



Statistics Sweden

Statistiska centralbyrån

# Alternativa datainsamlingsmetoder i ULF, fas 2

En jämförelse mellan två olika  
datainsamlingsmetoder

2010:1

I serien Bakgrundsfakta presenteras bakgrundsmaterial till den statistik som SCB producerar inom området befolkning och välfärd. Det kan röra sig om produktbeskrivningar, metodredovisningar samt olika sammanställningar av statistik som kan ge en överblick och underlätta användandet av statistiken.

## Utgivna publikationer från 2001 i serien Bakgrundsfakta till Befolknings- och välfärdsstatistik

- 2001:1 Hälsindex
- 2001:2 The role of official statistics in a democratic society: statistics as an investment-free for users
- 2001:3 Co-operations project "Population Statistics"
- 2001:4 Uppdatering av immigrationsuppgifter med anpassning till den nya RTB-strukturen
- 2001:5 Flergenerationsregistret – The Multi-Generation Register
- 2001:6 Utredning om komplettering och utvidgning av Flergenerationsregistret
- 2001:7 Konferens om välfärdsstatistik den 27 mars 2001
- 2002:1 Komplettering av Flergenerationsregistret – resultat från provregistrering av personakter över avlidna 1947 – 1967
- 2002:2 Ett nytt system för Register över totalbefolkningen. Bättre kvalitet och nya möjligheter – A new population register system. Better quality and new possibilities.
- 2002:3 Slutrapport: Översyn av SCB:s Intäkts- och kostandsundersökning för flerbostadshus. Bilaga: Minnesanteckningar från programmet boende. Byggnad och bebyggelse om IKU den 22 maj 2001.
- 2002:4 Befolkningens boende – idé till årlig sammanställning över nationell och internationell boendestatistik
- 2002:5 Översyn av SCB:s Bostads- och Hyresundersökning (BHU)
- 2002:6 Kvalitetssäkringsprogram av postenkätundersökningar. BV/ENK:s kvalitetssäkringsprogram
- 2003:1 Longitudinellt barnregister
- 2003:2 Översyn av HBS (HUT, HBU, Utgiftsbarometern) inför 2003 års undersökning
- 2003:3 Manuell registrering eller skanning
- 2003:4 BV-avdelningens Utvecklingsplan 2003
- 2003:5 Flergenerationsregistret 2002, En beskrivning av innehåll och kvalitet
- 2003:5.1 Multi-Generation Register 2002, A description of contents and quality
- 2003:6 Barns hälsa – Förslag till utbyggnad av SCB:s årliga statistik om Barn och deras familjer
- 2003:7 Översyn av Outhyrda lägenheter i flerbostadshus 2002 - 2003-09-12
- 2003:8 Studie av bortfallet i 2000-års undersökning av levnadsförhållanden (ULF)
- 2003:9 Alternativa datainsamlingsmetoder för ULF
- 2003:10 Slutrapport från projektet för översyn av produktionssystemet för framställning av tabeller till publikationer
- 2003:11 Ensamstående och sammanboende i folkbokföringen, taxeringen och i verkligheten

Fortsättning på omslagets tredje sida!

Ovannämnda rapporter, liksom övriga SCB-publikationer, kan beställas från:  
Statistiska centralbyrån, SCB, Publikationstjänsten, 701 89 ÖREBRO,  
telefon 019-17 68 00 eller fax 019-17 64 44.

Du kan också köpa SCB:s publikationer i **Statistikbutiken**:  
Karlavägen 100, Stockholm

**Bakgrundsfakta**

# **Alternativa datainsamlingsmetoder i ULF, fas 2**

**En jämförelse mellan två olika  
datainsamlingsmetoder**

**Befolknings- och välfärdsstatistik 2010:1**

**Statistiska centralbyrån  
2010**

## Background Facts

Population and Welfare Statistics 2010:1

# Alternative modes of data collection in ULF, phase 2

## A comparison between two different modes of data collection

Statistics Sweden  
2009

---

Tidigare publicering      se omslagets insida  
*Previous publication*      listed at the inside of the cover

Producent                      SCB, avdelningen för befolkning och välfärd  
*Producer*                      Statistics Sweden, Population and Welfare Department  
701 89 Örebro

Förfrågningar                Martin Axelson +46 19 17 61 18  
*Inquiries*                      marin.axelson@scb.se

Det är tillåtet att kopiera och på annat sätt mångfaldiga innehållet i denna publikation.  
Om du citerar, var god uppge källan på följande sätt:  
Källa: SCB, Bakgrundsfakta, Befolknings- och välfärdsstatistik 2010:1, *Alternativa datainsamlingsmetoder i ULF, fas 2. En jämförelse mellan två olika datainsamlingsmetoder*

It is permitted to copy and reproduce the contents in this publication.  
When quoting, please state the source as follows:  
Source: Statistics Sweden, Background Facts, Population and Welfare Statistics 2010:1, *Alternative modes of data collection in ULF, phase 2. A comparison between two different modes of data collection*

ISSN 1654-4331 (online)

URN:NBN:SE:SCB-2010-BE96BR1001\_pdf (pdf)

Denna publikation finns enbart i elektronisk form på [www.scb.se](http://www.scb.se).

This publication is only available in electronic form on [www.scb.se](http://www.scb.se).

## Förord

I föreliggande rapport redovisas resultaten från den andra fasen av den metodstudie som genomfördes inom ramen för projektet *Alternativa datainsamlingsmetoder för Undersökningen av Levnadsförhållanden (ULF)*, som påbörjades 2002. Innehållet i rapporten utgjorde en del av det underlag som användes när SCB 2006 fattade beslut om en övergång till telefonintervju som huvudsaklig insamlingsmetod i ULF.

En preliminär version av rapporten användes som underlag för diskussioner i programrådet för social välfärdsstatistik den 2 februari 2007 och i SCB:s Vetenskapliga råd den 8 mars 2007. Den version som här presenteras är till stor del identisk med den preliminära versionen. Avsnitt 3 har dock genomgått en relativt stor förändring, till följd av att kommentarer som framkom vid diskussionen i Vetenskapliga rådet beaktats. Även delar av avsnitt 8 har reviderats, delvis som en direkt följd av omarbetningen av avsnitt 3. För att möjliggöra en jämförelse, har avsnitten 3 och 8 från den preliminära versionen bifogats som bilagor.

Den preliminära rapporten utarbetades av Martin Axelson, Hassan Mirza och Michael Carlson, samtliga då verksamma vid BV/MET. Martin Axelson, U/ARK, har ansvarat för omarbetandet av den preliminära versionen till den version som härmed avrapporteras. Karin Andersson, PCA/MIH, har bidragit med nytt material till avsnitt 3.

Statistiska centralbyrån i januari 2010

Stina Andersson

Anders Sundström



## Innehåll

Förord .....	3
<b>Sammanfattning.....</b>	<b>7</b>
<b>1 Inledning.....</b>	<b>9</b>
1.1 Kort om ULF.....	9
1.2 Bakgrund till metodstudien .....	9
1.3 Metodstudien, fas 0 och 1 .....	10
<b>2 Metodstudien, fas 2 – en överblick.....</b>	<b>13</b>
2.1 Design av ordinarie ULF 2006.....	13
2.2 Design av återintervjustudien .....	13
2.3 Beslutskriterium för övergång till telefonintervju.....	14
<b>3 Reliabilitetsstudie.....</b>	<b>17</b>
3.1 Metod.....	17
3.2 Resultat och slutsatser .....	20
<b>4 Mätfelsmodeller (Latent Class Analysis).....</b>	<b>23</b>
4.1 Metod.....	23
4.1.1 Teori och modeller .....	23
4.1.2 Identifierbarhet och grupperingsvariabler .....	25
4.1.3 Modellvalskriterier.....	26
4.2 Urval och dikotomisering av frågor .....	28
4.3 Resultat.....	28
<b>5 En jämförelse mellan MM och CATI med avseende på bortfallsskevhetens storlek.....</b>	<b>31</b>
5.1 Upplägg.....	31
5.1.1 Teoretisk ansats .....	31
5.1.2 Praktiska aspekter .....	32
5.2 Resultat.....	33
<b>6 Barn-ULF i MM och CATI – en bortfallsjämförelse.....</b>	<b>35</b>
<b>7 Motivationen hos uppgiftslämnarna .....</b>	<b>37</b>
<b>8 Slutsatser.....</b>	<b>39</b>
<b>9 Källor.....</b>	<b>41</b>
Tryckta källor.....	41
Elektroniska källor .....	41
<b>Bilagor .....</b>	<b>43</b>
Bilaga A. Resultat reliabilitetsstudie .....	44
Bilaga B. Resultat mätfelsmodeller / LCA.....	61
Bilaga C. Resultat från studien rörande differenser i bortfallsskevhet.....	76
Bilaga D. Resultat motivationsfrågor .....	86
Bilaga E: Avsnitt 3, preliminära versionen .....	88
Bilaga F: Avsnitt 8, preliminära versionen .....	91
Bilaga G. Återintervjublanketten för MM.....	93





## Sammanfattning

I föreliggande rapport redovisas resultaten från den andra, och avslutande, fasen av den metodstudie som genomförts inom ramen för projektet *Alternativa datainsamlingsmetoder för Undersökning av Levnadsförhållanden (ULF)*, som påbörjades 2002.

Metodstudiens huvudsakliga syfte var att jämföra två olika metoder för datainsamling: en mixed mode ansats (MM) med en kombination av insamlingsmetoderna besök- och telefonintervju utan datorstöd respektive datorstödd telefonintervju (CATI). Jämförelser avseende huvudsakligen fyra olika kvalitetsaspekter redovisas: (1) mätkvaliteten, (2) bortfallsfelets storlek och inverkan på skattningar, (3) svarsandelen i Barn-ULF och (4) viljan hos uppgiftslämnarna för att delta i undersökningen.

Den allmänna slutsatsen från studien är att den systematiska felkomponenten (mät- och bortfallsfel) i tillförlitlighetskomponenten bedöms vara oförändrad vid en övergång till CATI. I kombination med att övergången skulle frigöra resurser för en urvalsökning, innebär detta att medelkvadratfelet (MSE) i skattningarna skulle minska vid en övergång från den tidigare ansatsen med MM till CATI som primär insamlingsmetod.



# 1 Inledning

I föreliggande rapport redovisas resultaten från den andra, och avslutande, fasen av den metodstudie som genomförts inom ramen för projektet *Alternativa datainsamlingsmetoder för ULF* som påbörjades 2002. Rapporten består, utöver sammanfattningen, av åtta avsnitt.

Avsnitt 1 innehåller, utöver en beskrivning av ULF, en bakgrund till den metodstudie som genomförts samt en sammanfattning av resultaten från metodstudiens första fas. I avsnitt 2 redovisas på ett övergripande plan upplägget av metodstudiens andra fas, medan avsnitten 3-7 innehåller mer detaljerade information, inklusive resultat och slutsatser, om var och en av de fem delstudier som tillsammans utgör den andra fasen. Rapporten avslutas med en kort diskussion om den sammanfattande slutsats som dras med utgångspunkt i resultaten från de fem delstudierna.

## 1.1 Kort om ULF

ULF, som har genomförts årligen sedan 1975, används för att belysa olika aspekter av den svenska välfärden framför allt genom lägesbeskrivningar och sambandsanalyser samt utvecklingstendenser. I allt väsentligt gäller att den undersökningsdesign som i dag ligger till grund för genomförandet av ULF fastslogs redan under första halvan av 1980-talet.

Utgångspunkten för den årliga datainsamlingen är ett slumpmässigt urval från Sveriges vuxna befolkning (16 – 84 år, ingen övre gräns fr.o.m. 2002). Urvalet utgörs av en paneldel, som återkommer vart åttonde år, och en tvärsnittsdel. Undersökningen innehåller en stor mängd undersökningsvariabler. Dessa består dels av ett antal årligen återkommande variabler som representerar alla de olika välfärdsområden som ULF omspannar, dels ett antal variabler av fördjupningskaraktär inom ett eller ett par områden. En stor del av variablerna i den sistnämnda kategorin följer samma rotationsmönster som paneldelen i urvalet, vilket gör det möjligt att långsiktigt följa utvecklingen område för område över tid. Undersökningsdesignen utgår från att datainsamlingen genomförs i form av besöksintervjuer utan datorstöd (PAPI), men i praktiken sker datainsamlingen med en kombination av insamlingsmetoderna besök- och telefonintervju utan datorstöd (PAPI respektive PATI). I den fortsatta framställningen kommer denna kombinerade metod att benämnas mixed mode (MM). Under 2003 genomfördes exempelvis drygt 24 procent av alla genomförda intervjuer med PATI som insamlingsmetod.

## 1.2 Bakgrund till metodstudien

Behoven av statistik inom välfärdspolitikens verksamhetsområden har den senaste tioårsperioden ökat påtagligt. Under senare år har efterfrågan av underlagsdata för indikatorer varit ett särskilt efterfrågat område. Dagens datakällor konstateras ofta otillräckliga för pågående indikatoraktiviteter. Indikator- och uppföljningssystem måste, på ett långsiktigt hållbart sätt, försörjas med regelbundet återkommande, kvalitetssäkrad statistik.

ULF utgör en viktig källa för kunskap om den svenska välfärden och användningen av statistik och mikrodata från undersökningen är mycket

omfattande. Trots att efterfrågan på statistik har ökat har av olika skäl resurserna för datainsamlingsdelen av ULF minskat. Det har tvingat fram hårda prioriteringar samt även urvalsneddragningar som minskat möjligheterna till många av de fördjupade analyser som nu efterlyses.

SCB:s programråd för social välfärdsstatistik tog för några år sedan initiativ till en analys av vad som skulle krävas för att ULF bättre skulle tillmötesgå de ökade krav på statistik som kommer från olika håll. En viktig slutsats var att urvalet snarast borde utökas och bl.a. framfördes ett förslag som, med bibehållet upplägg av undersökningen, förutsatte ett årligt ekonomiskt tillskott med ca 13 miljoner kronor<sup>1</sup>. Samtidigt restes frågan om i vilken utsträckning datainsamlingen skulle kunna rationaliseras, t.ex. genom att övergå till datorstödd telefonintervju (CATI) som huvudsaklig datainsamlingsmetod, för att göra det möjligt att redan inom oförändrad budgetram kunna genomföra en urvalsökning. I den efterföljande diskussionen framkom ett antal viktiga synpunkter och farhågor:

- 1) Eftersom undersökning har ett så stort omfång, är det över huvud taget inte möjligt att samla in ULF-data via telefonintervjuer.
- 2) Den rationaliseringspotential som kan förväntas uppstå vid en övergång till exempelvis telefonintervju som huvudsaklig insamlingsmetod är inte tillräckligt stor. En utredning av telefonintervju som huvudsaklig datainsamlingsmetod blir därmed inte ett lönsamt projekt, eftersom återbetalningstiden för investeringen blir alldeles för lång.
- 3) En övergång till en annan datainsamlingsmetod kan äventyra kvaliteten i ULF, eftersom produktens nuvarande mätinstrument är utvecklat, och därmed även anpassat, till besöksintervjuer.
- 4) En förändring av datainsamlingsmetoden kan leda till brott i tidsserierna.

I syfte att närmare utreda frågan om användandet av alternativa datainsamlingsmetoder i ULF påbörjades ett metodprojekt vid årsskiftet 2002/2003.

### 1.3 Metodstudien, fas 0 och 1

Metodstudien inleddes med en litteraturgenomgång och en analys av historiska ULF-data<sup>2</sup>. Utredningen fann följande omständigheter som talade för att en effektivare undersökningsdesign kan uppnås:

- Telefonintervjuer och enkätundersökningar är betydligt billigare än besöksintervjuer.
- ULF är redan idag en mixed mode-undersökning. Mätkvaliteten i telefonintervjuerna och dess effekt på slutprodukten är oklara. De höga vägrar- och kompletteringsandelar som observerats kan vara indikationer på brister i nuvarande metod.

---

<sup>1</sup> Avser 2002 års prisnivå

<sup>2</sup> Fas 0, se SCB (2003).

- Skillnaderna mellan de tre olika datainsamlingsmetoderna (enkät, telefon och besök) med avseende på mätkvaliteten tycks inte vara dramatiska. Förmodligen kan man förhindra uppkomsten av allvarliga brister genom att i frågeutformningen och designen i övrigt ta hänsyn till de speciella förhållanden som de olika datainsamlingsmetoderna ger.

Mot bakgrund av dessa resultat beslutade SCB, bl.a. efter diskussioner i programrådet för social välfärdsstatistik, att gå vidare med en metodstudie i två faser, innefattande ett antal prov och kontrollerade experiment. Syftet med metodstudien skulle vara att insamla information för att, genom en allsidig belysning av de fyra farhågorna i avsnitt 1.2, slutligen kunna fatta ett välgrundat beslut rörande valet av framtida datainsamlingsmetod.

För att CATI skall kunna betraktas som en möjlig datainsamlingsmetod för ULF, krävs att den efterfrågade datamängden kan samlas in. Dessutom måste kostnadsrelationen mellan MM och CATI vara till den senares fördel, dvs. kostnaden för att använda CATI måste signifikant understiga kostnaden för MM. Provet i metodstudiens första fas, som genomfördes under hösten 2004, syftade till att belysa dessa två aspekter. För ett urval på drygt 800 personer i åldern 16-84 år användes CATI som huvudsaklig insamlingsmetod. Resultaten från detta prov visade att:

- Andelen svarande i CATI var i paritet med andelen svarande i MM.
- Den uppmätta kostnaden för CATI indikerade att en övergång till telefonintervju som huvudsaklig insamlingsmetod förväntas medföra en kostnadsbesparing som motsvarar en ökning av ULF-urvalet med 60-70 procentenheter.

Resultaten från fas 1 har bl.a. behandlats i programrådet för social välfärdsstatistik och vid ett möte med SCB:s Vetenskapliga råd. På basis av resultaten från fas 1 samt den efterföljande diskussionen i och rekommendationerna från rådet beslutades att gå vidare med fas 2.



## 2 Metodstudien, fas 2 – en överblick

### 2.1 Design av ordinarie ULF 2006

En viktig utgångspunkt vid planeringen av fas 2 var att designen skulle vara sådan att den i efterhand möjliggjorde jämförelser av nivåskattningar baserade på data insamlade via MM respektive CATI på relativt detaljerad redovisningsgruppsnivå. Vidare skulle designen vara sådan att det i efterhand fanns underlag för

- en jämförelse mellan MM och CATI med avseende på mätkvaliteten.
- en jämförelse mellan MM och CATI med avseende på bortfallsfelets storlek
- en jämförelse mellan MM och CATI med avseende på svarsandelen i Barn-ULF
- en studie av uppgiftslämnarnas motivation att delta i en undersökning av ULF:s karaktär

Mot bakgrund av ovanstående, beslutades att inom ramen för ULF 2006 genomföra ett s.k. inbäddat experiment. Det ordinarie ULF-urvalet i åldersgruppen 16-84 år om ca 7300 personer delades slumpmässigt i två delar. Åldersgruppen 85- ingick inte i experimentet. För den första halvan användes MM som huvudsaklig metod för datainsamling, medan CATI användes för den andra halvan. Detta upplägg säkerställde tillgången till information som möjliggör ett (visst) användande av ULF:s tidsserier även efter en eventuell övergång till CATI som huvudsaklig insamlingsmetod. Dessutom säkerställde upplägget att önskemålen i den andra respektive fjärde strecksatsen kan tillgodoses inom ramen för ordinarie ULF. För att samtidigt tillgodose önskemålen i den första och den tredje strecksatsen ovan, beslutades vidare att genomföra en återintervjustudie för en slumpmässigt utvald delmängd av respondenterna i ordinarie ULF.

### 2.2 Design av återintervjustudien

Utgångspunkten vid valet av experimentdesign var att dataunderlaget skulle kunna användas till två, metodmässigt olika, studier av mätkvaliteten: dels en "klassisk" reliabilitetsstudie, dels en modellbaserad studie baserad på teori om analys av latent klasser. (Se avsnitt 3 och 4 för mer information om dessa studier och resultaten från dem.)

Mot bakgrund av detta beslutades att varje respondent vid återintervjun skulle undersökas med hjälp av samma datainsamlingsmetod – PAPI, PATI eller CATI – som användes vid genomförandet av datainsamlingen i ordinarie ULF. Vidare beslutades att återintervjun, om möjligt, skulle genomföras 14 – 28 dagar efter den ordinarie ULF-intervjun. Denna tidsperiod bedömdes vara tillräckligt lång för att reducera risken för starka minnes effekter, samtidigt som den bedömdes vara tillräckligt kort för att frågorna i återintervjun i allt väsentligt skulle kunna sägas avse samma referensperiod som frågorna i ordinarie ULF. Av flera skäl (t.ex. hänsyn till respondenterna, kostnadsskäl) ingick i återintervjustudien endast en delmängd av de frågor som ställs i ordinarie ULF. Frågorna som utvaldes att ingå i

experimentet bedömdes vara av centralt intresse i ULF. Urvalet av frågor skedde i nära samråd med komponentansvariga.

Återintervjublanketten återges i sin helhet i Bilaga E. Av blanketten framgår också att speciella kontrollfrågor infogats för att ta reda på om vissa förhållanden hade förändrats under tiden mellan de två intervjutillfällena. Totalt omfattade återintervjun 36 frågor inklusive kontrollfrågor (delfrågor ej räknade).

Storleken av suburvalet planerades utifrån ett önskemål om ca 450 genomförda återintervjuer per datainsamlingsmetod (MM resp. CATI). I samband med fältplaneringen för ordinarie ULF fastställdes, utifrån antaganden om svarsfrekvensen i såväl ordinarie ULF som återintervjustudien, vilka urvalspersoner som skulle återintervjuas. För att underlätta den praktiska hanteringen, samordnades urvalen till återintervjustudien med specifika undersökningsomgångar inom MM- respektive CATI-delen av ordinarie ULF.

*Kommentar:* För att ytterligare reducera risken för minneseffekter, särskilt för respondenter för vilka data insamlades via PAPI i ordinarie ULF, beslutades att ingen respondent skulle riskera att återintervjuas av samma intervjuare som genomförde intervjun i ordinarie ULF. Redan på planeringsstadiet förutsågs att detta villkor, för MM-delen av ULF-urvalet, skulle förorsaka problem vid återintervju av personer bosatta inom Gotlands, Jämtlands, Västerbottens eller Norrbottens län. Därför designades återintervjustudien, oavsett datainsamlingsmetod, till att gälla personer bosatta i Sverige exklusive personer bosatta inom Gotlands, Jämtlands, Västerbottens eller Norrbottens län.

*Kommentar:* Av försiktighetsskäl beslutades att Barn-ULF och återintervjustudien skulle samordnas negativt inom CATI-delen av ordinarie ULF. Praktiskt innebar detta att Barn-ULF ej genomfördes inom ramen för ordinarie ULF för de undersökningsomgångar inom CATI-delen som ingick i återintervjustudien. För dessa undersökningsomgångar drogs i stället ett tilläggsurval, negativt samordnat med ordinarie ULF-urvalet, för vilket Barn-ULF genomfördes.

### **2.3 Besluts Kriterium för övergång till telefonintervju**

Utgångspunkten för de diskussioner som förts gällande upplägget av de genomförda experimenten har varit kvaliteten för slutprodukten. Tidigt fastslogs att ett eventuellt beslut om en övergång från nuvarande datainsamlingsmetod, MM, till CATI som huvudsaklig insamlingsmetod, skall baseras på en jämförelse av undersökningens kvalitet under de båda metoderna.

Med kvalitet för en statistikprodukt (statistik) avses alla de egenskaper hos statistiken som har betydelse för hur väl denna tillfredsställer användarnas informationsbehov. SCB:s kvalitetsbegrepp för officiell statistik anger följande egenskaper:

- 1) Statistikens innehåll
- 2) Statistikens tillförlitlighet
- 3) Statistikens aktualitet



- 4) Jämförbarhet och sammanvändbarhet, som avser möjligheten
  - a) att göra jämförelser mellan grupper
  - b) att göra jämförelser över tiden
  - c) att använda statistiken tillsammans med annan statistik
- 5) Statistikens tillgänglighet och förståelighet

En väl planerad och genomförd undersökning kan sägas avspegla de användarkrav och den kostnadsfunktion som gällde vid den tidpunkt då undersökningsdesignen fastställdes. Med förändringar i samhället och kunskapsstillväxt följer dock ökade och/eller ändrade användarkrav. Den relativa kostnadsrelationen mellan olika produktions- och framställningsmetoder förändras också över tiden. Dessa omständigheter leder till krav på förbättringar och förändringar vad gäller statistikens upplägg och innehåll. Här kan en målkonflikt urskiljas. Förändringar i en undersökningsdesign som syftar till förbättringar i kvalitetsdimensionerna innehåll, aktualitet, och tillförlighet kan samtidigt försämra möjligheten till jämförelser över tiden. Det föreligger således en viss motsättning mellan komponenten jämförbarhet över tiden och de övriga kvalitetskomponenterna som ingår i punkterna 1-4 ovan. För att mildra denna motsättning strävar statistikproducenter vid förbättrings- och omläggningsarbete efter att mäta effekterna av de förändringar som genomförs.

För den genomförda metodstudien fastställdes följande beslutskriterium:

”I samband med översynsarbete av en undersökning väljs det upplägg som ger den högsta möjliga kvaliteten till en given kostnad. I begreppet kvalitet ingår egenskaperna 1 – 4 exklusive 4.b (jämförbarhet över tiden). Om valet innebär en omläggning av undersökningen kan eventuella tidsseriebrott hanteras och avhjälpas genom att framställa produkten på två olika sätt, ny och gammal metod, vid övergången.”

Att frågan om tidsseriebrott exkluderats från beslutskriteriet innebär inte att jämförbarhet över tiden är en oviktig kvalitetsaspekt. Utgångspunkten är dock att detta är ett problem man i viss utsträckning tvingas räkna med, och acceptera, för att kunna uppnå andra viktiga kvalitetsvinster. För att mildra effekterna för användarna är det dock viktigt att stora förändringar genomförs under kontrollerade former, samt att användarna ges utförlig information om effekterna av de genomförda förändringarna.



## 3 Reliabilitetsstudie

### 3.1 Metod

Den klassiska mätmodellen brukar formuleras som  $Y = T + E$ , där  $Y$  är ett observerat (manifest) värde,  $E$  ett slumpfel och  $T$  ett hypotetiskt sant (latent) värde. Under antagandet om att mätfelet och det sanna värdet är oberoende stokastiska variabler följer att  $Y$  är en stokastisk variabel med varians lika med summan av varianserna hos  $T$  och  $E$ . Reliabiliteten  $R$  i mätningen av  $T$  via  $Y$  definieras som kvoten mellan dessas varianser enligt

$$R = \frac{\text{Var}(T)}{\text{Var}(Y)} = \frac{\text{Var}(T)}{\text{Var}(T) + \text{Var}(E)}$$

Det följer av antagandet och definitionen att  $0 \leq R \leq 1$ . Om effekten av mätfelet är ringa (liten relativ varians i  $E$ ) ligger reliabiliteten nära 1, om den är betydande (stor relativ varians i  $E$ ) ligger det nära noll. Följande antaganden kan användas för att illustrera en situation med upprepade mätningar:

- Varje individ i populationen har en hypotetisk fördelning av möjliga responser.
- Ett obundet slumpmässigt urval av storlek  $n$  individer väljs ur en population av storlek  $N$ .
- Varje individ betraktas som ett kluster av potentiella responser som observeras (eller dras med återläggning) oberoende av varandra.

En undersökning där det föreligger mätfel kan med andra ord ses som ett två-steps klusterurval med OSU i varje steg (utan återläggning i första, med återläggning i andra). Låt  $m$  beteckna antalet mätningar som görs hos varje individ. Med en intervju-återintervju-design är  $m = 2$  och man observerar två slumpmässiga och oberoende utfall från varje individs hypotetiska fördelning av möjliga responser. Dessutom antas oberoende mellan olika individers respektive responser. Antag att man är intresserad av populationsparametern

$$\bar{Y} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M y_{ij}$$

Då det är fråga om urval med återläggning i andra steget bör  $M$  egentligen betraktas som oändligt stort, varför man kan betrakta parametern ovan som genomsnittet av varje individs väntevärde över respektive fördelning av responser. Populationsgenomsnittet skattas med

$$\bar{y} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m y_{ij}$$

med varians

$$\text{Var}(\bar{y}) = (1-f) \frac{S_1^2}{n} + \frac{S_2^2}{nm} = (1-f) \frac{SV}{n} + \frac{SRV}{nm}$$

där  $f$  betecknar urvalsfraktionen i första steget och

$$S_1^2 = SV, \text{ samplingvarians, "between primary sampling units"}$$

$$S_2^2 = SRV, \text{ responsvarians, "within primary sampling units"}$$

I fallet med endast en mätning per individ ( $m = 1$ ) och en försumbar urvalsfraktion ( $f \approx 0$ ) gäller att

$$\text{Var}(\bar{y}) = \frac{SV + SRV}{n}$$

Om variansen för de enskilda responserna  $SRV$  är liten blir den totala variansen i skattningen liten. Reliabilitetskvoten (*reliability ratio*) definieras alltså analogt med den inledande beskrivningen enligt

$$R = \frac{SV}{SV + SRV}$$

och den omvända s.k. inkonsistenskvoten (*ratio of inconsistency*) enligt

$$I = 1 - R$$

Inkonsistensen (alternativt reliabiliteten) i en variabel säger oss med andra ord något om hur mycket av den totala variansen (osäkerheten) i en skattning som kan hänföras till mätfelet. Ju lägre reliabilitet, desto större osäkerhet.

För en dikotom responsvariabel i en intervju-återintervju-studie gäller givetvis samma resultat som ovan, men med något enklare formler i slutändan. Antag att vi har följande fyrfältstabell med observationer:

		Återintervju, $y_{i2}$		
		0	1	
Intervju, $y_{i1}$	0	$a$	$b$	
	1	$c$	$d$	
				$n = a+b+c+d$

där cellerna representerar frekvenser av respektive kombination av responserna 0 och 1. Definiera  $g$  som andelen respondenter som svarar olika från en gång till nästa (*gross different rate*), dvs.

$$g = \frac{\sum_{i=1}^n (y_{i1} - y_{i2})^2}{n} = \frac{b+c}{n}$$

Definiera vidare

$$p_1 = \frac{1}{n} \sum_{i=1}^n y_{i1} = \frac{c+d}{n} \quad \text{och} \quad p_2 = \frac{1}{n} \sum_{i=1}^n y_{i2} = \frac{b+d}{n}$$

dvs. andel ettor från respektive intervjutillfälle 1 och 2, samt låt

$$q_1 = 1 - p_1 = \frac{a+b}{n} \quad \text{och} \quad q_2 = 1 - p_2 = \frac{a+c}{n}$$

Då kan det relativt enkelt visas att  $SRV$  och  $SV+SRV$  kan skattas med

$$srv = \frac{g}{2} = \frac{b+c}{2n}$$

respektive

$$sv + srv = \frac{p_1(1-p_2) + p_2(1-p_1)}{2} = \frac{(c+d)(a+c) + (b+d)(a+b)}{2n^2}$$

varför inkonsistensen  $I$  och reliabiliteten  $R$  skattas med

$$\hat{I} = \frac{g}{p_1q_2 + p_2q_1} = \frac{n(b+c)}{(c+d)(a+c) + (b+d)(a+b)}$$

respektive

$$\hat{R} = 1 - \hat{I}$$

Den förra,  $\hat{I}$ , brukar benämnas inkonsistensindex (*index of inconsistency*) och är ett mått på hur stor andel av den totala variansen som kan hänföras till mätfelet, medan  $\hat{R}$  mäter resterande andel av variansen.

För en fyrfältstabell kan man lätt visa att  $\hat{R}$  är identisk med Cohens kappas, ett mått på överensstämmelse (*interrater agreement*) som introducerades i början 1960-talet. Således kan den estimator som används för att skatta den asymptotiska variansen för Cohens kappas även användas för att skatta den asymptotiska variansen för såväl  $\hat{I}$  som  $\hat{R}$ . Denna varians estimator ges av

$$avar(\hat{R}) = avar(\hat{I}) = \frac{1}{n} \frac{B+C-D}{(1-A)^2}$$

där

$$A = \frac{(a+b)(a+c) + (c+d)(b+d)}{n^2}$$

$$B = \frac{a}{n} \left[ 1 - \frac{(2a+b+c)\hat{I}}{n} \right]^2 + \frac{d}{n} \left[ 1 - \frac{(b+c+2d)\hat{I}}{n} \right]^2$$

$$C = \hat{I}^2 \left[ \frac{b(a+2c+d)^2 + c(a+2b+d)^2}{n^3} \right]$$

och

$$D = (1 - \hat{I} - A\hat{I})^2$$

För att testa om inkonsistensen är densamma i  $q$  strata kan följande teststatistika användas:

$$Q = \sum_{h=1}^q \frac{(\hat{I}_h - \hat{I}_{overall})^2}{avar(\hat{I}_h)}$$

där  $\hat{I}_{overall} = \sum_{h=1}^q w_h \hat{I}_h$ , med

$$w_h = \frac{1/\text{avar}(\hat{I}_h)}{\sum_{h'=1}^q 1/\text{avar}(\hat{I}_{h'})}$$

Under nollhypotesen, dvs. om inkonsistensen är densamma i samtliga  $q$  strata, har  $Q$  en asymptotisk  $\chi^2$ -fördelning med  $q-1$  frihetsgrader. För mer information om Cohens kappas, samt referenser för de ovan använda resultaten, se avsnittet "The FREQ Procedure: Tests and Measures of Agreement" i online-dokumentationen för SAS 9.2. (För webbadress, se referenslistan.).

### 3.2 Resultat och slutsatser

Från återintervjublanketten valdes 17 frågor ut. Dessa dikotomiserades enligt beskrivningarna i Bilaga A varefter  $g$  (*gross difference rate*) och inkonsistensindex  $\hat{I}$  för respektive mode (MM respektive CATI) beräknades. Resultaten redovisas i Bilaga A och sammanfattas i Tabell 3.1 nedan.

**Tabell 3.1 Gross difference rate ( $g$ ) i %-enheter och inkonsistensindex ( $\hat{I}$ ) för ett urval av frågor i ULF för MM och CATI**

Fråga	MM			CATI			$p$
	$g$	$\hat{I}$	$se(\hat{I})$	$g$	$\hat{I}$	$se(\hat{I})$	
Frg004d	2.88	0.060	0.017	1.47	0.030	0.011	0.143
Frg005	1.21	0.030	0.013	2.53	0.061	0.017	0.152
Frg057	11.61	0.280	0.037	9.79	0.229	0.032	0.293
Frg071	2.81	0.493	0.129	3.08	0.349	0.087	0.358
Nyfrg072	10.20	0.205	0.031	6.61	0.135	0.024	0.072
Frg100e	2.99	0.151	0.049	2.33	0.052	0.015	0.054
Frg122a	10.19	0.218	0.040	12.09	0.272	0.041	0.343
Frg122b	15.29	0.305	0.077	18.29	0.363	0.083	0.611
Frg122c	32.94	0.665	0.102	18.29	0.366	0.085	0.025
Frg123c	21.59	0.449	0.052	21.57	0.444	0.047	0.941
Frg124a	10.57	0.218	0.039	7.84	0.164	0.032	0.281
Frg221b	1.43	0.133	0.053	2.08	0.211	0.065	0.350
Frg224a	14.69	0.360	0.041	9.79	0.234	0.032	0.016
Frg250a	0.24	(1)	(1)	0.21	(1)	(1)	(1)
Frg251a	2.61	0.490	0.128	1.67	0.275	0.093	0.175
Frg252a	2.38	0.429	0.122	2.29	0.326	0.092	0.502
Frg253a	5.46	0.406	0.077	2.71	0.260	0.069	0.156

(1) För få observationer av den ena kategorin för att tillförlitligt kunna beräknas

se= asymptotiskt medelfel

P= p-värdet i test för lika kappakoefficient

För flertalet av frågorna kan ingen signifikant skillnad mellan CATI och MM påvisas. Vid test på 5 % signifikansnivå erhålls signifikans för Frg122c och Frg224a, för vilka CATI resulterar i ett signifikant lägre inkonsistensindex än MM. Utökas signifikansnivån till 10 %, erhålls signifikans även för frågorna Nyfrg072 och Frg100. Även för dessa frågor gäller att CATI resulterar i ett signifikant lägre inkonsistensindex än MM. Frg250a låter sig inte jämföras då data inte innehåller tillräckligt med observationer i kategorin  $Y = 0$  för att teststatistikan tillförlitligt ska kunna beräknas.

MM uppvisar en större variation i inkonsistens, som lägst 0.03 och som högst 0.66. Motsvarande tal för CATI är 0.03 respektive 0.44. Allmänt måste ett inkonsistensindex som är så högt som 0.66 betraktas som alarmerande; det innebär att ca 66 % av variansen i en skattning kan hänföras till mätfelet.





## 4 Mätfelsmodeller (Latent Class Analysis)

### 4.1 Metod

Idén bakom *Latent Class Analysis* (LCA), dvs. analys av latent klasser, sammanfaller i stort med teorin som beskrevs i förra kapitlet om reliabilitet. Man antar att det bakom varje observerad variabel  $Y$  finns ett hypotetiskt sant eller latent värde  $T$  som på grund av mätfel inte alltid sammanfaller med  $Y$ . LCA är en metod för att modellera mätfelsstrukturen när både  $Y$  och  $T$  är kategoriska variabler. Allmänt kan LCA hantera flera kategorier per variabel men i föreliggande studie har samtliga undersökta variabler antingen varit dikotoma från början eller dikotomiserats i efterhand (se avsnitt 4.2). Därför kommer beskrivningen i det följande att koncentreras helt till dikotoma variabler. Detta avsnitt inleds med en kort beskrivning av LCA och hur mätfel i kategoriska variabler kan modelleras. Därefter följer en kortfattad beskrivning av skattningen av modellens parametrar samt de kriterier och mått för modellval som användes i studien.

- Biemer (2001) utgjorde en viktig källa för såväl upplägget som genomförandet av LCA-studien. För en utmärkt beskrivning av teorin för LCA med empiriska exempel, se t.ex. Magidson och Vermunt (2004). I studien har programvaran Panmark (2000) genomgående använts för parameterskattningarna.

#### 4.1.1 Teori och modeller

Låt  $\pi$  beteckna andelen i en population med en viss egenskap. Analogt med förra kapitlet betecknas detta med  $T = 1$  och motsatsen med  $T = 0$ . Enligt samma resonemang som i förra avsnittet kommer, på grund av mätfel,  $Y$  att observeras istället för  $T$ . De möjliga utfallen kan sammanfattas i följande fyrfältstabell, med en parametrering av sannolikheterna för felklassificering enligt

	Observerat, $Y$		
Sant, $T$	0	1	Total
0	$1 - \phi$	$\phi$	1
1	$\theta$	$1 - \theta$	1

där parametrarna  $\phi$  och  $\theta$  är definierade som *betingade* sannolikheter enligt

$$1 - \phi = \Pr(Y = 0 \mid T = 0), \text{ specificitet}$$

$$\phi = \Pr(Y = 0 \mid T = 1), \text{ falsk negativ respons}$$

$$\theta = \Pr(Y = 1 \mid T = 0), \text{ falsk positiv respons}$$

$$1 - \theta = \Pr(Y = 1 \mid T = 1), \text{ sensitivitet}$$

Då  $\pi$  anger sannolikheten att en slumpmässigt vald individ har egenskapen i fråga kan de *obetingade* sannolikheterna att observera respektive en nolla eller en etta beräknas enligt

$$\Pr(Y = 0) = \theta\pi + (1 - \phi)(1 - \pi)$$

respektive

$$\Pr(Y_i = 1) = (1 - \theta)\pi + \phi(1 - \pi)$$

I situationer där en jämförande kontrollmätning kan göras utan mätfel (dvs. mer eller mindre felfri relativt den ursprungliga mätningen  $Y$ , som kanske genomförs med en billigare mätmetod) kan parametrarna  $\phi$  och  $\theta$  skattas direkt. I situationer med upprepade mätningar med mer eller mindre identiska mätsituationer krävs däremot en mer komplex ansats. Beteckna sannolikheterna att observera olika kombinationer av 0 och 1 i en design med intervju-återintervju enligt

Intervju, $Y_1$	Återintervju, $Y_2$	
	0	1
0	$P_{00}$	$P_{01}$
1	$P_{10}$	$P_{11}$

Antag att svaren vid respektive intervjutillfälle är oberoende betingat på det latenta värdet är  $T = t$ , dvs. antag att mätfelen i de två mätningarna lokalt oberoende. Antag vidare att parametrarna  $\pi$ ,  $\phi$  och  $\theta$  är konstanta över mätningarna. Då kan sannolikheterna  $P_{kl}$  i tabellen ovan formuleras i termer av parametrarna enligt

$$P_{00} = \sum_{t=0}^1 P(Y_1 = Y_2 = 0 | T = t)P(T = t) = \theta^2\pi + (1 - \phi)^2(1 - \pi)$$

$$P_{01} = \sum_{t=0}^1 P(Y_1 = 0, Y_2 = 1 | T = t)P(T = t) = \theta(1 - \theta)\pi + \phi(1 - \phi)(1 - \pi)$$

$$P_{10} = \sum_{t=0}^1 P(Y_1 = 1, Y_2 = 0 | T = t)P(T = t) = \theta(1 - \theta)\pi + \phi(1 - \phi)(1 - \pi)$$

och

$$P_{11} = \sum_{t=0}^1 P(Y_1 = 1, Y_2 = 1 | T = t)P(T = t) = (1 - \theta)^2\pi + \phi^2(1 - \pi)$$

Observera att antagandena medför att  $P_{01} = P_{10}$ . Notera även att antagandena om lokalt oberoende och konstanta parametrar i praktiken kan vara svåra att validera och att de i vissa fall är orimliga, beroende t.ex. på hur lång tid som har förflutit mellan mättillfällena.

Med ledning av parametriseringen ovan möjliggörs vidare beräkningar (i praktiken skattningar) av t.ex. varianskomponenterna  $SV$  och  $SRV$ , reliabilitetskvoten  $R$  och skevhet. I föreliggande studie har dock huvudsyftet varit att studera om sannolikheterna för felklassificering, dvs.  $\phi$  och  $\theta$ , är signifikant olika mellan MM och CATI.

#### 4.1.2 Identifierbarhet och grupperingsvariabler

I en intervju-återintervju-studie erhålls en  $2 \times 2$ -kontingenstabell med frekvenserna för de olika kombinationerna av 0 och 1 enligt exemplet nedan

	Återintervju		
Intervju	0	1	Total
0	$a$	$b$	
1	$c$	$d$	
Total			$n=a+b+c+d$

Betingat på det totala antalet observationer  $n$  är cellfrekvenserna  $a$ ,  $b$ ,  $c$  och  $d$  multinomialfördelade med sannolikheter  $P_{00}$ ,  $P_{01}$ ,  $P_{10}$  och  $P_{11}$  definierade enligt föregående avsnitt. Med tre parametrar och endast två frihetsgrader är dock inte modellen identifierbar (estimerbar). För att komma runt detta problem kan ytterligare restriktioner införas, t.ex. genom att anta att  $\phi = \theta$ ; detta medför dock i praktiken ofta orealistiska antaganden. Alternativt kan antalet frihetsgrader ökas genom att införa grupperingsvariabler,  $G_1$ ,  $G_2$  osv. i kombination med mer realistiska (eller åtminstone mer intressanta) parameterrestriktioner.

I föreliggande studie användes tre bivariata grupperingsvariabler:  $M$  = mode (MM eller CATI),  $K$  = kön (man eller kvinna) och  $\hat{A}$  = ålder (dikotomiserad enligt  $< 50$  eller  $\geq 50$ ). Den enkla kontingenstabellen delas således upp i åtta stycken deltabeller. Nedan ges ett exempel på hur data struktureras med denna gruppering.

**Tabell 4.1 Exempel på datastruktur**

	Mixed mode				CATI							
	Unga		Gamla		Unga		Gamla					
	0	1	0	1	0	1	0	1				
Män	0	11	8	0	20	7	0	8	10	0	18	5
	1	7	93	1	4	69	1	1	132	1	10	61
Kvinnor	0	15	8	0	13	7	0	16	5	0	34	5
	1	7	113	1	10	51	1	6	106	1	9	54

Den huvudsakliga anledningen till att just ålder och kön valdes som grupperingsvariabler var att de bedömdes som möjliga orsaker till förändringar i svarsbeteendet; t.ex. kan unga och gamla eventuellt bete sig olika beroende på minneseffekter. Med samma antaganden som tidigare (lokalt oberoende och konstanta parametrar) utökas antalet parametrar med en uppsättning för varje kombination  $M \times K \times \hat{A}$ , dvs.

$$\pi_{MK\hat{A}}, \phi_{MK\hat{A}} \text{ och } \theta_{MK\hat{A}}$$

till totalt 24 stycken. Dessvärre blir inte heller denna modell identifierbar utan ytterligare restriktioner, t.ex. genom att anta att  $\phi$  och  $\theta$  är lika över kön men olika över mode och ålder. Totalt är sju modeller med liknande restriktioner identifierbara med en hierarkisk struktur i komplexitet (nästlade modeller):

Nivå 0:  $\phi$  och  $\theta$  är lika över samtliga grupper, tom modell

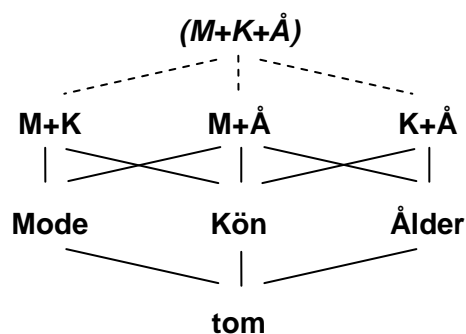
Nivå 1:  $\phi$  och  $\theta$  är olika över en vald faktor (mode, kön eller ålder)

Nivå 2:  $\phi$  och  $\theta$  är olika över två valda faktorer (M+K, M+Å el. K+Å)

Notera dock att samtliga modeller bygger på antagandet att  $\pi_{MK\dot{A}}$ , dvs. sannolikheterna att man har den studerade egenskapen, är olika över grupper. Detta antagande är egentligen det som gör modellen identifierbar. Men det medför dessutom att bland annat reliabiliteten kan skilja sig åt i olika grupper av kombinationerna av mode, kön och ålder.

Hierarkin för de olika modellerna och sättet varpå de är nästlade kan illustreras på följande sätt (notera att den högsta nivån M+K+Å inte är identifierbar):

**Figur 4.1. Hierarkisk struktur (nästling) av LCA modeller med grupperingsvariablerna Mode, Kön och Ålder**



Härmed har man en möjlighet att testa om mode har en avgörande effekt på parametrarna, dvs. om sannolikheterna för felklassificering skiljer sig åt mellan MM och CATI och samtidigt kontrollera för kön eller ålder.

#### 4.1.3 Modellvalskriterier

Programmet Panmark beräknar en mängd olika statistikor som kan användas vid modellval och för diagnostik av modellerna. Bland annat beräknas tre stycken anpassningsmått (*goodness-of-fit statistics*): likelihoodkvoten (LRT), Read-Cressie och Pearsons  $\chi^2$ . Dessa är samtliga  $\chi^2$  fördelade med antal frihetsgrader (fg) bestämda av antalet fria parametrar som skattas;  $p$ -värdena tolkas som sannolikheten att data har genererats av den skattade modellen och låga värden indikerar en dålig modellenpassning. När modellen skattas kan i vissa fall en del fria parametrar fixeras (till 0 eller 1) och några extra frihetsgrader kan "vinnas över". I samband med hypotesprövningar ska man dock utgå ifrån den a priori bestämda modellen med det ursprungliga antalet frihetsgrader.

Vid jämförelser av nästlade (*nested*) modeller kan differenser mellan LRT-statistikan användas som ett modellvalskriterium. Differensen  $LRT_1 - LRT_2 =$

$\Delta\chi^2$  mellan två nästlade modeller 1 och 2 är  $\chi^2$ -fördelad med  $fg_1 - fg_2 = \Delta fg$  frihetsgrader. Om differensen  $\Delta\chi^2$  är signifikant större än ett kritiskt värde kan det vara motiverat att välja den större modellen. Tabell 4.2 nedan kan användas som en snabb referens vid jämförelser av nästlade modeller via LRT-statistikan. Observera att Read-Cressie och Pearsons  $\chi^2$  inte kan användas på motsvarande sätt.

**Tabell 4.2. Kritiska värden för differenser i LRT-statistikan**

Jämförelse	$\Delta fg$	kritiskt värde vid $\alpha = 0.05$
Ingen faktor jmfrt med 1 faktor	2	5.99
1 faktor jmfrt med 2 faktorer	4	9.49
Ingen faktor jmfrt med 2 faktorer	6	12.6

Kriterier baserade på informationsteori, Akaike (AIC) och Schwartz-Bayesian (BIC) beräknas också av programvaran och kan användas för att välja mellan såväl icke-nästlade som nästlade modeller. De beräknas enligt

$$AIC = -2\log L + 2p$$

$$BIC = -2\log L + p\log n$$

där  $\log L$  betecknar loglikelihooden,  $p$  är antalet fria parametrar som skattas och  $n$  antalet observationer. Bägge måtten är avtagande i likelihooden  $L$  (som är ett mått på anpassning) och växande i  $p$  (som är ett mått på överanpassning), och i valet mellan två modeller ska modellen med det lägsta värdet på AIC och/eller BIC väljas. BIC straffar hårdare för antalet fria parametrar jämfört med AIC då den även tar  $n$  i beaktande.

Slutligen beräknas för varje modell även ett s.k. dissimilaritetsindex (*dissimilarity index*) som tolkas som den minsta andel av observationerna som måste reallokeras till andra celler för att modellen ska passa perfekt. Indexet bör vara mindre än 0.05 för en välartad modell och definieras enligt

$$D = \sum_{g=1}^G \frac{|n_g - \hat{n}_g|}{2n}$$

där summeringen sker över samtliga  $G$  grupper under den valda modellen,  $n_g$  betecknar antalet observationer i grupp  $g$  och  $\hat{n}_g$  betecknar det förväntade antalet observationer i grupp  $g$  under den skattade modellen.

Även om samtliga mått beaktades i analysen och valet av modell användes framförallt LRT, LRT-differenser och BIC som modellvalskriterier (se Magidson och Vermunt, 2004). I Bilaga B redovisas samtliga mått och index för varje fråga och modell.

## 4.2 Urval och dikotomisering av frågor

Totalt valdes 15 frågor ur återintervjublanketten för analys med LCA. Ett viktigt kriterium för valet av frågor var att i möjligaste mån undvika att ta med villkorsbundna frågor, dvs. sådana frågor som besvaras under villkor av hur man svarat i en tidigare fråga. Anledningen till detta var att man ville undvika att oavsiktligt "ärva" ett hypotetiskt mätfel från en annan fråga. Det är inte helt klart hur sådana beroenden ska hanteras och med den givna tidsramen för analysen ansågs det att de frågor som uppfyllde kriteriet var tillräckligt många. Vidare valdes endast sådana frågor som antingen från början var dikotoma eller enkelt kunde dikotomiseras. Modeller med fler än två kategorier är förvisso möjliga att analysera med LCA tekniken, däri ligger ingen begränsning, men då antalet observationer i vissa kategorier riskerar att bli väldigt litet med fler kategorier förutsågs problem med skattningen av modellerna. De frågor som kodades om var i vissa fall på en ordinalskala eller också skapades nya variabler utifrån frågor (eller kombinationer av frågor) som på ett enkelt sätt definierade räknevariabler (t.ex. NyBesv utifrån antalet kroniska besvär). I Bilaga B redovisas kortfattat hur kodningen (dikotomiseringen) definierades för samtliga frågor som ingick i studien. I stort täcktes hela frågebatteriet in av studien, endast utbildningsfrågorna saknas bland de utvalda.

## 4.3 Resultat

I den ursprungliga analysen skattades endast modellerna Mode+Ålder (nivå 2), Ålder (nivå 1) och den tomma modellen (nivå 0). Tanken var att man direkt skulle avgöra om det fanns en skillnad i felklassificerings-sannolikheter som berodde på mode och samtidigt kontrollera för ålder (gruppering efter kön avsåg endast att garantera identifierbarhet på nivå 2). Det insågs dock ganska snart att detta inte skulle täcka alla eventualiteter; t.ex. missar man möjligheten av olikhet över mode men likhet över ålder och kön, ett resultat som skulle vara ytterst intressant. Därför utökades analysen till att omfatta samtliga sju modeller på alla tre (möjliga) nivåer. Med detta möjliggörs en mer fullständig modellvalsstrategi med ett stegvis förfarande som kan börja valfritt uppifrån och ner eller nerifrån och upp i modellhierarkin. Man bör dock hålla i minnet att alla stegvisa förfaranden för modellval givetvis påverkar statistisk styrka och signifikans. Resultaten sammanfattas i all korthet, med en redovisning av LRT-statistikan och BIC kriteriet, i Tabell 4.3 nedan. En mer detaljerad redovisning med samtliga kriterier och index ges i Bilaga B.

Av analysen framgick det att de allra flesta frågorna inte kunde uppvisa någon tydlig effekt med avseende på någon av grupperingsvariablerna. En fråga uppvisade en (svag) modeeffekt och två frågor gav inte någon god anpassning till någon av modellerna. Alla övriga tolv frågor i analysen uppvisade tydliga tecken på att den tomma modellen var den mest hållbara, dvs. en modell utan skillnader mellan mode, kön och ålder. Några kommentarer rörande enskildheter i resultaten är som följer:

- Frg 057: Modellen för Ålder gav ett signifikant värde på  $\Delta\chi^2$  jämfört med den tomma modellen, men anpassningen i övrigt var dålig.
- NyFrg090: Den enda fråga som kunde påvisa en effekt av mode. Modellen för Mode ger en godtagbar anpassning ( $p>0.05$ ) medan de övriga är dåligt anpassade ( $p<0.05$ );  $\Delta\chi^2$  är dock inte signifikant jämfört med den

tomma modellen. AIC indikerar Mode som den bästa modellen men BIC indikerar den tomma modellen. I Tabell 4.4 nedan redovisas punktskattningarna och medelfelen för felklassificeringssannolikheterna  $\phi_M$  och  $\theta_M$  för respektive mode under modellen. Trots olikheterna i dessa skattningar är det svårt att dra någon definitiv slutsats, i synnerhet som ingen av nollhypoteserna  $\theta_{MM} = \theta_{CATI}$  eller  $\phi_{MM} = \phi_{CATI}$  förkastas vid test på 5 % signifikansnivå.

- Frg224b: Ingen av de skattade modellerna gav en godtagbar anpassning och ingen slutsats kunde dras.
- Frg250a: Den tomma modellen är den starkaste men modellerna för Mode+Kön och Mode+Ålder kunde inte skattas (konvergenskriterierna uppfylldes inte). Ingen definitiv slutsats kan dras.
- Emp, Frg251a - 253a uppvisade samtliga värden på AIC som indikerar Mode som den bästa modellen, men samtliga övriga mått och index indikerar på den tomma modellen.

Den allmänna slutsatsen måste således bli att insamlingsmetod inte har en avgörande effekt på felklassificeringssannolikheterna i den specificerade mätfelsmodellen i någon fråga utom möjligen en.

**Tabell 4.3. Sammanfattning av skattade mätfelsmodeller (LCA) avseende anpassningsmåten likelihoodkvot (LRT) och Bayes-Schwarz informationskriterium (BIC)**

Antal faktorer	0		1						2					
	14		12						8					
Faktor/er	Ingen		Mode		Kön		Ålder		Mode + Kön		Mode + Ålder		Kön + Ålder	
	LRT	BIC	LRT	BIC	LRT	BIC	LRT	BIC	LRT	BIC	LRT	BIC	LRT	BIC
Frg004d	15.647	5346.5	13.418	5357.9	14.642	5359.1	15.411	5359.9	13.249	5385.0	9.7619	5381.6	14.194	5386.0
Frg005	15.548	5241.3	12.754	5252.1	14.624	5254.0	12.734	5252.1	9.7992	5276.5	9.5328	5276.2	12.255	5279.0
Frg057	37.621	5653.0	33.728	5662.8	33.762	5662.8	29.648	5658.7	32.238	5688.6	28.847	5685.2	27.871	5684.2
Frg071	20.303	5412.9	18.211	5424.4	20.214	5426.4	17.734	5424.0	13.259	5446.8	16.098	5449.6	14.520	5448.1
NyFrg072	14.552	5645.3	10.404	5654.9	14.086	5658.5	10.242	5654.7	8.8975	5680.7	8.9911	5680.8	9.9905	5681.8
NyFrg090	24.079	5566.3	18.576	5574.5	23.229	5579.1	23.532	5579.4	18.176	5601.4	18.207	5601.4	22.350	5605.6
NyBesv	19.172	5794.7	18.903	5808.1	17.927	5807.1	16.199	5805.4	10.987	5827.5	12.552	5829.0	13.175	5829.7
Emp	20.554	4950.7	15.647	4959.3	17.930	4961.6	18.332	4962.0	12.729	4983.5	12.907	4983.7	15.747	4986.5
Frg100e	14.745	4272.4	13.516	4284.7	13.417	4284.6	11.757	4282.9	11.250	4309.4	10.571	4308.8	9.012	4307.2
Frg221b	13.773	4442.0	13.528	4455.4	12.030	4453.9	13.029	4454.9	10.743	4479.9	12.361	4481.6	11.170	4480.4
Frg224a	40.661	5652.2	37.136	5662.3	37.388	5662.6	42.601	5667.8	30.155	5682.6	35.000	5687.5	42.363	5694.8
Frg250a	9.260	4058.1	8.630	4071.2	7.271	4069.8	7.271	4069.8	(1)	(1)	(1)	(1)	5.702	4095.6
Frg251a	19.969	4392.8	15.210	4401.7	19.204	4405.7	13.295	4427.1	12.431	4426.2	18.705	4405.2	18.158	4431.9
Frg252a	21.924	4453.9	17.595	4463.3	19.523	4465.2	19.677	4465.4	13.485	4486.5	17.063	4490.1	18.377	4491.4
Frg253a	12.947	4637.6	7.609	4645.9	12.556	4650.8	12.338	4650.6	4.209	4669.8	1.468	4667.1	7.027	4672.6

(1) Modellen ej estimerbar, konvergenskriterier ej uppfyllda.

**Tabell 4.4. Felklassificeringsannolikheter ( $\phi_M$  och  $\theta_M$ ), över mode, för variabel NyFrg90, modell Mode (nivå 1); skattningar och standardfel inom parentes**

MM				CATI			
Sant	Observerat		Total	Sant	Observerat		Total
	0	1			0	1	
0	0.989	0.011	1.000	0	0.899	0.101	1.000
	(0.057)	(0.057)			(0.018)	(0.018)	
1	0.296	0.704	1.000	1	0.076	0.924	1.000
	(0.152)	(0.152)			(0.073)	(0.073)	



## 5 En jämförelse mellan MM och CATI med avseende på bortfallsskevhetens storlek

### 5.1 Upplägg

Detta avsnitt inleds med en kort definition av begreppet bortfallsskevhet utifrån ett designbaserat samplingteoretiskt perspektiv. Därefter följer en kort beskrivning av en metodik som kan användas för att skatta bortfallsskevheten under olika undersökningsdesigner. Avsnittet avslutas med en beskrivning av hur olika undersökningsdesigner kan jämföras med avseende på bortfallsskevhetens storlek.

#### 5.1.1 Teoretisk ansats

Låt  $U$  beteckna en population bestående av  $N$  objekt, samt låt  $y$  beteckna en icke-stokastisk variabel. Populationstotalen för variabeln  $y$  ges av

$$t_y = \sum_{k \in U} y_k$$

där  $y_k$  betecknar värdet på  $y$  för objekt  $k$  i populationen. För att skatta  $t_y$  dras ett sannolikhetsurval  $s$  från  $U$ , enligt en på förhand bestämd urvalsdesign. I samband med datainsamlingen uppstår bortfall, varför information om  $y$  endast erhålls för svarsmängden  $r$ . Låt  $\hat{t}_{y,r}$  beteckna den estimator som används för att utifrån svarsmängden  $r$  skatta  $t_y$ . I den fortsatta framställningen antas att  $\hat{t}_{y,r}$  kan skrivas på formen

$$\hat{t}_{y,r} = \sum_{k \in r} w_k y_k / \hat{\theta}_k$$

där  $w_k$  är en uppräkningsvikt som beror av den använda urvalsdesignen och  $\hat{\theta}_k$  är en skattning av sannolikheten att under den använda undersökningsdesignen få svar från objekt  $k$ . Bortfallsskevheten för estimatorn  $\hat{t}_{y,r}$  definieras som

$$B(\hat{t}_{y,r}) = E(\hat{t}_{y,r}) - t_y$$

där  $E(\hat{t}_{y,r})$  betecknar väntevärdet med avseende på såväl den använda urvalsdesignen som den okända svarsmekanism som genererar den slutliga svarsmängden. Den relativa bortfallsskevheten definieras som

$$RB(\hat{t}_{y,r}) = 100B(\hat{t}_{y,r}) / t_y$$

Låt

$$\hat{t}_{y,s} = \sum_{k \in s} w_k y_k$$

dvs. låt  $\hat{t}_{y,s}$  betecknar den estimator som  $\hat{t}_{y,r}$  resulterar i då  $r = s$ . I den fortsatta framställningen i detta avsnitt antas att  $E(\hat{t}_{y,s}) \approx t_y$ , dvs. att  $\hat{t}_{y,s}$  är approximativt väntevärdesriktig för  $t_y$  under den använda urvalsdesignen. Givet att värdet på  $y$  är känt för alla objekt i urvalsmängden  $s$ , gäller således att approximativt väntevärdesriktiga estimatorer för  $B(\hat{t}_{y,r})$  och  $RB(\hat{t}_{y,r})$  ges av

$$\hat{B}(\hat{t}_{y,r}) = \hat{t}_{y,r} - \hat{t}_{y,s}$$

respektive

$$\hat{RB}(\hat{t}_{y,r}) = 100[\hat{t}_{y,r} / \hat{t}_{y,s} - 1]$$

Under explicita antaganden om den parametriska formen för den okända svarsmekanismen, kan dessutom variansestimatorer  $\hat{V}[\hat{B}(\hat{t}_{y,r})]$  och

$\hat{V}[\hat{RB}(\hat{t}_{y,r})]$  härledas. Genom att låta variabeln  $y$  vara en registervariabel, som antas samvariera starkt med en eller flera viktiga undersökningsvariabler, kan man utifrån ovanstående ansats studera hur det använda estimationsförfarandet påverkas av den okända svarsmekanismen.

Antag nu att valet står mellan två olika undersökningsdesigner,  $U$  och  $V$ , för att genomföra en undersökning. Om båda designerna implementeras och registerinformation finns tillgänglig, kan bortfallsskevhetens storlek, såväl absolut som relativt, under respektive design skattas utifrån ovanstående ansats. Hypotesen

$$H_0 : B_U(\hat{t}_{y,r}) - B_V(\hat{t}_{y,r}) = 0$$

där subskriptet indikerar vilken undersökningsdesign som avses, kan användas för att testa om de båda undersökningsdesignerna ger upphov till samma bortfallsskevhet absolut sett. För att testa hypotesen kan teststatistikan

$$\hat{Z}_B = \frac{\hat{B}_U(\hat{t}_{y,r}) - \hat{B}_V(\hat{t}_{y,r})}{\sqrt{\hat{V}[\hat{B}_U(\hat{t}_{y,r}) - \hat{B}_V(\hat{t}_{y,r})]}}$$

användas. Under praktiskt rimliga förutsättningar gäller att  $\hat{Z}_B$  approximativt följer standardnormalfördelningen under  $H_0$ . Ett analogt testförfarande kan användas för att testa om de två undersökningsdesignerna är behäftade med samma relativa bortfallsskevhet.

### 5.1.2 Praktiska aspekter

Ansatsen i avsnitt 5.1.1. har använts för att studera om användandet av MM respektive CATI i ULF resulterar i samma bortfallsskevhet. I studien har följande registervariabler använts:

- Utbildningsnivå (källa: Utbildningsregistret)
- Sysselsättningsstatus (källa: RAMS)
- Lönesummeklass (källa: LSUM)

De skattningar som beräknats baseras på ett dataunderlag som svarar mot de delar av MM- respektive CATI-urvalen, samt motsvarande svarsmängder, som svarar mot första halvåret 2006. Estimatorerna  $\hat{t}_{y,r}$  och  $\hat{t}_{y,s}$  bygger, oavsett mätmetod, på en simultan poststratifiering efter variablerna kön, civilstånd, åldersklass samt region. Antalet poststrata uppgick till 27 stycken.

## 5.2 Resultat

De numeriska resultaten, vilka presenteras i Bilaga C, kan sammanfattas i följande kommentarer:

- Båda insamlingsmetoderna tenderar att underskatta antalet personer med låg utbildning. I åldersgruppen 35-64 år är den bortfallsskevhet som CATI medför signifikant skild från den bortfallsskevhet som MM medför; tendensen är att CATI, i jämförelse med MM, resulterar i en större underskattning av antalet personer med låg utbildning.
- Båda insamlingsmetoderna tenderar att överskatta antalet sysselsatta personer. I åldersgruppen 35-64 år är den bortfallsskevhet som CATI medför signifikant skild från den bortfallsskevhet som MM medför; tendensen är att CATI, i jämförelse med MM, resulterar i en större överskattning av antalet sysselsatta.
- Båda insamlingsmetoderna tenderar att underskatta både antalet utländska medborgare och antalet utlandsfödda, såväl inom olika sysselsättningskategorier som totalt. För utrikes födda är den bortfallsskevhet som CATI medför signifikant skild från den bortfallsskevhet som MM medför; tendensen är att användandet av CATI, i jämförelse med MM, resulterar i en större underskattning av såväl antalet ej sysselsatta utrikes födda som det totala antalet utrikes födda.
- Båda insamlingsmetoderna tenderar att underskatta antalet låginkomsttagare, för att istället överkatta antalet medel- och höginkomsttagare. Vid skattning av antalet medelinkomsttagare (229000 – 325500 kr per år) i åldersgruppen 20 – 64 år gäller att den bortfallsskevhet som CATI medför är signifikant skild från den bortfallsskevhet som MM medför; tendensen är att CATI, i jämförelse med MM, resulterar i en större överskattning av antalet medelinkomsttagare.

Med utgångspunkt från de variabler och de redovisningsgrupper som ingick i studien, blir således den sammanfattande slutsatsen att CATI och MM inte i någon väsentlig utsträckning skiljer sig vad gäller den bortfallsskevhet som respektive metod ger upphov till. Man kan observera att i de fall där signifikanta skillnader mellan MM och CATI föreligger, så uppvisar CATI störst skevhet i absoluta termer. Riktningen i skevhet (positiv eller negativ) är dock den samma för båda metoderna.

De påvisade skillnaderna kan dock till viss del antas bero på diverse störningar i insamlingsarbetet för CATI. Bland annat var CATI blanketten fortfarande under utveckling under de första insamlingsomgångarna, vilket medförde ett betydande bortfall (ca 50 %). Detta, i sin tur, påverkade de efterföljande omgångarna, eftersom intervjuarna då fick ägna en icke oväsentlig tid åt uppföljning av det tidigare bortfallet. Vidare fanns det skillnader mellan intervjuarkårerna. Datainsamlingen för MM utfördes av den ordinarie "ULF-kåren" med i många fall flera års erfarenhet av under-

sökningen. CATI utfördes å andra sidan av den centrala telefonintervjuarkåren som saknar denna erfarenhet. Även om det är svårt att uppskatta effekterna, bör dessa omständigheter rimligen ha påverkat utfallet av analysen och då till nackdel för CATI. Denna typ av problem bör minska i takt med att intervjuarna får en ökad erfarenhet av undersökningen.

Vid framtagandet av resultaten ovan har hypotesen

$$H_0 : B_{CATI}(\hat{t}_{y,r}) - B_{MM}(\hat{t}_{y,r}) = 0$$

utgjort utgångspunkt. Samtliga test är genomförda på ca 5 % signifikansnivå; beslutsregeln "Förkasta  $H_0$  om  $|\hat{Z}_B| > 2$ " motsvarar en teoretisk signifikansnivå på 4.56 %. Det kan vara värt att notera att i princip samma resultat skulle erhållits om man istället utgått från nollhypotesen om likhet mellan  $RB_{CATI}(\hat{t}_{y,r})$  och  $RB_{MM}(\hat{t}_{y,r})$ , då denna hypotes är ekvivalent med den ovan.

## 6 Barn-ULF i MM och CATI – en bortfallsjämförelse

Barn-ULF (BU) är en undersökning där barnen själva får komma till tals genom att besvara frågor om sina egna förhållanden. Målpopulationen för undersökningen är barn i åldersgruppen 10-18 år bosatta i Sverige med vårdnadshavare folkbokförda i Sverige. En viktig frågeställning som undersöktes var i vilken utsträckning och riktning insamlingsmode påverkade bortfallet av respondenter i BU.

Undersökningen genomfördes i anknytning till ordinarie "vuxen"-ULF (VU). Urvalspersonerna för BU identifieras vid fältarbetet av VU. Urvalet kan därför betraktas som en form av nätverksurval. Den tänkta insamlingsmetoden i MM bygger på att barnen besvarar frågor som de hör i en hörlur, vilka spelas upp av en bandspelare eller CD-spelare. Svaren antecknas av barnet i ett svarshäfte. Detta sker samtidigt som VU intervjun äger rum. Det finns alltså en intervjuare på plats vid BU som vid behov kan assistera barnet. Bland det viktigaste i sammanhanget är dels tillgången till intervjuaren, dels att den är konfidentiell gentemot föräldrarna.

Vidare förutsätter metoden att VU genomförs som *face-to-face*, troligen hemma hos urvalspersonen. Avvikelser från den tilltänkta modellen har dock observerats. Cirka 1/5 av alla barnintervjuer sker nämligen som telefonintervju. Detta har sin grund i följande förhållande:

- 1) VU-intervjun sker hemma men barnet/barnen anträffas ej
- 2) VU-intervjun är face-to-face men utanför hemmet
- 3) VU-intervjun är telefonintervju

En övergång till en alternativ datainsamlingsmetod ställer inte bara kvalitetskrav på VU-delen, utan övergången till en ny metod måste också kunna garantera att BU är genomförbar. I diskussioner med programrådet för social välfärdsstatistik, speciellt med företrädare för Institutet för social forskning vid Stockholms universitet, SOFI, framkom det att genomförbarhet i detta sammanhang likställdes med en ej försämrade svarsandel bland barn i fältarbetet. I detta syfte utvecklades i samarbete med SOFI en blankett för datorstött telefonintervju.

Som redan nämnts i avsnitt 2.1 delades ULF-urvalet av vuxna (VU) i två delar, där MM användes som huvudsaklig metod för datainsamling för den första halvan, medan CATI användes för den andra halvan. Barn som identifierades via MM-delen av VU intervjuades barnen enligt det gamla sättet, medan övriga barn intervjuades med den nya CATI-baserade metoden.

Låt  $P_{NR,MM}$  och  $P_{NR,CATI}$  beteckna bortfallsandelarna i BU för de två olika metoderna för datainsamling. Av Tabell 6.1 nedan, som redovisar resultatet av fältarbetet med avseende på svarsbortfallet i BU, framgår att hypotesen

$$H_0 : P_{NR,MM} - P_{NR,CATI} \leq 0$$

förkastas vid test på 5 % signifikansnivå, dvs. användandet av den nya CATI-baserade metoden leder till ett större deltagande än den gamla metoden. Ett 95 % konfidensintervall för differensen  $P_{NR,MM} - P_{NR,CATI}$  ges av 3.6 – 16.2 %.

**Tabell 6.1 Fältarbetsresultatet avseende Barn-ULF, första halvåret 2006**

	MM		CATI		Signifikanstest		
	antal	%	antal	%	Diff %	Z	p-värde
Svarande	264	70.59	297	80.49			
Bortfall totalt	110	29.41	72	19.51	9.90	3.1372	<0.001
- varav förälder/rar	(1)	-	37	10.03			
- varav barn	(1)	-	35	9.49			
Totalt	374		369				

(1) Uppgift saknas

## 7 Motivationen hos uppgiftslämnarna

Denna studie tillkom efter sammanträdet i SCB:s Vetenskapliga råd den 11 februari 2004, där resultatet från fas 1 presenterades och planerna för fas 2 av metodstudierna diskuterades. En av rådets rekommendationer var att i metodstudien inkludera en undersökning av uppgiftslämnarmotivationen. Detta eftersom resultat från en tidigare genomförd undersökning indikerat att det redan fanns vissa motivationsproblem i ULF (se Brage, R., och Bergman, L., 2004) och problemet befarades bli förstärkt vid en övergång till telefonintervjuer. Återintervjublanketten kompletterades därför med ytterligare tre frågor som utarbetades tillsammans med professor Lars Bergman, ledamot av Vetenskapliga rådet:

- Frg38 Hur gärna eller hur motvilligt deltog du i den här undersökningen?
- Frg39 Var det jobbigt att vara med eller gick det lätt?
- Frg40 Skulle du kunna tänka dig att vara med i en liknande undersökning om "LF" om t.ex. ett år?

De olika svarsalternativen, samt dessas respektive fördelningar för respektive MM och CATI, redovisas i detalj i Bilaga D. Den genomgående slutsatsen är att det finns en liten skillnad mellan MM och CATI, där tendensen är att respondenterna tycks uppleva MM som något lättare att delta i.

Resultatet skall dock inte överdrivas. För att illustrera skillnaderna dikotomiserades svarsalternativen för respektive fråga i positiva och negativa alternativ (bortfall ej inkluderat i beräkningarna). Med ett enkelt signifikanstest avgörs om det finns några skillnader mellan MM och CATI grupperna, se Tabell 7.1 nedan. Vid test på 5 % signifikansnivå erhålls en signifikant skillnad för FRG39, men för FRG38 och FRG40 kan inga signifikanta skillnader påvisas. Ett 95 % konfidensintervall för skillnaden mellan MM och CATI vad gäller andelen personer som angivit svaret lätt på Frg39 ges av 0.2 – 9.0 %. Vidare testades för oberoende mellan svarsfördelning och mode vilket redovisas i Tabell 7.2. Med detta test kan inga signifikanta skillnader (beroenden) påvisas vid test på 5 % signifikansnivå, medan test på 10 % signifikansnivå medför att en modeeffekt kan påvisas för två av de tre frågorna.

**Tabell 7.1. Signifikanstest för skillnader mellan MM och CATI avseende andelen positiva efter dikotomisering av svarsalternativen.**

Fråga	Svarsalternativ		Andel positiva %		Diff %	Signifikanstest	
	Neg	Pos	MM	CATI		Z	p-värde
Frg38	3, 4, 5	1, 2	77.878	74.739	3.139	1.11901	0.26313
Frg39	1, 2	3	89.391	84.760	4.631	2.08783	0.03681
Frg40	1, 2	3, 4	92.955	92.178	0.777	0.44679	0.65502

**Tabell 7.2. Test för oberoende mellan svaralternativ och mode.**

Fråga	Kommentar	$\chi^2$	df	p-värde
Frg38	svarsalt. 4 och 5 sammanslagna	7.2036	3	0.06568
Frg39		4.7478	2	0.09312
Frg40		4.9267	3	0.17722



## 8 Slutsatser

Med ledning av resultaten från de fem olika delstudierna kan följande slutsatser dras:

- För flertalet av frågorna i reliabilitetsstudien kunde ingen signifikant skillnad mellan CATI och MM påvisas. Vid test på 5 % signifikansnivå erhöles signifikans för två frågor, för vilka CATI uppvisade ett lägre inkonsistensindex än MM. Med en utökning av signifikansnivån till 10 %, erhöles signifikans för ytterligare två frågor. Även för dessa frågor uppvisade CATI ett lägre inkonsistensindex än MM. En av de frågor som ursprungligen var tänkt att ingå i reliabilitetsstudien exkluderades på grund av numeriska problem i samband med analysen.
- Av modelleringen och analysen av mätfelsmodeller med LCA metoden framgick det att i tolv av femton fall inte kunde påvisas någon effekt med avseende på insamlingsmetod. En fråga uppvisade en (svag) modeeffekt och två frågor gav inte någon god anpassning till någon av modellerna. Den allmänna slutsatsen är att insamlingsmetod inte har en avgörande effekt på felklassificeringssannolikheterna i mätfelsmodellen i någon fråga utom möjligen en.
- CATI och MM skiljer sig inte i någon väsentlig utsträckning vad gäller den bortfallsskevhet som respektive metod ger upphov till. Där statistiskt signifikanta skillnader föreligger, tenderar dock användandet av CATI att resultera i störst bortfallsskevhet. Oavsett val av insamlingsmetod bör en del av problemet med bortfallsskevhet kunna åtgärdas i löpande produktion, t.ex. genom ett mer effektivt utnyttjande av tillgänglig registerinformation.
- Skillnaden i bortfallet i Barn-ULF är signifikant skild från noll, till fördel för CATI som insamlingsmetod. Ett 95 % konfidensintervall för differensen  $P_{NR,MM} - P_{NR,CATI}$  ges av 3.6 – 16.2 %. En förklaring till det större deltagandet bland barn i CATI är att respondenterna gavs en större möjlighet att själva välja när intervjun skulle äga rum. Det är inte utrett om deltagandet i MM skulle öka i motsvarande grad om respondenterna fick samma möjlighet att själva välja tidpunkt för barnen, med t.ex. erbjudande om ett återbesök.
- Motivationsstudien indikerar att det kan finnas en liten skillnad mellan MM och CATI, där tendensen är att respondenterna tycks uppleva MM som något lättare att delta i. Vid test baserade på dikotomiserade svarsalternativ erhöles en mode skillnad (5 % signifikansnivå) för Frg39, som avser hur lätt det var att delta i undersökningen. Ett 95 % konfidensintervall för skillnaden i andel personer som angivit svaret lätt på Frg39 i MM respektive CATI ges av 0.2 – 9.0 %. Vid test för oberoende mellan svarsfördelning och mode erhöles dock inga signifikanta resultat vid test på 5 % signifikansnivå. Den sammanfattande slutsatsen från motivationsstudien är att en övergång till CATI inte i någon större utsträckning påverkar respondenternas motivation att delta i undersökningen.

Sammanfattningsvis gäller att den systematiska felkomponenten (mät- och bortfallsfel) i tillförlitlighetskomponenten bedöms förbli oförändrad vid en övergång till telefonintervjuer. I kombination med att övergången frigör

resurser för en urvalsökning, innebär detta att medelkvadratfelet (MSE) i skattningarna skulle minska signifikant vid en övergång till telefon.

Övergången ger, för en given budget, ökat intervjuutrymme och därmed ökad utrymme för att förbättra statistikens innehåll. Vidare medför en urvalsökning ökade möjligheter för redovisningar av ytterligare undergrupper. En övergång från nuvarande metod (MM) till datorstödd telefonintervju (CATI) leder dessutom till ökad aktualitet i och med att framställningstiden minskar.

**Tabell 8.1. Översikt av effekterna av en övergång från MM till CATI**

	Sämre	Lika	Bättre	Fas
<b>Innehåll</b>			<b>X</b>	
- Variabler			X	
- Redovisningsgrupper			X	
<b>Tillförlitlighet</b>			<b>X</b>	
- Urval/varians			X	1
- Mätning		X		2
- Svartsbortfall/bortfallsfel		(X)		2
<b>Aktualitet</b>			<b>X</b>	
- Framställningstid			X	
- Punktlighet		X		
<b>Jämförbarhet och sammanvändbarhet</b>				
- Jmf över tiden	(X)		-	
- Jmf mellan grupper	(X)		(X)	
- Samanvändbarhet			(X)	
<b>Tillgänglighet och förståelighet</b>		<b>X</b>		
<b>Övrigt</b>				
- Barn-ULF			X	2
- Motivation		X		2

## 9 Källor

### Tryckta källor

Biemer, P.P. (2001) Nonresponse Bias and Measurement Bias in a Comparison of Face to Face and Telephone Interviewing. *Journal of Official Statistics*, **17**, 2, 295-320.

Brage, R., och Bergman, L. (2004) Attitudes toward Surveys and Survey Experiences in the Swedish General Population: Basic findings. Rapport nr 28 inom projektet "Modern statistical survey methods", Statistiska institutionen, Stockholms universitet.

Magidson, J., och Vermunt, J. (2004) Latent Class Models i D.Kaplan (ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences*, Chapter 10, sidor 175-198. Thousand Oaks: Sage Publications.

Van de Pol, F, Langeheine, R., och de Jong, W. (2001) *PANMARK 3, User's Manual. Panel Analysis Using Markov Chains, A Latent Class Analysis Program.*

SCB (2003) Alternativa datainsamlingsmetoder för ULF. *Bakgrundsfakta 2003:9*. SCB, Stockholm.

### Elektroniska källor

SAS 9.2, dokumentation av PROC FREQ:

[http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/statug\\_freq\\_a0000000647.htm](http://support.sas.com/documentation/cdl/en/statug/63033/HTML/de fault/statug_freq_a0000000647.htm)



# Bilagor

- A. Resultat reliabilitetsstudie
- B. Resultat mätfelsmodeller/LCA
- C. Resultat från studien rörande differenser i bortfallskevheter
- D. Resultat motivationsfrågor
- E. Avsnitt 3, ursprunglig version
- F. Avsnitt 8, ursprunglig version
- G. Återintervjublanketten för MM

## Bilaga A. Resultat reliabilitetsstudie

### Frg004d

#### Ursprungsvariabelns beteckning:

- MM, ordinarie intervju: frg004d
- MM, återintervju: 1
- CATI: b4a

#### Ungefärlig frågelydelse i MM, återintervjun:

Anteckna typen av hus up bor i

#### Selektion:

Endast Up med samma insamlingsmetod vid båda intervjutillfällena.

Ej ändrade bostadsförhållanden enligt fråga 3 i återintervjublanketten

#### Kodning / dikotomisering:

Variabelvärdet = 1 om svarsalternativ 3, 4 eller 5, annars = 0

**Tabell A.1 Svartsfördelning över insamlingsmode, Frg004d**

		MM			CATI		
		Återintervju, $y_{i2}$			Återintervju, $y_{i2}$		
		0	1	$\Sigma$	0	1	$\Sigma$
Intervju, $y_{i1}$	0	246	7	253	269	3	272
	1	5	159	164	4	199	203
	$\Sigma$	251	166	417	273	202	475

**Frg005****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg005
- MM, återintervju: 2
- CATI: b6

**Ungefärlig frågelydelse i MM, återintervjun:**

På vilket sätt disponerar du/ni bostaden?

**Selektion:**

Endast Up med samma insamlingsmetod vid båda intervjutillfällena.

Ej ändrade bostadsförhållanden enligt fråga 3 i återintervjublanketten.

**Kodning / dikotomisering:**

Variabelvärdet = 1 om svarsalternativ 3, 4, 6 eller 7, annars = 0

**Tabell A.2 Svartsfördelning över insamlingsmode, Frg005**

		MM			CATI		
		Återintervju, $y_{12}$		$\Sigma$	Återintervju, $y_{12}$		$\Sigma$
		0	1		0	1	
Intervju, $y_{11}$	0	292	5	297	330	4	334
	1	0	115	115	8	133	141
	$\Sigma$	292	120	412	338	137	475

**Frg057****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg057
- MM, återintervju: 5
- CATI: f21

**Ungefärlig frågelydelse i MM, återintervjun:**

Har du under de senaste 12 månaderna på din fritid gjort en semesterresa som varade minst en vecka?

**Selektion:**

Endast Up med samma insamlingsmetod vid båda intervjutillfällena.

Konsistens med frågor 6 och 7 i återintervjublanketten.

**Kodning / dikotomisering:**

Variabelvärdet = 1 om svarsalternativ 2 eller 8, annars = 0

**Tabell A.3 Svartsfördelning över insamlingsmode, Frg057**

		MM			CATI		
		Återintervju, $y_{2}$			Återintervju, $y_{2}$		
		0	1	$\Sigma$	0	1	$\Sigma$
Intervju $y_{1}$	0	274	28	302	308	29	337
	1	21	99	120	18	125	143
	$\Sigma$	295	127	422	326	154	480



**Frg071****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg071
- MM, återintervju: 8
- CATI: h41a

**Ungefärlig frågelydelse i MM, återintervjun:**

Hur bedömer du ditt allmänna hälsotillstånd?

**Selektion:**

Endast Up med samma insamlingsmetod vid båda intervjutillfällena.

Ej förändrat hälsotillstånd enligt fråga 16 i återintervjublanketten.

**Kodning / dikotomisering:**

Variabelvärdet = 1 om svarsalternativ 4 eller 5, annars = 0

**Tabell A.4 Svartsfördelning över insamlingsmode, Frg071**

		MM			CATI		
		Återintervju, $y_{12}$			Återintervju, $y_{12}$		
		0	1	$\Sigma$	0	1	$\Sigma$
Intervju, $y_{11}$	0	374	6	380	426	8	434
	1	5	6	11	6	14	20
	$\Sigma$	379	12	391	432	22	454

**NyFrg072****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg073
- MM, återintervju: 10
- CATI: h42, h45 och nykoll

**Ungefärlig frågelydelse i MM, återintervjun:**

Ursprunglig frågelydelse saknas. Variabeln NyFrg072 härleds utifrån ursprungsvariablerna ovan.

**Selektion:**

Endast Up med samma insamlingsmetod vid båda intervjutillfällena.

Ej förändrat hälsotillstånd enligt fråga 16 i återintervjublanketten.

**Kodning / dikotomisering:**

Variabelvärdet = 1 om variabeln "Antal kroniska besvär från frågor om ej specificerade besvär" > 0, annars = 0.

**Tabell A.5 Svartsfördelning över insamlingsmode, NyFrg072**

		MM			CATI		
		Återintervju, $y_{2}$		$\Sigma$	Återintervju, $y_{2}$		$\Sigma$
		0	1		0	1	
Intervju, $y_{1}$	0	180	20	200	244	11	255
	1	20	171	191	19	180	199
$\Sigma$		200	191	391	263	191	454

**Frg100e****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg100e
- MM, återintervju: 18e
- CATI: arbl

**Ungefärlig frågelydelse i MM, återintervjun:**

Vad var aktuellt under förra veckan:

- Arbetslös, sökte eller väntade på jobb hela veckan

**Selektion:**

Frågan besvaras endast om Up är född 1932 eller senare (74 år eller yngre)

Endast Up med samma insamlingsmetod vid båda intervjutillfällena.

Ej förändrat hälsotillstånd enligt fråga 19 i återintervjublanketten.

**Kodning / dikotomisering:**

Variabelvärdet = 1 om ursprungsvariabeln besvaras med Ja (1), dvs. arbetslös, annars = 0.

**Tabell A.6 Svartsfördelning över insamlingsmode, Frg100e**

		MM			CATI		
		Återintervju, $y_{12}$			Återintervju, $y_{12}$		
		0	1	$\Sigma$	0	1	$\Sigma$
Intervju, $y_{11}$	0	263	3	266	306	6	312
	1	6	29	35	5	156	161
$\Sigma$		269	32	301	311	162	473

**Frg122a****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg122a
- MM, återintervju: 20
- CATI: s122a

**Ungefärlig frågelydelse i MM, återintervjun:**

Är det bullrigt där du arbetar?

**Selektion:**

Endast Up med samma insamlingsmetod vid båda intervjutillfällena.

Har ej bytt arbete eller fått arbete.

**Kodning / dikotomisering:**

Variabelvärdet = 0 om svarsalternativ Ja, annars = 1.

**Tabell A.7 Svartsfördelning över insamlingsmode, Frg122a**

		MM			CATI		
		Återintervju, $y_{2}$			Återintervju, $y_{2}$		
		0	1	$\Sigma$	0	1	$\Sigma$
Intervju, $y_{1}$	0	85	16	101	82	30	112
	1	11	153	164	7	187	194
	$\Sigma$	96	169	265	89	217	306

**Frg122b****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg122b
- MM, återintervju: 21
- CATI: s122b

**Ungefärlig frågelydelse i MM, återintervjun:**

Bullrar det för jämnan eller bara ibland?

**Selektion:**

Endast Up med samma insamlingsmetod vid båda intervjutillfällena.

Har ej bytt arbete eller fått arbete samt svarat Ja på Frg122a i båda intervjuerna.

**Kodning / dikotomisering:**

Variabelvärdet = 0 om svarsalternativ Ja, annars = 1.

**Tabell A.8 Svartsfördelning över insamlingsmode, Frg122b**

	MM			CATI			
	Återintervju, $y_{12}$			Återintervju, $y_{12}$			
	0	1	$\Sigma$	0	1	$\Sigma$	
Intervju, $y_{11}$	0	33	10	43	31	12	43
	1	3	39	42	3	36	39
	$\Sigma$	36	49	85	34	48	82

**Frg122c****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg122c
- MM, återintervju: 22
- CATI: s122c

**Ungefärlig frågelydelse i MM, återintervjun:**

Bullrar det öronbedövande?

**Selektion:**

Endast Up med samma insamlingsmetod vid båda intervjutillfällena.

Har ej bytt arbete eller fått arbete samt svarat Ja på Frg122a i båda intervjuerna.

**Kodning / dikotomisering:**

Variabelvärdet = 0 om svarsalternativ Ja, annars = 1.

**Tabell A.9 Svartsfördelning över insamlingsmode, Frg122c**

		MM			CATI		
		Återintervju, $y_{2}$			Återintervju, $y_{2}$		
		0	1	$\Sigma$	0	1	$\Sigma$
Intervju, $y_{1}$	0	33	16	49	33	7	40
	1	12	24	36	8	34	42
	$\Sigma$	45	40	85	41	41	82

**Frg123c****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg123c
- MM, återintervju: 23
- CATI: s123c

**Ungefärlig frågelydelse i MM, återintervjun:**

Är ditt arbete psykiskt ansträngande?

**Selektion:**

Endast Up med samma insamlingsmetod vid båda intervjutillfällena.

Har ej bytt arbete eller fått arbete.

**Kodning / dikotomisering:**

Variabelvärdet = 0 om svarsalternativ Ja, annars = 1.

**Tabell A.10 Svartsfördelning över insamlingsmode, Frg123c**

		MM			CATI		
		Återintervju, $y_{12}$		$\Sigma$	Återintervju, $y_{12}$		$\Sigma$
		0	1		0	1	
Intervju, $y_{11}$	0	77	35	112	91	47	138
	1	22	130	152	19	149	168
	$\Sigma$	99	165	264	110	196	306

**Frg124a****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg124a
- MM, återintervju: 24
- CATI: s124a

**Ungefärlig frågelydelse i MM, återintervjun:**

Kräver ditt arbete tunga lyft?

**Selektion:**

Endast Up med samma insamlingsmetod vid båda intervjutillfällena.

Har ej bytt arbete eller fått.

**Kodning / dikotomisering:**

Variabelvärdet = 0 om svarsalternativ Ja, annars = 1.

**Tabell A.11 Svartsfördelning över insamlingsmode, Frg124a**

		MM			CATI		
		Återintervju, $y_{2}$			Återintervju, $y_{2}$		
		0	1	$\Sigma$	0	1	$\Sigma$
Intervju, $y_{1}$	0	95	19	114	110	11	121
	1	9	142	151	13	172	185
	$\Sigma$	104	161	265	123	183	306



**Frg221b****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg221b
- MM, återintervju: 27
- CATI: m221b

**Ungefärlig frågelydelse i MM, återintervjun:**

Är du medlem i något politiskt parti?

**Selektion:**

Endast Up med samma insamlingsmetod vid båda intervjutillfällena.

Ej förändrat tillstånd enligt fråga 28 i återintervjublanketten.

**Kodning / dikotomisering:**

Variabelvärdet = 0 om svarsalternativ Ja, annars = 1.

**Tabell A.12 Svartsfördelning över insamlingsmode, Frg221b**

		MM			CATI		
		Återintervju, $y_{12}$			Återintervju, $y_{12}$		
		0	1	$\Sigma$	0	1	$\Sigma$
Intervju, $y_{11}$	0	21	2	23	20	4	24
	1	4	394	398	6	450	456
	$\Sigma$	25	396	421	26	454	480

**Frg224a****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg224a
- MM, återintervju: 29
- CATI: m224a

**Ungefärlig frågelydelse i MM, återintervjun:**

Skulle du själv kunna författa en skrivelse och överklaga ett beslut av en myndighet?

**Selektion:**

Endast Up med samma insamlingsmetod vid båda intervjutillfällena.

**Kodning / dikotomisering:**

Variabelvärdet = 0 om svarsalternativ Ja, annars = 1.

**Tabell A.13 Svartsfördelning över insamlingsmode, Frg224a**

		MM			CATI		
		Återintervju, $y_{22}$			Återintervju, $y_{22}$		
		0	1	$\Sigma$	0	1	$\Sigma$
Intervju, $y_{21}$	0	272	15	287	314	16	330
	1	47	88	135	31	119	150
	$\Sigma$	319	103	422	345	135	480

**Frg250a****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg250a
- MM, återintervju: 30a
- CATI: t250a

**Ungefärlig frågelydelse i MM, återintervjun:**

Har du under de senaste 12 månaderna blivit utsatt för våld som lämnade synliga skador och som resulterade i läkarbesök?

**Selektion:**

Endast Up med samma insamlingsmetod vid båda intervjutillfällena.

**Kodning / dikotomisering:**

Variabelvärdet = 1 om ursprungsvariabeln ej är besvarad med Ja, annars = 0.

**Tabell A.14 Svartsfördelning över insamlingsmode, Frg250a**

		MM			CATI		
		Återintervju, $y_{2}$		$\Sigma$	Återintervju, $y_{2}$		$\Sigma$
		0	1		0	1	
Intervju, $y_{1}$	0	5	1	6	2	0	2
	1	0	415	415	1	477	478
	$\Sigma$	5	416	421	3	477	480

**Frg251a****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg251a
- MM, återintervju: 32a
- CATI: t251a

**Ungefärlig frågelydelse i MM, återintervjun:**

Har du under de senaste 12 månaderna blivit utsatt för våld som lämnade synliga skador men som inte resulterade i läkarbesök?

**Selektion:**

Endast Up med samma insamlingsmetod vid båda intervjutillfällena.

**Kodning / dikotomisering:**

Variabelvärdet = 1 om ursprungsvariabeln ej är besvarad med Ja, annars = 0.

**Tabell A.15 Svarsfördelning över insamlingsmode, Frg251a**

		MM			CATI		
		Återintervju, $y_{2}$		$\Sigma$	Återintervju, $y_{2}$		$\Sigma$
		0	1		0	1	
Intervju, $y_{1}$	0	6	1	7	11	5	16
	1	10	404	414	3	461	464
	$\Sigma$	16	405	421	14	466	480

**Frg252a****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg252a
- MM, återintervju: 34a
- CATI: t252a

**Ungefärlig frågelydelse i MM, återintervjun:**

Har du under de senaste 12 månaderna blivit utsatt för våld som inte lämnade synliga skador?

**Selektion:**

Endast Up med samma insamlingsmetod vid båda intervjutillfällena.

**Kodning / dikotomisering:**

Variabelvärdet = 1 om ursprungsvariabeln ej är besvarad med Ja, annars = 0.

**Tabell A.16 Svartsfördelning över insamlingsmode, Frg252a**

		MM			CATI		
		Återintervju, $y_{2}$		$\Sigma$	Återintervju, $y_{2}$		$\Sigma$
		0	1		0	1	
Intervju, $y_{1}$	0	7	4	11	12	3	15
	1	6	404	410	8	457	465
	$\Sigma$	13	408	421	20	460	480

**Frg253a****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg253a
- MM, återintervju: 36a
- CATI: t253a

**Selektion:**

Endast Up med samma insamlingsmetod vid båda intervjutillfällena.

**Ungefärlig frågelydelse i MM, återintervjun:**

Har du under de senaste 12 månaderna blivit utsatt för hot om våld som föranledde rädsla?

**Kodning / dikotomisering:**

Variabelvärdet = 1 om ursprungsvariabeln ej är besvarad med Ja, annars = 0.

**Tabell A.17 Svarsfördelning över insamlingsmode, Frg253a**

		MM			CATI		
		Återintervju, $y_{2}$			Återintervju, $y_{2}$		
		0	1	$\Sigma$	0	1	$\Sigma$
Intervju, $y_{1}$	0	19	10	29	20	6	26
	1	13	379	392	7	447	454
	$\Sigma$	32	389	421	27	453	480

**Bilaga B. Resultat mätfelsmodeller / LCA****Frg004d****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg004d
- MM, återintervju: 1
- CATI: b4a

**Ungefärlig frågelydelse i MM, återintervjun:**

Anteckna typen av hus up bor i

**Kodning / dikotomisering:**

Variabelvärdet = 1 om ursprungsvariabeln besvaras med Friliggande villa eller Radhus/kedjehus (1 eller 2), 0 annars.

**Data**

Antal observationer: 923

	Mixed mode						CATI					
	Unga		Gamla		Unga		Gamla		Unga		Gamla	
	0	1	0	1	0	1	0	1	0	1	0	1
Män	0	52	1	0	31	1	0	65	4	0	31	0
	1	2	64	1	2	66	1	0	82	1	0	63
Kvinnor	0	61	1	0	25	2	0	57	1	0	47	0
	1	1	80	1	2	52	1	2	73	1	1	54

**Resultat av modellering**

	Ingen faktor	1 faktor			2 faktorer		
		Mode	Kön	Ålder	M+K	M+Å	K+Å
Frihetsgrader	14	12	12	12	8	8	8
- extra fg	1	2	2	2	3	4	3
LRT	15.647	13.418	14.642	15.411	13.249	9.7619	14.194
- p-värde	0.335	0.339	0.262	0.220	0.104	0.282	0.077
Read-Cressie	12.868	11.870	11.895	12.708	11.575	8.4164	11.255
- p-värde	0.537	0.456	0.454	0.391	0.171	0.394	0.188
Pearson $\chi^2$	12.584	12.005	11.518	12.417	11.622	8.3799	10.806
- p-värde	0.560	0.445	0.485	0.413	0.169	0.397	0.213
AIC	5264.4	5266.2	5267.4	5268.2	5274.0	5270.5	5275.0
BIC	5346.5	5357.9	5359.1	5359.9	5385.0	5381.6	5386.0
Diss.index	0.00912	0.00877	0.00858	0.00895	0.00860	0.00621	0.00802

**Frg005****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg005
- MM, återintervju: 2
- CATI: b6

**Ungefärlig frågelydelse i MM, återintervjun:**

På vilket sätt disponerar du/ni bostaden?

**Kodning / dikotomisering:**

Variabelvärdet = 1 om ursprungsvariabeln besvaras med Äganderätt, Bostadsrätt eller Tjänstebostad (1, 2 eller 5), 0 annars.

**Data**

Antal observationer: 923

	Mixed mode				CATI				
	Unga		Gamla		Unga		Gamla		
	0	1	0	1	0	1	0	1	
Män	0	38	2	0	19	2	0	44	3
	1	2	77	1	1	78	1	2	102
Kvinnor	0	53	0	0	21	0	0	41	4
	1	1	89	1	1	59	1	3	85

**Resultat av modellering**

	Ingen faktor	1 faktor			2 faktorer		
		Mode	Kön	Ålder	M+K	M+Å	K+Å
Frihetsgrader	14	12	12	12	8	8	8
- extra fg	-	2	2	2	3	4	3
LRT	15.548	12.754	14.624	12.734	9.7992	9.5328	12.255
- p-värde	0.342	0.387	0.263	0.389	0.279	0.299	0.140
Read-Cressie	12.053	9.4619	11.977	9.8197	7.3510	7.3243	9.6170
- p-värde	0.602	0.663	0.447	0.632	0.499	0.502	0.293
Pearson $\chi^2$	11.408	8.7602	11.682	9.2739	6.8487	6.8852	9.1456
- p-värde	0.654	0.723	0.471	0.680	0.553	0.549	0.330
AIC	5159.2	5160.4	5162.3	5160.4	5165.5	5165.2	5167.9
BIC	5241.3	5252.1	5254.0	5252.1	5276.5	5276.2	5279.0
Diss.index	0.01092	0.00936	0.00927	0.00876	0.00636	0.00706	0.00807



**Frg057****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg057
- MM, återintervju: 5
- CATI: f21

**Ungefärlig frågelydelse i MM, återintervjun:**

Har du under de senaste 12 månaderna på din fritid gjort en semesterresa som varade minst en vecka?

**Kodning / dikotomisering:**

Variabelvärdet = 1 om ursprungsvariabeln besvaras med Ja (1), 0 annars.

**Data**

Antal observationer: 923

	Mixed mode				CATI							
	Unga		Gamla		Unga		Gamla					
	0	1	0	1	0	1	0	1				
Män	0	31	0	0	21	12	0	30	3	0	23	8
	1	6	82	1	8	59	1	13	105	1	3	60
Kvinnor	0	28	7	0	24	3	0	39	3	0	33	4
	1	10	98	1	7	47	1	7	84	1	6	59

**Resultat av modellering**

	Ingen faktor	1 faktor			2 faktorer		
		Mode	Kön	Ålder	M+K	M+Å	K+Å
Frihetsgrader	14	12	12	12	8	8	8
- extra fg	1	2	2	2	2	4	4
LRT	37.621	33.728	33.762	29.648	32.238	28.847	27.871
- p-värde	0.001	0.001	0.001	0.003	0.000	0.000	0.000
Read-Cressie	33.422	28.585	28.811	25.464	27.186	24.516	24.409
- p-värde	0.003	0.005	0.004	0.013	0.001	0.002	0.002
Pearson $\chi^2$	32.710	27.454	27.775	24.583	26.0617	23.549	23.813
- p-värde	0.003	0.007	0.006	0.017	0.001	0.003	0.002
AIC	5571.0	5571.1	5571.1	5567.0	5577.6	5574.2	5573.2
BIC	5653.0	5662.8	5662.8	5658.7	5688.6	5685.2	5684.2
Diss.index	0.03537	0.02994	0.02950	0.03087	0.02626	0.02791	0.02827

**Frg071****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg071
- MM, återintervju: 8
- CATI: h41a

**Ungefärlig frågelydelse i MM, återintervjun:**

Hur bedömer du ditt allmänna hälsotillstånd?

**Kodning / dikotomisering:**

Variabelvärdet = 1 om ursprungsvariabeln besvaras Mycket gott eller Gott (1 eller 2), 0 annars.

**Data**

Antal observationer: 923

	Mixed mode				CATI							
	Unga		Gamla		Unga		Gamla					
	0	1	0	1	0	1	0	1				
Män	0	11	8	0	20	7	0	8	10	0	18	5
	1	7	93	1	4	69	1	1	132	1	10	61
Kvinnor	0	15	8	0	13	7	0	16	5	0	34	5
	1	7	113	1	10	51	1	6	106	1	9	54

**Resultat av modellering**

	Ingen faktor	1 faktor			2 faktorer		
		Mode	Kön	Ålder	M+K	M+Å	K+Å
Frihetsgrader	14	12	12	12	8	8	8
- extra fg	-	-	-	1	2	2	3
LRT	20.303	18.211	20.214	17.734	13.259	16.098	14.520
- p-värde	0.121	0.109	0.063	0.124	0.103	0.041	0.069
Read-Cressie	18.912	17.343	18.878	16.659	12.272	15.029	13.529
- p-värde	0.168	0.137	0.092	0.163	0.139	0.059	0.095
Pearson $\chi^2$	18.569	17.214	18.559	16.412	12.038	14.771	13.303
- p-värde	0.182	0.142	0.100	0.173	0.150	0.064	0.102
AIC	5330.8	5332.7	5334.7	5332.2	5335.8	5338.6	5337.0
BIC	5412.9	5424.4	5426.4	5424.0	5446.8	5449.6	5448.1
Diss.index	0.02943	0.02513	0.02893	0.02577	0.01578	0.01961	0.01975

**NyFrg072****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg073
- MM, återintervju: 10
- CATI: h42, h45 och nykoll

**Ungefärlig frågelydelse i MM, återintervjun:**

Ursprunglig frågelydelse saknas. Variabeln NyFrg072 härleds utifrån ursprungsvariablerna ovan.

**Kodning / dikotomisering:**

Variabelvärdet = 1 om variabeln "Antal kroniska besvär från frågor om ej specificerade besvär" > 0, 0 annars.

**Data**

Antal observationer: 923

	Mixed mode				CATI							
	Unga		Gamla		Unga		Gamla					
	0	1	0	1	0	1	0	1				
Män	0	70	7	0	38	3	0	110	2	0	29	3
	1	6	36	1	5	54	1	8	31	1	5	57
Kvinnor	0	70	8	0	18	5	0	80	7	0	36	2
	1	11	54	1	3	55	1	3	43	1	5	59

**Resultat av modellering**

	Ingen faktor	1 faktor			2 faktorer		
		Mode	Kön	Ålder	M+K	M+Å	K+Å
Frihetsgrader	14	12	12	12	8	8	8
- extra fg	-	-	-	2	-	1	3
LRT	14.552	10.404	14.086	10.242	8.8975	8.9911	9.9905
- p-värde	0.409	0.581	0.295	0.595	0.351	0.343	0.266
Read-Cressie	13.548	10.034	13.033	9.6830	8.6128	8.6666	9.3804
- p-värde	0.484	0.613	0.367	0.644	0.376	0.371	0.311
Pearson $\chi^2$	13.225	9.9294	12.665	9.4918	8.5365	8.5726	9.1613
- p-värde	0.509	0.622	0.394	0.660	0.383	0.380	0.329
AIC	5563.3	5563.1	5566.8	5563.0	5569.6	5569.7	5570.7
BIC	5645.3	5654.9	5658.5	5654.7	5680.7	5680.8	5681.8
Diss.index	0.02484	0.01774	0.02440	0.01663	0.01246	0.01306	0.01609

**NyFrg090****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg090a-f
- MM, återintervju: 15a-f
- CATI: h70a1, h70d1, h70f1, h70g1, h70i1, h70k1

**Ungefärlig frågelydelse i MM, återintervjun:**

Ursprunglig frågelydelse saknas. Variabeln NyFrg090 härleds utifrån ursprungsvariablerna ovan.

**Kodning / dikotomisering:**

Variabelvärdet = 1 om variabeln "Antal svåra besvär från frågor om specificerade besvär" > 0, 0 annars.

**Data**

Antal observationer: 923

	Mixed mode				CATI							
	Unga		Gamla		Unga		Gamla					
	0	1	0	1	0	1	0	1				
Män	0	94	9	0	76	4	0	116	6	0	66	6
	1	4	12	1	7	13	1	21	8	1	12	10
Kvinnor	0	102	6	0	52	5	0	94	11	0	51	8
	1	14	21	1	9	15	1	11	17	1	9	34

**Resultat av modellering**

	Ingen faktor	1 faktor			2 faktorer		
		Mode	Kön	Ålder	M+K	M+Å	K+Å
Frihetsgrader	14	12	12	12	8	8	8
- extra fg	-	-	1	1	-	1	3
LRT	24.079	18.576	23.229	23.532	18.176	18.207	22.350
- p-värde	0.045	0.099	0.026	0.024	0.020	0.020	0.004
Read-Cressie	24.085	18.159	23.231	23.317	17.625	17.659	22.364
- p-värde	0.045	0.111	0.026	0.025	0.024	0.024	0.004
Pearson $\chi^2$	24.321	18.079	23.450	23.415	17.476	17.512	22.564
- p-värde	0.042	0.113	0.024	0.024	0.026	0.025	0.004
AIC	5484.3	5482.8	5487.4	5487.7	5490.4	5490.4	5494.5
BIC	5566.3	5574.5	5579.1	5579.4	5601.4	5601.4	5605.6
Diss.index	0.03680	0.02605	0.03478	0.03588	0.02275	0.02310	0.03228

**NyBesv****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: -
- MM, återintervju: -
- CATI: -

**Ungefärlig frågelydelse i MM, återintervjun:**

Ursprunglig frågelydelse saknas. Variabeln NyBesv härleds från variablerna NyFrg072 och NyFrg090.

**Kodning / dikotomisering:**

Variabelvärdet = 1 om NyFrg072 + NyFrg090 > 0, 0 annars.

**Data**

Antal observationer: 923

	Mixed mode						CATI					
	Unga			Gamla			Unga			Gamla		
	0	1		0	1		0	1		0	1	
Män	0	65	8	0	37	2	0	92	7	0	25	4
	1	5	41	1	5	56	1	19	33	1	6	59
Kvinnor	0	58	9	0	16	5	0	69	7	0	30	5
	1	17	59	1	4	56	1	8	49	1	4	63

**Resultat av modellering**

	Ingen faktor	1 faktor			2 faktorer		
		Mode	Kön	Ålder	M+K	M+Å	K+Å
Frihetsgrader	14	12	12	12	8	8	8
- extra fg	-	-	-	1	1	4	4
LRT	19.172	18.903	17.927	16.199	10.987	12.552	13.175
- p-värde	0.158	0.091	0.118	0.182	0.202	0.128	0.106
Read-Cressie	19.522	19.117	18.373	16.680	10.727	12.613	13.317
- p-värde	0.146	0.0858	0.105	0.162	0.218	0.126	0.101
Pearson $\chi^2$	19.882	19.414	18.732	17.036	10.652	12.710	13.463
- p-värde	0.134	0.079	0.095	0.148	0.222	0.122	0.0969
AIC	5712.6	5716.3	5715.4	5713.6	5716.4	5718.0	5718.6
BIC	5794.7	5808.1	5807.1	5805.4	5827.5	5829.0	5829.7
Diss.index	0.03290	0.03190	0.03291	0.03081	0.01702	0.02362	0.02493

**Emp****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg100a-d
- MM, återintervju: 18a-d
- CATI: ak1-4

**Selektion:**

Frågan besvaras endast om Up är född 1932 eller senare (74 år eller yngre)

**Ungefärlig frågelydelse i MM, återintervjun:**

Vad var aktuellt under förra veckan:

- Anställd på heltid
- Anställd på deltid
- Arbetade eller var ledig från familj jordbruk
- Arbetade eller var ledig från familj företag

**Kodning / dikotomisering:**

Variabelvärdet = 1 om "Antal Ja på 18a-d i ÅI-MM" > 0, dvs. om sysselsatt, 0 annars.

**Data**

Antal observationer: 864

	Mixed mode				CATI							
	Unga		Gamla		Unga		Gamla					
	0	1	0	1	0	1	0	1				
Män	0	26	2	0	25	4	0	31	1	0	28	1
	1	1	90	1	0	57	1	2	117	1	0	55
Kvinnor	0	42	6	0	23	1	0	36	2	0	29	1
	1	5	90	1	3	41	1	4	91	1	0	50

**Resultat av modellering**

	Ingen faktor	1 faktor			2 faktorer		
		Mode	Kön	Ålder	M+K	M+Å	K+Å
Frihetsgrader	14	12	12	12	8	8	8
- extra fg	1	2	1	2	4	3	4
LRT	20.554	15.647	17.930	18.332	12.729	12.907	15.747
- p-värde	0.114	0.208	0.118	0.106	0.122	0.115	0.046
Read-Cressie	17.158	12.828	15.949	16.909	10.349	10.674	14.280
- p-värde	0.248	0.382	0.194	0.153	0.241	0.221	0.075
Pearson $\chi^2$	16.652	12.368	16.142	17.316	9.968	10.249	14.608
- p-värde	0.275	0.417	0.185	0.138	0.267	0.248	0.067
AIC	4869.8	4868.9	4871.2	4871.6	4874.0	4874.1	4877.0
BIC	4950.7	4959.3	4961.6	4962.0	4983.5	4983.7	4986.5
Diss.index	0.01621	0.01293	0.01454	0.01539	0.00893	0.01047	0.01174

**Frg100e****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg100e
- MM, återintervju: 18e
- CATI: arbl

**Selektion:**

Frågan besvaras endast om Up är född 1932 eller senare (74 år eller yngre)

**Ungefärlig frågelydelse i MM, återintervjun:**

Vad var aktuellt under förra veckan:

- Arbetslös, sökte eller väntade på jobb hela veckan

**Kodning / dikotomisering:**

Variabelvärdet = 1 om ursprungsvariabeln besvaras med Ja (1), dvs. arbetslös, 0 annars.

**Data**

Antal observationer: 864

	Mixed mode						CATI					
	Unga			Gamla			Unga			Gamla		
	0	1		0	1		0	1		0	1	
Män	0	111	4	0	81	1	0	135	1	0	82	1
	1	3	1	1	2	2	1	6	9	1	0	1
Kvinnor	0	125	5	0	66	1	0	115	6	0	77	0
	1	6	7	1	0	1	1	6	6	1	1	2

**Resultat av modellering**

	Ingen faktor	1 faktor			2 faktorer		
		Mode	Kön	Ålder	M+K	M+Å	K+Å
Frihetsgrader	14	12	12	12	8	8	8
- extra fg	1	2	1	1	3	2	4
LRT	14.745	13.516	13.417	11.757	11.250	10.571	9.012
- p-värde	0.396	0.333	0.340	0.465	0.188	0.227	0.341
Read-Cressie	11.924	11.000	10.869	8.670	9.457	8.628	7.434
- p-värde	0.612	0.529	0.540	0.731	0.305	0.375	0.491
Pearson $\chi^2$	11.150	10.299	10.321	9.327	9.0733	8.118	7.067
- p-värde	0.674	0.590	0.588	0.675	0.336	0.422	0.529
AIC	4191.4	4194.2	4194.1	4192.4	4199.9	4199.2	4197.7
BIC	4272.4	4284.7	4284.6	4282.9	4309.4	4308.8	4307.2
Diss.index	0.01091	0.01152	0.01099	0.01240	0.00789	0.00899	0.00732

**Frg221b****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg221b
- MM, återintervju: 27
- CATI: m221b

**Ungefärlig frågelydelse i MM, återintervjun:**

Är du medlem i något politiskt parti?

**Kodning / dikotomisering:**

Variabelvärdet = 1 om ursprungsvariabeln besvaras med Ja (1), 0 annars.

**Data**

Antal observationer: 923

	Mixed mode						CATI					
	Unga			Gamla			Unga			Gamla		
	0	1		0	1		0	1		0	1	
Män	0	115	1	0	85	4	0	147	1	0	85	1
	1	1	2	1	0	11	1	1	2	1	2	6
Kvinnor	0	137	1	0	75	0	0	126	2	0	92	2
	1	1	4	1	0	6	1	1	4	1	0	8

**Resultat av modellering**

	Ingen faktor	1 faktor			2 faktorer		
		Mode	Kön	Ålder	M+K	M+Å	K+Å
Frihetsgrader	14	12	12	12	8	8	8
- extra fg	-	-	1	1	2	3	2
LRT	13.773	13.528	12.030	13.029	10.743	12.361	11.170
- p-värde	0.467	0.332	0.443	0.367	0.217	0.136	0.192
Read-Cressie	10.961	10.947	9.548	10.498	8.168	10.288	9.436
- p-värde	0.689	0.533	0.655	0.572	0.417	0.245	0.307
Pearson $\chi^2$	10.597	10.691	9.212	10.159	7.675	10.181	9.400
- p-värde	0.717	0.556	0.684	0.602	0.466	0.252	0.310
AIC	4359.9	4363.7	4362.2	4363.2	4368.9	4370.5	4369.3
BIC	4442.0	4455.4	4453.9	4454.9	4479.9	4481.6	4480.4
Diss.index	0.00605	0.00617	0.00503	0.00560	0.00394	0.00510	0.00422



**Frg224a****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg224a
- MM, återintervju: 29
- CATI: m224a

**Ungefärlig frågelydelse i MM, återintervjun:**

Skulle du själv kunna författa en skrivelse och överklaga ett beslut av en myndighet?

**Kodning / dikotomisering:**

Variabelvärdet = 1 om ursprungsvariabeln besvaras med Ja (1), 0 annars.

**Data**

Antal observationer: 923

	Mixed mode				CATI							
	Unga		Gamla		Unga		Gamla					
	0	1	0	1	0	1	0	1				
Män	0	23	11	0	16	7	0	35	16	0	17	4
	1	3	82	1	4	73	1	5	95	1	3	70
Kvinnor	0	25	20	0	24	10	0	39	4	0	28	7
	1	7	91	1	1	46	1	4	86	1	4	63

**Resultat av modellering**

	Ingen faktor	1 faktor			2 faktorer		
		Mode	Kön	Ålder	M+K	M+Å	K+Å
Frihetsgrader	14	12	12	12	8	8	8
- extra fg	1	1	2	1	3	2	3
LRT	40.661	37.136	37.388	42.601	30.155	35.000	42.363
- p-värde	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Read-Cressie	42.336	36.579	37.795	42.708	28.556	34.403	42.447
- p-värde	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Pearson $\chi^2$	43.870	36.984	38.561	43.392	28.218	34.651	43.144
- p-värde	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AIC	5570.1	5570.6	5570.8	5576.0	5571.6	5576.4	5583.8
BIC	5652.2	5662.3	5662.6	5667.8	5682.6	5687.5	5694.8
Diss.index	0.04705	0.04191	0.04223	0.04396	0.03310	0.03781	0.04250

**Frg250a****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg250a
- MM, återintervju: 30a
- CATI: t250a

**Ungefärlig frågelydelse i MM, återintervjun:**

Har du under de senaste 12 månaderna blivit utsatt för våld som lämnade synliga skador och som resulterade i läkarbesök?

**Kodning / dikotomisering:**

Variabelvärdet = 1 om ursprungsvariabeln ej är besvarad med Nej, 0 annars.

**Data**

Antal observationer: 923

	Mixed mode						CATI					
	Unga		Gamla		Unga		Gamla		Unga		Gamla	
	0	1	0	1	0	1	0	1	0	1	0	1
Män	0	113	0	0	100	0	0	148	1	0	94	0
	1	2	4	1	0	0	1	0	2	1	0	0
Kvinnor	0	143	0	0	79	0	0	133	0	0	100	1
	1	0	0	1	1	1	1	0	0	1	1	0

**Resultat av modellering**

	Ingen faktor	1 faktor			2 faktorer		
		Mode	Kön	Ålder	M+K	M+Å	K+Å
Frihetsgrader	14	12	12	12	8	8	8
- extra fg	5	6	6	6	-	-	6
LRT	9.260	8.630	7.271	7.271	-	-	5.702
- p-värde	0.814	0.734	0.839	0.839	-	-	0.681
Read-Cressie	6.708	6.464	5.371	5.371	-	-	4.203
- p-värde	0.945	0.891	0.944	0.944	-	-	0.838
Pearson $\chi^2$	6.190	6.067	5.000	5.000	-	-	3.909
- p-värde	0.961	0.913	0.958	0.958	-	-	0.865
AIC	3976.1	3979.4	3978.1	3978.1	-	-	3984.5
BIC	4058.1	4071.2	4069.8	4069.8	-	-	4095.6
Diss.index	0.00204	0.00200	0.00149	0.00149	-	-	0.00152

**Frg251a****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg251a
- MM, återintervju: 32a
- CATI: t251a

**Ungefärlig frågelydelse i MM, återintervjun:**

Har du under de senaste 12 månaderna blivit utsatt för våld som lämnade synliga skador men som inte resulterade i läkarbesök?

**Kodning / dikotomisering:**

Variabelvärdet = 1 om ursprungsvariabeln ej är besvarad med Nej, 0 annars.

**Data**

Antal observationer: 923

	Mixed mode						CATI					
	Unga			Gamla			Unga			Gamla		
	0	1		0	1		0	1		0	1	
Män	0	109	5	0	98	1	0	142	1	0	92	1
	1	2	3	1	1	0	1	4	4	1	0	1
Kvinnor	0	129	8	0	79	0	0	127	0	0	98	2
	1	3	3	1	1	1	1	1	5	1	1	1

**Resultat av modellering**

	Ingen faktor	1 faktor			2 faktorer		
		Mode	Kön	Ålder	M+K	M+Å	K+Å
Frihetsgrader	14	12	12	12	8	8	8
- extra fg	1	2	2	3	4	5	5
LRT	19.969	15.210	19.204	13.295	12.431	18.705	18.158
- p-värde	0.131	0.230	0.084	0.102	0.133	0.0959	0.020
Read-Cressie	17.500	13.331	17.059	11.642	10.740	16.467	16.104
- p-värde	0.231	0.345	0.147	0.168	0.217	0.171	0.041
Pearson $\chi^2$	17.209	13.228	16.843	11.408	10.465	16.639	15.864
- p-värde	0.245	0.352	0.156	0.180	0.234	0.164	0.044
AIC	4310.7	4309.9	4313.9	4316.0	4315.2	4313.4	4320.9
BIC	4392.8	4401.7	4405.7	4427.1	4426.2	4405.2	4431.9
Diss.index	0.01174	0.01162	0.01262	0.00993	0.00907	0.01214	0.01191

**Frg252a****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg252a
- MM, återintervju: 34a
- CATI: t252a

**Ungefärlig frågelydelse i MM, återintervjun:**

Har du under de senaste 12 månaderna blivit utsatt för våld som inte lämnade synliga skador?

**Kodning / dikotomisering:**

Variabelvärdet = 1 om ursprungsvariabeln ej är besvarad med Nej, 0 annars.

**Data**

Antal observationer: 923

	Mixed mode						CATI					
	Unga		Gamla		Unga		Gamla		Unga		Gamla	
	0	1	0	1	0	1	0	1	0	1	0	1
Män	0	105	8	0	97	2	0	139	4	0	90	2
	1	4	2	1	1	0	1	2	6	1	1	1
Kvinnor	0	130	8	0	79	0	0	128	2	0	98	1
	1	1	4	1	1	1	1	0	3	1	1	2

**Resultat av modellering**

	Ingen faktor	1 faktor			2 faktorer		
		Mode	Kön	Ålder	M+K	M+Å	K+Å
Frihetsgrader	14	12	12	12	8	8	8
- extra fg	1	1	2	3	3	4	3
LRT	21.924	17.595	19.523	19.677	13.485	17.063	18.377
- p-värde	0.080	0.129	0.077	0.073	0.096	0.029	0.019
Read-Cressie	19.029	14.664	17.990	17.295	11.661	14.374	16.686
- p-värde	0.164	0.260	0.116	0.139	0.167	0.073	0.034
Pearson $\chi^2$	18.492	13.991	17.880	16.865	11.277	13.775	16.471
- p-värde	0.185	0.301	0.119	0.155	0.186	0.088	0.036
AIC	4371.9	4371.6	4373.5	4373.7	4375.5	4379.0	4380.4
BIC	4453.9	4463.3	4465.2	4465.4	4486.5	4490.1	4491.4
Diss.index	0.01300	0.01121	0.01504	0.01348	0.00952	0.01140	0.01248

**Frg253a****Ursprungsvariabelns beteckning:**

- MM, ordinarie intervju: frg253a
- MM, återintervju: 36a
- CATI: t253a

**Ungefärlig frågelydelse i MM, återintervjun:**

Har du under de senaste 12 månaderna blivit utsatt för hot om våld som föranledde rädsla?

**Kodning / dikotomisering:**

Variabelvärdet = 1 om ursprungsvariabeln ej är besvarad med Nej, 0 annars.

**Data**

Antal observationer: 923

	Mixed mode						CATI					
	Unga			Gamla			Unga			Gamla		
	0	1		0	1	0	1	0	1	0	1	
Män	0	106	5	0	94	2	0	137	2	0	91	1
	1	6	2	1	1	3	1	2	10	1	1	1
Kvinnor	0	118	7	0	76	1	0	122	2	0	95	3
	1	5	13	1	2	2	1	2	7	1	2	2

**Resultat av modellering**

	Ingen faktor	1 faktor			2 faktorer		
		Mode	Kön	Ålder	M+K	M+Å	K+Å
Frihetsgrader	14	12	12	12	8	8	8
- extra fg	-	1	1	1	2	4	3
LRT	12.947	7.609	12.556	12.338	4.209	1.468	7.027
- p-värde	0.531	0.815	0.402	0.419	0.838	0.993	0.534
Read-Cressie	12.772	7.582	12.085	11.748	4.227	1.460	6.958
- p-värde	0.545	0.817	0.439	0.466	0.836	0.993	0.541
Pearson $\chi^2$	12.834	7.650	11.976	11.612	4.271	1.459	6.977
- p-värde	0.540	0.812	0.448	0.477	0.832	0.993	0.539
AIC	4555.5	4554.2	4559.1	4558.9	4558.8	4556.0	4561.6
BIC	4637.6	4645.9	4650.8	4650.6	4669.8	4667.1	4672.6
Diss.index	0.01627	0.01401	0.01526	0.01347	0.00691	0.00431	0.01361

## Bilaga C. Resultat från studien rörande differenser i bortfallskevheter

Skuggad cell indikerar att hypotesen  $B_{CATI} - B_{MM} = 0$  förkastas på 5% signifikansnivå ( $|\hat{Z}| > 2$ ).

**Tabell C.1a: Befolkningen fördelad efter utbildningsnivå samt efter kön och åldersklasser (CATI)**

Kön Ålder, år	$\hat{B}_{CATI}$		
	Utbildningsnivå enligt utbildningsregistret		
	Förgymnasial utbildning	Gymnasial utbildning	Eftergymnasial utbildning
<u>Män</u>			
16-34	-16967	-4625	50277
35-64	-105607	-2388	72246
65-84	-17890	1155	16735
16-84	-140464	-5859	139258
<u>Kvinnor</u>			
16-34	687	-48794	18553
35-64	-50919	25611	61927
65-84	-24863	18318	6545
16-84	-75095	-4866	87026
<u>Båda könen</u>			
16-34	-16280	-53420	68831
35-64	-156527	23223	134173
65-84	-42753	19472	23280
16-84	-215559	-10725	226284

**Tabell C.1b: Befolkningen fördelad efter utbildningsnivå samt efter kön och åldersklasser (MM)**

Kön Ålder, år	$\hat{B}_{MM}$		
	Utbildningsnivå enligt utbildningsregistret		
	Förgymnasial utbildning	Gymnasial utbildning	Eftergymnasial utbildning
<u>Män</u>			
16-34	-15008	7592	25678
35-64	-36683	-9441	27690
65-84	-39429	24432	14997
16-84	-91119	22583	68364
<u>Kvinnor</u>			
16-34	-16297	-29550	43275
35-64	-24296	-25532	52573
65-84	-17642	4166	13477
16-84	-58236	-50916	109324
<u>Båda könen</u>			
16-34	-31305	-21958	68952
35-64	-60979	-34972	80262
65-84	-57071	28597	28474
16-84	-149355	-28333	177688

**Tabell C.1c: Befolkningen fördelad efter utbildningsnivå samt efter kön och åldersklasser (differens mellan CATI och MM). Skattat medelfel inom parentes**

Kön Ålder, år	$\hat{B}_{CATI} - \hat{B}_{MM}$		
	Utbildningsnivå enligt utbildningsregistret		
	Förgymnasial utbildning	Gymnasial utbildning	Eftergymnasial utbildning
<u>Män</u>			
16-34	-1959 (28350)	-12217 (32610)	24599 (22631)
35-64	-68925 (33348)	7052 (41210)	44556 (32621)
65-84	21539 (18380)	-23277 (17178)	1738 (8989)
16-84	-49345 (46764)	-28442 (50363)	70894 (38869)
<u>Kvinnor</u>			
16-34	16984 (28152)	-19244 (31860)	-24721 (23348)
35-64	-26623 (28133)	51143 (37398)	9355 (32080)
65-84	-7220 (19355)	14152 (17122)	-6932 (11692)
16-84	-16859 (43692)	46050 (47909)	-22298 (39329)
<u>Båda könen</u>			
16-34	15025 (34604)	-31462 (43153)	-122 (32392)
35-64	-95548 (43629)	58195 (55649)	53911 (45751)
65-84	14318 (26691)	-9125 (24254)	-5193 (14748)
16-84	-66205 (60803)	17609 (67937)	48596 (55222)



**Tabell C.2a: Befolkningen fördelad efter sysselsättning samt efter kön och åldersklasser (CATI)**

Kön Ålder, år	$\hat{B}_{CATI}$		
	Sysselsättningsstatus enligt RAMS		
	Sysselsatt	Ej sysselsatt	Befolkning
<u>Män</u>			
16-34	39429	-10745	28684
35-64	61805	-97555	-35749
65-84	3260	-3260	0
16-84	104494	-111559	-7065
<u>Kvinnor</u>			
16-34	-13910	-15643	-29554
35-64	95579	-58960	36619
65-84	-856	856	0
16-84	80813	-73748	7065
<u>Båda könen</u>			
16-34	25518	-26388	-870
35-64	157385	-156515	870
65-84	2404	-2404	0
16-84	185307	-185307	0

**Tabell C.2b: Befolkningen fördelad efter sysselsättning samt efter kön och åldersklasser (MM)**

Kön Ålder, år	$\hat{B}_{MM}$		
	Sysselsättningsstatus enligt RAMS		
	Sysselsatt	Ej sysselsatt	Befolkning
<u>Män</u>			
16-34	7215	11047	18261
35-64	23799	-42233	-18434
65-84	3269	-3269	0
16-84	34283	-34455	-172
<u>Kvinnor</u>			
16-34	27053	-29626	-2572
35-64	4694	-1949	2745
65-84	10019	-10019	0
16-84	41766	-41594	172
<u>Båda könen</u>			
16-34	34268	-18579	15689
35-64	28493	-44182	-15689
65-84	13288	-13288	0
16-84	76049	-76049	0

**Tabell C.2c: Befolkningen fördelad efter sysselsättning samt efter kön och åldersklasser (differens mellan CATI och MM)**

Kön Ålder, år	$\hat{B}_{CATI} - \hat{B}_{MM}$		
	Sysselsättningsstatus enligt RAMS		
	Sysselsatt	Ej sysselsatt	Befolkning
<b>Män</b>			
16-34	32214 (34082)	-21791 (31177)	10423 (38840)
35-64	38006 (38189)	-55322 (31677)	-17316 (28948)
65-84	-9 (12209)	9 (12209)	0 (0)
16-84	70211 (42348)	-77104 (45098)	-6893 (25895)
<b>Kvinnor</b>			
16-34	-40964 (31707)	13982 (33237)	-26982 (38559)
35-64	90886 (36794)	-57011 (29111)	33874 (28570)
65-84	-10875 (12481)	10875 (12481)	0 (0)
16-84	39047 (41395)	-32154 (44685)	6893 (25895)
<b>Båda könen</b>			
16-34	-8750 (45121)	-7809 (37710)	-16559 (40672)
35-64	128891 (53030)	-112333 (43022)	16559 (40672)
65-84	-10884 (17459)	10884 (17459)	0 (0)
16-84	109258 (58103)	-109258 (58103)	0 (0)

**Tabell C.3a: Befolkningen fördelade efter sysselsättningsstatus enligt RAMS, medborgarskap och födelseland (CATI)**

Medborgarskap/ Födelseland Ålder, år	$\hat{B}_{CATI}$		
	Sysselsättningsstatus enligt RAMS		
	Sysselsatt	Ej sysselsatt	Befolkning
Svenska medborgare 16-84	193496	-107278	86218
Utländska medborgare 16-84	-8189	-78029	-86218
Inrikes födda 16-84	193511	-20538	172973
Utrikes födda 16-84	-8204	-164769	-172973
Samtliga 16-84	185307	-185307	0

**Tabell C.3b: Befolkningen fördelade efter sysselsättningsstatus enligt RAMS, medborgarskap och födelse land (MM)**

Medborgarskap/ Födelse land Ålder, år	$\hat{B}_{MM}$		
	Sysselsättningsstatus enligt RAMS		
	Sysselsatt	Ej sysselsatt	Befolkning
Svenska medborgare 16-84	76718	-14328	62390
Utländska medborgare 16-84	-669	-61721	-62390
Inrikes födda 16-84	87086	-30409	56677
Utrikes födda 16-84	-11037	-45641	-56677
Samtliga 16-84	76049	-76049	0

**Tabell C.3c: Befolkningen fördelade efter sysselsättningsstatus enligt RAMS, medborgarskap och födelse land (differens mellan CATI och MM)**

Medborgarskap/ Födelse land Ålder, år	$\hat{B}_{CATI} - \hat{B}_{MM}$		
	Sysselsättningsstatus enligt RAMS		
	Sysselsatt	Ej sysselsatt	Befolkning
Svenska medborgare			
16-84	116778 (59146)	-92950 (56564)	23828 (35184)
Utländska medborgare			
16-84	-7520 (19669)	-16308 (29963)	-23828 (35184)
Inrikes födda			
16-84	106425 (61440)	9870 (56168)	116296 (52146)
Utrikes födda			
16-84	2833 (34867)	-119128 (41699)	-116296 (52146)
Samtliga			
16-84	109258 (58103)	-109258 (58103)	0 (0)

**Tabell C.4a: Befolkningen fördelad efter lönesummeklasser och åldersgrupper (CATI)**

Lönesumma, 100-tal kronor	$\hat{B}_{CATI}$			
	Ålder, år			
	16-19	20- 64	65-84	16-84
0 -19	16910	-215702	-9642	-208433
20 - 229 0	21932	-80344	6992	-51421
229 1 - 325 5	0	137364	-2551	134813
325 6 -	0	119841	5200	125041

**Tabell C.4b: Befolkningen fördelad efter lönesummeklasser och åldersgrupper (MM)**

Lönesumma, 100-tal kronor	$\hat{B}_{MM}$			
	Ålder			
	16-19	20- 64	65-84	16-84
0 -19	11850	-121643	-32746	-142539
20 - 229 0	-264	12795	21016	33547
229 1 - 325 5	0	26549	6975	33523
325 6 -	0	70714	4755	75469

**Tabell C.4c: Befolkningen fördelad efter lönesummeklasser och åldersgrupper (differens mellan CATI och MM)**

Lönesumma, 100-tal kronor	$\hat{B}_{CATI} - \hat{B}_{MM}$			
	Ålder			
	16-19	20- 64	65-84	16-84
0 -19	5060 (23168)	-94059 (50859)	23104 (19001)	-65894 (57763)
20 - 229 0	22196 (17739)	-93139 (60054)	-14024 (17982)	-84968 (62661)
229 1 - 325 5	0 (0)	110815 (50451)	-9525 (6045)	101290 (50812)
325 6 -	0 (0)	49127 (41414)	446 (3776)	49572 (41586)

## Bilaga D. Resultat motivationsfrågor

### Frg38

#### Ursprungsvariabelns beteckning:

– MM frgny38\_M2

– CATI: SLUT38\_C2

#### Ungefärlig frågelydelse:

Hur gärna eller hur motvilligt deltog du i den här undersökningen?

**Tabell D.1 Svartalternativ och svarsfördelning, Frg38**

Kod	Svartalternativ	MM		CATI	
		Frekvens	%	Frekvens	%
1	Mycket gärna	173	39.05	147	30.63
2	Ganska gärna	172	38.83	211	43.96
3	Varken gärna eller motvilligt	81	18.28	98	20.42
4	Ganska motvilligt	15	3.39	23	4.79
5	Mycket motvilligt	2	0.45	0	0
9	ej svar	0	0	1	0.21
	Summa	443	100	480	100

### Frg39

#### Ursprungsvariabelns beteckning:

– MM frgny39\_M2

– CATI: SLUT39\_C2

#### Ungefärlig frågelydelse:

Var det jobbigt att vara med eller gick det lätt?

**Tabell D.2 Svartalternativ och svarsfördelning, Frg39**

Kod	Svartalternativ	MM		CATI	
		Frekvens	%	Frekvens	%
1	Jobbigt	12	2.71	15	3.13
2	Varken lätt eller jobbigt	35	7.90	58	12.08
3	Lätt	396	89.39	406	84.58
9	ej svar	0	0	1	0.21
	Summa	443	100	480	100



**Frg40****Ursprungsvariabelns beteckning:**

- MM frgny40\_M2
- CATI: SLUT40\_C2

**Ungefärlig frågelydelse:**

Skulle du kunna tänka dig att vara med i en liknande undersökning om "LF" om t.ex. ett år?

**Tabell D.3 Svartalernativ och svarsfördelning, Frg40**

Kod	Svartalernativ	MM		CATI	
		Frekvens	%	Frekvens	%
1	Ja, absolut	228	51.47	261	54.38
2	Ja, kanske	181	40.86	175	36.46
3	Troligen inte	19	4.29	30	6.25
4	Absolut inte	12	2.71	7	1.46
8	ej svar	2	0.45	0	0
9	ej svar	1	0.23	7	1.46
	Summa	443	100	480	100

## Bilaga E: Avsnitt 3, preliminära versionen

### Metod

Den klassiska mätmodellen brukar formuleras som  $Y = T + E$  där  $Y$  är ett observerat (manifest) värde,  $E$  ett slumpfel och  $T$  ett hypotetiskt sant (latent) värde. Under antagandet om att mätfelet och det sanna värdet är oberoende stokastiska variabler följer att  $Y$  är en stokastisk variabel med varians lika med summan av varianserna hos  $T$  och  $E$ . Reliabiliteten  $R$  i mätningen av  $T$  via  $Y$  definieras som kvoten mellan dessas varianser enligt

$$R = \frac{\text{Var}(T)}{\text{Var}(Y)} = \frac{\text{Var}(T)}{\text{Var}(T) + \text{Var}(E)}$$

Det följer av antagandet och definitionen att  $0 \leq R \leq 1$ . Om effekten av mätfelet är ringa (liten relativ varians i  $E$ ) ligger reliabiliteten nära 1, om den är betydande (stor relativ varians i  $E$ ) ligger det nära noll. Följande antaganden kan användas för att illustrera en situation med upprepade mätningar:

- Varje individ i populationen har en hypotetisk fördelning av möjliga responser.
- Ett obundet slumpmässigt urval av storlek  $n$  individer väljs ur en population av storlek  $N$ .
- Varje individ betraktas som ett kluster av potentiella responser som observeras (eller dras med återläggning) oberoende av varandra.

En undersökning där det föreligger mätfel kan med andra ord ses som ett två-steps klusterurval med OSU i varje steg (utan återläggning i första, med återläggning i andra). Låt  $m$  beteckna antalet mätningar som görs hos varje individ. Med en intervju-återintervju design är  $m = 2$  och man observerar två slumpmässiga och oberoende utfall från varje individs hypotetiska fördelning av möjliga responser. Dessutom antas oberoende mellan olika individers respektive responser. Antag att man är intresserad av populationsparametern

$$\bar{Y} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M y_{ij}$$

(då det är fråga om urval med återläggning i andra steget bör  $M$  egentligen betraktas som oändligt stort varför man istället hellre bör betrakta parametern ovan som genomsnittet av varje individs väntevärde över respektive fördelning av responser). Populationsgenomsnittet skattas med

$$\bar{y} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m y_{ij}$$

med varians

$$\text{Var}(\bar{y}) = (1-f) \frac{S_1^2}{n} + \frac{S_2^2}{nm} = (1-f) \frac{SV}{n} + \frac{SRV}{nm}$$

där  $f$  betecknar urvalsfraktionen i första steget och

$$S_1^2 = SV, \quad \text{samplingvariens, "between primary sampling units"}$$

$$S_2^2 = SRV, \quad \text{responsvariens, "within primary sampling units"}$$

I fallet med endast en mätning per individ ( $m = 1$ ) och en försumbar urvalsfraktion ( $f \approx 0$ ) gäller att

$$\text{Var}(\bar{y}) = \frac{SV + SRV}{n}$$

Om variansen för de enskilda responserna  $SRV$  är liten blir den totala variansen i skattningen liten. Reliabilitetsknoten (*reliability ratio*) definieras alltså analogt med den inledande beskrivningen enligt

$$R = \frac{SV}{SV + SRV}$$

och den ekvivalenta men omvända s.k. inkonsistenskvoten (*ratio of inconsistency*) enligt

$$I = 1 - R$$

Inkonsistensen (alternativt reliabiliteten) i en variabel säger oss med andra ord något om hur mycket av den totala variansen (osäkerheten) i en skattning som kan hänföras till mätfelet. Ju lägre reliabilitet, desto större osäkerhet.

För en dikotom responsvariabel i en intervju-återintervjustudie gäller givetvis samma resultat som ovan men med något enklare formler i slutändan. Antag att vi har följande fyrfältstabell med observationer:

		Återintervju, $y_{i2}$		
		0	1	
Intervju, $y_{i1}$	0	$a$	$b$	
	1	$c$	$d$	
				$n = a+b+c+d$

där cellerna representerar frekvenser av respektive kombination av responserna 0 och 1. Definiera  $g$  som andelen respondenter som svarar olika från en gång till nästa (*gross different rate*), dvs.

$$g = \frac{b+c}{n} = \frac{\sum_{i=1}^n (y_{i1} - y_{i2})^2}{n}$$

Definiera vidare

$$p_j = \frac{1}{n} \sum_{i=1}^n y_{ij}, \quad j = 1, 2$$

dvs. andel ettor från respektive intervjutillfälle 1 och 2. Då kan det relativt enkelt visas att  $SRV$  och  $(SV+SRV)$  kan skattas med

$$srv = \frac{g}{2} = \frac{b+c}{2n}$$

respektive

$$sv + srv = \frac{p_1(1-p_2) + p_2(1-p_1)}{2} = \frac{(c+d)(a+c) + (b+d)(a+b)}{2n^2}$$

och därmed kan inkonsistensen  $I$  och reliabiliteten  $R$  skattas med respektive

$$\hat{I} = \frac{g}{p_1q_2 + p_2q_1} = \frac{n(b+c)}{(c+d)(a+c) + (b+d)(a+b)}$$

$$\hat{R} = 1 - \hat{I}$$

Den förra brukar benämnas inkonsistensindex (*index of inconsistency*) och är ett mått på hur stor andel av den totala variansen som kan hänföras till mätfelet, den senare som kan visas vara identisk med Cohens kappas mäter resterande andel av variansen.

### Resultat och slutsatser

Från återinterrvyblanketten valdes 17 frågor ut. Dessa dikotomiserades enligt beskrivningarna i Bilaga A varefter  $g$  (*gross difference rate*) och inkonsistensindex  $\hat{I}$  för vardera mode (MM respektive CATI) beräknades. Resultaten redovisas i Bilaga A och sammanfattas i Tabell 3.2 nedan.

**Tabell 3.2 Gross difference rate i %-enheter och Inkonsistensindex  $\hat{I}$  för ett urval av frågor i ULF för respektive MM och CATI**

Fråga	Gross difference rate %		Inkonsistensindex	
	MM	CATI	MM	CATI
Frg004d	2.88	1.47	0.060	0.030
Frg005	1.21	2.53	0.030	0.061
Frg057	11.61	9.79	0.280	0.229
Frg071	2.81	3.08	0.493	0.349
Nyfrg072	10.84	7.29	0.217	0.149
Frg100	2.99	2.33	0.151	0.052
Frg122a	10.19	12.09	0.218	0.272
Frg122b	15.29	18.29	0.305	0.363
Frg122c	32.94	18.29	0.665	0.366
Frg123c	21.59	21.57	0.449	0.444
Frg124a	10.57	7.84	0.218	0.164
Frg221b	1.43	2.08	0.133	0.211
Frg224a	14.69	9.79	0.360	0.234
Frg250a	0.24	0.21	(1)	(1)
Frg251a	2.61	1.67	0.490	0.275
Frg252a	2.38	1.88	0.429	0.282
Frg253a	5.46	2.71	0.406	0.260

(1) För få observerade av den ena kategorin för att tillförlitligt kunna beräknas.

CATI uppvisar en lägre inkonsistens (högre reliabilitet) i 11 av 17 frågor, MM är bättre i 4 av 17. Frg123c får sägas vara oavgjord och Frg250a låter sig inte jämföras då data inte innehåller tillräckligt med observationer i kategorin  $Y = 0$  för att tillförlitligt kunna beräknas. Notera att jämförelserna inte är genomförda med något formellt test. Med Frg 250a borttagen blir det genomsnittliga indexvärdet ca 0.31 för MM och ca 0.23 för CATI. MM uppvisar också en större variation i inkonsistens, som lägst 0.03 och som högst 0.66. Motsvarande tal för CATI är 0.03 resp 0.44. Allmänt måste ett inkonsistensindex som är så högt som 0.66 betraktas som illa; det innebär att ca 66 % av variansen i en skattning kan hänföras till mätfelet.

## Bilaga F: Avsnitt 8, preliminära versionen

Med ledning av resultaten från de fem olika delstudierna kan följande slutsatser dras:

- Av reliabilitetsstudien framgick det att CATI uppvisar en högre reliabilitet i 11 av 17 frågor och att MM är bättre i fyra fall. Två frågor förblev oavgjorda. Med en fråga borttagen blev den genomsnittliga reliabiliteten ca 0.69 för MM och ca 0.77 för CATI. MM uppvisar också en större variation i reliabilitet med som lägst värden på ca 0.34. Den allmänna slutsatsen är att reliabiliteten är bättre i CATI.
- Av modelleringen och analysen av mätfelsmodeller med LCA metoden framgick det att i tolv av femton fall inte kunde påvisas någon effekt med avseende på insamlingsmetod. En fråga uppvisade en (svag) modeffekt och två frågor gav inte någon god anpassning till någon av modellerna. Den allmänna slutsatsen är att insamlingsmetod inte har en avgörande effekt på felklassificeringssannolikheterna i mätfelsmodellen i någon fråga utom möjligen en.
- CATI och MM skiljer sig inte i någon väsentlig utsträckning vad gäller den bortfallsskevhet som respektive metod ger upphov till. I den utsträckning statistiskt signifikanta skillnader föreligger, tenderar dock användandet av CATI att resultera i störst bortfallsskevhet, men skillnaderna är inte alarmerande. Dessutom bör en del av problemet med bortfallsskevhet, oavsett val av insamlingsmetod, kunna åtgärdas i löpande produktion, t.ex. genom ett effektivt utnyttjande av tillgänglig registerinformation.
- Skillnaden i bortfallet i Barn-ULF är helt klart signifikant ( $< 0.1\%$ ) och till fördel för CATI som insamlingsmetod med en ökning i deltagandet på nära 10 %-enheter. En förklaring till det större deltagandet bland barn i CATI är att respondenterna gavs en större möjlighet att själva välja när intervjun skulle äga rum. Det är inte utrett om deltagandet i MM skulle öka i motsvarande grad om respondenterna fick samma möjlighet att själva välja tidpunkt för barnen, med t.ex. erbjudande om ett återbesök.
- Motivationstudien kan inte påvisa någon signifikant skillnad utom möjligen i en av tre frågor. Den faktiska skillnaden kan dock inte betraktas som särskilt stor i detta fall (ca 89 % positiva i MM jämfört med ca 85 % i CATI). Slutsatsen är att en övergång till CATI inte i någon större utsträckning påverkar respondenternas motivation att delta i undersökningen.

Sammanfattningsvis gäller således att den systematiska felkomponenten (mät- och bortfallsfel) i tillförlitlighetskomponenten bedöms vara oförändrad vid en övergång till telefonintervjuer. I kombination med att övergången frigör resurser för en urvalsökning, innebär detta att medelkvadratfelet (MSE) i skattningarna skulle minska signifikant vid en övergång till telefon.

Övergången ger, för en given budget, ökat intervjuutrymme och därmed ökad utrymme för att förbättra statistikens innehåll. Vidare medför en urvalsökning ökade möjligheter för redovisningar av ytterligare undergrupper. En övergång från nuvarande metod (mixed mode) till datorstödd telefonintervju (CATI) leder dessutom till ökad aktualitet i och med att framställningstiden minskar.

Tabell 8.1. Översikt av effekterna av en övergång från MM till CATI

	Sämre	Lika	Bättre	Fas
<b>Innehåll</b>			<b>X</b>	
- Variabler			X	
- Redovisningsgrupper			X	
<b>Tillförlitlighet</b>			<b>X</b>	
- Urval/variants			X	1
- Mätning		X		2
- Svartsbortfall/bortfallsfel		(X)		2
<b>Aktualitet</b>			<b>X</b>	
- Framställningstid			X	
- Punktlighet		X		
<b>Jämförbarhet och sammanvändbarhet</b>				
- Jmf över tiden	(X)		-	
- Jmf mellan grupper	(X)		(X)	
- Samanvändbarhet			(X)	
<b>Tillgänglighet och förståelighet</b>		<b>X</b>		
<b>Övrigt</b>				
- Barn-ULF			X	2
- Motivation		X		2



**Boende**

- Fråga 1            *Anteckna typen av hus up bor i:*
- 1            FRILIGGANDE EN- ELLER TVÅFAMILJSVILLA
  - 2            SAMMANBYGGT RADHUS ELLER KEDJEHUS
  - 3            FLERBOSTADSHUS MED FÄRRE ÄN 10 LÄGENHETER
  - 4            FLERBOSTADSHUS MED 10 ELLER FLER LÄGENHETER
  - 5            ANNAT    *Vad?*
- .....

- 
- Fråga 2            På vilket sätt disponerar du/ni bostaden - äger du/ni eller någon annan familjemedlem huset, har du/ni lägenheten med bostadsrätt eller hyr du/ni bostaden?
- 1            ÄGANDERÄTT/ANDELSLÄGENHET/ARRENDE
  - 2            BOSTADSRÄTT
  - 3            HYRESRÄTT, FÖRSTA HAND
  - 4            HYRESRÄTT, ANDRA HAND
  - 5            TJÄNSTEBOSTAD
  - 6            INNEBOENDE (*Ej vuxna barn som bor hemma hos föräldrarna*)
  - 7            ANNAT, *ange vad:*
- .....

- 
- Fråga 3            Har din/era bostadsförhållanden ändrats på något sätt sedan förra intervjun den .... /... ?
- 1            JA
  - 2            NEJ → **Fråga 5**
- 

- Fråga 4            På vilket sätt?
- .....
- .....

**Fritid**

- 
- Fråga 5            Har du under de senaste 12 månaderna på **din fritid** gjort någon semesterresa eller nöjesresa i Sverige eller utomlands som varat minst en vecka? Vi tänker även på sådant som campingsemester, vara ute i fritidsbåt eller besöka släkt och vänner.
- 1            JA
  - 2            NEJ → **Fråga 8**



Fråga 6 Har du gjort någon sådan resa efter förra intervjun (den ....)?

- 1 JA  
2 NEJ → Fråga 8

Fråga 7 Hade du gjort någon sådan resa även före förra intervjun (den ....),  
(men under de senaste 12 månaderna)?

- 1 JA  
2 NEJ

### Hälsa

Fråga 8 Hur bedömer du ditt allmänna hälsotillstånd? Är det ...

- 1 mycket gott  
2 gott  
3 någorlunda  
4 dåligt  
5 mycket dåligt?

Fråga 9 Har du några kroniska eller långvariga sjukdomar eller besvär?

- 1 JA  
2 NEJ → Fråga 12

Fråga 10 Vad är det?

*Anteckna varje besvär och sjukdom up nämner så noga som möjligt*

*Besvär 1: .....*

.....

*Besvär 2: .....*

.....

*Besvär 3: .....*

.....

*Besvär 4: .....*

.....

*Besvär 5: .....*

.....

*Besvär 6: .....*

.....

#### Uppföljningsfrågor vid behov:

- a) Kan du förklara detta lite närmare?      b) Vad sade läkaren att det var?      c) Var i kroppen har du dessa besvär?

Fråga 11	Har du därutöver någon annan kronisk eller långvarig sjukdom eller besvär?
1	JA → Fråga 10
2	NEJ

Fråga 12	Tar du regelbundet medicin för någonting (annat)?
1	JA → Fråga 10
2	NEJ

**Om up angett besvär vid fråga 10 fyll i fråga 13 kolumnvis.  
Annars → Fråga 15**

Besvär nr: (enligt fråga 10)

Fråga 13 a)	När hade du första gången besvär av (...)	1	2	3	4	5	6
<b>Svarskort 76 a</b>							
	FÖR MINDRE ÄN 3 MÅNADER SEDAN .....	1	1	1	1	1	1
	FÖR 3 MÅNADER SEDAN MEN MINDRE 1 ÅR SEDAN ..	2	2	2	2	2	2
	FÖR 1 ÅR SEDAN MEN MINDRE ÄN 5 ÅR SEDAN .....	3	3	3	3	3	3
	FÖR 5 ÅR SEDAN MEN MINDRE ÄN 10 ÅR SEDAN..	4	4	4	4	4	4
	FÖR 10 ÅR SEDAN MEN MINDRE ÄN 20 ÅR SEDAN	5	5	5	5	5	5
	FÖR 20 ÅR SEDAN ELLER TIDIGARE .....	6	6	6	6	6	6
	SEDAN FÖDELSEN .....	7	7	7	7	7	7
b)	Hur ofta har du besvär av (...)? Är det ...						
	hela tiden .....	1	1	1	1	1	1
	ofta .....	2	2	2	2	2	2
	då och då eller .....	3	3	3	3	3	3
	ganska sällan eller inte alls .....	4	4	4	4	4	4
c)	Är besvären ...						
	obetydliga eller inga alls .....	1	1	1	1	1	1
	måttliga.....	2	2	2	2	2	2
	svåra eller .....	3	3	3	3	3	3
	mycket svåra .....	4	4	4	4	4	4
d)	Får du regelbunden medicinsk behandling eller står du under medicinsk kontroll för ( ... )?						
	JA	1	1	1	1	1	1
	NEJ	2	2	2	2	2	2

**Om up har besvär enligt fråga 10. Annars → Fråga 15**

- Fråga 14 a) Medför den/de sjukdom(ar) du har att din arbetsförmåga är nedsatt .....
- (Till **alderspensionärer**: Hindrar den/de sjukdom(ar) du har dig i några sysselsättningar ...)
- 1 ... i hög grad  
2 ... i någon mån eller  
3 ... inte alls? → Fråga 14 c (OBS!)

**Om up har endast ett besvär → Fråga 14 c**

- b) Vilken av sjukdomarna påverkar din arbetsförmåga/dina sysselsättningar mest?  
Besvär (nr) ..... (enligt fråga 10)
- c) Medför den/de sjukdom(ar) du har att du behöver hjälp med något för att klara av ditt dagliga liv t.ex. sköta hygien, laga mat, handla, gå till banken, etc.?
- 1 JA  
2 NEJ

- Fråga 15 Jag har redan frågat om sjukdomar men det är några sjukdomar eller besvär som jag för säkerhets skull vill ställa några särskilda frågor om.  
Har du möjligen något av följande?

**Svarskort 90****Läs upp besvären för up i tur och ordning.****Om up svarar JA - fråga om besvären är svåra eller lätta**

	JA, SVÅRA	JA, LÄTTA	NEJ
a) Värk i skuldror, nacke eller axlar?	1	2	3
b) Ryggsmärtor, ryggvärk, höftsmärtor eller ischias?	1	2	3
c) Värk eller smärtor i händer, armbågar, ben eller knän?	1	2	3
d) Ängslan, oro eller ångest?	1	2	3
e) Dyslexi?	1	2	3
f) Astma, allergi?	1	2	3

- Fråga 16 Har ditt hälsotillstånd förändrats på något sätt – försämrats eller förbättrats - sedan förra intervjun (den ..../...)?
- 1 JA  
2 NEJ → Fråga 18

- Fråga 17 På vilket sätt? Förklara närmare:
- .....

**Sysselsättning** .....  
**Om Up är född 1931 eller tidigare (75 år eller äldre) → Fråga 25**

Fråga 18 Nästa avsnitt gäller sysselsättning och arbetsplatsförhållanden.  
 Vilka av följande sysselsättningar var aktuella för dig under förra veckan, dvs från måndagen den.....till och med söndagen den.....? Tänk också på sådant som endast varit aktuellt under en del av veckan.  
**Svarskort 100**

Läs upp **hela** listan och marker svar för **samtliga** alternativ

JA NEJ

- |          |   |   |   |
|----------|---|---|---|
| <b>A</b> | 1 | 2 | Anställd på heltid (inkl. semester, sjukledig, vård av barn)              |
| <b>B</b> | 1 | 2 | Anställd på deltid (inkl. semester, sjukledig, vård av barn)              |
| <b>C</b> | 1 | 2 | Arbetade eller var ledig från arbete i eget eller familjemedlems jordbruk |
| <b>D</b> | 1 | 2 | Arbetade eller var ledig från arbete i eget eller familjemedlems företag  |

**Till up som har flera arbeten (dvs JA på mer än ett av alt. A,B,C,D):**

Ringa in en bokstav

<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	Vilket är ditt huvudsakliga arbete?
1	2	3	4	

**Till alla som svarat JA på A, B, C eller D:**

- |          |   |   |  |
|----------|---|---|--|
| <b>I</b> | 1 | 2 | Hade du ytterligare något arbete som anställd eller egen företagare/jordbr.? |
| <b>J</b> | 1 | 2 | Var föräldraledig  |
| <b>K</b> | 1 | 2 | Var tjänstledig av andra skäl  |

<b>E</b>	1	2	Var arbetslös, sökte eller väntade på arbete (hela eller del av veckan)
----------	---	---	---

<b>F 1</b>	1	2	Har pension (även delpension)
------------	---	---	-------------------------------

<b>F 2</b>	1	2	Har sjuk- eller aktivitetsersättning (f.d. förtidspension resp. sjukbidrag)
------------	---	---	---

<b>G</b>	1	2	Studerade (inkl. vuxenutbildning och omskolning, exkl. hobbykurser)
----------	---	---	---

<b>H</b>	1	2	Var sjuk
----------	---	---	----------

<b>H</b>	1	2	Skötte eget hushåll (ej som löneanställd)
----------	---	---	---

<b>H</b>	1	2	Gjorde militärtjänst (inkl. civilförsvaret)
----------	---	---	---

<b>H</b>	1	2	Övrigt, nämligen .....
----------	---	---	------------------------

**Om A B C D. Annars → Fråga 25**

- Fråga 19 Har du bytt arbete eller fått arbete sedan förra intervjun (den ....!... )?
- 1 JA → **Fråga 25**  
2 NEJ

---

**Frågorna 20 – 24 gäller det huvudsakliga arbetet**

- Fråga 20 Är det bullrigt där du arbetar?

- 1 JA  
2 NEJ → **Fråga 23**

- 
- Fråga 21 Bullrar det för jämnan eller bara ibland?

- 1 FÖR JÄMNAN  
2 IBLAND

- 
- Fråga 22 Bullrar det öronbedövande?

- 1 JA  
2 NEJ

- 
- Fråga 23 Är ditt arbete psykiskt ansträngande?

- 1 JA  
2 NEJ

- 
- Fråga 24 Kräver ditt arbete tunga lyft?

- 1 JA  
2 NEJ
-

**Utbildning**

Nästa fråga gäller deltagande i olika slags utbildningar eller kurser.

## Fråga 25 a)

Deltog du i någon utbildning eller kurs under vårterminen 2005 och/eller höstterminen 2005 som anordnades av något studieförbund eller genom arbetsgivaren?

- 1 JA, STUDIEFÖRBUND
  - 1 JA, ARBETSGIVAREN
  - 1 NEJ → **Fråga 27**
- 

## b)

**Om studieförbund:**

Deltog du i en eller flera kurser/utbildningar anordnade av studieförbund under vårterminen 2005 eller höstterminen 2005?

- 1 EN KURS/UTBILDNING
- 2 FLERA KURSER/UTBILDNINGAR

Var kursen(-erna)/utbildningen(-arna) ett led i din yrkeskvalifikation eller var det som hobby eller för nöjes skull som du gick?

- 1 LED I YRKESKVALIFIKATION:

Hur många kurser? (under VT och/eller HT 2005)

- 1 HOBBY/NÖJE/INTRESSE:

Hur många kurser? (under VT och/eller HT 2005)

---

## c)

**Om arbetsgivare:**

Deltog du i en eller flera kurser/utbildningar genom arbetsgivaren under vårterminen 2005 eller höstterminen 2005?

- 1 EN KURS/UTBILDNING
- 2 FLERA KURSER/UTBILDNINGAR

### Medborgerliga aktiviteter

Fråga 27	Är du medlem i något politiskt parti (även politiskt ungdoms- eller kvinnoförbund)?
1	JA
2	NEJ → <b>Fråga 29</b>
<hr/>	
Fråga 28	Var du medlem även vid förra intervjun (den ..../... )?
1	JA
2	NEJ, HAR BLIVIT MEDLEM EFTER FÖRSTA INTERVJUN
<hr/>	
Fråga 29	Skulle du själv kunna författa en skrivelse och överklaga ett beslut av en myndighet?
1	JA
2	NEJ

### Trygghet

Fråga 30	a)	Har du själv under de <b>senaste 12 månaderna</b> , alltså sedan ....., varit utsatt för våld som ledde till sådana skador att det krävde besök hos läkare, tandläkare eller sjuksköterska?
		<b>Om JA:</b> Hur många gånger?
		ANTAL TILLFÄLLEN: .....
97		NEJ → <b>Fråga 32</b>

b) Var inträffade det?	Fyll i tablan kolumnvis.					
	NR 1	NR 2	NR 3	NR 4	NR 5	NR 6
EGEN BOSTAD .....	1	1	1	1	1	1
ANNANS BOSTAD .....	2	2	2	2	2	2
EGEN ARBETSPLATS .....	3	3	3	3	3	3
TÅG, BUSS ELLER TUNNELBANA ELLER T-BANE-, BUSS-, JÄRNVÄGSSTATION	4	4	4	4	4	4
RESTAURANG, DANSSTÄLLE, FOLKPARK ELLER ANNAT NÖJESSTÄLLE	5	5	5	5	5	5
GATA, TORG ELLER ANNAN ALLMÄN PLATS .....	6	6	6	6	6	6
ANNAN PLATS, <i>nämigen</i> .....	7	7	7	7	7	7

Fråga 31 Händelse detta/någon av de här tillfällena efter förra intervjun (den .... /...)?

- 1 JA, TILLFÄLLE NR .....
- 2 NEJ

Fråga 32 a) Har du (därutöver) under de senaste 12 månaderna, alltså sedan ....., blivit utsatt för våld som ledde till synliga märken eller kroppsskada men som inte krävde besök hos läkare, tandläkare eller sjuksköterska?

**Om JA:** Hur många gånger?

ANTAL TILLFÄLLEN: .....

97 NEJ → **Fråga 34**

**Fyll i tablan kolumnvis.**

b) Var inträffade det?	TILLFÄLLE					
	NR 1	NR 2	NR 3	NR 4	NR 5	NR 6
EGEN BOSTAD .....	1	1	1	1	1	1
ANNANS BOSTAD .....	2	2	2	2	2	2
EGEN ARBETSPLATS .....	3	3	3	3	3	3
TÅG, BUSS ELLER TUNNELBANA ELLER T-BANE-, BUSS-, JÄRNVAGSSTATION	4	4	4	4	4	4
RESTAURANG, DANSSTÄLLE, FOLK PARK ELLER ANNAT NÖJESSTÄLLE	5	5	5	5	5	5
GATA, TORG ELLER ANNAN ALLMÄN PLATS	6	6	6	6	6	6
ANNAN PLATS, <i>nämigen</i> .....	7	7	7	7	7	7

Fråga 33 Händelse detta/någon av de här tillfällena efter förra intervjun (den .... /...)?

- 1 JA, TILLFÄLLE NR .....
- 2 NEJ

Fråga 34 a) Har du (därutöver) under de senaste 12 månaderna, alltså sedan ....., blivit utsatt för våld som **inte** ledde till synliga märken eller kroppsskada?

**Om JA:** Hur många gånger?

ANTAL TILLFÄLLEN: .....

97 NEJ → **Fråga 36**



**Fyll i tablån kolumnvis.**  
TILLFÄLLE

b) Var inträffade det?	NR 1	NR 2	NR 3	NR 4	NR 5	NR 6
EGEN BOSTAD .....	1	1	1	1	1	1
ANNANS BOSTAD .....	2	2	2	2	2	2
EGEN ARBETSPLATS .....	3	3	3	3	3	3
TÅG, BUSS ELLER TUNNELBANA ELLER T-BANE-, BUSS-, JÄRNVÄGSSTATION	4	4	4	4	4	4
RESTAURANG, DANSSTÄLLE, FOLKPARK .....	5	5	5	5	5	5
ELLER ANNAT NÖJESSTÄLLE						
GATA, TORG ELLER ANNAN ALLMÄN PLATS .....	6	6	6	6	6	6
ANNAN PLATS, <i>nämigen</i> .....	7	7	7	7	7	7

Fråga 35 Händelse detta/någon av de här tillfällena efter förra intervjun (den .... /... )?

- 1 JA, TILLFÄLLE NR .....
- 2 NEJ

Fråga 36 a) Har du (därutöver) under de senaste 12 månaderna, alltså sedan ....., blivit utsatt för hot eller hotelser om våld som var farliga eller så allvarliga att du blev rädd?

**Om JA:** Hur många gånger?

ANTAL TILLFÄLLEN: .....

97 NEJ → **Fråga 38**

**Fyll i tablån kolumnvis.**  
TILLFÄLLE

b) Var inträffade det?	NR 1	NR 2	NR 3	NR 4	NR 5	NR 6
EGEN BOSTAD .....	1	1	1	1	1	1
ANNANS BOSTAD .....	2	2	2	2	2	2
EGEN ARBETSPLATS .....	3	3	3	3	3	3
TÅG, BUSS ELLER TUNNELBANA ELLER T-BANE-, BUSS-, JÄRNVÄGSSTATION	4	4	4	4	4	4
RESTAURANG, DANSSTÄLLE, FOLKPARK .....	5	5	5	5	5	5
ELLER ANNAT NÖJESSTÄLLE						
GATA, TORG ELLER ANNAN ALLMÄN PLATS .....	6	6	6	6	6	6
ANNAN PLATS, <i>nämigen</i> .....	7	7	7	7	7	7

Fråga 37 Hände detta/någon av de här tillfällena efter förra intervjun (den ..../....)?

- 1 JA, TILLFÄLLE NR .....
- 2 NEJ

---

Till sist har jag bara ett par frågor om hur du tyckte det var att bli intervjuad i Undersökningen om levnadsförhållanden, alltså den stora undersökning du var med i den ..../.....

Fråga 38 Hur gärna eller hur motvilligt deltog du i den undersökningen?

**Svarkort till fråga 38**

- 1 MYCKET GÄRNA
- 2 GANSKA GÄRNA
- 3 VARKEN GÄRNA ELLER MOTVILLIGT
- 4 GANSKA MOTVILLIGT
- 5 MYCKET MOTVILLIGT
  
- 8 VET INTE
- 9 VILL INTE SVARA

Fråga 39 Var det jobbigt att vara med eller gick det lätt?

- 1 JOBBIGT
- 2 VARKEN LÄTT ELLER JOBBIGT
- 3 LÄTT
  
- 8 VET INTE
- 9 VILL INTE SVARA

Fråga 40 Skulle du kunna tänka dig att vara med igen i en liknande undersökning om levnadsförhållande om t.ex. ett år?

**Svarkort till fråga 40**

- 1 JA, ABSOLUT
- 2 JA, KANSKE
- 3 TROLIGEN INTE
- 4 ABSOLUT INTE
  
- 8 VET INTE
- 9 VILL INTE SVARA

---

**STORT TACK FÖR HJÄLPEN!**

Intervjun slutar:

timmar		minuter	

2004:1 Örestat II, Utvidgad och integrerad flyttningsstatistik, Förstudie

2004:2 Frågebank över bakgrundsfrågor i postenkätundersökningar

2004:3 Samordning av BHU-H och HEK – Samordning 2003

2004:4 Könsuppdelad statistik – Återrapportering av ett uppdrag i regleringsbrevet

2004:5 Flergenerationsregistret 2003 – En beskrivning av innehåll och kvalitet

2004:6 Den nya HUT:en, Hushållens utgifter

2004:7 Multi-Generation Register 2003 - A description of contents and quality

2005:1 Flergenerationsregistret 2004 – En beskrivning av innehåll och kvalitet

2005:2 Multi-Generation Register 2004 - A Description of contents and quality

2005:3 Datafångst av personakter för komplettering av Flergenerationsregistrer

2005:4 Historiska befolkningsregistret (HBR)

2005:5 Migrations- och integrationsstatistik vid SCB – nuläge och utveckling

2006:1 Publikationsprojektet – slutrapport

2006:2 Översyn av forskarutbildningsstatistiken

2006:3 Historic population register

2006:4 Evalvering av utbildningsregistret

2006:5 Flergenerationsregistret 2005 – En beskrivning av innehåll och kvalitet

2006:6 Multi-generation register 2005. A description of contents and quality

2006:7 To measure and monitor internal migration based on national population register. IAOS Conference Ottawa, Canada

2007:1 Personnummer– dess konstruktion och hantering inom Statistiska centralbyrån

2007:2 Flergenerationsregistret 2006. En beskrivning av innehåll och kvalitet

2007:3 Multi-generation register 2006. A description of contents and quality

2008:1 Flergenerationsregistret 2007. En beskrivning av innehåll och kvalitet

2008:2 Multi-generation register 2007. A description of contents and quality

2009:1 Flergenerationsregistret 2008. En beskrivning av innehåll och kvalitet

2009:2 Multi-generation register 2008- A description of contents and quality

ISSN 1654-4331 (online)

All officiell statistik finns på: **www.scb.se**  
Kundservice: tfn 08-506 948 01

All official statistics can be found at: **www.scb.se**  
Customer service, phone +46 8 506 948 01