6
	Relativ difference	-0,9	-11,6	-7,5	-1,0	3,7	14,9
	Variansreduktion	2,1	-10,4	-1,4	0,8	10,2	21,1
K 16-24	Difference	0,0	-0,2	-0,1	0,0	0,1	0,1
	Relativ difference	-0,3	-5,2	-1,0	0,0	0,3	1,3
	Variansreduktion	2,3	0,8	1,8	2,2	2,8	5,1
K 25-44	Difference	0,0	-0,5	-0,2	0,0	0,4	0,8
	Relativ difference	0,5	-2,4	-1,0	0,1	2,3	4,3
	Variansreduktion	-0,4	-3,1	-1,8	-0,2	0,5	2,0
K 45-64	Difference	0,0	-0,4	-0,2	0,0	0,2	0,5
	Relativ difference	0,5	-1,1	-0,5	0,2	1,9	3,3
	Variansreduktion	-1,1	-3,4	-2,2	-0,9	-0,2	0,2
K 65-79	Difference	0,0	-0,5	-0,2	0,0	0,2	0,5
	Relativ difference	0,3	-1,7	-0,7	0,0	1,9	3,2
	Variansreduktion	-1,5	-3,9	-2,6	-1,3	-0,6	1,5
K 80+	Difference	0,0	-0,4	-0,2	0,0	0,1	0,4
	Relativ difference	-0,1	-5,5	-1,5	-0,1	1,5	4,3
	Variansreduktion	-1,6	-4,7	-3,3	-1,7	-0,1	4,1
M 16-24	Difference	0,0	-0,1	-0,1	0,0	0,1	0,1
	Relativ difference	0,0	-1,2	-0,6	0,0	0,5	0,8
	Variansreduktion	2,2	1,2	1,7	2,1	2,7	3,8
M 25-44	Difference	0,0	-0,5	-0,2	0,0	0,3	0,6
	Relativ difference	0,0	-5,8	-1,6	-0,2	2,1	3,5
	Variansreduktion	0,2	-2,9	-1,0	0,3	1,3	6,0
M 45-64	Difference	0,0	-0,5	-0,4	0,0	0,2	0,7
	Relativ difference	0,3	-2,9	-1,4	-0,3	2,6	8,2
	Variansreduktion	-0,9	-9,0	-2,4	-0,5	0,5	2,1
M 65-79	Difference	0,0	-0,6	-0,3	0,0	0,2	0,5
	Relativ difference	0,1	-2,1	-1,0	-0,2	1,7	3,2
	Variansreduktion	-0,9	-4,2	-2,1	-0,7	0,2	1,4
M 80+	Difference	0,0	-1,0	-0,4	0,0	0,3	0,9
	Relativ difference	0,0	-6,3	-2,2	-0,3	2,7	6,1
	Variansreduktion	-1,0	-7,2	-3,2	-0,9	1,1	5,8

Af tabel 7.3 ses, at forskellen mellem kalibreret og ukalibreret forekomst i gennemsnit henover 69 indikatorer er -0,1 % (rækken 'I alt' - Difference, søjlen 'Mean'). Minimumsværdien er -2,4 %, 10 % af værdierne er -0,9 % eller mindre (søjlen Q10), halvdelen af værdierne er -0,1 % eller mindre (søjlen Median), 90% af værdierne er 0,4 % eller mindre og den maksimale difference er 1,3 %. Således ses, at kalibrering i gennemsnit gør de i SUSY-rapporten givne forekomster en anelse mindre, idet der dog kan ske både øgning og formindskning. Tilsvarende ses, at den gennemsnitlige relative difference er -0,6 således at de i SUSY-rapporten givne forekomster totalt ved kalibrering i gennemsnit gøres 0,6 gange mindre.

Som forventet mindskes usikkerheden på de i SUSY-rapporten givne totale forekomster – standard error bliver i gennemsnit 2,1 % mindre ved kalibrering (søjlen Variansreduktion). Det skal bemærkes, at der både sker variansøgning og variansformindskning, og at der synes at være nogle få indikatorer, hvor der sker en markant reduktion af usikkerheden (max-værdi for variansreduktion er 20,6 og medianen er 1,2, hvilket er mindre end den gennemsnitlige variansreduktion).

De tilsvarende mål er opgjort for kvinder og mænd hver for sig. Det ses, at kalibrering for alder, civilstand og region synes at have den største betydning for mænd, idet både difference og relativ difference er mindre og varierer mere for mænd end for kvinder og variansreduktionen er også større for mænd (2,1 vs. 1,2).

Endelig er opgørelserne lavet for køn og alder (K=kvinder, M=mænd). For både mænd og kvinder ses en gennemsnitlig variansøgning efter kalibrering for civilstand og region for alder 45 år og opefter.

7.1 Diskussion

Der er i dette afsnit påvist visse forskelle i forekomst, når forekomster som publiceret i SUSY-2005 rapporten (Ekholm et.al.2006) sammenlignes med forekomster, der er kalibreret for bortfaldet i forhold til køn/alder, civilstand og region. Forskellen i forekomst ligger generelt indenfor +/- 1 %.

Der er nogen kommentarer at knytte til analysen:

For det første er den her anvendte metode et forsøg på at korrigere for en ukendt bias i forekomsten af den betragtede indikator. Ved at bruge den i afsnit 6 angivne vægtning forventer vi af estimere den ukendte forekomst rigtigt, men pga. bortfaldet vil vi forvente, at det angivne estimat er 'skævt' (biased) i forhold

til det rigtige. Den anvendte metode er et forsøg på en fornuftig og veldokumenteret korrektion.

For det andet må man sige, at de anvendte socio-demografiske oplysninger er nogen relativt svage prædiktorer for de angivne indikatorer for sundhed og sygelighed. Dermed skal forstås, at de som regel har en sammenhæng med den betragtede indikator, men at der er mange, mange andre faktorer, der har betydning for forekomsten af indikatoren, og disse kan der ikke kalibreres for.

For det tredje fremgår det af tabel 4.2, at det største bortfald ligger i de grupper, der ikke 'fylder' særlig meget i den samlede stikprøve. Det må derfor forventes, at effekten fra disse grupper ikke kan være særlig markant.

Den brugte metode er velegnet, fordi den kan anvendes på mange indikatorer og tager udgangspunkt i en sammenligning mellem den aktuelle stikprøve og aggregerede oplysninger fra offentlige danske registre. Der er dog det problem med metoden, at der kan optræde negative vægte (Särndal & Lundström), og at valget af usikkerhedsberegning ikke er ligetil.

8. Dataindsamlingsmetoder og deres betydning

Dataindsamlingen i SUSY-2005 svarer på mange punkter til dataindsamlingen i de tidligere SUSY-undersøgelser, men på nogle punkter er der forskelle. Nedenfor redegøres for dette, og der søges givet et overblik over de undersøgelser, der er lavet for at belyse konsekvenser af de valgte dataindsamlingsmetoder.

8.1 Metoder

Dataindsamlingen i SUSY-2005 er som i tidligere sundheds- og sygelighedsundersøgelser baseret på personligt interview. Hver person i stikprøven fik tilsendt et introduktionsbrev med en invitation til at deltage i et interview med en professionel interviewer i eget hjem. Kontakten til de udtrukne blev forestået af en interviewer, som skulle forsøge at få fat i hver person mindst 4 gange. Hver person, der gennemførte et interview, blev efter dette spurgt, om vedkommende ville modtage et spørgeskema, som skulle udfyldes af personen selv (det selvadministrerede spørgeskema) og derefter returneres til SFI-SURVEY.

I forhold til tidligere undersøgelser var der følgende ændringer i dataindsamlings-proceduren:

- Ved interviewet blev udfyldelsen af spørgeskemaet gennemført ved hjælp af CAPI (computer-assisted personal interview), dvs. intervieweren skrev svar direkte ind på en bærbar computer. I tidligere undersøgelser

er der blevet brugt PAPI (paper-and-pencil interviewing), dvs. intervieweren har krydset af på et skema.

- I SUSY-2005 foregik dataindsamlingen kontinuerligt fra 1. maj 2005 til 17. februar 2006. I tidligere undersøgelser foregik den i tre perioder af hver ca. to måneders varighed.
- Grundet den lave deltagelsesprocent (se afsnit 3.2) blev der til slut i dataindsamlingsperioden også brugt telefoninterview (CATI – computer assisted telephone interview) omfattende personer, der havde afvist et personligt interview.
- Grundet den lave svarprocent på det selvadministrerede spørgeskema (se afsnit 3.2) blev der i april 2006 iværksat to rykkere for dette skema.

Betydningen af flere af disse forhold er forsøgt analyseret før og efter starten på dataindsamlingen. Resultatet af disse undersøgelser er beskrevet i afsnit 8.2.

8.2 Betydningen af dataindsamlingsmetode

I dette afsnit vil betydningen af følgende forhold blive omtalt:

- overgang af interviewerens spørgeskemaudfyldelse fra papir til computer (PAPI til CAPI).
- betydning af den årstid hvori data indsamles.
- betydning af supplerung med telefoninterview (CATI).
- betydning af interviewere.

De 3 førstnævnte forhold repræsenterer nye tiltag i dataindsamlingsmetode i SUSY-2005, mens det sidste forhold er af generelt tilsnit i forhold til SUSY-undersøgelserne.

8.2.1 Overgang af interviewerens spørgeskemaudfyldelse fra papir til computer (konvertering fra PAPI til CAPI)

I SUSY-2005 er alle interviewdata indsamlet ved, at intervieweren har udfyldt et skema direkte på en medbragt bærbar PC. Dette er indført for at effektivisere dataindsamlingen, og sikre at data hurtigere bliver klar til brug.

Der findes nogen litteratur omkring effekten af en sådan overgang, men ikke megen om, hvad det betyder i helbredsundersøgelser. Derfor er der fra SIF dels holdt foredrag om dette emne og dels udarbejdet et udkast til en engelsksproget artikel. Derfor bringes der her en oversættelse af abstract fra artikeludkastet. Ønskes nærmere oplysninger kan henvendelse ske til Ola Ekholm, e-mail: ola@ssi.dk.

Abstract fra artiklen: 'A Comparison of Paper-And-Pencil Interviewing (PAPI) with Computer-Assisted Personal Interviewing (CAPI) in a Health Interview Survey'.

Formålet med undersøgelsen er at sammenligne de to dataindsamlingsmetoder Paper-And-Pencil Interviewing (PAPI) og Computer-Assisted Personal Interviewing (CAPI) mht. respons-rate, samlet interviewtid og en række velkendte indikatorer for sundhed og sygelighed. Data stammer fra en dansk sundheds- og sygelighedsundersøgelse foretaget i 2003. Fire stikprøver hver bestående af 1.000 voksne danske statsborgere blev tilfældigt udtrukket fra CPR-registeret. I en stikprøve blev der brugt PAPI og i de øvrige 3 CAPI. Respons-raten var 67,5% i PAPI-stikprøven og 64,4% i CAPI-stikprøven. Resultaterne indikerer, at kvinder og ældre er mindre tilbøjelige til at deltage når der bruges CAPI. Resultaterne viser, at det er problematisk at gennemføre en overgang fra PAPI til CAPI i helbredsundersøgelser.

8.2.2 Betydning af den årstid hvori data indsamles

Som nævnt i afsnit 3.1 er dataindsamlingen i SUSY-2005 sket kontinuerligt i løbet af 2005. I undersøgelseerne fra før 2005 er dataindsamlingen sket i 3 perioder (februar/marts, maj/juni og september/ oktober). Inden SUSY-2005 gik i gang blev det forsøgt vurderet, hvilken effekt en sådan ændring kunne få. Dette udmøntedes i et såkaldt 'associated paper', der er vedlagt i bilag 10.2, samt et internt notat, der kan fås ved henvendelse til dette notats forfatter.

Undersøgelsen blev foretaget på baggrund af SUSY-2000, og kan sammenfattes som følger:

'Formålet med dette paper er at undersøge sæsonmæssige forskelle i helbredsindikatorer og forsøge at vurdere, om der findes en optimal periode at foretage helbredsundersøgelser i. Den relative afvigelse mellem forekomster i de fire sæsoner (vinter, forår, sommer, efterår) og den totale forekomst ligger generelt mellem -5% og +5 % - dog ses større og mindre afvigelser ikke så sjældent. Der ses ikke noget klart billede mht. hvilken sæsonforekomst, der bedst svarer til den totale forekomst. Der ses en svag tendens til at forekomster i efteråret ligner den totale forekomst mest, men dette skyldes i stor udstrækning, at der er blevet interviewet flest personer i efteråret'.

Om forekomsten af en given indikator stiger eller falder med stigende eller faldende uddannelse (den sociale gradient i en indikator) påvirkes ikke af sæson – men størrelsen af den sociale gradient gør. Dette betyder, at hvis man udelukkende er interesseret i retningen af den sociale gradient fås samme resultat uanset sæson, men hvis man også ønsker at sige noget om størrelsen af gradienten, kan ingen sæson anbefales.

Det er således svært at lave en entydig konklusion, men der ses en svag tendens til at efteråret giver den bedste tilnærmelse til det totale resultat.

8.2.3 Betydning af supplering med telefoninterview (CATI)

I forbindelse med det i afsnit 2.2 nævnte krav fra SIF om en opnåelsesprocent på 70 %, blev det i november 2006 klart, at dette mål ikke kunne nås. I den forbindelse blev det aftalt med SFI Survey, der stod for dataindsamlingen, at man fra medio december 2005 måtte forsøge et telefoninterview med de personer, der på en ikke alt for bastant måde afslog et interview i eget hjem – dvs. de 'bløde nægttere'. Formålet var primært at gøre opnåelsesprocenten på det personlige interview højere.

Nogle konsekvenser af denne beslutning er belyst ved et foredrag på den internationale konference 'European Conference on Quality and Methodology in Official Statistics (Q2006)'. I denne forbindelse blev der fremstillet et engelsksproget paper vedlagt i bilag 10.3, der fortalte om resultaterne. Nedenfor bringes konklusionen på dansk:

'Det var ikke fra starten planlagt at inkludere telefoninterview i SUSY-2005 og det blev indført for at øge respons-raten og herved opfylde et prædefineret krav. Respons-raten blev øget, men kravet blev ikke opfyldt. Det er vist, at vi med CATI har opnået interview med nogen personer, vi ville forvente havde større tendens til ikke at ville deltage (ugifte personer), og dette er et positivt resultat. Det at personer interviewet med CATI rapporterer dårligere selv vurderet helbred end resten, støtter dette, mens det at personer interviewet med CATI rapporterer lavere alkoholforbrug, synes at tale imod. Alt i alt synes indførelsen af CATI at øge responset og mindske bias i undersøgelsen.'

8.2.4 Betydning af interviewere

Som nævnt i afsnit 2.2.1 er denne undersøgelse såvel som tidligere SUSY-undersøgelser gennemført ved besøgsinterview i svarpersonernes hjem, og det er således interviewerens, der registrerer svarpersonernes oplysninger. Et nærliggende spørgsmål er, om dette forhold kan have betydning for de afgivne oplysninger, dvs. om interviewerens på trods af alle bestræbelser på at være neutral kan have en betydning. I hvor høj grad dette er tilfældet er belyst i et speciale ved Københavns Universitet, Afdeling for Matematisk Statistik og Operationsanalyse, med titlen 'Interviewereffekt i SUSY-2000' og skrevet af Esben Meulengracht Flachs. Her bringes specialets resumé. Hvis nogen ønsker nærmere oplysninger, henvises til Københavns Universitet eller Esben Meulengracht Flachs (e-mail: emf@si-folkesundhed.dk).

Resumé af specialet 'Interviewereffekt i SUSY-2000'

'Sundheds- og sygelighedsundersøgelsen år 2000 er en del af det program for overvågning af folkesundheden i Danmark, som Statens Institut for Folkesundhed varetager. Undersøgelsen omfatter ca. 23.000 personer udtrukket tilfældigt ved hjælp af CPR-nummer. Heraf er omtrent 16.000 blevet interviewet i en af tre runder i februar, maj og september 2000. Disse interview er foretaget i respondenternes eget hjem af ca. 300 interviewere, således at respondent og interviewer er parret tilfældigt. Interviewerne er ansat hos SFI Survey, der tilstræber en i forhold til befolkningen repræsentativ interviewerskare.

Statens Institut for Folkesundhed har med udgangspunkt i den tilfældige parring af respondent og interviewer arbejdet udfra den antagelse, at designet minimerer effekt af interviewere på respondenternes svar. Således har udgangspunktet for studier med baggrund i disse data været, at interviewerne ikke bidrager med yderligere usikkerhed. Dette speciale har som formål at undersøge om denne antagelse er korrekt, eller om de benyttede interviewere har en effekt på undersøgelsen, og hvis dette er tilfældet, at beskrive hvorledes og i hvor stor udstrækning interviewerne påvirker sundheds- og sygelighedsundersøgelsen.

Den anvendte metode er at forsøge at modellere respondenternes svar med udgangspunkt i en antagelse om, at interviewerne påvirker data med både en systematisk og en tilfældig effekt. De benyttede modeller er på grund af de fleste spørgsmåls diskrete svarmuligheder og den påståede tilfældige interviewer-effekt alle varianter af generaliserede mixede lineære modeller. I specialet behandles ni spørgsmål fra forskellige centrale problemstillinger i undersøgelsen. Disse er tre spørgsmål om interviewsituationen: "Bortfald blandt respondenter", "antallet af kontaktforsøg" og "accept af spørgeskema", tre spørgsmål om helbredstilstand og brug af behandlingstilbud: "selvvurderet helbred", "langvarig sygdom" og "medicinforbrug", og tre spørgsmål til respondentens livsstil: "alkoholforbrug", "rygevaner" og "body mass index".

De ni undersøgte spørgsmål deler sig i tre kategorier. En gruppe, der omfatter de spørgsmål, hvorpå der konstateres både tilfældig og systematisk interviewereffekt.

Denne omfatter "bortfald blandt respondenter", "antallet af kontaktforsøg", "selvvurderet helbred", "langvarig sygdom" og "medicinforbrug". Den anden gruppe er spørgsmål helt uden effekt af interviewere: "alkoholforbrug", "rygevaner" og "body mass index". Den tredje gruppe kun bestående af spørgsmålet om "accept af spørgeskema", idet den valgte klasse af modeller ikke kan modellere dette tilfredsstillende.

Den anden gruppe, gruppen af spørgsmål helt uden interviewereffekt, udmærker sig ved at bestå af spørgsmål, der enten er interviewomstændigheder eller er mere faktisk prægede i sammenligning med den første gruppe, der overvejende består af mere holdningsprægede spørgsmål.

Specialet munder ud i en række anbefalinger både til det fortsatte arbejde med data fra sundheds- og sygelighedsundersøgelsen år 2000 og til dataindsamling ved fremtidige undersøgelser.

9. Referencer

Andersson C, Nordberg L: A users guide to CLAN 97. 1998; Statistics Sweden (incl. supplement).

Davidson M, Kjølner M: The Danish Health and Morbidity Survey 2000 - Design and Analysis. *Statistics in Transition*, 2002; 5(6), pp. 901-25

Davidson M, Helweg-Larsen K, Kjølner M: DANCOS - en registeropfølgning af de nationale danske sundheds- og sygelighedsundersøgelser. *Ugeskrift for Læger*. 2004. 166(15-16), pp. 1441-4

Ekholm O, Kjølner M, Davidson M, Hesse U, Eriksen L, Christensen AH, Grøn-bæk M: Sundhed og sygelighed i Danmark & udviklingen siden 1987. København: Statens Institut for Folkesundhed, 2006.

Kjølner M, Rasmussen NK, Keiding L, Petersen HC, Nielsen GA: Sundhed og sygelighed i Danmark 1994 – og udviklingen siden 1987. København; Statens Institut for Folkesundhed. 1995.

Kjølner M, Rasmussen NK, eds. Sundhed og sygelighed i Danmark år 2000 ... & udviklingen siden 1987. København: Statens Institut for Folkesundhed, 2002.

Rasmussen NK, Groth MV, Bredkjær SR: Sundhed og sygelighed i Danmark 1987. København; Statens Institut for Folkesundhed. 1988.

Särndal CE, Lundström S: *Estimation in Surveys with Nonresponse*. 2005; Wiley, New York

Särndal CE, Swensson B and Wretman J: *Model Assisted Survey Sampling*. 1992; Springer, New York

10. Bilag

10.1 Invitation af tidligere inviterede (SUSY-1994 og SUSY-2000)

Beslutning om hvorvidt tidligere inviterede personer skal opsøges igen

Må SIF opsøge personen igen? (egen angivelse)

1994	2000	Forskerbeskyttelse (CPR)	Inviteres 2005
Ja	Ja	Ja	Ja
Ja	Ja	Nej	Ja
Ja	Nej	Ja	Ja
Ja	Nej	Nej	Ja
Ja	Uoplyst	Ja	Ja
Ja	Uoplyst	Nej	Ja
Nej	Ja	Ja	Ja
Nej	Ja	Nej	Ja
Nej	Nej	Ja	Nej
Nej	Nej	Nej	Nej
Nej	Uoplyst	Ja	Nej
Nej	Uoplyst	Nej	Nej
Uoplyst	Ja	Ja	Ja
Uoplyst	Ja	Nej	Ja
Uoplyst	Nej	Ja	Nej
Uoplyst	Nej	Nej	Nej
Uoplyst	Uoplyst	Ja	Ja
Uoplyst	Uoplyst	Nej	Nej

10.2 'Associated paper' om årstidsvariation

Seasonal differences in a Danish Health Interview Survey

Michael Davidsen¹, Ola Ekholm¹, Mette Kjølner¹

Associated paper

European Conference on Quality and Methodology in Official Statistics (Q2004), Mainz, Germany, May 24-26, 2004.

¹: National Institute of Public Health
Øster Farimagsgade 5A
DK-1399 Copenhagen K
Denmark

Correspondence to:
Michael Davidsen
Phone: +45 39 20 77 77
E-mail: md@niph.dk

Introduction

Design of population-based health surveys is a complicated task involving consideration of many complex problems. Examples of these are selecting the population and sample, ensuring financial support, defining the purpose of the survey and developing the questionnaire. One issue that rarely is considered in health surveys is the timing of the data collection in the survey – timing here meaning in the design phase deciding the time of year to perform the survey.

The Danish Health Interview Survey programme was initiated in 1986 with the purpose of monitoring health and morbidity in the adult Danish population and

determinants hereof. Health Interview Surveys have been carried out in 1986/1987, 1994 and 2000 and a new survey to be conducted in 2005 is presently in the planning phase. The previous surveys have all been carried out in three waves in February/March, May/June and September/October, the idea being to minimize seasonal differences. A data collection phase of at least 9 months is very expensive and some might claim that results are outdated when published.

In the 2005 survey it has been suggested to restrict the period of data collection to approximately 3 months.

The overall purpose of this paper thus is to investigate seasonal differences in survey data on public health indicators in order to evaluate if there exist an optimal period of the year in which to perform health survey. In order to approach this we

- quantify and assess seasonal differences in prevalence
- assess the impact of season when describing social gradients in health.

Methods

Data in the present paper derives from the Danish Health and Morbidity Survey 2000. This survey was conducted by the Danish National Institute of Public Health among adult Danish citizens aged 16 years or more mainly in year 2000 but in order to fulfill the required sampling scheme to be described below the data collection period was prolonged to include 2 months in year 2001.

Design of the Danish Health and Morbidity Survey 2000

The design this survey has previously been described in detail [1], here only a brief summary will be presented.

Adult Danish citizens (age 16 or more) were drawn at random using the Danish Civil Person Register. All Danes have a unique personal identification number stored in this register and this number is used in all contacts with civil authorities including e.g. hospitalization.

The sample consisted of three subsamples: the *national sample*, the *follow-up sample* and the *supplementary county sample*.

The *national sample* is a close parallel to previous surveys as the sample was selected by stratification proportional to the size of each of 15 Danish counties.

The *follow-up sample* consists of three parts: (a) all persons invited to the survey in 1994 and still alive in 2000 (b) a supplement of persons aged 16-21 in year 2000 and (c) a supplement of persons who have obtained Danish citizenship between 1994 and 2000.

The *supplementary county sample* was drawn with the purpose of ensuring at least 1,000 completed interviews in each of the 15 counties.

As in the previous surveys, the primary sampling unit was individual persons.

Persons were randomly chosen to each of the three waves conducted in February/March, May/June and September/October.

Each selected person received a written invitation and if accepted, a face-to-face interview was carried out in the respondent's home. At the completion of the interview all respondents were asked to complete a self-administered questionnaire and return it at convenience

For each person a date of last contact was established, for persons accepting an invitation this date is the date of interview. This date was related to each of the four seasons winter (December, January, and February), spring (March, April, May), summer (June, July, August) and autumn (September, October, November).

Questions

In the Danish Health Interview Survey programme the concept of health is broadly defined [2]. Issues addressed in the surveys can be divided into core elements included in every survey and ad hoc issues of special interest.

Adapted to subsample, mode (interview or self-administered) and wave a set of questionnaires was developed in the design phase of the survey. An English translation of all questions asked can be found on <http://www.sifolkesundhed.dk/susy>.

The following socio-demographic variables is considered in this paper: age, sex, education measured by ISCED (International Standardized Classification of Education, grouped as 0-10 years, 11-12 years, 13-14 years and 15 years or more), being economically active (yes/no, only age 16-66) and cohabital status (grouped as married, cohabitant, single or not married – in contrast to ‘single’ ‘not married’ never have been married).

We chose the following indicators in our analysis: Good/really good and bad/very bad self-rated health (based on the question ‘How do you rate your present state of health in general? – 5 categories), longstanding illness, daily smoking, heavy smoking (at least 15 cigarettes pr. day), alcohol (drinking more than recommended by the Danish Board of Health – for men 21 drinks pr. week, for women 14), functional limitations for persons aged 60 or more (based on ability to read a newspaper, hear what is said, walk 400 m, walk a staircase, carry 5 kg and speak without difficulties), obesity (having a self-reported BMI (Body Mass Index) of 30 kg/m² or more) and contact to a general practitioner (GP) within the past 3 months. All indicators were dichotomized into two categories ‘yes’ and ‘no’.

Statistical methods

The weighting scheme and statistical analysis in the Danish Health and Morbidity Survey 2000 has previously been described [1].

Prevalence of socio-demographic variables and public health indicators were compared over the four seasons using a simple chi-square test. Adjustment for multiple comparison was done by controlling the false discovery rate [3] using a sequential Bonferroni procedure.

To assess the magnitude of differences we calculated the (percentage) relative difference (RD) in the following way:

Let p_s be the prevalence in season s and p be the total prevalence pooling the four seasons. Then RD is

$$(1) \quad RD = 100 * (p_s / p - 1) = 100 * (p_s - p) / p$$

Thus RD measures the percentage difference between the prevalence in each season and the overall prevalence.

Education (ISCED) was used to measure social gradients. Expected (season-specific) odds-ratios adjusted for age and sex were calculated in a logistic regression model [1].

Results

Season and response

All in all 22,486 persons were invited, of these 16,690 (74.2%) had an interview.

Shown in figure 1 is the number of persons in the Danish Health and Morbidity Survey 2000 distributed according to interview / non-response and subsample.

The interview scheme with invitation in February/March, May/June and September/October is clearly seen. The data collection period was almost one year in order to ensure at least 1,000 interview in each county - this also is the reason for the increased number of persons invited in the last wave and especially in the supplementary county subsample. 15,579 persons (93.3%) had an interview within the three waves.

Table 1 shows the response proportion according to the four seasons. A markedly higher response proportion was seen in spring and summer than in winter and autumn ($p < 0.0001$). The relative difference (RD) from the overall response proportion did not exceed 10%. The pattern could not be explained by age, gender and subsample. Among persons having an interview 18.4% were interviewed in winter, 23.5% in spring, 16.2% in summer and 42.0% in autumn

Seasonal differences in socio-demographic and public health indicators

Shown in table 2 is the distribution of socio-demographic variables according to season. For many of the variables statistically significant differences between the four seasons were found but the RD in most cases ranges between -10% and +10%, exceptions being in winter. No clear pattern is seen in RD.

In table 3 seasonal differences in prevalence of the selected public health indicators are shown. For alcohol the overall prevalence of exceeding recommended alcohol intake is 11.7%, the prevalence is 8.8% smaller in winter, 7.4% smaller in spring, 25.7% higher in summer and 1.8% smaller in autumn. These differences cannot be assumed to be negligible. This pattern is not surprising, it is expected that Danes drink more alcohol in warm weather. Likewise an increase in moderate or heavy leisure time physical activity during summer is to be expected but it is a bit surprising that the same pattern was not seen in spring. Adjustment for multiple testing did not alter any conclusions regarding statistically significant differences.

The general impression is that prevalence in autumn are most like the overall prevalence.

Shown in table 4 are age- and sex-adjusted odds-ratios in the four seasons and in total in different educational groups (defined by ISCED). In all analysis 13-14 years was chosen as the reference group and thus the group is not displayed, the rationale being to simulate a situation in which the survey had taken place in that season only. For instance for 'really good or good self-rated health' a clear overall pattern of increasing occurrence of good self-rated health with increasing education was seen (OR=0.47, 0.65 and 1.39 going from <=10 years to 15 or more years). If data had been available for summer only, the OR would have been 0.38, 0.51 and 1.05 in the educational groups. Thus the direction of the gradient (increase with increasing ISCED) is the same but the slope is different. The general impression from table 4 is as just described, i.e. the direction of the gradient is preserved in each season but the magnitude of the increase differ somewhat. Inspection of RD (based on the OR, not shown) reveals that they are lowest in the autumn season and that there are some large deviations in the group with 15 years or more of education especially in winter. This in part may be due to small numbers in these groups.

Discussion

The results indicate that seasonal differences exist with respect to prevalence of socio-demographic and public health variables. The results show no clear pattern making conclusions difficult. The direction of the social gradient was not influenced by season but the size of increase (or decrease) along educational groups was. Several of the chosen public health indicators are included in other European health surveys.

Significance testing is not of major importance in this setting because even moderate deviations may turn out to be statistically significant due to the sample size. Therefore inspection of the size of deviations may be more relevant.

The difficulty with the RD-measure clearly is how to decide when the size of RD is noteworthy. A global cut-point cannot be derived based on the results presented here – and in general does not seem feasible. Presumably the decision will depend heavily on the purpose of the survey and how crucial the indicator is.

The interpretation of the results is to some extent hampered by the choice of the four seasons, the reason for this being unequal number of persons – 42% of all interviews was conducted in autumn and only 16% in summer. This implies that the overall prevalence depends more strongly on the autumn season and thus it is not surprising that deviations in the autumn season in many instances are the smallest. If the seasons had been defined by the three waves, the relative differences of odds-ratios in table 4 would be somewhat diminished and more scattered across educational groups but RD would be lowest in the September/October wave. Also the result on direction of the gradient would be the same.

The low response proportion in autumn may be partly due to a rather complex re-invitation scheme in which persons unable to accept an invitation (e.g. due to lack of time) in one of the first waves were offered an interview in the last wave. Thus more persons actually not intending to participate may be found in the third wave (September/October) than in the first two. However this does not in any way explain the low response proportion in winter.

Other more practical considerations are highly relevant when considering taking season into the planning phase of a survey. The month of December in Denmark is dominated by lots of Christmas preparations and arrangement and people usually are very busy. Many Danes have their holiday in July especially among blue-collar workers and thus are not home in this month.

Conclusion

The overall purpose of this paper is to investigate seasonal differences in survey data on public health indicators in order to evaluate if there exist an optimal period of the year in which to perform health survey. The magnitude of difference in prevalence comparing each season to the total in general is between -5% and +5% (measure by percentage relative difference, RD) but deviations outside these limits are seen and not rarely. Is not clear which season resembles the overall result the most and no clear pattern is seen. Although prevalence from the autumn season may seem to resemble the overall prevalence the most this is to a large extent influenced by the fact that most persons were interviewed during autumn.

The direction of social gradient was not affected by season but the magnitude of increase (or decrease) was. If interest solely is in the direction of the social gradient this indicate that all seasons are equally bad. Thus it is hard to make conclusions but there seem to be a weak tendency of autumn season most resembling the overall results.

Reference List

[1] Davidsen M Kjøller M. The Danish Health and Morbidity Survey 2000 - Design and Analysis. SiT, 2002;5(6):927-42.

[2] Rasmussen NK Kjøller M. SIF's program for sundheds- og sygelighedsundersøgelser (in Danish, English Summary). Ugeskrift for Læger 2004;

[3] Benjamin Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J Royal Stat Soc B 1995;57(1):289-300.

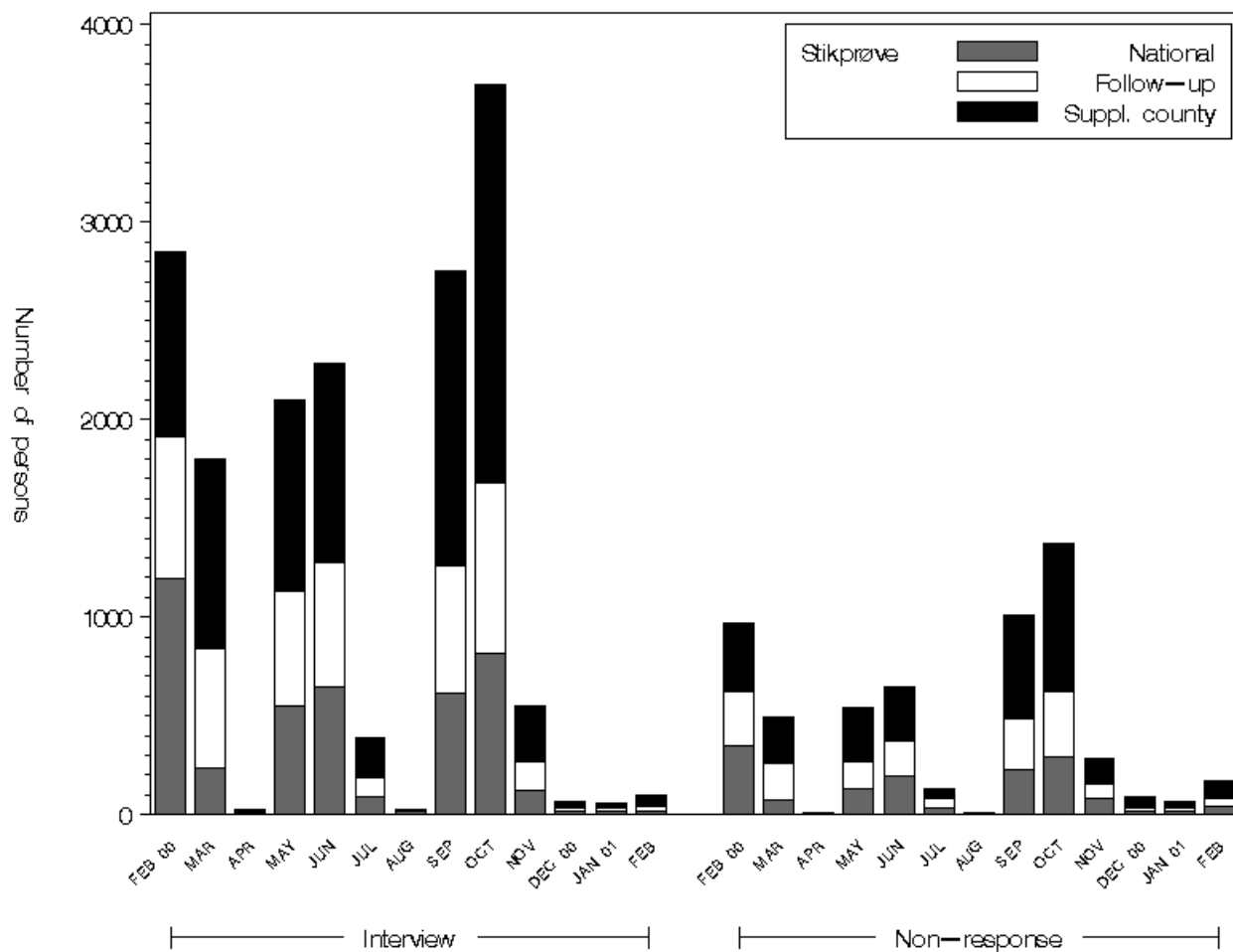


Figure 1: Time of last contact/interview in the Danish Health and Morbidity Survey 2000

Table 1
Response proportion in the four seasons

Season	Interview (n / %)		Total (n)
	Yes	No	
Winter (December, January, February)	3065 70.3	1298 29.8	4363
Spring (March, April, May)	3923 79.0	1044 21.0	4967
Summer (June, July, August)	2700 77.5	785 22.5	3485
Autumn (September, October, November)	7002 72.4	2669 27.6	9671
Total	16690 74.2	5796 25.8	22486

Table 2

Prevalence and relative differences (RD) of socio-demographic variables.

Variable	Group	RD (%)				Total (%)	p
		Winter	Spring	Summer	Autumn		
Age	16-24	10.2	-7.4	-3.6	1.0	13.4	<0.0001
	25-44	-5.5	-2.5	-4.6	5.6	35.4	
	45-66	-3.3	1.6	9.5	-3.1	35.5	
	67+	11.0	8.2	-8.0	-6.4	15.8	
Sex	Men	2.4	-1.7	1.9	-0.8	49.1	0.25
	Women	-2.3	1.6	-1.8	0.8	50.9	
ISCED	<=10 years	1.1	2.3	0.5	-2.0	24.9	0.0005
	11-12 years	5.6	2.3	-2.1	-3.0	24.1	
	13-14 years	2.7	-2.9	0.2	0.4	29.6	
	15+ years	-14.1	-3.2	3.8	6.6	19.4	
	Other	16.0	18.5	-22.5	-8.9	2.0	
Employed*	Yes	-3.6	-1.7	2.2	1.6	74.7	0.0007
	No	10.6	5.1	-6.5	-4.8	25.3	
Cohabital status	Married	-2.2	3.8	2.2	-2.0	51.6	0.0141
	Cohabitant	-1.1	-6.0	-6.2	6.3	15.9	
	Single	4.2	0.6	2.9	-3.3	13.5	
	Not married	3.9	-5.8	-2.7	2.6	19.0	

*: Only persons aged 16-66 years

Table 3

Seasonal difference in selected public health indicators

Indicator	RD (%)					Prevalence (%)	p	Adjust. p
	1.1.1.1	Winter	Spring	Summer	Autumn			
Good / really good self-rated health		-2.3	-0.9	2.7	0.5	77.9	*	*
Bad / very bad self-rated health		-4.1	10.1	-5.0	-2.0	6.1		
Longstanding illness		-0.6	1.6	2.7	-1.7	41.1		
Functional limitations ^a		-4.0	2.6	12.1	-4.3	6.7		
Current smoking		0.6	1.3	-2.7	0.0	34.0		
Heavy smoking		-3.3	-0.5	-3.5	3.1	18.5		
Alcohol		-8.8	-7.4	25.7	-1.8	11.7	*	*
Moderate/heavy leisure time physical activity		-11.3	-4.5	5.1	5.6	23.0	*	*
Sedative leisure time physical activity		7.9	6.1	-8.2	-3.9	16.3	*	*
Obesity (BMI≥30)		5.7	-3.7	-6.5	2.1	9.5		
Contact GP last 3 months		3.6	1.1	-3.6	-0.8	39.7		

*: p<0.05

^a: Only persons aged 60 or more

Table 4- Impact of season on social gradient in selected public health indicators.

Variable	ISCED *	OR				OR
		Winter	Spring	Summer	Autumn	
Good / really good self-rated health	<=10 years	0.51	0.45	0.38	0.48	0.47
	11-12 years	0.76	0.63	0.51	0.66	0.65
	15+ years	1.26	1.69	1.05	1.44	1.39
Bad / very bad self-rated health	<=10 years	2.66	2.27	2.81	2.38	2.45
	11-12 years	1.90	1.68	1.50	1.69	1.69
	15+ years	1.35	0.52	0.62	0.65	0.71
Longstanding illness	<=10 years	1.31	1.77	1.68	1.45	1.53
	11-12 years	1.27	1.60	1.44	1.34	1.39
	15+ years	0.98	0.94	0.90	0.93	0.93
Functional limitations ^a	<=10 years	1.76	1.45	1.83	1.84	1.71
	11-12 years	1.34	1.66	1.55	1.67	1.58
	15+ years	1.21	0.34	0.81	0.76	0.72
Current smoking	<=10 years	1.30	1.96	2.12	1.84	1.78
	11-12 years	1.28	1.41	1.56	1.40	1.40
	15+ years	0.54	0.54	0.75	0.57	0.58
Heavy smoking	<=10 years	1.43	2.05	1.96	1.95	1.86
	11-12 years	1.36	1.48	1.42	1.32	1.38
	15+ years	0.52	0.58	0.47	0.45	0.49
Alcohol	<=10 years	0.72	0.69	0.67	0.78	0.73
	11-12 years	1.08	0.90	0.97	1.10	1.02
	15+ years	1.45	1.05	1.23	1.03	1.14
Moderate/heavy leisure time physical	<=10 years	0.69	0.64	0.64	0.60	0.63
	11-12 years	0.94	0.78	0.80	0.77	0.81

Variable	ISCED *	OR				OR
		Winter	Spring	Summer	Autumn	
physical activity	15+ years	1.37	1.30	1.32	1.22	1.28
Sedative leisure time physical activity	<=10 years	1.39	1.92	1.76	2.10	1.82
	11-12 years	1.16	1.19	1.09	1.33	1.22
	15+ years	0.72	0.69	0.93	0.76	0.76
Obesity	<=10 years	2.03	2.11	1.53	1.84	1.87
	11-12 years	1.35	1.45	1.34	1.39	1.38
	15+ years	0.92	0.76	0.71	0.64	0.72
Contact GP last 3 months	<=10 years	1.00	1.12	0.96	1.14	1.08
	11-12 years	0.95	1.12	0.86	1.05	1.01
	15+ years	1.00	0.95	0.85	0.75	0.85

OR: Odds-ratio

^a: Only persons aged 60 or more

*: 13-14 years reference group and not displayed (RD=0, OR=1) and therefore the category does not appear in the table.

10.3 Paper om effekt af telefon-interview

Supplementing CAPI with CATI in a National Health Survey – effect on non-response

Daidsen M¹ and Ekholm O¹

Associated paper

European Conference on Quality and Methodology in Official Statistics (Q2006),
Cardiff, England - 2006

¹: National Institute of Public Health, Øster Farimagsgade 5A 2., DK-1399 Copenhagen K, Denmark

Correspondence to: Michael Daidsen, National Institute of Public Health, Øster Farimagsgade 5A 2., DK-1399 Copenhagen K, Denmark. E-mail: md@niph.dk

Keywords: HIS, interview-mode, non-response, CAPI, CATI

1. Introduction

1.1 Description

In recent years response rates has decreased all over Europe and in many different surveys. In order to avert this problem several possibilities exists (Biemer 2003). One possibility is to use different modes of data collection. When applying this technique it would be desirable if the different modes were able to capture slightly different parts of the sample. For example it would in many countries be expected that younger persons more easily are contactable by telephone than by a visit in their home and thus it would be expected that more younger persons would answer by telephone than by personal interviewing.

In Denmark nationally representative Health Interview Surveys on health and morbidity have been conducted regularly since 1986. In 2005 the Danish Health and Morbidity Survey 2005 was planned and carried out. In this survey a response rate of 70% was required and it was attempted to supplement the main data collection mode of computer-assisted personal interview (CAPI) by computer-assisted telephone interview (CATI), both initiatives in order to secure satisfactory response rate. In this paper we describe the effect of this strategy both with regard to non-response and with regard to consequences on a few selected health indicators. Further we compare the sex- age- and legal marital status-distribution among the two types of interviewing mode and non-responders.

2. The Danish Health and Morbidity Survey 2005 (DHMS-2005)

2.1 The Danish Health and Morbidity Survey Program

The DHMS-2005 is part of the Danish Health and Morbidity Survey Program that was initiated in 1986 (Rasmussen 2004). All surveys in this program are Health Interview Surveys where each selected individual receives a face-to-face interview in their home. In Denmark each inhabitant is uniquely registered in the Danish Civil Register. Each survey is based on a random sample from this register and consists of Danish citizens aged 16 or above. Each selected individual receives a written invitation and an introduction to the survey.

A broad range of questions on e.g. health, health behaviour, life-style, contact to general practitioner, morbidity and symptoms are posed in each survey. See www.si-folkesundhed.dk/sugy for more details.

Up to year 2005 general surveys had been conducted in year 1987, 1994 and 2000.

2.2 The DHMS-2005 survey

Data collection started May 1. 2005 and ended February 17. 2006. In the design phase of this survey it was decided that all interviews should be conducted using CAPI only. However as shown in section 3.1 this decision was revised during the data **collection** period.

The response rate in the DHMS-1987 survey was 79.9%, in the DHMS-1994 survey 77.2% and in the DHMS-2000 survey 74.2% (Davidsen 2004). Therefore a response rate of 70% was decided as goal for the overall response rate in the DHMS-2005 survey.

In Denmark a new administrative structure with impact on health care and prevention has been decided. A consequence of this decision is the introduction of 5 new regions and to take account of this it was decided that a stratified sample should be drawn. The sample should be constructed in such a way that

- a) 3,000 interviews were obtained in each region
- b) the overall response rate should be 70%
- c) the regional response rate should be 70% except in region 1 where it should be 64%.

Based on these requirements a sample of 21,832 individuals was constructed. The sample comprised 4,688 individuals from region 1 and 4,286 individuals from region 2, 3, 4 and 5.

The data collection was conducted by SFI-SURVEY at the Danish National Institute of Social Research. All in all 176 interviewers took part in the data collection. Contact to individuals eligible for an interview was conducted by the interviewer who were obliged to make at least 4 attempts. Each interviewer was free to decide the order in which to contact persons appointed to her by SFI-SURVEY.

3. Response rate and CATI supplement

3.1 CATI supplement

Shown in table 1 is the response rate overall and in each region on December 8. 2005. The response rate was quite low (58.7%) especially in region 1 (50.2%) and 2 (50.3%). Furthermore due to the way contact between interviewer and respondent was made it was expected that response rate would be

lower during the last 2 months of data collection. Based on these considerations it was decided to supplement by interviewing by telephone (CATI).

Table 1: Regional response rate December 8 2005.

Region	Interviews	Response rate (%)
1	2371	50.6
2	2276	53.1
3	2806	65.5
4	2652	61.9
5	2700	63.0
Total	12805	58.7

CATI interviewing was attempted with the following groups of persons:

- 'soft' refusers. - persons who not unambiguously had refused to participate
- persons out of town
- persons who had not been met
- persons who had been hospitalized

Prior to start of CATI interviewing it was decided that only 34 experienced interviewers should conduct the interviews, the interviews should be performed primarily in region 1 and 2 and performed in January and February 2006.

Prior to a full-scale CAI interviewing a small pilot study of 40 persons was established. This study showed that it was possible to contact the above mentioned group of persons again.

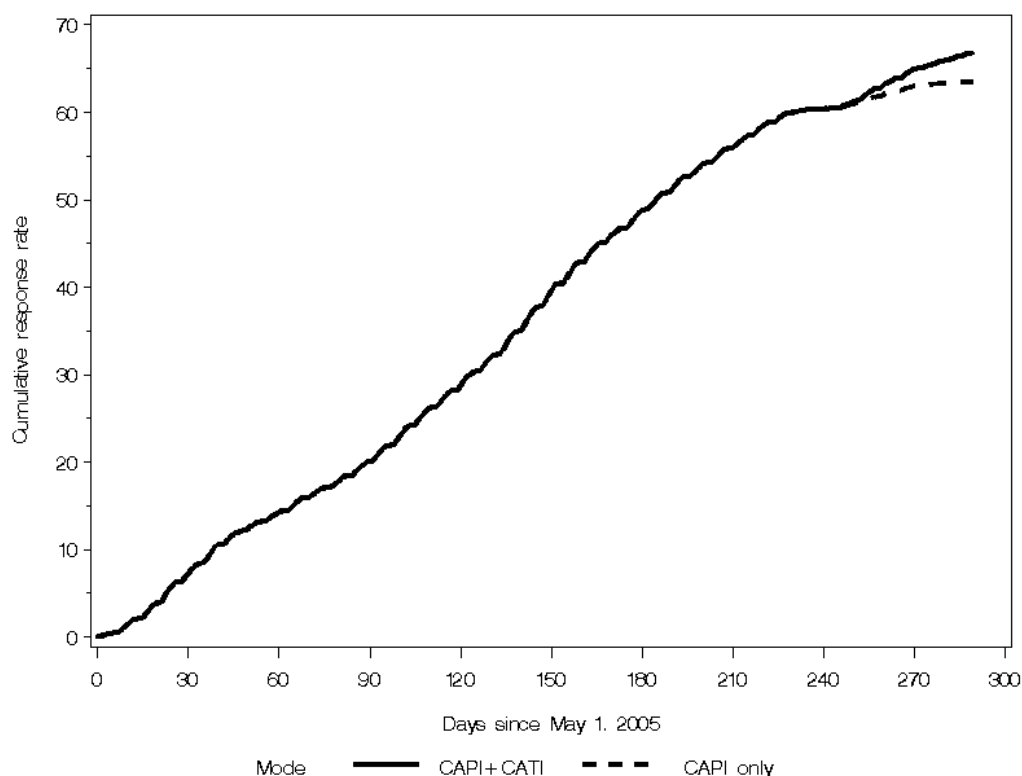
Unfortunately among persons refusing to have an interview at December 8 2005 it was not registered who were 'soft refusers' and who were not. Therefore it is not possible to report CATI response rates in this paper. Personal communication with SFI-SURVEY indicate a CATI response rate of approximately 20%. In further sections we therefore focus on final interviewing status.

Response	CAPI	CATI sup	Non-response	n
December 8.				
Refused	2.1	7.6	90.3	4912
Out of town	2.0	12.8	85.1	148
Not met	3.2	13.8	83.0	814
Not contacted	42.6	6.0	51.4	2030
Other	6.7	5.7	87.6	1123
Total	11.9	7.6	80.5	9027

3.2 Final response rates

Shown in figure 1 is the cumulative response rate as a function of days elapsed since May 1. 2005. The solid line shows the response rate for both CAPI and CATI interviews while the dotted shows response rate by CAPI alone – i.e. regarding CATI as non-response. Thus it can be seen that the introduction of CATI increased the response rate by 3.2%. The two curves start to split at December 8 2005.

Figure 2 Cumulative response rate from May 1. 2005.



Shown in table 2 is the final interviewing status for persons who had not been interviewed at December 8 2005 where CATI interviewing was decided. It is seen that interview has been obtained among all persons who were selected for contact. Note that invitations were sent out continuously during the data collection period and therefore not all persons had been contacted December 8. The high CAPI response rate in the 'Not yet contacted'-group shows CATI was indeed regarded as a supplement in January and February 2006.

Table 2: Final response status according to status when CATI was decided

Response status	CAPI	CATI sup	Non-response	n
December 8.				
Refused	2.1	7.6	90.3	4912
Out of town	2.0	12.8	85.1	148
Not met	3.2	13.8	83.0	814
Not contacted	42.6	6.0	51.4	2030
Other	6.7	5.7	87.6	1123
Total	11.9	7.6	80.5	9027

Shown in table 3 is the final interviewing status distributed according to the 5 Danish regions. It is seen that all in all 13,875 CAPI and 691 CATI interviews were performed leading to an overall response rate of 66.7%. Thus the required response rate of 70% was not attained. Furthermore it is seen that 3,000 interviews were obtained in only one region. Thus although CATI interview seemed to increase the response rate in especially region 1 and 2 the initial demands on response rate was not met.

Table 3: Final response status according to region

Region	CAPI		CATI sup		Interview		Total
	n	%	n	%	n	%	
1	2731	58.3	192	4.1	2923	62.4	4688
2	2499	58.3	319	7.4	2818	65.7	4286
3	3077	71.8	12	0.3	3089	72.1	4286
4	2776	64.8	98	2.3	2874	67.1	4286
5	2792	65.1	70	1.6	2862	66.8	4286
Total	13875	63.6	691	3.2	14566	66.7	21832

4. Comparison of groups

4.1 Comparison of interviewing status by basic characteristics

In the DHMS-2000 survey it was demonstrated that non-response was highest among men, among older persons and among unmarried persons (Kjøller 2002). In order to see how CATI interviewing influence this pattern the distribution according to sex, age and legal marital status among CAPI, CATI and non-responders is shown in table 4. As is seen, the CATI group resembles the non-responders with regard to sex (51.3% respectively 51.5% men) and legal marital status (37.6% respectively 35.5% unmarried) more than it resembles the CAPI groups (corresponding figures: 48.3% men and 28.3% unmarried). With regard to age it would seem that the CATI group is somewhat younger than both the CAPI- and non-responders.

Table 4: Distribution according to sex, age and legal marital status among CATI-, CAPI and non-responders.

%	CAPI	CATI sup	Non-resp.
Sex			
Men	48.3	51.3	50.5
Women	51.7	48.7	49.5
Age			
16-24	9.5	14.1	10.8
25-44	33.2	34.6	31.3
45-66	39.3	35.5	36.7
67-	17.9	15.8	21.1
Marital status			
Widow	7.3	7.5	9.7
Divorced	8.7	9.3	10.7
Married	55.7	45.7	44.0
Unmarried	28.3	37.6	35.6

4.2 Comparison of selected health indicators.

To further assess the impact of the CATI supplement we compared the prevalence of 3 selected health indicators among persons interviewed. These are self-perceived health, obesity and alcohol consumption. Obesity is defined based on a body mass index ($\text{weight (kg)} / (\text{height (m)}^2)$) of 30 or more. Alcohol consumption is defined as exceeding recommendations set up by the Danish National Board of Health. The results are shown in table 5 in the form of age-standardised prevalences. It is seen that CATI-respondents have a considerable worse self-perceived health than the CAPI-respondents (prevalence of good/very good health 78.0% vs. 74.7%) Likewise it is seen that there is no difference with regard to obesity but that CATI-responders has a markedly lower alcohol intake than CAPI-responders.

Table 5: Effect of CAPI and CATI interviewing on three selected health indicators.

Age-standardised prevalence	CAPI	CATI supp.	Total
<i>Self-rated health</i>			
Very good/good	79.0	74.7	78.8
Bad/very bad	5.8	6.6	5.8
<i>BMI</i>			
Obese	11.8	12.1	11.8
<i>Alcohol consump.</i>			
Abstainer	22.6	31.2	23.0
Above recommended	14.5	9.7	14.2

5. Conclusion

Including CATI interviewing in the Danish Health and Morbidity Survey 2005 was not planned from the start of the survey and was a result of trying to improve response rates in order to accomplish pre-defined requirements. Although an increase in response rate has been demonstrated it has not lead to a fulfillment of these requirements. It has been shown that CATI-interviewing has the effect of obtaining interview with groups of persons that we would expect would have more tendency to non-response (men and unmarried persons) which is a positive result. The fact that persons who accept an interview by CATI reports worse self-perceived health supports this claim but the fact that they report lower alcohol consumption does not. Thus it would seem that the CATI supplement has had a positive effect.

Litterature

Biemer P. P., Lyberg L. E. (2003): Introduction to Survey Quality. New York, Wiley

Daidsen M., Helweg-Larsen K., Kjøllet M. (2004): "DANCOS – a registry-based follow-up of Danish National Health Surveys", Ugeskrift for Læger, 166,,1441-1444 (in Danish)

Kjøllet M., Rasmussen N. K. (ed) (2002): Sundhed og sygelighed i Danmark år 2000 ... & udviklingen siden 1987. Statens Institut for Folkesundhed; Copenhagen

Rasmussen N. K., Kjøllet M. (2004): "The National Institute of Public Health program for health and morbidity surveys". Ugeskrift for Læger, 166, 1438-1641 (in Danish)